

**Noise Reduction Algorithms and Performance Metrics
for Improving Speech Reception in Noise
by Cochlear-Implant Users**

by

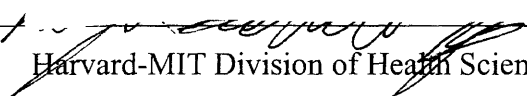
Raymond Lee Goldsworthy
B.S., University of Kentucky (1997)

Submitted to the Harvard-MIT Division of
Health Sciences and Technology
In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy
at the

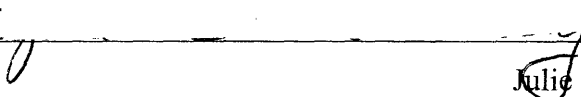
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[February 2005]
January 2005

© Massachusetts Institute of Technology, 2005. All rights reserved.


Signature of Author


Harvard-MIT Division of Health Sciences and Technology
January 10, 2005

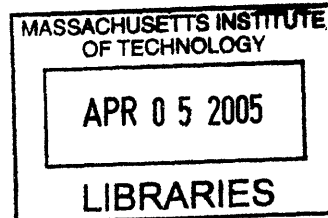
Certified by


Julie E. Greenberg, Ph.D.
Principal Research Scientist, Research Laboratory of Electronics

Accepted by


Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical & Electrical Engineering
Director, Harvard-MIT Division of Health Sciences and Technology

ARCHIVES





Noise Reduction Algorithms and Performance Metrics for Improving Speech Reception in Noise by Cochlear-Implant Users

by
Raymond Lee Goldsworthy

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on January 10th, 2005, in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Abstract

This thesis addresses the design and evaluation of algorithms to improve speech reception for cochlear-implant (CI) users in adverse listening environments. We develop and assess performance metrics for use in the algorithm design process; such metrics make algorithm evaluation efficient, consistent, and subject independent. One promising performance metric is the Speech Transmission Index (STI), which is well correlated with speech reception by normal-hearing listeners for additive noise and reverberation. We expect the STI will effectively predict speech reception by CI users since typical CI sound-processing strategies, like the STI, rely on the envelope signals in frequency bands spanning the speech spectrum. However, STI-based metrics have proven unsatisfactory for assessing the effects of nonlinear operations on the intelligibility of processed speech. In this work we consider modifications to the STI that account for nonlinear operations commonly found in CI sound-processing and noise reduction algorithms.

We consider a number of existing speech-based STI metrics and propose novel metrics applicable to nonlinear operations. A preliminary evaluation results in the selection of three candidate metrics for extensive evaluation. In four central experiments, we consider the effects of acoustic degradation, N-of-M processing, spectral subtraction, and binaural noise reduction on the intelligibility of CI-processed speech. We assess the ability of the candidate metrics to predict speech reception scores. Subjects include CI users as well as normal-hearing subjects listening to a noise-vocoder simulation of CI sound-processing.

Our results show that: 1) both spectral subtraction and binaural noise reduction improve the intelligibility of CI-processed speech and 2) of the candidate metrics, one method (the normalized correlation metric) consistently predicts the major trends in speech reception scores for all four experiments.

Thesis Supervisor: Julie E. Greenberg, Ph.D.
Title: Principal Research Scientist, Research Laboratory of Electronics



Acknowledgements

This work was supported by the National Institute of Health and the National Institute on Deafness and Other Communication Disorders (grant numbers 1-R01-DC00117 and 5-T32-DC0038) and the Harvard-MIT Division of Health Sciences and Technology Training Fellowship.

I thank my thesis committee for the guidance I have received over the years. Julie Greenberg has been integral to all aspects of this thesis from genesis to finale. She has been an exemplar of advising through the emphasis of academic rigor and scientific method. I thank Louis Braida, the chair of the thesis committee, for continual insight into the realm of psychoacoustics. I thank Donald Eddington for his insight into cochlear implants as well as his guidance towards becoming an adept research scientist. I thank Karen Payton for her insight into the nature of STI and the many conversations that took place as the novel metrics presented in this thesis were developed.

I have had a wonderful time along the way. I am very grateful for the 7th floor crew who brought sanity when least expected (and most needed) over the years. I'm thankful for everyone who has made these years an adventure.

I thank my family. I thank my brothers who are a source of inspiration. I thank my grandmothers who have shaped my spirit. I thank my wife who is my best friend and a wonderful companion on this fantastic voyage.

I dedicate this thesis to my parents. They have always simply wished for me to be happy, and they will be pleased to know that I am.



Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 10 |
| 2 | Background | 15 |
| 2.1 | Cochlear Implants | 16 |
| 2.2 | Speech Transmission Index | 22 |
| 2.3 | Noise Reduction Algorithms | 29 |
| 3 | STI Modifications | 34 |
| 3.1 | Modifications of the STI for Nonlinear Operations | 35 |
| 3.2 | CI-Specific Intelligibility Metrics | 39 |
| 4 | Experimental Design | 45 |
| 4.1 | Stimuli | 46 |
| 4.2 | Subjects | 46 |
| 4.3 | Experimental Conditions | 48 |
| 4.4 | Experimental Procedure for Main Experiments | 55 |
| 4.5 | Calculation of Intelligibility Metrics | 58 |
| 4.6 | Psychometric Model | 62 |
| 5 | Preliminary Experiments | 71 |
| 5.1 | Preliminary Study of Binaural Noise Reduction Algorithms | 72 |
| 5.2 | Evaluation of Speech-Based STI for Nonlinear Operations | 75 |
| 5.3 | Selection of Candidate Metrics | 83 |

| | | |
|----------|---|------------|
| 6 | Experiment 1: Acoustic Degradation | 95 |
| 6.1 | Introduction | 96 |
| 6.2 | Conditions | 97 |
| 6.3 | Results of the Listening Experiment | 99 |
| 6.4 | Results of the Intelligibility Predictions | 103 |
| 6.5 | Discussion | 105 |
| 6.6 | Conclusions | 109 |
| | | |
| 7 | Experiment 2: N-of-M Processing | 122 |
| 7.1 | Introduction | 123 |
| 7.2 | Conditions | 124 |
| 7.3 | Results of the Listening Experiment | 125 |
| 7.4 | Results of the Intelligibility Predictions | 126 |
| 7.5 | Frequency-Band Analysis | 128 |
| 7.6 | Discussion | 129 |
| 7.7 | Conclusions | 132 |
| | | |
| 8 | Experiment 3: Spectral Subtraction | 140 |
| 8.1 | Introduction | 141 |
| 8.2 | Conditions | 142 |
| 8.3 | Results of the Listening Experiment | 143 |
| 8.4 | Results of the Intelligibility Predictions | 145 |
| 8.5 | Frequency-Band Analysis | 147 |
| 8.6 | Discussion | 148 |
| 8.7 | Conclusions | 150 |
| | | |
| 9 | Experiment 4: Binaural Noise Reduction | 164 |
| 9.1 | Introduction | 165 |
| 9.2 | Conditions | 166 |
| 9.3 | Results of the Listening Experiment | 168 |
| 9.4 | Results of the Intelligibility Predictions | 170 |
| 9.5 | Frequency-Band Analysis | 172 |
| 9.6 | Discussion | 173 |
| 9.7 | Conclusions | 176 |

| | |
|---|------------|
| 10 Discussion | 190 |
| 10.1 Summary of Intelligibility Predictions | 191 |
| 10.2 Evaluation of the Performance Metrics across Experiments | 192 |
| 10.3 Future Work | 194 |
| 10.4 Final Conclusions | 196 |
| | |
| Appendix A: Selection of Candidate Metrics | 204 |
| | |
| Appendix B: Theoretical Derivations Concerning the STI and NCM Methods | 208 |
| B.1 Stochastic Reformulation of the Envelope regression method | 209 |
| B.2 Normalized Correlation Expressed as Energy-Weighted MTF | 210 |
| B.3 Modulation Metric (M) Expressed as Energy-Weighted MTF | 211 |
| B.4 Relation of Energy-Weighted to One-Third Octave Averaging | 214 |
| B.5 Relation of Apparent SNR to True SNR | 215 |
| B.6 Effect of Using ρ^2 Directly as the Transmission Index | 217 |
| | |
| Appendix C: Repeated Measures Analysis of Variance Tables | 222 |
| | |
| References | 229 |

Chapter 1

Introduction

The long-term goal of this research is to design and evaluate algorithms that improve speech reception for cochlear-implant (CI) users in adverse listening environments such as additive noise and reverberation. Much attention has been given to the related problem of improving speech reception in adverse environments for hearing-aid users. This work differs from previous efforts in that we consider the noise reduction problem with respect to the CI sound-processing strategy as part of the design process. Because the CI sound-processor encodes a subset of the available acoustic information, it should be possible to design algorithms specifically tailored to improve the intelligibility of the coded signal.

Towards this end, we wish to determine a physical performance metric that is specifically tailored to CI sound-processing strategies. A physical performance metric is a predictor of speech reception that is derived solely from acoustical analysis of the speech signal and does not require measurements of speech reception by human subjects. Since speech reception will be limited to the information coded by the speech processor, it makes sense to evaluate algorithms based on analysis of the coded information. The advantages of determining a relevant physical performance metric for CI users is that it makes algorithm evaluation efficient, consistent, and subject independent. Evaluations can be made across general algorithm classes to screen for beneficial candidate algorithms. For a particular algorithm, the performance metric can be used to optimize

performance by guiding selection of parameter values. The performance metric can also be used to perform preliminary evaluations to determine if subject testing is warranted.

The speech transmission index (STI) is a physical performance metric that is well correlated to speech reception in normal-hearing listeners. The STI is based on the envelope signals in a number of frequency bands that span the relevant spectrum for speech. We hypothesize that the STI is particularly suited as a predictor of CI user performance since typical CI sound-processing strategies also extract and transmit the envelope signals in a number of frequency bands that span the speech spectrum. Existing methods for calculating the STI, as well as novel methods proposed in this thesis, will be tailored to specific CI sound-processing strategies and will be evaluated to select the metric that best serves as a predictor of speech reception for CI-processed speech.

In Chapter 2 we review the background material relevant to cochlear implants, STI, and noise reduction algorithms. The background material stresses similarities between CI sound-processing strategies and STI calculation procedures that allow the STI procedures to be tailored to a specific CI sound-processing strategy. Because the use of noise vocoders as a simulation of CI sound-processing strategies is integral to the work described in this thesis, several issues related to these simulations are addressed in Section 2.1.2. A key point raised in Chapter 2 is that current STI procedures are not capable of assessing the effects of nonlinear operations on speech reception. In Chapter 3 we describe how STI may be modified to address this problem. We also propose a novel metric, termed the normalized correlation metric (NCM) that is better suited for nonlinear operations. Specific procedures for tailoring the performance metrics to a given CI sound-processing strategy are also given in Chapter 3.

In Chapter 4 we describe the experimental methods for the experiments presented in this thesis. In Chapter 5 we describe the results of our preliminary studies. These preliminary studies include an initial evaluation of binaural noise reduction algorithms (Section 5.1), an evaluation of the speech-based STI in the context of nonlinear operations (Section 5.2), and an evaluation of the speech-based STI metrics to select a

subset that comprise the candidate metrics considered for detailed evaluation in this thesis (Section 5.3).

The two main goals of this thesis are:

- Identify physical performance metrics that predict speech reception for CI-processed speech in adverse listening conditions.
- Design and evaluate signal processing strategies to improve speech reception in adverse listening conditions for CI users.

There are four main experiments that support these goals. Experiments 1 and 2 address our goal of developing a physical performance metric that serves as an accurate predictor of speech reception for CI-processed speech for acoustic degradations and nonlinear operations. Experiments 3 and 4 evaluate the metrics, but also address our long-term goal of developing noise reduction algorithms to improve speech reception in noise for CI users. Two limitations that CI users face are a reduction in fine spectral information and a lack of binaural information. The noise reduction strategies implemented and tested, spectral subtraction and binaural noise reduction, attempt to address these limitations.

Experiment 1, presented in Chapter 6, considers the effect of acoustic degradation (i.e. additive noise and reverberation) on the intelligibility of CI-processed speech. We investigate whether or not differences in speech reception exist between normal-hearing listeners (*not* listening to a vocoder simulation of CI processing) and CI users for different types of degradations. We are interested not only in overt differences, such as measures of reception in quiet, but also more subtle differences, such as how speech reception degrades in different environments. One particular interest is the effect of noise source modulation on speech reception. We wish to determine if CI subjects perform differently for a highly modulated (i.e. time-varying spectrum) noise source such as a competing talker compared to an unmodulated (i.e. stationary spectrum) noise source. We hypothesize that such a difference may exist since normal-hearing listeners may be able to capitalize on cues that CI users cannot resolve.

We assess the ability of the candidate metrics to quantify the effects of acoustic degradations on speech reception. By evaluating the candidate metrics for a range of acoustic degradation conditions including additive noise and reverberation, we establish a baseline comparison for novel metrics with more traditional STI approaches. Such a baseline is important since the traditional STI is an accurate predictor (for normal-hearing listeners) of speech reception for additive stationary noise and reverberation and it is important that new metrics retain this property.

Experiment 2, presented in Chapter 7, considers the effect of N-of-M processing on speech reception. N-of-M processing is an operation that is employed in some CI sound processors. The N-of-M operation selects a subset of the envelope signals to transmit to the implanted electrode array per stimulation cycle. The rationale behind N-of-M processing is that a subset of the envelope signals can be used to transmit the essential signal energy. One motivation for studying N-of-M processing is to quantify how speech reception is affected by coding only a subset of the envelope information. A second motivation is that N-of-M processing highlights inadequacies in certain STI approaches. The effect of N-of-M processing is analyzed for various noise types and numbers of active channels. A performance metric that accounts for the effects of N-of-M processing on speech reception will be applicable to a broader class of CI users.

Experiment 3, presented in Chapter 8, considers the effect of spectral subtraction on the intelligibility of CI-processed speech. Previous studies have shown (see Section 2.3.1) that spectral subtraction does not improve speech reception for normal-hearing listeners but does improve speech reception for CI users. We discuss the limited spectral resolution of CI systems as a cause of this performance difference. It is argued that the performance metrics are better suited for CI users precisely because they are based on wider frequency bands. We also consider the possibility of using the successful candidate metrics to optimize selection of parameters within the spectral subtraction algorithm.

Experiment 4, presented in Chapter 9, considers whether binaural noise reduction can improve the intelligibility of CI-processed speech. The majority of CI users have a

single implant and therefore do not have access to binaural information. Binaural noise reduction algorithms capitalize on two microphone inputs—one over each ear—and the corresponding binaural cues to improve the intelligibility of speech in noise. Thus, these algorithms attempt to enhance the signal before delivery to the implant. We investigate the benefit of this approach for a variety of acoustic degradations and consider the utility of the candidate metrics in predicting the results.

These four experiments are designed to investigate the effects of different degradations and processing algorithms on speech reception for CI-processed speech. The various types of degradations and processing conditions considered yield insight into basic speech reception psychoacoustics for CI-processed speech. In addition, the development and selection of the best candidate metric provide a framework for analyzing these results. Chapter 10 is a general discussion of the successes (and failures) of the candidate metrics. We analyze the performance of the most promising metric across experiments and suggest future work and adaptations.

Chapter 2

Background

This thesis is concerned with developing a physical performance metric specifically tailored to CI sound-processing strategies in order to design and evaluate noise reduction algorithms. As such, it brings together three fields of auditory science: cochlear implants, speech reception metrics, and noise reduction algorithms. This chapter reviews the relevant background material in each of these areas. A primary hypothesis of this thesis is that the STI will serve as a reliable performance metric for CI users. As the background material is developed in this section, the reader should begin to appreciate the similarities between STI computation and CI sound-processing strategies.

2.1 Cochlear Implants

A cochlear implant is a prosthetic device that can restore a degree of hearing to profoundly impaired individuals. Cochlear implants generate a sound sensation by directly stimulating the auditory nerve with electric currents. In this manner, a cochlear implant bypasses damaged components of the external, middle, and inner ears. Cochlear implants are appropriate for profoundly impaired individuals who receive little or no benefit from conventional hearing aids or from corrective surgery. There are roughly 25,000 CI users in the United States and over 250,000 hearing-impaired individuals who would be good candidates for cochlear implantation (NIDCD, 2004).

The key components of a cochlear implant are the microphone, the speech processor, and the electrode array that stimulates surviving auditory nerve fiber. The role of the CI sound-processing strategy is to transform the signal obtained by the microphone to electric stimuli delivered to the auditory nerve via the electrode array.

The benefit CI users receive from current devices is limited by electrophysiological constraints. Consider the schematic of the internal apparatus of a cochlear implant as given in Figure 2.1. The CI sound-processing strategy might attempt

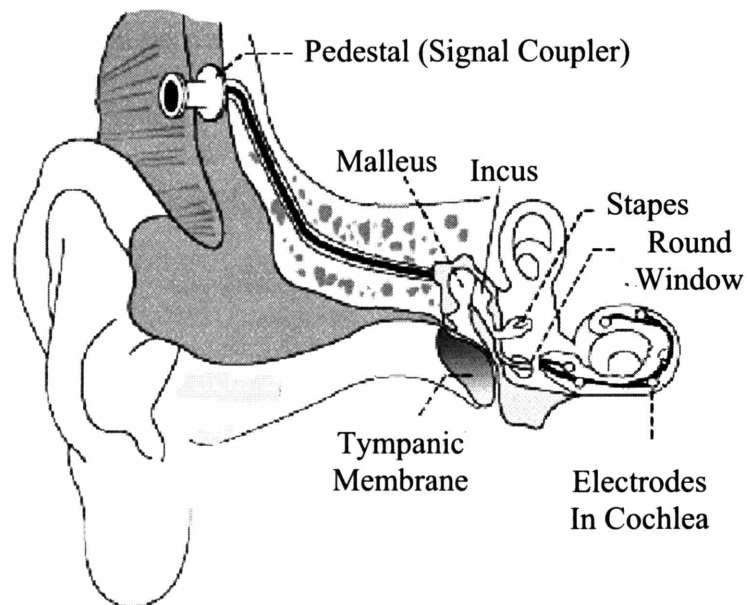


Figure 2.1: Schematic of cochlear implant (after Eddington and Pierschalla, 1994 with permission).

to use electric stimuli to recreate the auditory nerve response that occurs in a normal-hearing listener. However, the electrodes stimulated, their position and stimulation rate are limited by electrophysiological safety concerns, electrode technology, and by surgical techniques. Furthermore, the integrity of the stimulated auditory nerve varies widely between implant recipients. Common implants today have between 6 and 22 electrodes that can be stimulated at rates of 250-8000 Hz per electrode.

2.1.1 CI Sound-Processing Strategies

With these constraints in place, only a subset of the acoustic information available to a normal-hearing listener can be coded for the CI user. Different attempts have been made for coding a subset of the acoustic information into electric stimuli. (The overview given here is based on Loizou, 1998). The various attempts can be classified into three types: feature extraction strategies, waveform strategies, and N-of-M strategies.

The first strategies developed for the Nucleus device (manufactured by Cochlear Corp.) was a feature extraction strategy—the F0/F2 strategy—that assumed that the incoming waveform was speech and attempted to extract relevant features for stimulus coding. In that strategy, the second formant was estimated and used to select a particular electrode; the fundamental frequency, F0, was estimated and used to control the stimulation rate. A subsequent strategy—the F0/F1/F2 strategy—estimated the first formant in addition to the fundamental and second formant. Further improvements led to the MPEAK strategy that estimated, and attempted to code, the energy associated with frequencies higher than the second formant. One problem with these strategies is that the formant trackers performed poorly in adverse listening environments.

The waveform strategies, in contrast to the feature extraction strategies, do not assume the incoming waveform is speech and, consequently, attempt to convey the general spectral properties of the incoming waveform. The compressed-analog (CA) approach, originally used in the Ineraid device, processes the incoming signal into a number of frequency bands. The output of each frequency band is compressed and delivered to a corresponding implanted electrode. A major concern with the CA strategy

is that the simultaneous stimulation of electrodes would produce unwanted interactions. The continuous interleaved sampling (CIS) strategy was developed to avoid these unwanted interactions. The CIS strategy processes the incoming waveform into a number of frequency bands, but then extracts the envelope for each band. This envelope is compressed and used to modulate electric pulse trains that are interleaved in time across electrodes. Subject testing showed that the CIS strategy produced substantial gains in speech reception over the CA strategy. Some researchers have argued that the advancement of electrode arrays with positioning systems will allow for simultaneous stimulation without electrode interactions, leading to new interest in CA strategies (Osberger and Fisher, 1990). However, stimulation procedures using interleaved pulses modulated by extracted envelopes are more commonly used in state-of-the-art processors.

The signal processing associated with CIS strategies is illustrated in Figure 2.2:

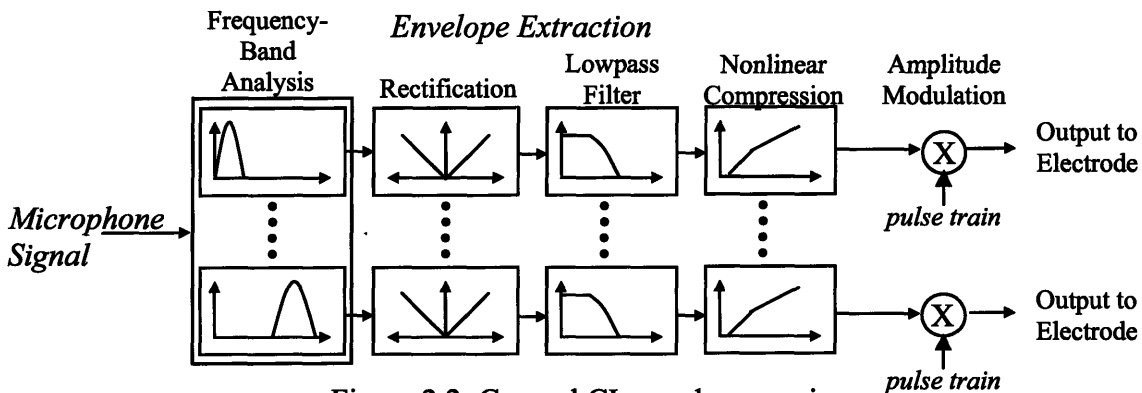


Figure 2.2: General CI sound-processing.

This is a very general diagram; the key point is that the stimulation of a given electrode is based on the envelope of a particular frequency band. In other words, a processor contains a number of bandpass filters with each frequency band corresponding to a particular electrode in a one-to-one fashion.

For the CIS strategy, the microphone signal is sometimes first pre-emphasized followed by processing through a filterbank. The bandpass filtered signals are processed to extract the envelopes by rectification followed by lowpass filtering. These envelope signals are compressed and used to amplitude modulate biphasic electric pulses. Figure 2.3 illustrates four channels of biphasic-pulse, CIS stimuli delivered to four electrode

contacts. Since the pulse trains are interleaved, no two electrodes are stimulated at the same time.

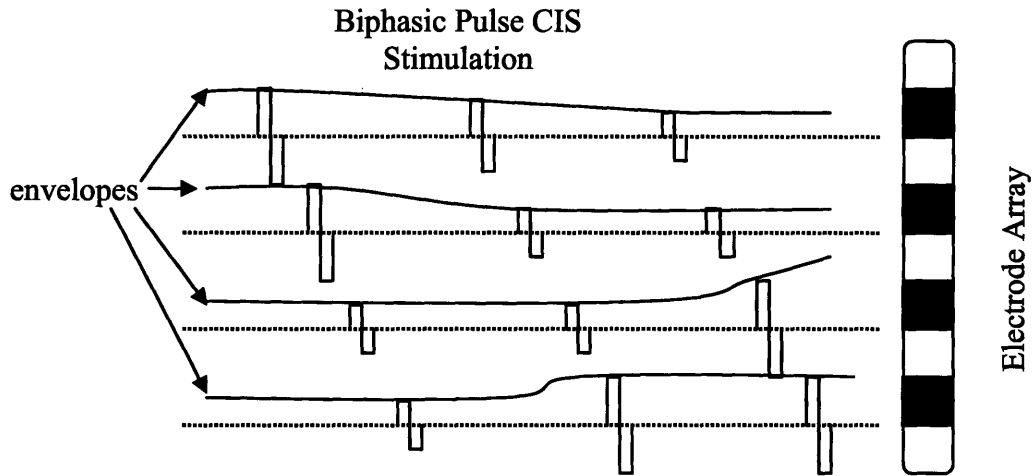


Figure 2.3: Illustration of 4-channel CIS stimulation.

The stimulation order can be varied to minimize electrode interaction. For example, if the electrodes in a six electrode array are numbered from base to apex, then the biphasic pulse may be delivered in order [1,2,3,4,5,6] as in Figure 2.3, or as [1,4,2,5,3,6] to minimize interaction. The term *stimulation cycle* will be used to define the period of time over which each electrode is stimulated once.

N-of-M processing strategies are quite similar to the CIS strategies except that only a subset of the electrodes are stimulated in each stimulation cycle. In particular, the frequency bands are analyzed per stimulation cycle to determine which N electrodes of M possible will be stimulated. The rationale behind this approach is that by coding the N frequency bands with the highest energy, most of the information will be transmitted. By only stimulating a subset of the channels, the algorithm allows the channels that are selected to be stimulated at higher pulse rates. A general CI sound-processing strategy that includes N-of-M processing is illustrated in Figure 2.4.

The operation of the N-of-M subsystem is to select a subset of the envelopes to code in each stimulation cycle based on some criterion. For example, N envelopes with the highest energy might be chosen. This process can be thought of as setting the other envelope signals to zero for this stimulation cycle. Figure 2.5 illustrates the effect of N-

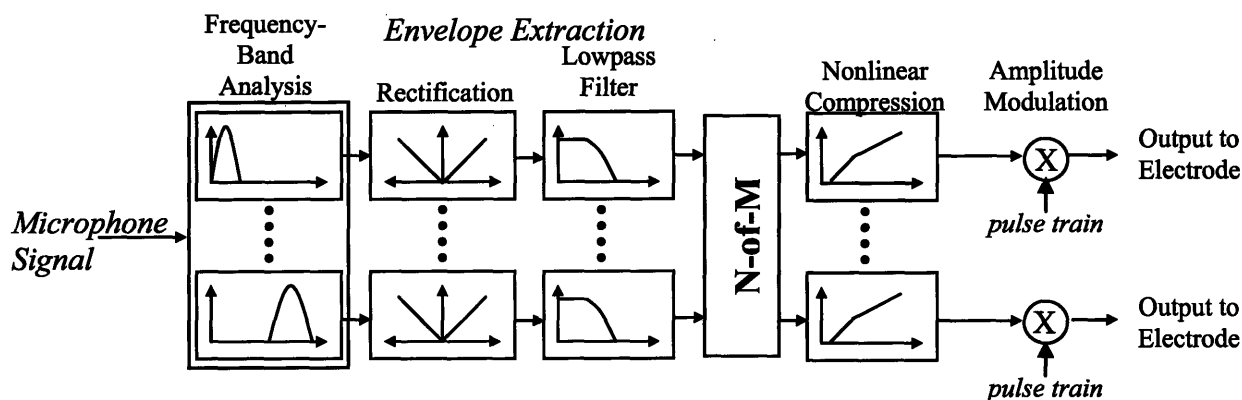


Figure 2.4: General CI sound-processing including the N-of-M operation.

of-M processing on the sentence “the birch canoe slid on the smooth planks” with $N = 2$ and $M = 6$. The example shown in Figure 2.5 illustrates that 2 out of 6 electrodes provides a fair representation of high-frequency and low-frequency bands, but that the mid-frequency bands (especially the mid-frequency band labeled on the figure) contain substantial degradation.

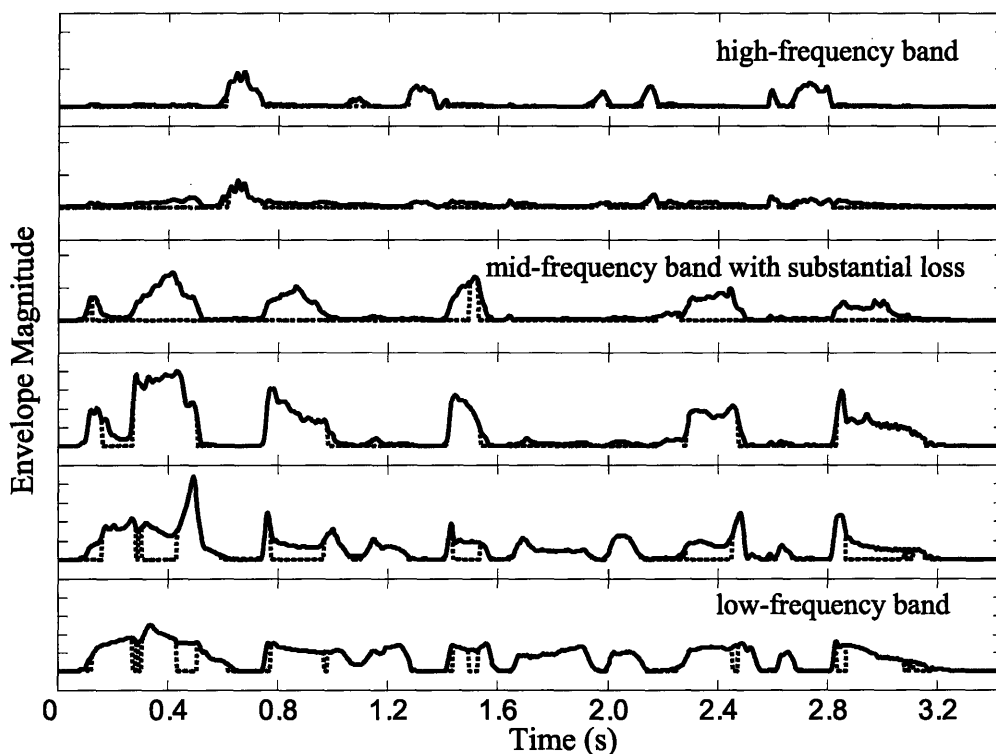


Figure 2.5: Illustration of the effect of N-of-M processing on envelope signals using 2-of-6. Solid and dotted lines illustrate envelopes with and without the N-of-M processing.

The SPEAK processing strategy used with the Spectra speech processor manufactured by Cochlear Corporation is an N-of-M processor with $M = 20$ and N varying from 6 to 10 depending on the spectral composition of the signal. The N chosen envelopes are used to modulate biphasic pulses as in the CIS strategy. A more recent N-of-M strategy designed for the Nucleus-24 system is referred to as ACE (advanced combination encoders). The ACE strategy is similar to the SPEAK strategy in that $M = 22$ and N varies from 6 to 10, but uses higher stimulation rates.

One goal of this thesis is to incorporate the effects of N-of-M processing into the STI model, as will be discussed in Section 3.2. A primary motivation for focusing on N-of-M processing is that the direct manipulation of envelope signals that occurs in the algorithm may prove insightful to gross failings of different STI approaches.

2.1.2 Vocoder Simulations of CI Sound-Processing Strategies

While the CI sound-processing strategy is important in determining an effective mapping of input signal to electric stimulation, the speech-reception benefit enjoyed by CI users of the same strategy is dependent on subject-specific factors (e.g. electrode placement, auditory-nerve survival, and language development). As a result of these subject-specific factors, subject performance varies widely even for subjects using the same sound-processing strategy.

The effect of certain elements of the CI sound-processing strategy on speech reception can be investigated using normal-hearing listeners. To this end, researchers have developed simulations of CI sound-processing strategies. These simulations attempt to capture certain elements of the CI sound-processing strategy while generally avoiding the subject-specific differences. As mentioned in the preceding section, CI sound-processing strategies are generally based on the envelope information in a number of frequency bands. The vocoder simulation extracts the speech envelope using the same procedure as the CI processor of interest. The envelope information is delivered to the normal-hearing listener by modulating a carrier (e.g. sinusoids or band-limited noise), then band-limiting and summing the bands. Normal-hearing subjects listening to a

vocoder simulation of CI sound-processing strategies have been used to investigate the effects of CI sound-processing strategies on speech reception for a variety of processing effects (Shannon et al., 1995; Loizou et al., 1999; Fu et al., 1998, 2000; Dorman et al., 1997a, 1997b, 1998a, 1998b). One relevant result from these studies is that the best CI users perform comparably to normal-hearing subjects listening to the vocoder simulation.

To avoid confusion between normal-hearing subjects listening to a vocoder simulation of CI sound-processing strategies and normal-hearing subjects listening to unprocessed speech, the former will be referred to as NH-CI_{sim}. Specific simulations will be identified using different subscripts. For example, normal-hearing subjects listening to an 8-channel CI sound-processing strategy will be referred to as NH-CI₈.

2.2 Speech Transmission Index

2.2.1 Development of the STI¹

Early attempts to predict speech reception led to the development of the articulation index (AI) (French and Steinberg, 1947; Kryter, 1962a, 1962b). A fundamental principle of the AI is that the intelligibility of speech depends on a weighted average of the signal to noise ratios (SNRs) in frequency bands spanning the speech spectrum. By accounting for the contribution of different regions of the spectrum to intelligibility, the AI successfully predicts the effects of additive noise and simple linear filters.

The STI (Houtgast and Steeneken, 1971; Steeneken and Houtgast, 1980; IEC, 1998) is an intelligibility metric that differs from the AI by using reduction in signal modulation rather than band-specific SNRs. By including modulation reduction in the frequency-band analysis, the STI can predict the effects of reverberation as well as additive noise. Calculation of the STI is based on changes in signal modulation when modulated probe stimuli are transmitted through a channel of interest. The degraded responses to the probe stimuli are measured in multiple frequency bands for a range of modulation frequencies relevant to speech. The STI successfully quantifies the effects of

¹ Section 2.2.1 is reproduced from Goldsworthy and Greenberg, 2004: Section I, "Introduction." Changes were made to section titles and numbers in order to be internally consistent with this thesis.

room acoustics and broadcast channels on speech reception (Steeneken and Houtgast, 1982). The STI has also been adapted for use with hearing-impaired subjects (Humes et al., 1986; Ludvigsen, 1987; Payton et al., 1994).

Steeneken and Houtgast (1980) suggest that applying the STI to nonlinear operations requires more sophisticated probe signals than used in their original procedure. They introduced complex test signals that combine modulated noise with artificial speech-like signals, allowing the STI to predict the effects of automatic gain control and peak clipping. Other researchers have developed variations that use speech, rather than an artificial probe, to investigate nonlinear operations. These speech-based methods have been used to analyze dynamic amplitude compression (Hohmann and Kollmeier, 1995; Payton et al., 2002; Drullman, 1995), spectral subtraction (Ludvigsen et al., 1993), and envelope processing (Drullman, 1994a, 1994b, and 1995). In addition, speech-based STI methods have been used to investigate the intelligibility differences between clear and conversational speech (Payton et al., 1994; Payton et al., 1999).

The speech-based STI methods have generally failed to predict performance for nonlinear operations. In some studies, STI intelligibility predictions have been qualitatively inconsistent with performance results. A study of envelope expansion found that “the prediction from STI is in the wrong direction for the expansion conditions” (Van Buuren et al., 1998). In an investigation of speech-based STI and spectral subtraction, researchers concluded “STI, even in its modified version, is an unreliable predictor when nonlinear processes are involved” (Ludvigsen et al., 1993). Other researchers (Drullman, 1995; Payton et al., 2002; Hohmann and Kollmeier, 1995; Goldsworthy and Greenberg, 2001, 2003, 2004) have also concluded that speech-based STI methods proposed thus far do not adequately predict the intelligibility of nonlinearly processed speech. This general failure of the STI methods in the context of nonlinear operations motivates our introduction of novel methods in Chapter 3.

2.2.2 STI Calculation²

Both the traditional and speech-based STI methods employ a frequency-band analysis as illustrated in Figure 2.6. A bank of bandpass filters splits the clean (probe) and degraded (response) signals into frequency bands, where i indicates the frequency band number. Typically, octave bands with center frequencies from 125 to 8000 Hz are used. For each

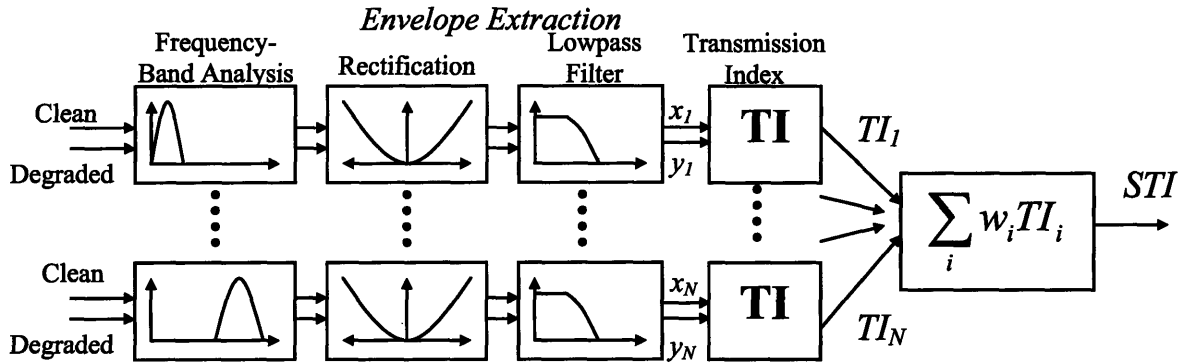


Figure 2.6: General STI calculation.

band, the clean and degraded envelope signals, $x_i(t)$ and $y_i(t)$, respectively, are computed by rectification and lowpass filtering and then compared to determine a transmission index, TI_i . The TI_i values are combined using a weighted average to determine the STI value. The various STI methods differ in how the envelope signals are computed and in how the TI_i values are computed from the envelopes.

Traditional Method of Computing the STI

For the traditional method (Steeneken and Houtgast, 1980), the TI_i values are computed from an intermediate function called the modulation transfer function (MTF). The MTF is a function of modulation frequency, f , calculated individually for each value of f . For each frequency band, the clean signal consists of speech-shaped noise that has been bandpass filtered (based on the analysis band) and then intensity (square-law rectification) modulated at a particular modulation frequency. The clean signal is passed

² Section 2.2.2 is reproduced from Goldsworthy and Greenberg, 2004: Section II, "Background." Changes were made to section titles and numbers, as well as equation and table numbers, in order to be internally consistent with this thesis.

through the system to be evaluated and the output is traditionally referred to as the “response” or degraded signal. The fractional change in modulation depth between clean (x) and degraded (y) intensity envelopes is quantified for that value of f , and the process is repeated for other modulation frequencies to determine the complete MTF for one frequency band. The MTF is typically characterized using modulation frequencies ranging from $f = 0.63$ Hz to $f = 12.7$ Hz in one-third octave intervals (IEC, 1998). As an alternative to artificial probe signals, Houtgast and Steeneken (1985) proposed determining the MTF for each frequency band from spectra of the intensity envelopes of running speech. Omitting the subscript i to simplify notation, this approach can be described as (Drullman, 1994b)

$$MTF(f) = \alpha \frac{|Y(f)|}{|X(f)|} = \alpha \sqrt{\frac{S_{yy}(f)}{S_{xx}(f)}} \quad (2.1)$$

where $\alpha = \mu_x / \mu_y$, $\mu_x = E\{x(t)\}$, $\mu_y = E\{y(t)\}$, and $E\{\cdot\}$ denotes expected value. $|X(f)|$ and $|Y(f)|$ are magnitude spectra, and $S_{xx}(f)$ and $S_{yy}(f)$ are power spectra, of the clean and degraded envelope signals, respectively.

The signal-to-noise ratio (SNR) in decibels as a function of f is calculated for each frequency band as

$$SNR_i(f) = 10 \log_{10} \left(\frac{MTF_i(f)}{1 - MTF_i(f)} \right). \quad (2.2)$$

An overall apparent SNR ($aSNR_i$) for each frequency band is determined by clipping the $SNR_i(f)$ values and then averaging across modulation frequencies, that is,

$$cSNR_i(f) = \begin{cases} -15 & SNR_i(f) < -15 \\ SNR_i(f) & -15 \leq SNR_i(f) \leq 15 \\ 15 & SNR_i(f) > 15 \end{cases} \quad (2.3)$$

$$aSNR_i = \text{mean}(cSNR_i(f)). \quad (2.4)$$

The transmission index is a linear function of the apparent SNR for each band, defined to be between zero and one,

$$TI_i = \frac{aSNR_i + 15}{30}. \quad (2.5)$$

Finally, the overall STI value is calculated as a weighted average of the TI_i values,

$$STI = \sum_i w_i TI_i, \quad (2.6)$$

where w_i is a psychoacoustically derived weighting (Pavlovic, 1987). The weights, w_i , are defined to sum to one, thereby restricting the STI values to a range between zero and one.

Speech-Based STI Methods

This section summarizes four speech-based methods proposed in the literature. The first three speech-based methods use intensity envelopes calculated by squaring and then smoothing, while the fourth uses magnitude envelopes. For each method, the description focuses on the calculation of TI_i for one frequency band. To simplify notation, the subscript i is omitted for intermediate variables such as $MTF(f)$ and $aSNR$.

1) Magnitude Cross-Power Spectrum Method

Payton and colleagues (2002) proposed a speech-based method where the MTF is based on the magnitude of the cross-power spectra as given by

$$MTF(f) = \alpha \frac{|S_{xy}(f)|}{|S_{xx}(f)|}, \quad (2.7)$$

where $S_{xy}(f)$ is the cross-power spectrum (CPS) of the clean and degraded envelopes. The MTF given by Eq. 2.7 is used in Eq. 2.2, and the STI is calculated from Eqs. 2.2 through 2.6.

2) Real Cross-Power Spectrum Method

Drullman and colleagues (1994b) introduced a phase-locked MTF in order to investigate the effects of reducing low-frequency modulations on the intelligibility of speech. The phase-locked MTF is defined as

$$MTF(f) = \alpha \operatorname{Re} \left(\frac{S_{xy}(f)}{S_{xx}(f)} \right), \quad (2.8)$$

where $\operatorname{Re}(\cdot)$ denotes taking the real part of the complex-valued function. Although they did not propose a corresponding STI calculation procedure, the MTF in Eq. 2.8 could be used to calculate the STI in conjunction with Eqs. 2.2 through 2.6.

3) Envelope Regression Method

Ludvigsen and colleagues (1990) proposed a method where the clean envelope signal, $x(t)$, and the degraded envelope signal, $y(t)$, are compared using linear regression analysis. In this method, the apparent SNR for each frequency band is defined as

$$aSNR = 10 \log_{10} \left(\frac{A\mu_x}{B} \right), \quad (2.9)$$

where A and B are the parameters that produce the best fit for the model $y(t) = Ax(t) + B$. This apparent SNR is clipped to values between ± 15 dB, and the STI is calculated via Eqs. 2.5 and 2.6.

4) Normalized Covariance Method

The normalized covariance method (Koch, 1992; Holube and Kollmeier, 1996) is based on the covariance between the clean and degraded envelope signals. For each frequency band, the apparent SNR is calculated as

$$aSNR = 10 \log_{10} \left(\frac{r^2}{1-r^2} \right) \quad (2.10)$$

where r is the normalized covariance between $x(t)$ and $y(t)$ given by

$$r^2 = \frac{\lambda_{xy}^2}{\lambda_x \lambda_y} \quad (2.11)$$

with

$$\lambda_{xy} = E\{(x(t) - \mu_x)(y(t) - \mu_y)\} \quad (2.12)$$

and

$$\lambda_x = E\{(x(t) - \mu_x)^2\}. \quad (2.13)$$

The apparent SNR of Eq. 2.10 is clipped to values between ± 15 dB and the STI is calculated via Eqs. 2.5 and 2.6.

Summary of Speech-Based Methods

The speech-based methods described above all compute the STI as a weighted sum of TI values determined from the envelopes of the clean and degraded signals in each frequency band. The key difference among the methods is how the TI values are calculated. Table 2.1 summarizes the intermediate modulation metrics used to calculate TI values for the different methods.

| Magnitude CPS | Real CPS | Envelope Regression | Normalized Covariance |
|---|---|---|--|
| $MTF(f) =$ | $MTF(f) =$ | $M =$ | $r^2 =$ |
| $\alpha \left \frac{S_{xy}(f)}{S_{xx}(f)} \right $ | $\alpha \operatorname{Re} \left(\frac{S_{xy}(f)}{S_{xx}(f)} \right)$ | $\alpha \frac{\lambda_{xy}}{\lambda_x}$ | $\frac{\lambda_{xy}^2}{\lambda_x \lambda_y}$ |

Table 2.1: Intermediate modulation metrics for speech-based STI methods proposed in the literature. These metrics use the normalization term $\alpha = \mu_x / \mu_y$. They are calculated for each frequency band and then combined to produce a single STI value as described in the text.

In the case of the envelope regression method, the modulation metric in Table 2.1 is an alternate form that is derived in Appendix B.1. For the two cross-power spectrum methods, the modulation metric is a function of modulation frequency; while for the other

two methods there is a single value for each frequency band. The implications of this fundamental difference are discussed in Section 5.2. In the following sections, these modulation metrics will be used to yield insight into the behavior of the speech-based STI methods.

2.3 Noise Reduction Algorithms

Cochlear implant technology has reached the point where the best CI users can understand speech in quiet without any visual cues. However, performance deteriorates rapidly in adverse listening environments. This is not surprising, since CI users receive only a subset of the information available to normal-hearing listeners. The CI sound-processing strategy reduces the information available in a number of ways including reducing spectral resolution, removing temporal fine structure, limiting the dynamic range of stimulus intensity, and limiting the range of frequencies available to the implant user. In addition, since current CI systems only use one microphone signal as input to the CI sound-processing strategy, no binaural information is available. This thesis will consider noise reduction strategies applied to the signal prior to the CI sound-processing strategy. In particular, binaural information and increased spectral resolution will be used in an attempt to improve the speech signal before the CI sound-processing strategy is applied.

2.3.1 Spectral Subtraction

Many noise reduction algorithms operate in the frequency domain and are based on estimates of the noise signal. This includes Wiener filtering, spectral subtraction, and subspace filtering (Lim and Oppenheim, 1979; Boll, 1979; and Yariv and Van Trees, 1995). Spectral subtraction was chosen for investigation in this thesis for two reasons. First, it is practical since it can be implemented in real-time (unlike subspace filtering) so that if the thesis shows that spectral subtraction increases intelligibility for CI users, then CI manufacturers might be motivated to consider incorporating such a strategy into their

processors. Second, previous research has already shown some benefit of spectral subtraction for CI users (Weiss, 1993; Hochberg et al., 1992).

Spectral subtraction is a noise reduction technique for reducing the effects of stationary noise. Most of the research conducted on spectral subtraction has been on normal-hearing listeners or hearing-impaired listeners. This research has shown that spectral subtraction improves the subjective quality of processed speech, but does not improve the intelligibility of the processed signal. Studies of spectral subtraction indicate that spectral subtraction does increase intelligibility for CI users (Weiss, 1993; Hochberg et al., 1992). It is possible that spectral subtraction might improve intelligibility for CI users while not improving intelligibility for normal hearing and hearing-impaired listeners because the spectral subtraction algorithm uses information to enhance speech that is available to normal hearing listeners but is lost after CI sound-processing.

A generalized spectral subtraction method was described by Lim and Oppenheim (1979), and is illustrated in Figure 2.7. The Φ block represents a signal transformation

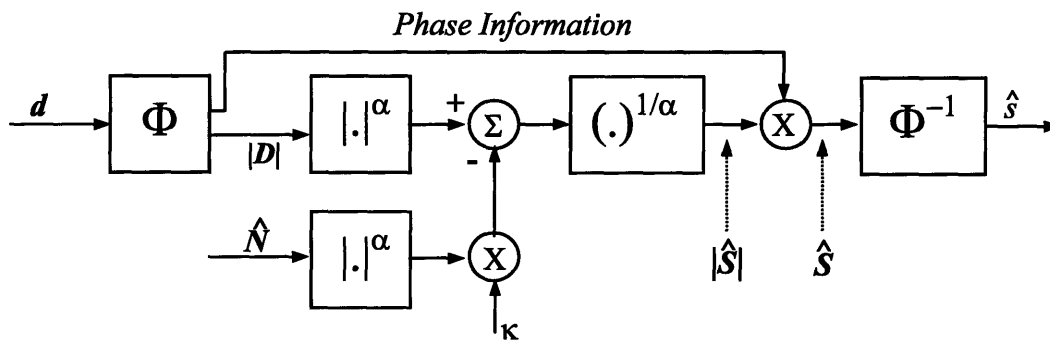


Figure 2.7: Generalized spectral subtraction.

such as the Fourier transform, and the Φ^{-1} block represents the inverse transformation. The transformation is typically conducted in a short-time manner on windowed sections of speech. The spectrum of the noise source, N , must either be known beforehand or estimated from an analysis of the microphone signal. The frequency domain estimate of the speech signal magnitude spectra (of a given short-time speech frame) for generalized spectral subtraction is given by:

$$|P(F, n)| = |D(F, n) - \kappa \hat{N}(F)|, \quad (2.14)$$

where $P(F, n)$ is the estimated speech spectrum of the n^{th} segment, $D(F, n)$ is the degraded speech spectrum, and $\hat{N}(F)$ is the estimated noise spectrum. The phase information is retained such that the phase of the output signal is the same as the input (degraded) speech signal. $|P(F)|$ is reconstructed using the phase of the original input signal and short-time reconstruction is performed to produce the time-domain output signal.

The control parameters α and κ can be used to vary the degree of noise reduction. When $\kappa = 1$ and $\alpha = 2$, the system corresponds to the power spectrum subtraction method previously studied with CI users (Weiss, 1993, Hochberg et al., 1992). In addition to these two control parameters, the window length in the short-time Fourier transform can be varied to adjust the spectral and temporal resolution.

This thesis will test a hypothesis regarding spectral subtraction, STI, and CI users. Previous work has shown that STI predicts an intelligibility improvement when speech degraded by additive noise is processed using spectral subtraction (Ludvigsen et al., 1990, 1993). They argue that this result is a shortcoming of STI prediction since neither normal-hearing listeners nor hearing-impaired listeners show improvements in intelligibility when listening to the processed speech. However, other research (Weiss, 1993, Hochberg et al., 1992) has shown intelligibility improvement for CI users after spectral subtraction as STI would predict. The hypothesis to be tested, then, is that STI may be a better indicator of performance for CI users than for normal-hearing listeners for speech processed using spectral subtraction.

2.3.2 Binaural Noise Reduction

The second noise reduction strategy uses binaural cues to enhance speech information to improve intelligibility. This approach naturally requires a second microphone to be worn over the opposite ear from the CI user's regular microphone. Of course, binaural

information cannot be given directly to the CI user without electrode arrays implanted in both cochleae, and even then, the delivery of binaural cues poses a great challenge. However, the binaural information can be used before CI sound-processing to enhance the speech signal.

The binaural noise reduction algorithm to be considered is motivated by algorithms previously considered by other researchers (Lockwood et al., 2004; Hamacher et al., 1997; Margo et al., 1997; Wittkopp et al., 1997; Schweitzer et al., 1996; Van Hoesel and Clark, 1995; Kollmeier et al., 1994, 1993) as well as preliminary studies performed as part of this thesis. This previous work has demonstrated that the binaural noise reduction approach can improve the intelligibility of speech in additive noise.

A generalized form of binaural noise reduction is shown in Figure 2.8. The vectors l and r represent windowed segments of the left and right microphone signals. Again, the Φ block represents a signal transformation, such as the FFT, and the Φ^{-1} block represents the inverse transformation. The Σ block represents the combination of

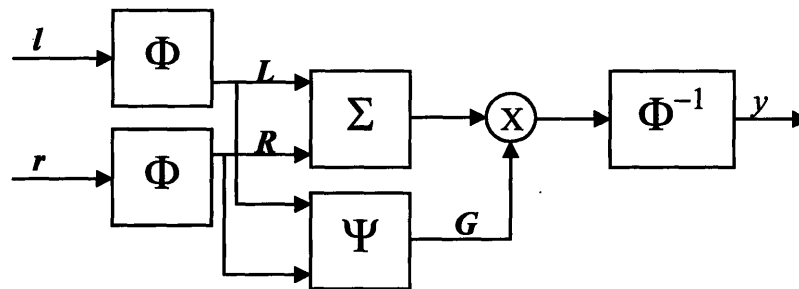


Figure 2.8: Generalized binaural noise reduction.

the two vectors to form a single output; a simple summation is generally used for this operation. Summing has the advantage of providing a fixed directional gain to a desired signal straight ahead of the listener. The Ψ block represents an adaptive determination of the frequency-dependent gain, G , based on a comparison of L and R . The gain is applied to the combined microphone signals in the frequency domain and then an inverse transform is applied.

One method of determining the applied gain is to compare the interaural phase information for low frequencies and interaural amplitude differences for high frequencies

from Fourier transform components. The inter-microphone phase difference (IPD) and the inter-microphone amplitude difference (IAD) can be used to calculate phase and amplitude related gain functions: $G_{phase}(F, IPD)$ and $G_{amplitude}(F, IAD)$. These gain functions can then be applied to the sum of the left and right spectral components. The resulting spectral representation is inversed transformed and combined to form the processed signal.

Chapter 3

STI Modifications

The goal of this thesis is to develop an intelligibility metric that is an accurate predictor of performance for CI users for a wide range of listening conditions. We consider speech-based STI methods as a starting point for developing such an intelligibility metric. However, as will be shown in Section 5.2, speech-based STI in its current form produces invalid predictions for nonlinear operations. Since we are primarily interested in nonlinear operations (e.g. noise reduction algorithms), we must modify STI for nonlinear operations. In this chapter, we first discuss methods for modifying STI to overcome the problems exhibited by existing speech-based STI methods. Second, we discuss a number of issues specifically related to developing intelligibility metrics for CI-processed speech.

3.1 Modifications of the STI for Nonlinear Operations³

In Section 5.2, we analyze the underlying calculation procedure of various speech-based STI methods to illustrate why they are poor predictors of intelligibility for certain nonlinear operations. In this section, we propose simple modifications to the STI calculation procedures to overcome problems with the existing methods. This results in five modified speech-based STI methods that are related to previously proposed methods. These modified STI methods are well correlated with the traditional STI for additive noise and reverberation and also exhibit qualitatively reasonable behavior for selected nonlinear operations. As a result, the modified STI methods are promising candidates to predict intelligibility of nonlinearly processed speech.

3.1.1 Normalization Based on Noise Envelope

Both CPS methods (Eqs. 2.7 and 2.8) include the term α , which normalizes the envelopes to account for the power of the clean and degraded signals. The alternate form of the envelope regression method derived in Appendix B.1 also depends on α ; for this method the apparent SNR in Eq. 2.9 can be expressed as

$$aSNR = 10 \log_{10} \left(\frac{M}{1-M} \right), \quad (3.1)$$

where M is a modulation metric defined as

$$M = \alpha \frac{\lambda_{xy}}{\lambda_x}. \quad (3.2)$$

Thus, the envelope regression method, as well as the two CPS methods, include the normalization term α . This term successfully normalizes the envelopes for the cases of additive noise and reverberation; however, for a large class of operations this

³ Most of Section 3.1 (through the end of Section 3.1.2) is reproduced from Goldsworthy and Greenberg, 2004: Section III, "Proposed Methods." Changes were made to section and equation numbers to be internally consistent with this thesis. Section 3.1.3 is an addendum that does not appear in Goldsworthy and Greenberg, 2004.

normalization ratio is not appropriate. In particular, when the processing reduces the overall amplitude of the degraded envelope, $y(t)$, α may increase without bound. As shown in Section 5.2, this leads to invalid values of the intermediate modulation metrics listed in Table 2.1.

An alternative normalization term is proposed here. The noise envelope is defined as

$$z(t) = |y(t) - x(t)|, \quad (3.3)$$

and a new normalization term is defined as

$$\beta = \frac{\mu_x}{\mu_x + \mu_z}. \quad (3.4)$$

For cases when $y(t) > x(t)$ for all t (as is typically the case for additive noise and reverberation) then $\mu_z = \mu_y - \mu_x$ and, consequently, $\beta = \alpha$. Thus, for certain operations, the proposed normalization term equals the original.

When the processing reduces the degraded envelope so that $y(t) < x(t)$ for some values of t , then μ_y decreases, causing α to increase. In some cases, high values of α may result in erroneously high values of apparent SNR for that frequency band. Since $\mu_z + \mu_x$ is always greater than μ_x , β will avoid characterizing reduced degraded envelopes as improved SNR.

3.1.2 Normalized Correlation STI

We hypothesize that the normalized covariance STI method (Sec. 2.2.2) is well suited to nonlinear operations. The normalized covariance defined in Eq. 2.11 is a metric that necessarily falls between zero and one, with a value of unity achieved only when the envelopes are identical. These constraints insure that the method always produces valid values of the intermediate metric. For the other speech-based methods, the intermediate metrics in Table 2.1 are not restricted to values between zero and one, and operations that

increase the modulation depth may cause the intermediate metrics to take on invalid values greater than one, as demonstrated in Section 5.2.

As a variation on the normalized covariance method, we consider the normalized correlation⁴, ρ , where

$$\rho^2 = \frac{\phi_{xy}^2}{\phi_x \phi_y} \quad (3.5)$$

with $\phi_{xy} = E\{x(t)y(t)\}$, $\phi_x = E\{x^2(t)\}$, and $\phi_y = E\{y^2(t)\}$. The STI is subsequently calculated by substituting ρ for r in Eq. 2.10, clipping to values between ± 15 dB, and applying Eqs. 2.5 and 2.6. The normalized correlation STI method differs from the normalized covariance STI method only in that the envelope means are included in the correlation terms.

| Magnitude CPS | Real CPS | Envelope Regression | Normalized Correlation |
|--|--|--|-------------------------------------|
| $MTF(f) =$ | $MTF(f) =$ | $M =$ | $\rho^2 =$ |
| $\beta \left \frac{S_{xy}(f)}{S_{xx}(f)} \right $ | $\beta \operatorname{Re} \left(\frac{S_{xy}(f)}{S_{xx}(f)} \right)$ | $\beta \frac{\lambda_{xy}}{\lambda_x}$ | $\frac{\phi_{xy}^2}{\phi_x \phi_y}$ |

Table 3.1: Intermediate modulation metrics for speech-based STI methods proposed in this work. These metrics are calculated for each frequency band and then combined to produce a single STI value as described in the text.

Table 3.1 summarizes the intermediate modulation metrics for the proposed speech-based methods. Comparing Table 3.1 to Table 2.1 reveals the key differences between the methods proposed in this work and those proposed previously.

⁴ Motivation for considering the normalized correlation comes in part from studies of binaural detection (Bernstein and Trahiotis, 1996), which have shown that the normalized correlation, ρ , is a better indicator of performance than the normalized covariance, r . By including the envelope means, the metric accounts for the average envelope power as well as the envelope fluctuations. While binaural detection is clearly different than speech intelligibility, it is possible that in both cases the auditory system utilizes the additional information about average envelope power provided by the normalized correlation.

3.1.3 Normalized Correlation Metric (NCM)

The normalized correlation STI method introduced above represents a strong departure from STI theory. The transmission index calculated for each band is based on the normalized correlation between clean and degraded envelope signals rather than a direct analysis of the corresponding modulation transfer function. We show in Appendix B.3 that the envelope regression method can be reformulated to express an underlying relationship to the modulation transfer function. A similar mathematical analysis is completed for the normalized correlation STI method (presented in Appendix B.2); however, the resulting dependency on the modulation transfer function is less transparent.

We will see in Section 5.2 that there is a one-to-one mapping between normalized correlation STI method and traditional STI. However, the mapping function is nonlinear, indicating that the normalized correlation STI method is not equivalent to the traditional STI method. This indicates that the normalized correlation STI method is a considerable departure from traditional STI and the underlying theory. The normalized correlation STI is related to traditional STI insofar as the metric is calculated from the envelope signals in a number of frequency bands. However, the calculation procedure for determining the transmission index based on these envelope signals is fundamentally different.

For the sake of continuity in this thesis, we will continue to refer to the normalized correlation STI method as such despite the dubious connection to traditional STI. However, we introduce another candidate metric closely related to normalized correlation STI, called the normalized correlation metric (NCM). The NCM is a further departure from STI theory in that it removes several intermediate steps in the calculation procedure.

In traditional STI procedures, an intermediate modulation metric is calculated and then transformed into an apparent SNR. This transformation to SNR (Eq. 2.2) is a logical and practical step to take since it represents the expected SNR for additive stationary noise (as shown in Appendix B.5). However, this property of the transformation—that the calculated apparent SNR corresponds to the expected SNR for the case of stationary

noise—does not hold for the normalized correlation STI. Thus, we suggest that the transformation from the normalized correlation to an apparent SNR be eliminated. The NCM is then defined simply as the psychoacoustically weighted average (Eq. 2.6) of the normalized correlation squared (Eq. 3.5) in each frequency band. The effect of bypassing the apparent SNR transformation on the calculated TI values is discussed in more detail in Appendix B.6.

3.2 CI-Specific Intelligibility Metrics

3.2.1 Tailoring the STI to CI sound-processing Strategies

The application of a physical performance metric to evaluating a processing system can be depicted as in Figure 3.1.

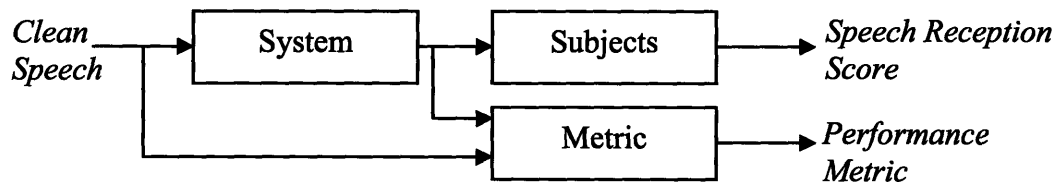


Figure 3.1: Block diagram of general problem.

This diagram encapsulates the three conceptual blocks relevant to using a performance metric to characterize subject performance. The system block can represent any processing of a speech signal including acoustic degradation, speech enhancement, noise reduction, compression, etc. The subject block represents a particular subject group such as normal-hearing listeners, CI users, or normal-hearing subjects listening to a CI simulation. The metric block represents the calculation of the performance metric based on the signal before and after processing by the system. The concept of tailoring the metric block to the subject group is important to consider in some detail.

The STI performance metric described in Section 2.2 originated as a metric for normal-hearing listeners and was eventually modified for hearing-impaired subjects. The performance metric can be viewed as a model for the subject group under consideration. Viewed as a model, the STI implies that the envelope signals in a number of frequency

bands contain the relevant information of speech signal that translates to intelligibility. Of course, the human auditory system is more complicated than the STI model implies; nevertheless, the STI has been successful in predicting the intelligibility of signals degraded by additive noise and reverberation for normal-hearing listeners. To predict results in hearing-impaired subjects, the STI model had to be specifically tailored to account for the subjects' impairment; in other words, the STI had to be tailored to the particular subject group. In the following discussion we consider how to tailor the STI model to account for particular details of the CI sound-processing strategy.

An essential contention of this thesis is that the STI is well suited as a performance metric for CI users due to the similarities between CI sound-processing and the STI calculation. Both use envelope signals from a number of frequency bands. A first step towards tailoring the performance metric calculation to CI users is to match the filter bank and envelope extraction procedure to those used in the CI sound-processing strategy. Two stages are considered here for tailoring. First, the filter bank used in the performance metric calculation is specified to be the same as that used in a particular CI processor. Second, the procedure used for extracting the envelope signal in the particular band is specified to be the same as that used in a particular CI processor.

Specifying that the envelope extraction procedure used be the same requires using the same rectification procedure (e.g. square or magnitude-law rectification) and lowpass-filter cut-off. Rectification used in STI procedures generally uses squaring, and this has important theoretical consequences for the STI; consequently, specifying the STI to be based on magnitude (full-wave rectification) envelopes rather than intensity (square-law rectification) envelopes would fundamentally change the results for the STI predictions. Consequently, we use intensity envelopes for the performance metrics that are closely related to traditional STI. This decision sacrifices one aspect of the tailoring process in order to remain consistent with STI theory. However, more freedom to tailor the performance metrics exists for those presenting a more substantial deviation from traditional STI. For those metrics, envelope extraction is based on magnitude envelopes.

The tailoring of the filter bank, and hence the number and allocation of the frequency bands used in STI analysis, requires specifying the corresponding psychoacoustic weights gauging the intelligibility contribution. Since we desire to define the frequency bands to be exactly the same in the performance metric calculations as in the CI sound-processing, then estimates of the weights must be determined either by psychoacoustic testing or by approximating new weights based on those used for conventional STI frequency bands. Further, since CI users will not, in general, have processors with filter bank specifications that match the conventional STI frequency bands, it is desired to have a warping function that allows estimation of the new weights.

Towards this end, consider the critical band weights for “average speech” (Pavlovic, 1987) given in the upper plot of Figure 3.2.

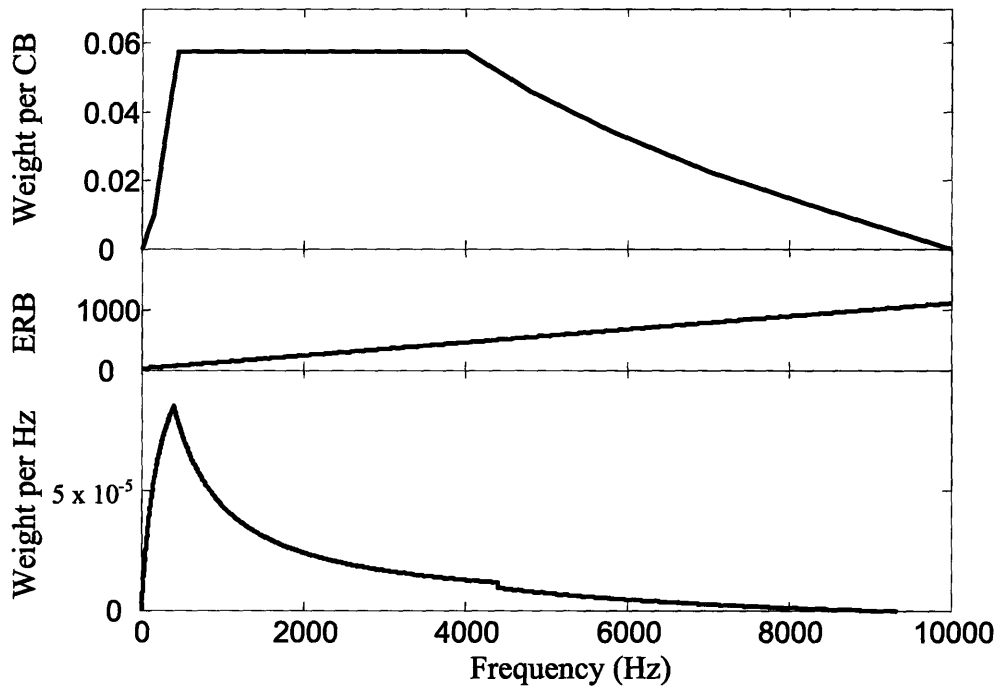


Figure 3.2: Specifying weights for arbitrary frequency bands.

The top plot in this figure gives the suggested weights as a function of center frequency for critical bands. To generalize this weighting function for arbitrary frequency bands,

the weighting function is divided by the following equivalent rectangular bandwidth function (Glasberg and Moore, 1990):

$$ERB = 24.7(4.37F_C + 1). \quad (3.6)$$

Where F_C is the center frequency of the critical band in kHz and ERB is the equivalent rectangular bandwidth. The lower plot is the result which is a per Hz weighting function. This per Hz weighting function can be summed over any arbitrary band to yield that band's weight.

The performance metric can also be modified to include the effects of N-of-M processing. To include the effects of N-of-M processing, the calculation of the performance metric contains an N-of-M processor for the calculation of the degraded envelopes. In other words, the N-of-M processing is considered as a degradation of the signal. In this way, Figure 2.6 can be redrawn as Figure 3.3 to include the effects of N-of-M processing.

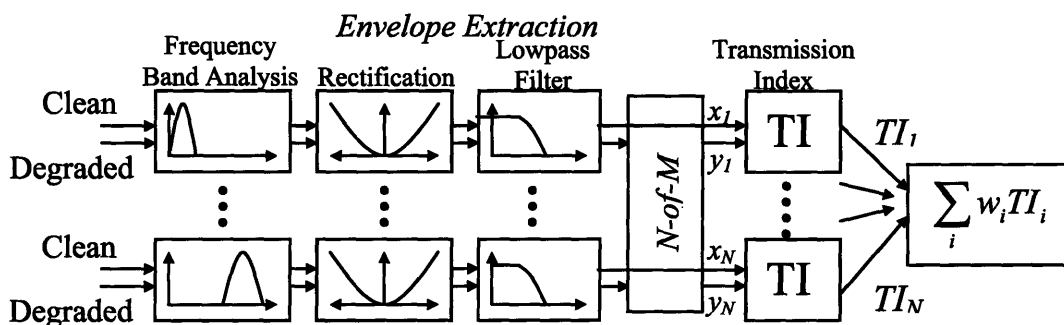


Figure 3.3: STI tailored to N-of-M processing.

Note that the line for the clean signal is drawn through the N-of-M block indicating that the reference signal envelopes are not transformed using the N-of-M processing. In other words, the clean reference envelopes, x_i , are based on the acoustically clean signal and do not suffer any distortions from the N-of-M algorithm.

3.2.2 Interpretation of CI-Specific STI Calculations

STI as a performance metric for CI users and for NH-CI_{Sim} has a subtle difference in meaning than for normal-hearing listeners. For normal-hearing listeners, an STI value of 1 corresponds to 100% intelligibility, while for CI users and for NH-CI_{Sim} an STI value of 1 may not translate to perfect intelligibility. However, in the CI-specific cases an STI value of 1 indicates that the subject should perform as well on the degraded signal as for clean speech. In other words, the CI sound-processing may limit the information available such that the subject does not have 100% intelligibility performance even when $STI = 1$.

STI is specific to the CI sound-processing strategy. Further, the same STI value for different CI processing strategies (e.g. 8 channels versus 20 channels) does not imply the same intelligibility results. Figure 3.4 illustrates hypothetical curves mapping STI to intelligibility performance. Each curve in Figure 3.4 represents a particular subject group. The curve hypothesized for 20 channel CI processing suggests perfect intelligibility for STI equal to one; however, larger STI values are required to obtain equivalent performance to normal-hearing subjects at other STI values. In other words, the CI users are hypothesized to have lower performance in the presence of the degradation. The curve hypothesized for the 8 channel CI processor reaches a maximal value of 80% intelligibility at STI equal to one. In other words, this subject group would not be expected to have 100% intelligibility when $STI = 1$.

In conclusion, it is suggested that the STI calculation is tailored to CI processing by specifying the calculation parameters related to bandpass filtering and envelope extraction to be identical to those used in a particular CI processing strategy. For each tailored STI calculation, experimental data can be used to determine a curve mapping STI to intelligibility. This curve is specific to a particular CI processing strategy. Such curves may prove useful for quantifying the effects of degradation and/or noise reduction for a particular speech processor.

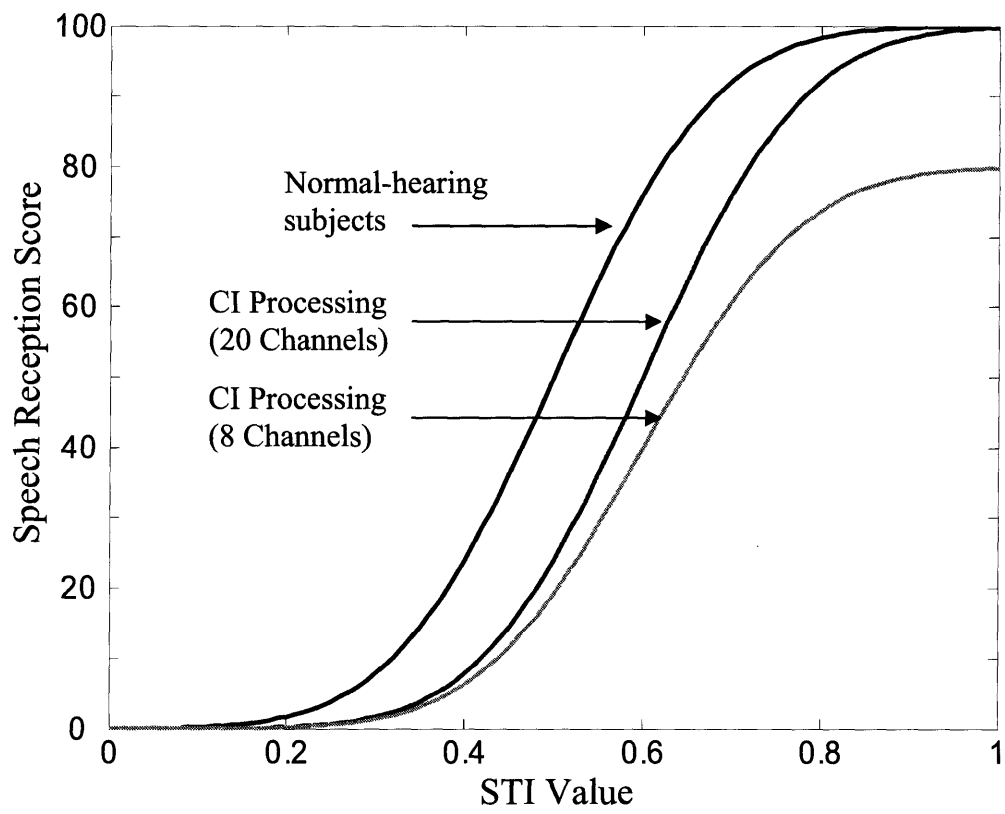


Figure 3.4: Hypothetical results, STI curves.

Chapter 4

Experimental Design

The experimental considerations that are common across experiments are described in this chapter. Amongst these are stimuli used, subjects, experimental conditions, experimental procedures, and performance metric analysis.

4.1 Stimuli

IEEE sentences were used for all preliminary experiments (IEEE, 1969). The sentences were spoken by one male talker and divided into 60 lists of 7 sentences each. All sentences are scored on 5 keywords, with a total of 35 keywords per list. The original sample frequency of the digitized IEEE sentences was 20,000 Hz. For the preliminary experiments of Section 5.1, the sentences were down-sampled to 16,000 Hz, and for the preliminary experiments of 5.2, the sentences were up-sampled to 22,050 Hz.

CUNY sentences were used for all main experiments (Boothroyd et al., 1985). The sentences were spoken by one female talker and divided into 60 lists of 12 sentences each. Sentence lengths range from three to fourteen words, with a total of 102 words per list. The sample frequency of the digitized CUNY sentences is 22,050 Hz.

Three noise types are used in the experiments: speech-shaped noise, multi-talker babble, and time-reversed speech. All noise types are designed to have the same long-term spectrum as the desired speech. The speech-shaped noise stimulus was generated by convolving white noise with an impulse response generated from the long-term spectrum of a concatenation of 2 lists selected from the corresponding (CUNY or IEEE) database. The multi-talker babble noise was generated by reshaping a 12-talker SPIN babble (Kalikow et al., 1977). This reshaping was accomplished by first whitening the babble and then convolving with an impulse response generated from the long-term spectrum of a concatenation of 2 lists selected from the CUNY database as above. The time-reversed speech was generated by randomly selecting a segment of speech from a concatenation of 2 lists selected from the CUNY database. Hence, the time-reversed speech interference is the same talker as the desired sentence but time-reversed. The two lists used for shaping the spectrum and for generating the time-reversed stimuli were not used as test sentences.

4.2 Subjects

Both normal-hearing listeners and CI users served as subjects for both the preliminary and main experiments.

4.2.1 Preliminary Experiments

Normal-Hearing Subjects

The normal-hearing subjects had audiometric thresholds less than 20 dB HL at octave frequencies between 125 and 8000 Hz. Their ages ranged from 18 to 25 and all were native speakers of American English. These subjects listened to degraded and processed speech that *did not incorporate* a noise-vocoder simulation of CI sound-processing.

Cochlear-Implant Subjects

Subjects tested were users of Nucleus devices using SPEAK processing strategies. The CI subjects were recruited from the Massachusetts Eye and Ear Infirmary and from personal contacts the author maintains with the CI community.

4.2.2 Main Experiments

Normal-Hearing Subjects

The normal-hearing subjects had audiometric thresholds less than 20 dB HL at octave frequencies between 125 and 8000 Hz. Their ages ranged from 18 to 29 and all were native speakers of American English. These subjects listened to degraded and processed speech that *did* incorporate a noise-vocoder simulation of CI sound-processing.

Cochlear-Implant Subjects

Two groups of CI users participated in these experiments: 1) subjects with Clarion devices using CIS processing strategies and 2) subjects with Nucleus devices using SPEAK processing strategies. The CI subjects were recruited from the Massachusetts Eye and Ear Infirmary and from personal contacts the author maintains with the CI community. Relevant audiological details for the CI subjects are summarized in Table 4.1. The duration of profound deafness in Table 4.1 refers to the pre-implantation duration. Subjects with Clarion and Nucleus devices used CIS and SPEAK (respectively) as the primary sound-processing strategy. The Clarion and Nucleus systems are 8 and 22 electrode systems, respectively. The speech reception in quiet was tested using 2 complete sentence lists from the CUNY database. If the subject's speech reception in

quiet was less than 30%, then the subject was excused from the remainder of the study since the testing conditions would prove too difficult.

| Subject | Age (years) | Duration of Profound Deafness (years) | CI Experience (years) | Etiology | Processor Type | Score in Quiet | Δ (dB) |
|---------|-------------|---------------------------------------|-----------------------|--------------------------|----------------|----------------|---------------|
| CI-1 | 55 | 18 | 9 | Infection | Clarion | 96.6 | 6 |
| CI-2 | 50 | 7 | 9 | Congenital (Progressive) | Clarion | 76.8 | 12 |
| CI-3 | 56 | 1 | 4 | Ototoxicity | Clarion | 63.1 | 12 |
| CI-4 | 29 | 1 | 16 | Ototoxicity | Nucleus | 94.6 | 6 |
| CI-5 | 49 | 2 | 2 | Ototoxicity | Nucleus | 96.6 | 0 |
| CI-6 | 53 | 10 | 7 | Progressive | Clarion | 97.1 | 3 |
| CI-7 | 44 | 14 | 10 | Congenital (Progressive) | Clarion | 95.1 | 9 |
| CI-8 | 27 | 2 | 10 | Infection | Nucleus | 46.1 | 9 |
| CI-9 | 69 | 60 | 8 | Usher's Syndrome | Clarion | 86.3 | 6 |

Table 4.1: Summary of CI subject information. Subjects with Clarion and Nucleus devices have 8 and 22 electrodes implanted, respectively.

4.3 Experimental Conditions

All stimulus processing was performed in MATLAB (Mathworks, Natick, MA) on a PC with an Intel Pentium III processor.

4.3.1 Acoustic Degradation

Acoustic degradation of the speech signal occurred in all preliminary and main experiments. The acoustic degradations investigated were additive noise and reverberation. The SNR of the acoustic degradation was defined as the ratio of desired speech power (at the desired speech's source) compared to noise power (at the noise's source). For main experiments 1 through 3 the speech and noise originated at the same location. For these experiments, the speech and noise are combined and then convolved

with the source-to-microphone transfer function resulting in the degraded speech. For experiment 4, the speech and noise originated from different locations. Consequently, the signals were convolved with their respective source-to-microphone transfer functions and then combined.

The source-to-microphone transfer functions for the preliminary experiment on binaural noise reduction (Section 5.1.2) were measured in an anechoic room using a Knowles experimental mannequin for acoustic research (KEMAR). These head-related transfer functions (HRTFs) have impulse responses that are 3 ms in duration and were measured with a microphone inside each ear with the source 1m away from the mannequin at angles from 0 to 180 degrees in 5-degree increments.

The source-to-microphone transfer functions for the preliminary STI experiments (Section 5.3) and for all main experiments (Chapters 6 through 9) were two-second long room impulse responses generated using a room simulation based on the image method (Allen and Berkley, 1979). The simulated room had dimensions of 5.2 by 3.4 by 2.8 meters, the listener was modeled as a rigid sphere of 12 cm radius at (2.7, 1.4, 1.6). For all experiments, the speech originated 1 meter in front (2.7, 2.4, 1.6) of the listener. The noise originated from the same location, except for the binaural experiments when the noise was specified to be 60 degrees to the right, and 1m away, from the listener (3.57, 1.9, 1.6). The walls, floor, and ceiling all had the same absorption coefficient, which was varied to produce three levels of reverberation with the resulting impulse responses corresponding to T_{60} times of 0, 0.15, and 1.2 seconds (anechoic, mild, and high, respectively).

Figures 4.1 and 4.2 illustrate the effect of speech-shaped noise and time-reversed speech (respectively) on speech envelopes and on phase-locked MTFs (Eq. 2.8). Both noise types have similar effects on the phase-locked MTFs despite having different effects on the speech envelopes. Figure 4.3 illustrates the effect of reverberation on speech envelopes and on phase-locked MTFs. The general effect of reverberation on the temporal envelope is to retard the dissipation of energy. The corresponding MTF illustrates that the effect of reverberation is less pronounced for lower modulation

frequencies. However, note that the phase-locked MTF actually takes on negative values for higher modulation frequencies. Such values, if inserted in the traditional method for calculating STI (see Eq. 2.2) would produce complex (in the mathematical sense) values. Complex values of the STI do not currently have any interpretational value. Therefore, a procedure must be introduced to account for these results. A simple solution is to limit all MTF values to the range between zero and one. This procedure avoids the generation of complex STI values. Other procedures for constraining the MTFs between zero and one could be introduced, or a novel interpretation of complex STI values could be sought; however, that is beyond the scope of this thesis. This topic is mentioned in the next section and discussed in more detail in Section 5.3.

4.3.2 Envelope Thresholding⁵

In the preliminary experiment presented in Section 5.2, envelope thresholding is used to analyze the effect of nonlinear processing. Envelope thresholding is a nonlinear operation that consists of setting to zero any samples of the original envelope that are below a threshold, that is

$$y[n] = \begin{cases} x[n] & x[n] \geq \tau \max(|x[n]|) \\ 0 & x[n] < \tau \max(|x[n]|) \end{cases} \quad (4.1)$$

where $x[n]$ and $y[n]$ are the clean and degraded envelopes, respectively, and τ is a fractional threshold relative to the maximum value of the clean envelope. Figure 4.4 illustrates the effect of the envelope thresholding on a speech envelope and shows that increasing the value of the threshold results in greater levels of modulation and increasingly distorted envelopes. Figure 4.5 illustrates the effect of envelope thresholding on a speech envelope and the corresponding MTF. It should be noted that

⁵ Section 4.3.2 is reproduced from Goldsworthy and Greenberg, 2004: Section IV.D. Changes were made to section, equation, and figure numbers to be internally consistent with this thesis. The final 5 sentences of this section, and the corresponding Figure 4.5, do not appear in Goldsworthy and Greenberg, 2004.

all values of the MTF are greater than one. These values would produce mathematically complex STI values unless limited to the range between zero and one. Even if values of the MTF greater than one were clipped to one, the metric would still imply that the thresholded speech is equally intelligible as the clean speech. As such, this represents a fundamental failing of the metric for this condition and is a caution towards blindly inserting the MTF into Eq. 2.2 even when the MTF is limited to the range between zero and one. This topic is addressed in detail in Section 5.3.

4.3.3 N-of-M Processing

In the main experiment presented in Chapter 7, the N-of-M algorithm operates on the envelope signals to select a subset of channels per analysis frame. The M envelopes are first down-sampled to 250 Hz. At 250 Hz the sample period is 4 ms which corresponds to the common analysis frame length (Loizou, 1998) used in the SPEAK processing strategy. For each frame, these M envelopes are then analyzed across channels to determine the N channels with the highest magnitude. The remaining $M-N$ channels are set to zero for that frame. The process is carried out for all time frames. The resulting envelopes are then up-sampled to the original envelope sample rate (22,050 Hz). The effect of N-of-M processing on the speech envelopes is comparable to envelope thresholding in that particular regions of the envelope signal are set to zero.

4.3.4 Spectral Subtraction⁶

Spectral subtraction attempts to reduce background noise by subtracting a noise spectral estimate from short-time magnitude spectra of the noisy signal. We investigate the general form given in Eq. 2.14 but set $\alpha = 1$ and focus on the effects of the control parameter κ . Thus the general frequency domain equation, for a given short-time segment, that we are interested in is given by

⁶ Section 4.3.4 is reproduced from Goldsworthy and Greenberg, 2004: Section IV.E. Changes were made to section, equation, and figure numbers to be internally consistent with this thesis. Minor textual changes were made and Eq. 4.2 was rewritten to emphasize that the spectral subtraction algorithm operates on magnitude spectra. The final paragraph of this section, and the corresponding Figure 4.7, do not appear in Goldsworthy and Greenberg, 2004.

$$|P(F, n)| = |D(F, n)| - \kappa |\hat{N}(F)|, \quad (4.2)$$

where $D(F, n)$ is the short-time magnitude spectrum of the input signal for the n^{th} segment, $\hat{N}(F)$ is the spectral estimate of the noise, $P(F, n)$ is the processed magnitude spectrum, and κ is a parameter that scales the noise estimate. $|P(F, n)|$ is reconstructed using the phase of the original input signal and short-time reconstruction is performed to produce the time-domain output signal.

The speech signal was degraded by noise with the same long-term spectrum as the clean speech (0 dB SNR) and then processed by the spectral subtraction algorithm using the overlap-add method with 25-ms Hamming windows. The control parameter, κ , was varied for investigation. A value of $\kappa = 0$ corresponds to no spectral subtraction processing and a value of $\kappa = 1$ corresponds to standard spectral subtraction. A value of $\kappa = 8$ corresponds to an extreme version where the spectral subtraction processing eliminates all but the highest spectral peaks. Figure 4.6 illustrates the effect of spectral subtraction for $\kappa = 0, 1, \text{ and } 8$ on a speech envelope in the time domain. The $\kappa = 1$ condition is potentially an improvement in that the noise in the speech envelope is suppressed. However, the $\kappa = 8$ condition clearly produces a distorted envelope.

The corresponding MTFs are illustrated in Figure 4.7 with the additive noise (no spectral subtraction processing) as a reference. The $\kappa = 1$ condition produces a higher MTF (relative to $\kappa = 0$) with many of the values close to, but not exceeding, one. However, the $\kappa = 8$ condition produces invalid results for the MTF with many values greater than one. This issue is addressed in more detail in Section 5.3.

4.3.5 Binaural Noise Reduction

A general description of the binaural noise reduction algorithm is given in Section 2.3.2, while the details of our particular implementation are specified in this section. The left and right microphone signals are transformed using a short-time Fourier transform. The Fourier analysis is completed using 31 ms long frames for the preliminary experiment

and 46 ms long frames for the main experiment. Both the preliminary and main experiments use overlapping windows with a half-window overlap. Values of the Fourier transform of the data in corresponding frames are compared in terms of their inter-microphone phase difference, IPD (radians), and their inter-microphone amplitude difference, IAD (dB). The microphone signals are summed to produce a single-channel signal that is then modified by the subsequent gain control, which is calculated from the IPD and IAD (and knowledge of the analysis frequency). The resulting signal is transformed by the inverse Fourier transform and the overlapping frames are recombined to produce the time domain output signal.

The general form of the dependence of attenuation on IPD and IAD is dictated by the assumption that the desired speech signal is straight ahead of the listener. An observation of IPD and IAD both near zero would indicate that the desired signal is much stronger than any off-axis source, leading to no attenuation. IPD and/or IAD very different from zero would indicate that off-axis non-desired signals are strong, leading to strong attenuation. In particular, the IPD is first converted to a predicted angle of arrival based on acoustic theory and then transformed into a phase-related gain function, $G_{phase}(F, IPD)$. The IAD is transformed into an amplitude-related gain function, $G_{amplitude}(F, IAD)$. The two parameters are combined as a weighted product,

$$G(F) = G_{phase}^{\alpha}(F, IPD) \cdot G_{amplitude}^{\beta}(F, IAD), \quad (4.3)$$

to form the final gain, $G(F)$, that is applied to the corresponding frequency component of the sum signal. In the calculation of the gain function, there are no dependencies across time or frequency; i.e., the gain function at each frequency and for each time frame is calculated only from the IPD and IAD for that frame and frequency.

As mentioned above, the IPD is converted into a phase-related gain parameter by first converting to a predicted angle (PA) using acoustic theory. The relation transforming the IPD to PA is

$$PA = \arcsin\left(\frac{v}{d} \cdot \frac{IPD}{2\pi F}\right) \quad (4.4)$$

where v is the velocity of sound, d is the inter-microphone distance, and F is the corresponding frequency. The PA is calculated for each frequency component and transformed into $G_{phase}(F)$ using

$$G_{phase}(F) = \cos\left(\frac{PA}{2}\right). \quad (4.5)$$

G_A is simply the IAD transformed to a linear factor, that is

$$G_{amplitude}(F, IAD) = 10^{(-|IAD|/20)} \quad (4.6)$$

For the preliminary binaural experiment (Section 5.2), these gain factors were used in conjunction with Eq. 4.3 with $\alpha = 16$ and $\beta = 0$ for frequencies less than 800 Hz and with $\alpha = 8$ and $\beta = 4$ for values above 800 Hz. For the main binaural experiment (Chapter 10), these gain factors were used in conjunction with Eq. 4.3 with $\alpha = 8$ and $\beta = 0$ for frequencies less than 800 Hz and with $\alpha = 8$ and $\beta = 4$ for values above 800 Hz.

4.3.6 Noise Vocoder

All main experiments performed with normal-hearing subjects involved the inclusion of a noise vocoder to simulate the effects of CI sound-processing. Both 8-channel and 20-channel noise vocoders were used in the experiments. The signal was first pre-emphasized using a first-order Butterworth (6 dB/Octave) highpass filter with cutoff frequency of 1200 Hz. The signal was then bandpass filtered into either 8 or 20 frequency bands using 8th-order Butterworth filters (96 dB/octave). The corner frequencies (3 dB down) for the 8-channel vocoder were at 250, 494, 697, 983, 1387, 1958, 2762, 3898, and 6800 Hz. These values were taken from the Clarion platinum sound processor filter table (Advanced Bionics, 1996). The corner frequencies (3 dB down) for the 20-channel vocoder were at 150, 350, 550, 750, 950, 1150, 1350, 1550, 1768, 2031, 2323, 2680, 3079, 3571, 4184, 4903, 5744, 6730, 7885, 9238, 9800 Hz.

These values were taken from the Nucleus sound processor filter table (Cochlear Corporation, 1996). Magnitude envelopes were extracted using full-wave rectification followed by lowpass filtering at 200 Hz. The lowpass filter used was a 4th order Butterworth design. The envelope signals were used to modulate a white noise carrier and then filtered through the same bandpass filters used in envelope calculation. The output of each band was normalized so that the RMS value at the output of each band equaled the RMS value before envelope extraction. Finally, the bands were summed to produce the NH-CI_{Sim} signal.

4.4 Experimental Procedure for Main Experiments

The experimental procedures for the main experiments are described in this section. The experimental procedure for preliminary experiments is described in the corresponding summary given in Chapter 5.

4.4.1 Normal-Hearing Subjects

The processed signal was converted to the analog domain using a soundcard (LynxStudio, LynxOne) at a 24-bit resolution. The signals were then passed through a headphone buffer (TDT HB6) and presented diotically via Sennheiser HD580 headphones to the subject, who was seated at a computer in a double-walled soundproof room. The subject controlled a computer interface using keyboard and mouse. The sound level was calibrated such that in the anechoic, no noise case, the speech signal had an average power of 65 dB SPL at the subject's ear.

All experiments were divided into three trials that were tested on three separate days. Each trial consisted of a complete set of 16 conditions. Conditions were partially counterbalanced across trials and across subjects as explained for each experiment in chapters 6 through 9. Within a trial, each condition was tested with a single list of twelve sentences from the CUNY database. The subjects' responses were scored as a percentage of words correct. A word was scored correct if they had the precise phonetic

pronunciation as the test word. The percent score for a trial is given as the total words correct to total words tested.

For *training* purposes, each trial began with the subject listening to two lists of sentences (quiet, anechoic) processed by the noise vocoder simulation of CI sound-processing. The subject heard each sentence once and typed what he/she heard; then the correct text was shown and the subject repeated the sentence as many times as desired. In order to prepare the subject, each condition begins with a *priming* sequence. The priming sequence consisted of six sentences degraded to correspond to the current condition and processed by the noise vocoder simulation of CI sound-processing. The subject heard the sentence once and typed as much as he/she could understand; the correct text was shown and the subject repeated the sentence as many times as desired. During *testing*, each of the 12 sentences was presented one at a time and the subject was instructed to “Type as much of the sentence as possible, then press ‘Okay’” without feedback. Subject responses typed during the training and priming sequences were disregarded. Sentence lists were reused in training and priming, but sentence lists used during testing were only presented once to each subject.

4.4.2 Cochlear-Implant Subjects

Cochlear-implant subjects were tested on a similar set of conditions as the normal-hearing subjects; however, since speech reception performance varies amongst CI users, a protocol was developed to shift the SNR of the conditions specific to each subject. Our task was to determine an SNR shift (Δ in Table 4.1) that would allow the CI users to perform comparably to the NH-CI_{sim} results on the corresponding conditions. The protocol that we use was based on the CI users speech reception threshold (SRT), defined here as the SNR at which the subject scores 50% of their speech reception in quiet (SRQ).

The subject’s SRQ is determined using two sentence lists from the CUNY database. Once the subject’s SRQ is determined, the following protocol is used to determine the appropriate Δ . Four sentence lists were set aside for this task. The initial

SNR tested is 6 dB. The decision tree given in Figure 4.8 was used to determine the subsequent SNRs tested. If the subject scores more than 50% of their SRQ, then the SNR was selected by moving down a row and to the left in the decision tree; otherwise, the SNR was selected by moving down a row and to the right. In this

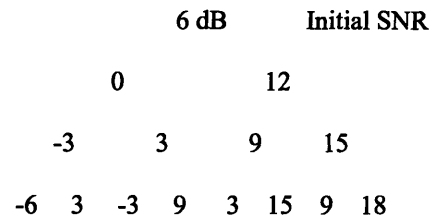


Figure 4.8: Decision tree for selecting SNR in CI protocol.

manner, four points were measured for the subject. The SNR with corresponding speech reception closest (linearly) to one-half of the SRQ was taken as the SRT. Thus if the subject scores 35% at 3 dB and 60% at 6 dB with a SRQ of 80%, then the SRT was taken to be 3 dB (since 40% is closer to 35% than to 60%). We decided to restrict the SRT to 3 dB increments for computational reasons.

Once the subject's SRT had been determined, it was used to set an appropriate Δ for the set of conditions tested. The value of Δ depends on whether the subject was tested for the acoustic degradation conditions or the noise reduction conditions. For the acoustic degradation conditions, the average speech reception of NH-CI₈ subjects for the speech-shaped noise condition at 0 dB SNR was 55.0 %. Thus the NH-CI₈ subjects scored near their SRT at 0 dB SNR. This suggests that we should adjust the SNR that the CI users are tested at for the anechoic SSN condition to achieve approximately one-half of the SRQ. This adjustment would require the shift to be defined as $\Delta = SRT$ dB. However, the SRT was determined in an anechoic environment, so we chose to define the shift conservatively as $\Delta = SRT + 3$ dB to avoid testing the CI users in too difficult of conditions for the reverberant case. Thus, if the CI user had an SRT of 6 dB, then the corresponding Δ would be 9 dB and the SNRs tested would be 6, 9, and 12 dB (shifted from the conditions of -3, 0, and 3 dB used for NH-CI₈). For the noise reduction conditions, the shift was referenced to the NH-CI₂₀ conditions. The average speech reception of NH-CI₂₀ subjects for the speech-shaped noise condition at -6 dB was just over 50% (it was 50.9 %). Thus, to match the CI users SRT to the -6 dB condition we defined the SNR shift as $\Delta = SRT + 6$ dB. Thus, if the CI user had an SRT of 6 dB, then

the corresponding Δ would be 12 dB and the SNR tested would be 9 dB (shifted from the conditions of -3 used for NH-CI₂₀) for the spectral subtraction conditions and 6 dB (shifted from the conditions of -6 used for NH-CI₂₀) for the binaural noise reduction conditions.

A couple other procedural differences exist for the CI subjects. First, since CI-processed speech is not a novelty for them, the training and priming sessions were omitted. The procedure for determining Δ described above allowed the subject to orient to the basic experimental task and interface. Of course the noise vocoder that simulates CI processing is also omitted. Second, the stimulus is delivered using a speaker within the soundproof room rather than headphones. The speaker is set 1 meter away—and on the same side as the implanted ear—from the subject. The sound level is calibrated such that quiet, anechoic speech produces a sound pressure level of 65 dB at the implanted ear.

4.5 Calculation of Intelligibility Metrics

4.5.1 Bandpass Filter and Envelope Extraction

The details for extracting the envelope signals for metric calculation are described below. This section is divided into details for the preliminary and main experiments since minor differences exist in the envelope extraction procedure used.

Preliminary Experiments

The bandpass filters were seven octave-band filters with center frequencies ranging from 125 Hz to 8 kHz. All filters were 8th-order Butterworth design. Intensity envelopes were calculated by squaring the bandpass-filtered signals and lowpass filtering. Magnitude envelopes were calculated by full-wave rectification of the bandpass-filtered signals followed by lowpass filtering. In both cases the lowpass filter was an 8th-order Butterworth with 50-Hz cutoff frequency. Envelopes were downsampled to 200 Hz before calculating the various metrics. The octave band weighting function used in Eq. 4.5 was taken from Houtgast and Steeneken (1985). The frequency band centered at 1 kHz is used for analysis of modulation metrics.

Main Experiments

The filter bank and envelope extraction were matched to the corresponding CI sound-processing as described in Section 3.2.1. Thus, the same envelope extraction procedure is used for the metric calculation as for the noise vocoder simulation given in Section 4.3.6. The exception is that intensity envelopes were used for the CPS and envelope regression methods. The decision to use intensity envelopes for the methods more closely tied to traditional STI was made based on the close association of intensity envelopes and STI theory as discussed in Section 3.2.1 and Appendix B.5. Consequently, only the normalized covariance, normalized correlation, and NCM methods capitalize on the additional tailoring of the metric regarding the envelope extraction procedure (in that they use *magnitude* envelopes).

The filter bank is matched to the CI sound-processing strategy using frequency bands given in Section 4.3.6. Since this results in frequency bands other than the standard octave or $1/3^{\text{rd}}$ octave bands, we determine appropriate weights to apply to the TI values in each band as described in Section 3.2.1.

4.5.2 Modulation Metric Calculation

The probe stimulus for the traditional method used in the preliminary experiment was a 60 second noise sequence with the same long-term spectrum as the speech. For the speech-based methods calculated in the preliminary experiments, the probe stimulus was a 120 second speech signal formed by concatenating 42 of the IEEE sentences described in Section 4.1.

For the speech-based method calculated for the main experiments, we computed STI values for the CUNY sentence materials. For each main experiment and each subject tested, we created a compact disc recording of 48 sentence lists, comprising the 16 partially counterbalanced conditions tested in each of the three trials. The metric values were calculated for each trial, resulting in three values per condition that were averaged to determine the overall metric value for each condition and disc. Chapters 6 through 9

report STI values derived from a single disc⁷. The following descriptions of the particular methods apply to both the preliminary and main experiments:

Traditional Method⁸

The traditional STI was calculated using fourteen modulation frequencies ranging from $f = 0.63$ Hz to 12.5 Hz in one-third octave increments. Because it requires the use of a probe noise sequence as the clean input, it was only practical to compute the traditional STI for the acoustic degradation conditions. For each modulation frequency, the noise sequence was amplitude modulated by $\sqrt{1 + \cos(2\pi(f/F_s)n)}$ to form the clean signal. The degraded response signal consisted of the clean signal combined with additive noise and/or reverberation. Both the clean and degraded signals were bandpass filtered into octave bands and intensity envelopes were computed by squaring followed by lowpass filtering. The modulation depth of each envelope was measured as the maximum value of the cross-covariance between the envelope and the function $\cos(2\pi(f/F_s)n)$ normalized by the envelope mean. The MTF value was determined from the ratio of the degraded envelope's modulation depth to the clean envelope's modulation depth.

Cross-Power Spectrum Methods

Both the magnitude and real CPS methods use intensity envelopes. Sample envelope means were calculated from the average of the envelope signals. The MTF for the two CPS methods requires estimating the auto- and cross-power spectra. This was accomplished using the periodogram method with 4096-point Hanning windows and 50% overlap. The resulting 0.05 Hz frequency bins were averaged into one-third octave intervals (Payton, 1999) centered from 0.63 to 12.7 Hz. This resulted in averaging of three bins for the lowest modulation frequency and 60 bins for the highest modulation

⁷ In order to justify using a single disc, rather than all discs, an analysis of the variance in STI values for the same conditions across discs was performed for the acoustic degradation conditions. The standard deviation of the means per conditions across discs was always less than 0.1% of the mean. Furthermore, the correlation coefficient between STI values compared across discs was always greater than 0.99. In other words, the mean STI values varied little across discs and therefore are always based on a single disc.

⁸ The remainder of Section 4.5.2 is reproduced from Goldsworthy and Greenberg, 2004: Section IV.B. Changes were made to section, equation, and figure numbers to be internally consistent with this thesis.

frequency. These quantities were used in the corresponding MTF (Eq. 2.7 or 2.8) for the original methods, and with β (Eq. 3.4) in place of α for the proposed methods. Then STI was calculated via Eqs. 2.2 through 2.6.

Envelope Regression Method

The envelope regression method was calculated from the intensity envelopes using the alternate form derived in the Appendix B.1. Sample envelope means were computed from the average of the envelope signals and the covariance was calculated as an unbiased estimate, that is,

$$\lambda_{xy} = E\{(x[n] - \mu_x)(y[n] - \mu_y)\} \approx \left(\frac{1}{N-1}\right) \sum_{i=1}^N (x[i] - \mu_x)(y[i] - \mu_y) \quad (4.7)$$

For each frequency band, the modulation metric, M , was calculated using Eq. 4.13 for the existing method and with β in place of α for the proposed method. The apparent SNR was then calculated from 2.2, clipped to values between ± 15 dB, and used in Eqs. 2.5 through 2.6.

Normalized Covariance and Normalized Correlation Methods

The normalized covariance and normalized correlation methods were calculated based on magnitude envelopes. For each frequency band, the normalized covariance, r , was calculated from Eq. 2.11, with estimates of the variance and covariance calculated as in Eq. 4.7. The normalized correlation, ρ^2 , was calculated according to Eq. 3.5 with the correlation estimated as

$$\phi_{xy} = E\{(x[n]y[n])\} \approx \left(\frac{1}{N-1}\right) \sum_{i=1}^N (x[i] \cdot y[i]). \quad (4.8)$$

The apparent SNRs were calculated from Eq. 2.10 (replacing r with ρ for the normalized correlation method), clipped to values between ± 15 dB, and used in Eqs. 2.5 through 2.6.

In addition, the NCM method was calculated using ρ^2 from Eq. 3.5 directly in place of the transmission index value in Eq. 2.6.⁹

4.6 Psychometric Model

A commonly used psychometric function for fitting an intelligibility metric, M , to observed speech reception scores, S , is the three-parameter integral of a Gaussian:

$$\hat{S} = \frac{S_{\max}}{\sqrt{2\pi}\sigma} \int_{-\infty}^M \exp\left(\frac{-(x - M_{50})^2}{2\sigma^2}\right) dx \quad (4.9)$$

where the three fitting parameters— S_{\max} , M_{50} , and σ —correspond to the maximum predicted speech reception score, the metric value at 50% of this maximum score, and a parameter controlling the slope of the function, respectively. A common procedure for selecting the fitting parameters is to choose parameters such that the mean-square-error (or other error criterion) between predicted and observed scores is minimized. A potential problem with this approach is that it does not account for the fact that the variance in observed scores is much smaller for scores below 15% and above 85%. It could be argued that when fitting the psychometric function, the subject scores below 15% and above 85% should receive more emphasis since they are known with more certainty.

One solution to this problem of emphasizing certain data points is to transform the observed scores to rationalized arcsin units (RAU) (Studebaker, 1985). This transformation has the desirable property that the scores expressed in RAU have approximately equal variance across the entire range thus avoiding the problem associated with unequal variance. Scores transformed to RAU have a range between -23 and 123 ; the psychometric function expressed in Eq. 4.9 can be specified to this range as

$$\hat{R} = R_{\min} + \frac{(R_{\max} - R_{\min})}{\sqrt{2\pi}\sigma} \int_{-\infty}^M \exp\left(\frac{-(x - M_{50})^2}{2\sigma^2}\right) dx, \quad (4.10)$$

⁹ This point concerning the calculation of the NCM does not appear in Goldsworthy and Greenberg, 2004.

where $R_{\max} = 123$ RAU and $R_{\min} = -23$ RAU.

Our procedure for fitting the various intelligibility metrics is as follows. The observed speech reception score averaged across subjects and trials is converted to RAU using

$$T = 2 \arcsin \left(\sqrt{\frac{S}{100}} \right), \quad (4.11)$$

and

$$R = 1.46(31.83T - 50) + 50. \quad (4.12)$$

For the NH-CI_{Sim} subjects, we assume that subjects score 100% in the quiet, anechoic condition corresponding to $R_{\max} = 123$ RAU that a minimum of 0% exists for some condition corresponding to $R_{\min} = -23$ RAU, thus the psychometric model only has two free parameters: σ and M_{50} . For the CI subjects, it is expected that maximum speech reception will vary, thus R_{\max} is also treated as a free parameter. The free parameters are selected to minimize the mean-square-error between predicted (\hat{R}) and observed (R) scores defined as

$$\varepsilon = \sqrt{\frac{1}{N-1} \sum_i (R_i - \hat{R}_i)^2}, \quad (4.13)$$

where the subscript i denotes condition number. This mean-square-error—which is in RAU—is used as an indication of the quality of fit between observed and predicted scores. ε serves only as a general indicator of the goodness of fit. It cannot be used to place confidence intervals on the predictions since the underlying probability density of the error is not known. If the probability density of the error were known, then corresponding confidence intervals could be determined. For example, if the error had a normal distribution, then (since ε is the standard deviation of the error), it follows that 70.7% of the time the model would be accurate to within $\pm\varepsilon$.

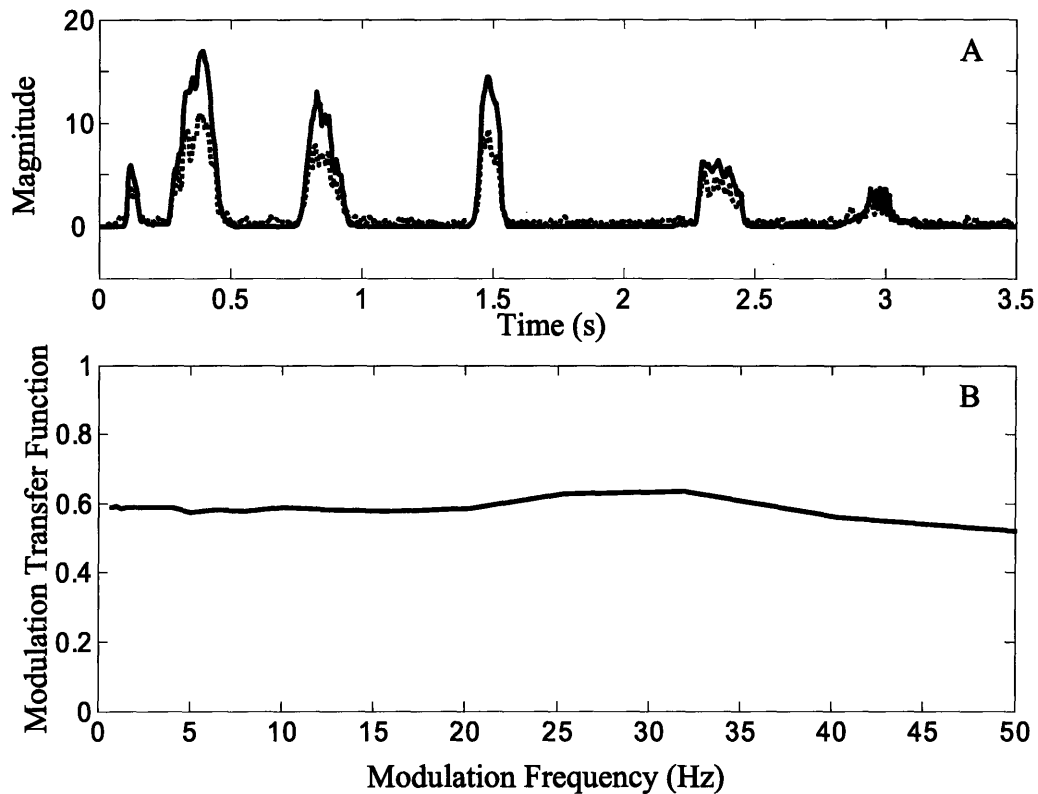


Figure 4.1: A) Effect of additive stationary noise on envelope signal (for an octave-band centered at 1 kHz) normalized by envelope mean. Solid line represents envelope of clean speech, and dotted line represents the same speech degraded by speech-shaped noise (0 dB SNR). B) Effect of additive stationary noise on the phase-locked MTF (Eq. 2.8) for envelopes shown in A.

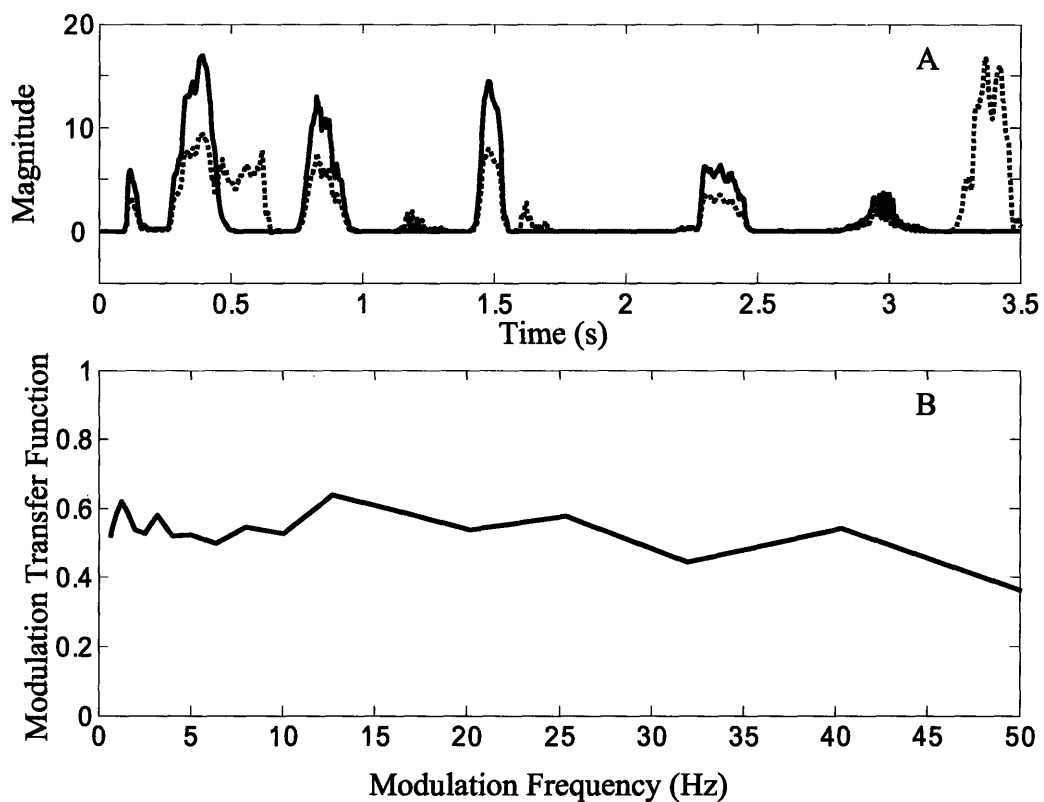


Figure 4.2: A) Effect of time-reversed speech on envelope signal (for an octave-band centered at 1 kHz) normalized by envelope mean. Solid line represents envelope of clean speech, and dotted line represents the same speech degraded by time-reversed speech (0 dB SNR). B) Effect of time-reversed speech on the phase-locked MTF (Eq. 2.8) for envelopes shown in A.

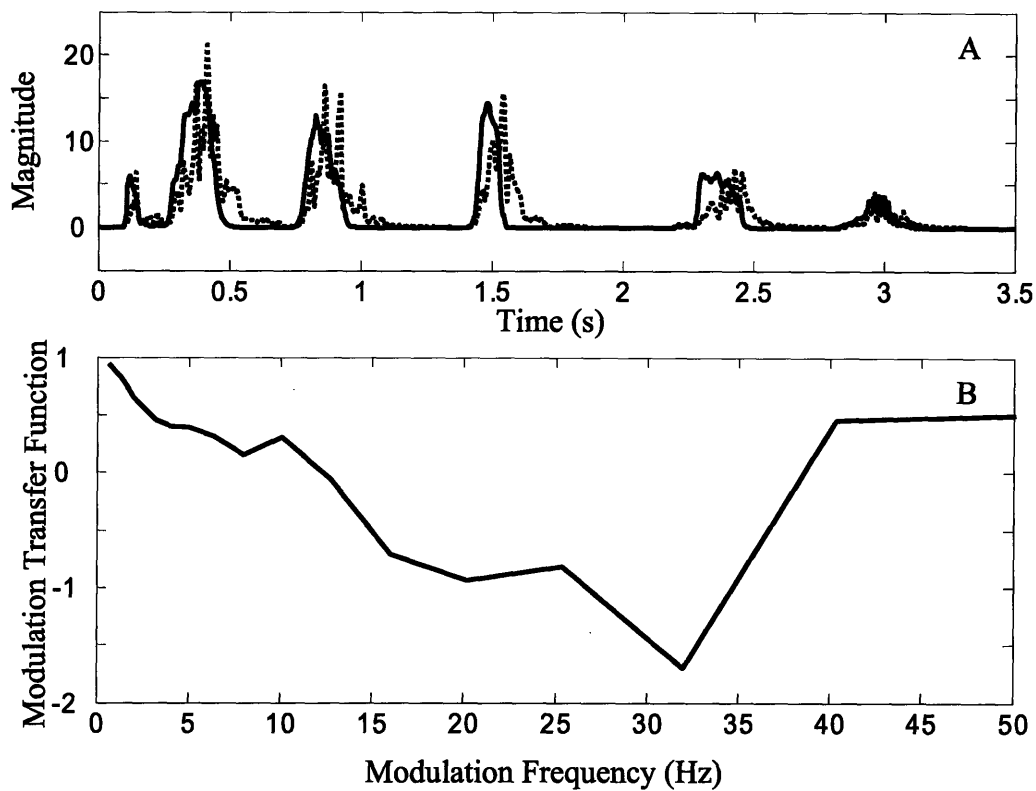


Figure 4.3: A) Effect of reverberation on envelope signal (for an octave-band centered at 1 kHz) normalized by envelope mean. Solid line represents envelope of clean speech, and dotted line represents the same speech in reverberation ($T_{60} = 1.2$ seconds). B) Effect of reverberation on the phase-locked MTF (Eq. 2.8) for envelopes shown in A.

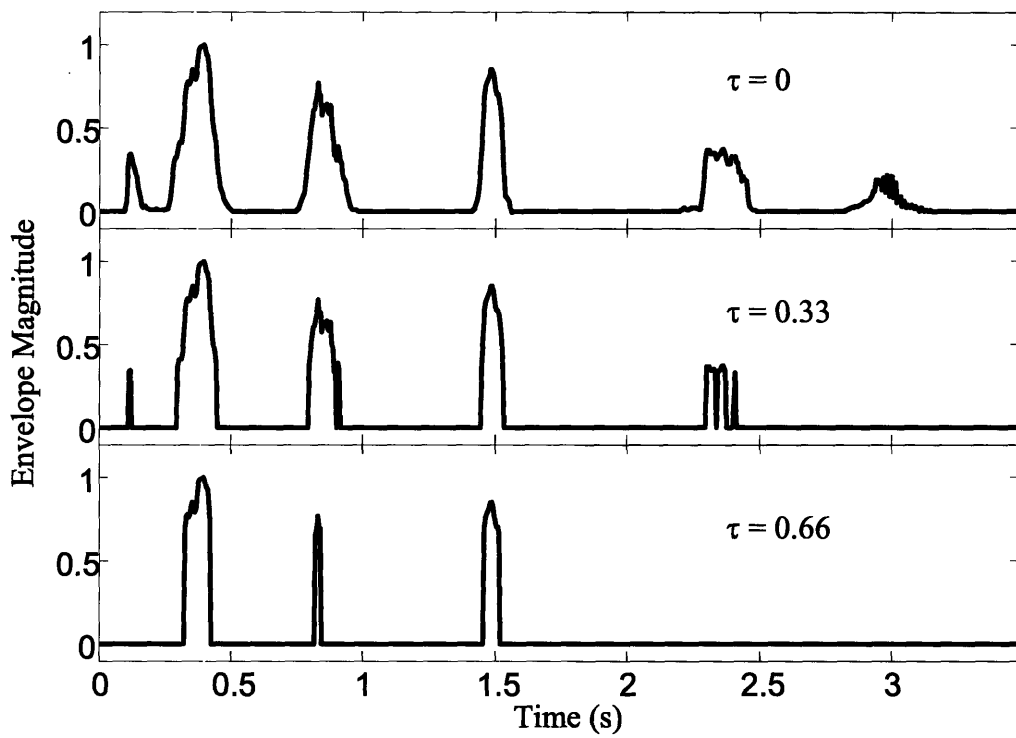


Figure 4.4: Effect of envelope thresholding on clean-speech envelope signal (for an octave-band centered at 1 kHz) for no processing ($\tau = 0$), $\tau = 0.33$, and 0.66.

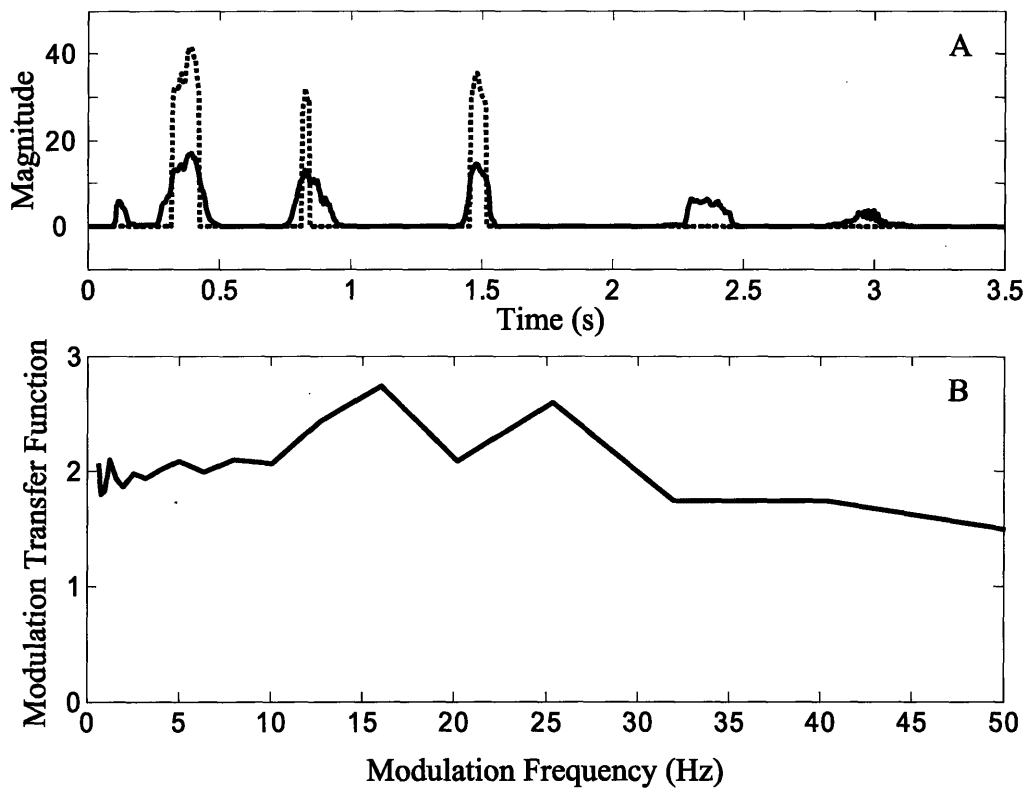


Figure 4.5: A) Effect of envelope thresholding on an envelope signal (for an octave-band centered at 1 kHz) normalized by envelope mean. Solid line represents envelope of speech in quiet, and dotted line represents the same envelope after applying thresholding of $\tau = 0.8$. B) Effect of envelope thresholding on the phase-locked MTF (Eq. 2.8) for envelopes shown in A.

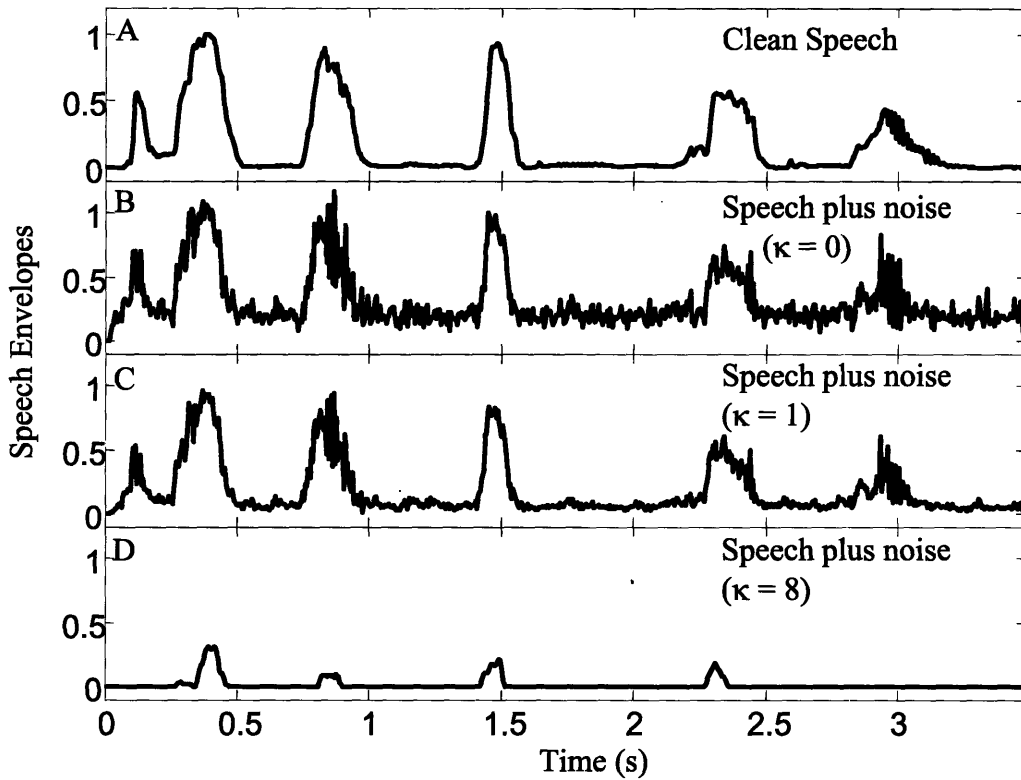


Figure 4.6: Effect of spectral subtraction on envelope signal (for an octave-band centered at 1 kHz). A) Envelope of clean speech. B) Envelope of speech plus noise (0 dB SNR). C) Envelope of speech plus noise after applying spectral subtraction with $\kappa = 1$. D) Envelope of speech plus noise after applying spectral subtraction with $\kappa = 8$.

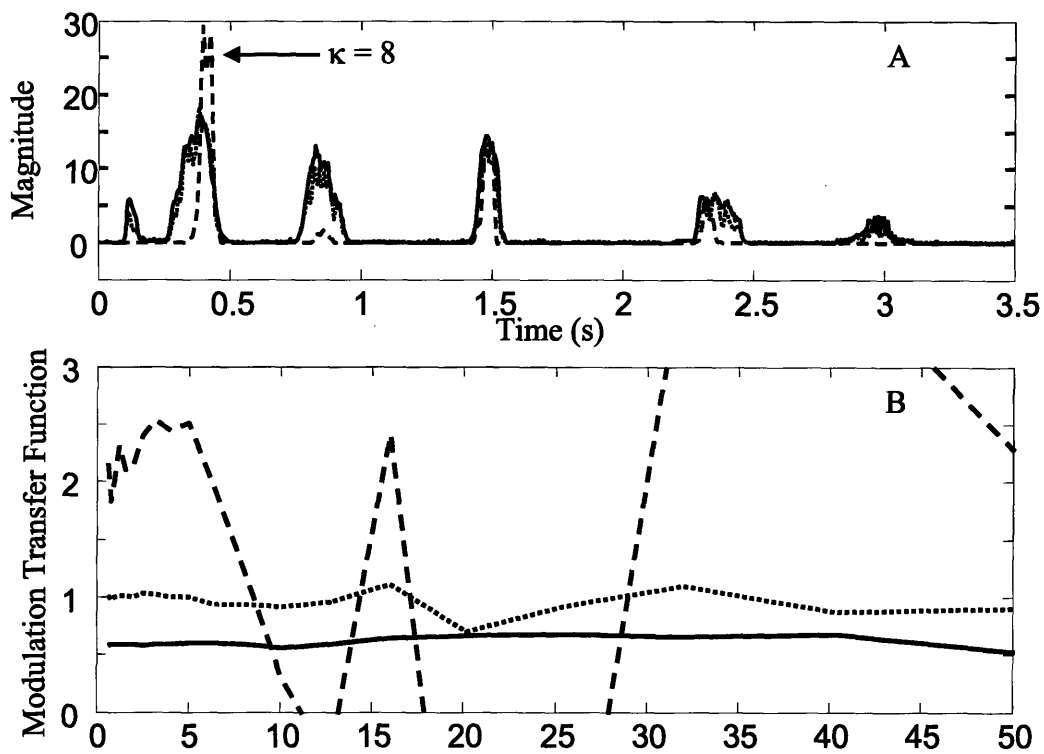


Figure 4.7: A) Effect of spectral subtraction on envelope signals (for an octave-band centered at 1 kHz as shown in 4.6 A, C, and D) normalized by envelope means. The solid line represents speech in quiet, the dotted line represents speech degraded by speech-shaped noise (0 dB SNR) and filtered using spectral subtraction ($\kappa = 1$), and the dashed line represents speech in noise (0 dB SNR) and filtered using spectral subtraction ($\kappa = 8$). B) Effect of spectral subtraction on the MTF. Dotted and dashed MTFs correspond to $\kappa = 1$ and $\kappa = 8$ respectively compared to clean reference. Solid line MTF corresponds to speech in noise without spectral subtraction (envelope shown in 4.6 B).

Chapter 5

Preliminary Experiments

Preliminary experiments for binaural noise reduction algorithms tested with CI users and normal-hearing subjects, as well as preliminary analytical studies of the STI are presented in this chapter. The binaural noise reduction evaluations include physical and subjective assessment of a commercial device—the Audallion BEAMformer—manufactured by Cochlear Corporation (Section 5.1.1), as well as evaluation of a novel binaural algorithm developed in this thesis (Section 5.1.2). The results identify weaknesses of the Audallion system and improvements in the novel algorithm. The results from the novel algorithm are promising and warrant further study. The analytical STI work considers four previously-proposed speech-based methods and four novel methods, studied under conditions of additive noise, reverberation, and two nonlinear operations (envelope thresholding and spectral subtraction). Analyzing intermediate metrics in the STI calculation reveals why some methods fail for nonlinear operations. Results (Section 5.2.1) indicate that none of the previously-proposed methods is adequate for all of the conditions considered, while the four novel methods produce qualitatively reasonable results and warrant further study. The discussion of 5.2.2 considers the relevance of this work to predicting the intelligibility of CI-processed speech. In Section 5.3 we justify the selection of three candidate metrics out of the pool of nine metrics developed in Chapters 2 and 3.

5.1 Preliminary Study of Binaural Noise Reduction Algorithms

5.1.1 Audallion BEAMformer Evaluation

We began research on intelligibility enhancing algorithms by evaluating the two-microphone Audallion BEAMformer developed by Cochlear Corporation (Goldsworthy and Greenberg, 1999, 2000). See Figure 2.8 for a general block diagram. The first stage in processing for the Audallion implementation is a short-time Fourier analysis. For each frequency component, the phase and magnitude differences between the two microphone signals are compared. Components are attenuated to a degree that depends on the inter-microphone phase and/or amplitude difference. The vector of attenuation values is applied to the sum of the two microphone signals. Frequency components that are dominated by a speaker in front of the listener should have small inter-microphone phase and amplitude differences and will not be attenuated, while frequency components dominated by noise from other directions will have larger inter-microphone differences and so will be attenuated more. Specifically, in the Audallion system, the phase difference between frequency components less than 1200 Hz is used to estimate the angle of arrival, and sounds estimated to arrive outside a specified beamwidth are attenuated. For frequencies above 1200 Hz, the intermicrophone amplitude difference is the determinant of degree of attenuation. [This description of the Audallion is based on Schweitzer et al., 1996].

The Audallion BEAMformer was evaluated using both physical measures and intelligibility tests. The physical measures were made while the Audallion BEAMformer was placed on a KEMAR manikin situated in the center of a soundproof (*not* anechoic) room with a single broadband noise source at 60 degrees. Figure 5.1 shows power spectra of the signals out of the Audallion preprocessor when it is operating in BEAMformer mode (Setting 4) and when it is simply summing the two microphone signals (an alternate mode of the device). Since the stimulus is presented alone at 60 degrees, the BEAMformer output should be attenuated at all frequencies. Yet, it is clear that the attenuation is weak for frequencies greater than approximately 1200 Hz, where the system attenuation is based on inter-microphone amplitude cues.

Figure 5.2 illustrates the directivity of the Audallion BEAMformer on Setting 4 for four particular frequencies. The stimulus for this particular measurement was a sinusoidal signal at the frequency of interest. The data for these plots were collected by measuring the relative power of the output of the Audallion BEAMformer as a function of angle. The relative power is normalized to the power of the signal generated from straight ahead of the listener. Note that the directivity is not as great for higher frequencies.

Three Nucleus 22 CI users with the Spectra 22 speech processor participated in intelligibility tests of the Audallion BEAMformer. Subjects were seated in a soundproof room while wearing the Audallion BEAMformer. Testing was done with the Audallion in BEAMformer mode and in sum mode. Phonetically-balanced sentences (IEEE, 1969) were played from a speaker directly in front of the listener while 8-talker babble was played from a speaker 60 degrees to the right of the subject; both loudspeakers were one meter away from center of the listener's head. SNR was varied by controlling the level of the noise. The results given in Figure 5.3 show the percent correct identification of key words as a function of the SNR. Each data point was determined using seven sentences (35 key words). For subject CI_p-1 (the subscript p distinguishes subjects taking part in this preliminary experiment from CI subjects taking part in the main experiment), multiple trials are plotted, with the mean value plotted as a line.

It is clear that the Audallion BEAMformer did not improve speech reception for the three subjects tested. In fact, for subject CI_p-1 and CI_p-3 , speech reception was consistently lower for the BEAMformer mode versus the sum mode. The physical measures suggest the limited BEAMformer attenuation at high frequencies may be partly responsible for this measured behavior and that the system could be improved by increasing the high-frequency attenuation of off-axis sources.

5.1.2 Further Algorithm Development and Evaluation

To explore this issue further, we developed software to implement binaural processing following the basic structure of the algorithm implemented in the Audallion

BEAMformer, and also described by Kollmeier et al. (1994) and Lindemann (1986), but with a modified attenuation control mechanism that was more strongly dependent on inter-microphone amplitude and phase differences than that of the original system. Our implementation operates using an overlap-add procedure with a 31 ms temporal window corresponding to a 32 Hz frequency resolution. No temporal frame or frequency bin averaging was employed.

Our implementation was run off-line using a simulation of an *anechoic environment* (Section 4.3.1). Physical performance was measured by first convolving a wide-band noise with KEMAR HRTFs at 60 degrees to simulate left and right microphone inputs for noise originating at 60 degrees. These simulated microphone signals were then used as inputs to the binaural algorithm as well as for a reference processing condition that was simply the sum of the two signals. The output spectra in response to a noise source at 60 degrees are given in Figure 5.4. The key result is that the 60-deg source is clearly attenuated due to the stronger weighting functions used in the binaural algorithm.

One CI user and three normal-hearing subjects were tested for speech reception using the binaural algorithm. Phonetically-balanced sentences (IEEE, 1969) were convolved with KEMAR HRTFs for a source at 0 degrees and 1m distant and 8-talker babble was added after convolving it with KEMAR HRTFs for a source at 60 degrees and 1m distant. Thus, the simulation was designed to model desired speech arriving from 0 degrees and a noise source at 60 degrees to the right of the listener. As in the evaluation of the Audallion BEAMformer, the comparison was made between the algorithm, our implementation in this case, and a simple summation of the left and right microphone signals. For the CI user, the signals were delivered directly to the speech processor. For the normal-hearing listeners, the processed or summed signals were presented diotically via headphones. An adaptive method was used to determine the speech reception threshold (SRT), the SNR at which the subject identified 50% of the keywords.

Figure 5.5 shows the SRTs measured with the binaural algorithm and for summation. The binaural algorithm improved the SRT of the CI user by 16 dB over that obtained with summation; the improvement for the normal-hearing listeners was 12-14 dB. The large improvement seen with our implementation of the binaural algorithm is in stark contrast to the poor results of the Audallion BEAMformer. We attribute the success of our implementation to the use of both high spectral resolution and strong attenuation functions (and also to the differences in acoustic environments). Both the resolution and attenuation function used in the Audallion system are proprietary information of Cochlear Corp. However, Figure 5.1 clearly indicates that the attenuation function is not very strong and in pre-commercial testing (Schweitzer et al., 1996), the researchers used frequency resolutions ranging from 56 to 76 Hz. In contrast, our implementation applies a stronger gain (Figure 5.4) and operates with a frequency resolution between 20 and 40 Hz.

5.2 Evaluation of Speech-Based STI for Nonlinear Operations

In this section, the existing speech-based STI methods are analyzed to determine why they fail to predict intelligibility for nonlinear operations. The modifications proposed in Section 3.1 are shown to overcome problems with the existing methods. These modified STI methods are well correlated with the traditional STI for additive noise and reverberation and also exhibit qualitatively reasonable behavior for selected nonlinear operations. As a result, the modified STI methods are promising candidates to predict intelligibility of nonlinearly processed speech.

5.2.1 Results¹⁰

Acoustic Degradation

The acoustic degradation was performed as described in Section 4.3.1. Speech-shaped noise was scaled to produce SNRs between -15 and 30 dB in 3-dB increments as well as

¹⁰ Beginning with the second paragraph, Section 5.2.1 is reproduced from Goldsworthy and Greenberg, 2004: Section V, "Results." Changes were made to section, equation, and figure numbers to be internally consistent with this thesis.

a no-noise condition. Reverberation times (T_{60}) ranged from 0 to 1.5 seconds in 0.3-second increments. The traditional and speech-based STIs were computed and compared for all combinations of SNR and reverberation time.

Since the traditional STI method is well established as an accurate predictor of speech reception for additive stationary noise and reverberation, any proposed speech-based method must produce similar values of STI under these conditions. Figure 5.6 compares the speech-based STI methods to the traditional STI for the acoustic degradation conditions of additive noise and reverberation. Figures 5.6A through D show the four previously proposed speech-based methods described in Section 2.2, while Figures 5.6E through H show the methods proposed in Section 3.1. Each curve represents STI values calculated over the 45-dB range of SNRs for one level of reverberation.

In Figure 5.6, complete agreement between the traditional STI method and a speech-based STI method would appear as a straight line from the bottom left to the top right of a particular plot. As seen in Figures 5.6A, B, and C, the original cross-power spectrum methods and the original envelope regression method all provide a reasonable match to the traditional method, although the real cross-power spectrum method is slightly less well-matched to the traditional than the other two.

Comparing Figures 5.6A, B, and C to Figures 5.6E, F, and G shows that for these acoustic degradation conditions, the modified methods using β as the normalization term are equivalent to the original methods using α . As described in Section 3.1.1, this equivalence is expected because the acoustic degradations increase the overall amplitude of the degraded envelopes relative to the clean envelopes.

The normalized covariance method (Figure 5.6D) and the proposed normalized correlation method (Figure 5.6H) are distinctly different from the other speech-based methods. The normalized covariance method does not exhibit a one-to-one relationship to the traditional method. The curves for different levels of reverberation are not superimposed, indicating that the normalized covariance method is not consistent with the traditional method in accounting for reverberation. Given the success of the

traditional STI, this implies that the normalized covariance method will not be a good predictor of intelligibility for additive noise and reverberation. The normalized correlation method comes closer to having a one-to-one relationship to the traditional method, with some divergence at high SNRs. This implies that the normalized correlation method may perform poorly when accounting for the effects of reverberation in quiet and low-noise environments.

While the relationship between the normalized correlation method and the traditional STI is approximately one-to-one, they are not equivalent metrics. In other words, some mapping is required to transform the values produced by the normalized correlation method to values corresponding to the traditional STI. To the extent that a unique mapping does exist for these conditions, the new metric will retain the predictive power of the traditional STI for additive noise and reverberation.

Envelope Thresholding

For the envelope thresholding and spectral subtraction conditions, the speech-based STI methods are characterized by intermediate modulation metrics for a single frequency band. The envelope thresholding is performed as described in Section 4.3.2. Clean speech is used as the input to the envelope thresholding algorithm. Intermediate modulation metrics were calculated for all speech-based STI methods for thresholds ranging from zero to the envelope maximum in 2% increments.

Figure 5.7 shows the effect of envelope thresholding on intermediate modulation metrics used to compute the various speech-based STI methods. Investigating these metrics, rather than the final STI values, is necessary to identify methods that produce invalid results. All of the intermediate modulation metrics have a valid range from zero to one, where zero indicates no preservation of the envelope modulations and one indicates perfect preservation. Values of the intermediate metric greater than one indicate a failure of the corresponding method.

Figures 5.7A, B, and C reveal that the original cross-power spectrum methods and the original envelope regression method fail for envelope thresholding. In all three plots, the modulation metrics increase above one as the threshold increases. These invalid

values of the intermediate metrics indicate that these methods are not applicable to the nonlinear operation of envelope thresholding. The remaining five plots reveal that all of the proposed methods (Figures 5.7E through H), as well as the normalized covariance method (Figure 5.7D), produce valid values of the intermediate metrics. As the threshold increases, all of the intermediate metrics monotonically decrease from an initial value of one.

The general effect of envelope thresholding is to emphasize peaks in the envelope by setting low-amplitude samples of the envelope to zero. As the threshold increases, more samples are set to zero. Because this increases the modulation depth of the envelope, most of the previously proposed speech-based STI methods erroneously interpret this operation as increasing intelligibility beyond the initial value of one for speech in quiet. These methods fail because envelope thresholding reduces the mean of the degraded envelope, μ_y . Since it is the denominator of the normalization term, α , small values of μ_y can lead to extremely large values of α . Although envelope thresholding also reduces the cross-spectrum, $S_{xy}(f)$, and cross-covariance, λ_{xy} , (which contribute to the numerator of the modulation metrics in Eqs. 2.7, 2.8, and 4.2), empirical observations indicate as the threshold increases, these terms decrease more gradually than μ_y , leading to invalid values of the modulation metrics.

The modified methods that use β as the normalization term do not fail in this way because, for envelope thresholding, μ_z varies from zero to μ_x as the threshold goes from 0 to 100%, corresponding to values of β ranging from 1 to 0.5 for the full range of envelope thresholding. This causes the intermediate metrics to decrease with increasing threshold.

The results for the three modified methods, as well as the normalized correlation and normalized covariance methods, are qualitatively consistent with the expected effect of envelope thresholding on the intelligibility of speech in quiet. The effect of increasing the threshold is to increase the distortion of the processed signal, thereby making it less intelligible. Increasing the threshold of a slightly different envelope manipulation has

been shown to decrease intelligibility (Drullman, 1995). Therefore, the methods that account for envelope thresholding by decreasing as the threshold increases are viable candidates for speech-based STI.

Spectral Subtraction

Spectral subtraction is performed as described in Section 4.3.4. The speech signal was degraded by speech-shaped noise (0 dB SNR) and then processed by the spectral subtraction algorithm. Intermediate modulation metrics were calculated for all speech-based STI methods for values of κ ranging from zero to eight in increments of 0.25.

Figure 5.8 shows the effects of spectral subtraction on intermediate modulation metrics used to compute the various speech-based STI methods. Figures 5.8A, B, and C reveal that the original cross-power spectrum methods and the original envelope regression method fail for spectral subtraction. In all three plots, the modulation metrics increase monotonically as the control parameter, κ , increases, eventually reaching invalid values greater than one. This indicates that these methods are not applicable to spectral subtraction. The remaining five plots reveal that all of the proposed methods (Figures 5.8E through H), as well as the normalized covariance method (Figure 5.8D), produce valid values of the intermediate metrics. As the control parameter increases, all of the intermediate metrics initially increase to a maximum and then decrease.

The proposed methods as well as the existing normalized covariance method exhibit behavior that is qualitatively consistent with a hypothetical trade-off between noise reduction and signal distortion. For each of these methods, the modulation metric initially increases, predicting slight improvements in intelligibility due to moderate levels of spectral subtraction ($\kappa \approx 1$) and predicting degradations in intelligibility for more severe processing ($\kappa > 2$). The modified cross-power spectrum methods and the modified envelope regression method predict the most benefit from spectral subtraction with $\kappa = 0.6$, while the normalized covariance and normalized correlation method favor $\kappa = 1.4$.

These results imply that spectral subtraction may improve the intelligibility of speech degraded by additive noise. A number of studies have shown that spectral

subtraction does not improve the intelligibility of speech for normal-hearing listeners (Lim and Oppenheim, 1979). However, spectral subtraction has been shown to improve intelligibility for cochlear implant listeners (Weiss, 1993; Hochberg et al., 1992). This is discussed in the next section.

5.2.2 Discussion¹¹

Candidate Speech-Based STI Methods

The results presented in the previous section indicate the suitability of the various speech-based STI methods for predicting intelligibility under conditions of acoustic degradation, envelope thresholding, and spectral subtraction. The long-term goal is to identify and validate a speech-based STI method that accurately predicts intelligibility of speech processed by a wide variety of linear and nonlinear operations. The immediate goal of this study is to identify speech-based STI methods that maintain a one-to-one relationship with the traditional STI for acoustic degradation while also producing qualitatively reasonable results for selected nonlinear operations.

Of the four original methods, only the normalized covariance method exhibited qualitatively reasonable behavior for the nonlinear operations considered in this study. However, this method does not have a one-to-one correspondence to the traditional STI for acoustic degradations. The other three previously-proposed methods produce invalid results for the nonlinear operations considered.

The four proposed speech-based STI methods exhibit one-to-one relationships with the traditional STI for acoustic degradations and produce qualitatively reasonable results for the nonlinear operations. Thus, all of the proposed methods are potential candidates to extend the STI to nonlinear operations while retaining their applicability to acoustic degradations. Additional work is required to determine if any of the proposed methods accurately predict speech reception for these and other nonlinear operations.

The normalized correlation method presents a substantial deviation from the traditional STI. The other proposed methods are equivalent to the traditional STI, that is,

¹¹ Section 5.2.2 is reproduced from Goldsworthy and Greenberg, 2004: Section VI, "Discussion." Changes were made to section titles and figure numbers to be internally consistent with this thesis.

the speech-base STI values correspond directly to traditional STI values. However, as seen in Figure 5.6, the normalized correlation method is not equivalent to the traditional STI, nor is it a linear transformation of traditional STI. A (nonlinear) function is required to map the normalized correlation STI values to the traditional STI. The normalized correlation metric is admittedly a departure from many of the principles of the traditional STI, and it may be preferable to consider it as a new intelligibility metric distinct from the STI except for the common elements of using frequency-band envelopes.

Predicting Intelligibility of CI-processed speech

The STI has already been adapted for use with hearing-impaired subjects (Humes et al., 1986; Payton et al., 1994), and it is a good candidate for predicting intelligibility of speech processed by cochlear implant speech processors. This expectation is based primarily on similarities between the STI calculation procedure and CI processing strategies; both the STI and conventional CI processing strategies use information from the envelopes in a number of frequency bands and neglect the fine structure. The STI calculation procedures can be tailored to match a particular CI sound-processing strategy by matching the frequency bands and method of envelope calculation.

Although the absolute performance of subjects listening to CI-processed speech differs from that of subjects listening to unprocessed speech, additive noise has relatively similar effects in both cases (Hochberg, 1992). Therefore, the STI methods that accurately predict the relative intelligibility among conditions of speech with additive noise (Figure 5.6) should also be valid for CI-processed speech with additive noise, although an alternate mapping from STI to percent correct scores may be required for CI-processed speech. It is expected that the same trends will exist for reverberant conditions, although there has been relatively little research assessing the intelligibility of CI-processed speech in reverberation.

The selection of envelope thresholding as a nonlinear operation was guided by our interest in CI-processed speech. Some CI processors use N-of-M processing, coding only a subset, N , of the total, M , frequency-band envelopes during each stimulation cycle (Loizou, 1998). The stimulation cycle is relatively short (a few milliseconds) compared

to the STI analysis frame (typically several seconds). The effect of N-of-M processing is comparable to setting the remaining $M - N$ envelopes to zero during intervals when the envelope is not selected. Although this is not identical to envelope thresholding, it has a similar effect on the shape of the envelope, preserving the envelope in intervals where its amplitude is relatively high and eliminating the envelope in intervals where its amplitude is low.

The envelope thresholding results in Figure 5.7 indicate that the four proposed methods are potential candidates for predicting the effect of N-of-M processing. If a frequency band is selected all of the time (equivalent to a threshold of 0%), then the intermediate modulation metric is one, contributing a transmission index value (TI_i) of one for that band. If a frequency band is never selected (equivalent to a threshold of 100%), then the intermediate modulation metric is zero and $TI_i = 0$. If a frequency band is selected intermittently, then the corresponding modulation metric will fall between zero and one, producing a transmission index that reflects that band's partial contribution to intelligibility. While all of the proposed methods are qualitatively correct in that they decrease monotonically from one to zero with increasing threshold, additional work is required to determine which methods, if any, are quantitatively accurate in predicting the effects of envelope thresholding and N-of-M processing on intelligibility.

While research indicates that spectral subtraction does not improve intelligibility for normal-hearing listeners (Lim and Oppenheim, 1979), it has been demonstrated to improve intelligibility for CI users (Weiss, 1993; Hochberg et al., 1992). We hypothesize that this may be related to the effective spectral resolution of the listeners; normal-hearing listeners have relatively fine spectral resolution that permits perception of narrow spectral peaks that rise above the background noise, while CI users are restricted to the relatively broad frequency bands used by their speech processors and therefore cannot perceive spectral peaks within a wider band of noise. As a result, normal-hearing listeners do not benefit from spectral subtraction, since they are already able to listen in relatively narrow bands. On the other hand, CI users benefit from spectral subtraction

algorithms that operate in frequency bins substantially narrower than the broader bands used by their speech processors.

The spectral subtraction results in Figure 5.8 indicate that the four proposed methods may be potential candidates for predicting the effect of spectral subtraction on CI-processed speech. The intermediate metrics indicate that the proposed STI methods will predict an improvement for speech processed with spectral subtraction algorithms using moderate values of the control parameter, κ . It appears that an appropriate speech-based STI may predict the effect of spectral subtraction on intelligibility more accurately for CI-users than for normal-hearing listeners precisely because it uses a broad frequency-band analysis similar to that used by CI sound-processing strategies. In fact, the success of the traditional STI for normal-hearing listeners may be due to the historic focus on broadband distortion such as reverberation and additive broadband noise. For example, consider the case of speech corrupted by a pure tone. This specialized interference would have little or no effect on intelligibility for normal-hearing listeners, but would have a detrimental effect on intelligibility when passed through a CI sound-processing strategy. In computing the STI, the effect of the pure tone would also show up in the apparent SNR for the corresponding frequency band, so that the STI would better predict the effect on intelligibility for CI-processed speech than for a normal-hearing listener.

5.3 Selection of Candidate Metrics

In Section 2.2.2 we summarized four distinct methods that exist in the literature for calculating STI based on speech signals. In Section 3.1 we introduced modifications of these methods resulting in five novel metrics. However, it is apparent from the data presented in the previous section that certain metrics produce similar results. In this section, we present an analysis of the nine candidate metrics justifying our selection of three metrics for further consideration.

First, we note that the real CPS, the magnitude CPS, and the envelope regression methods produce similar results for the experiment reported in Section 5.2 (see subplots

A, B, and C of Figures 5.6, 5.7, and 5.8. Not surprisingly, the resulting three modified version of those metrics also produce similar results (see subplots E, F, and G of Figures 5.6, 5.7, and 5.8). The similarity of these results suggests that we might be able to group the three existing and three modified methods into two classes.

This grouping of the real CPS, the magnitude CPS, and the envelope regression methods is reasonable since those methods have similar underlying mathematical structure. Both the real and magnitude CPS are calculated in identical manners with the exception that one uses the real part of the CPS while the other uses the magnitude. Furthermore, in Appendix B.3 we illustrate that the envelope regression method can be mathematically expressed as an energy-weighted average of the real CPS method. Hence, this method only differs from the real CPS method insofar as energy-weighted averaging differs from the traditional one-third octave weighting. A comparison shown in Appendix B.4 (see Figures B.1 and B.2) suggests that these two weighting strategies are quite similar.

It was also shown in the preceding section that the normalized covariance STI method does not produce a one-to-one mapping with traditional STI for additive noise and reverberation. Since traditional STI is well correlated to speech reception for these conditions, a candidate metric must have a one-to-one relationship with traditional STI if it will retain the success for those conditions. Consequently, we do not pursue analysis of the normalized covariance STI method.

It was shown that the normalized *correlation* STI method does produce a one-to-one mapping with traditional STI for additive noise and reverberation (Subplot H, Figure 5.6). However, the relationship between the normalized correlation STI method and the traditional STI is not linear. We feel that it is a stretch to classify this approach as an STI method since both the results for acoustic degradations and the underlying calculation suggest a fundamentally different metric. This is precisely why we developed the NCM, which is a variation on the normalized correlation method but a more substantial departure from the traditional STI methods. We did not analyze the NCM in the preceding section since this metric was developed after the preliminary experiment had

been conducted, however it is included in the following correlation analysis. We hypothesize that the performance of the NCM will be very similar to that of the normalized correlation STI method.

To further substantiate the classifications drawn above, we perform a correlation analysis among the nine metrics for all conditions examined in Chapters 6 through 9 of this thesis. The results are presented in Appendix A. The metrics are calculated as described in Section 4.5 for the conditions tested in each experiment. Thus, we calculated each of the nine metrics for 64 conditions (16 conditions in each of the 4 experiments). We calculated the correlation coefficients between pairs of metrics for each experiment. The correlation coefficients between the unmodified real CPS, magnitude CPS, and envelope regression methods were always at least 0.98. Similarly, The correlation coefficients between the modified real CPS, magnitude CPS, and envelope regression methods were always at least 0.98 with the exception that the correlation coefficient between the modified magnitude CPS and envelope regression methods was 0.91 for one of the four experiments. The high correlation coefficients between these metrics substantiate that these metrics form consistent groups for the conditions studied. Thus, for the remainder of the thesis we will focus on the envelope regression method (since it is the most efficient method of the three), both modified and unmodified. Similarly, the correlation coefficient calculated between the NCM and the normalized correlation STI was always at least 0.97 substantiating the grouping of those two methods. The normalized covariance STI method behaved similar to the NCM; however, the correlation coefficient between those methods dropped below 0.9 for the conditions tested in experiments 1 and 4. Thus, those methods should not necessarily be classified together. However, we don't consider the normalized covariance STI method further in this thesis based on its failure to map to the traditional STI in a one-to-one manner for additive noise and reverberation.

We have thus narrowed the candidate metrics to three: the envelope regression method, the modified envelope regression method, and the NCM. The unmodified envelope regression method is included in the selection of candidate metrics despite the

evidence presented in the previous section that it will produce invalid results for nonlinear operations. This method was included because of its similarity to more traditional STI methods and because we desire to establish for which nonlinear operations it fails to produce reasonable predictions.

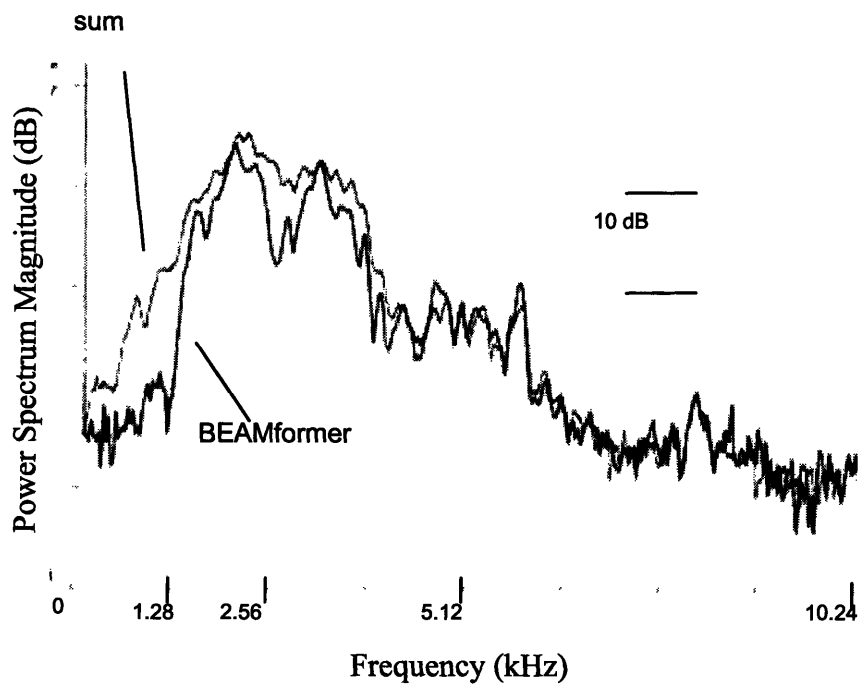


Figure 5.1: Physical response of Audallion BEAMformer.

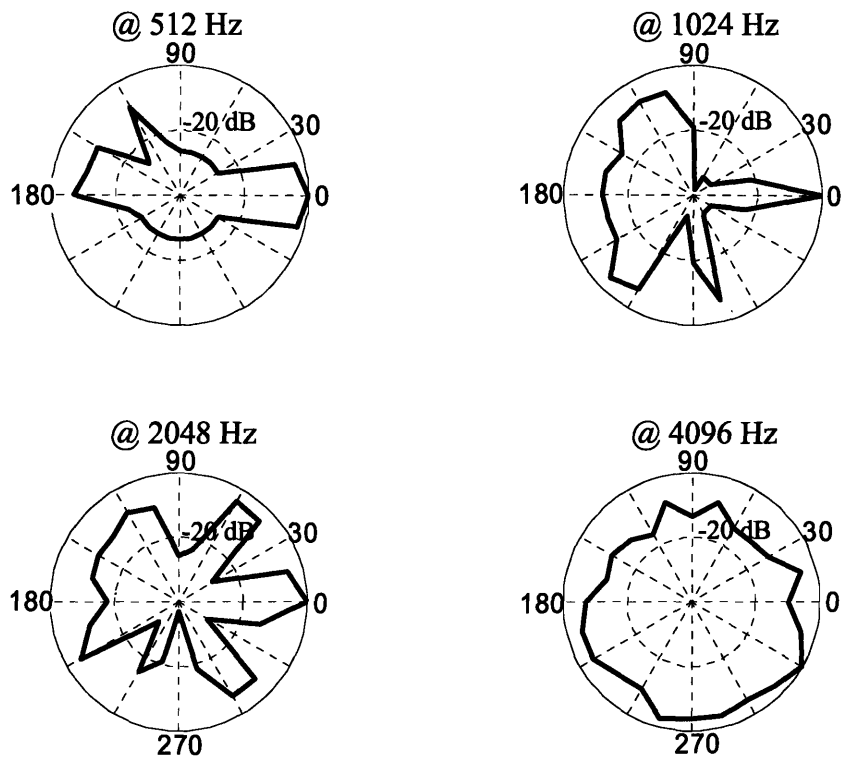


Figure 5.2: Physical response of Audallion BEAMformer. Desired direction at 0 degrees. The data is normalized such that the most intense response corresponds to 0 dB. This occurs at 0 degrees for 512, 1024, and 2048 Hz; but at 30 degrees for the 4096 Hz plot.

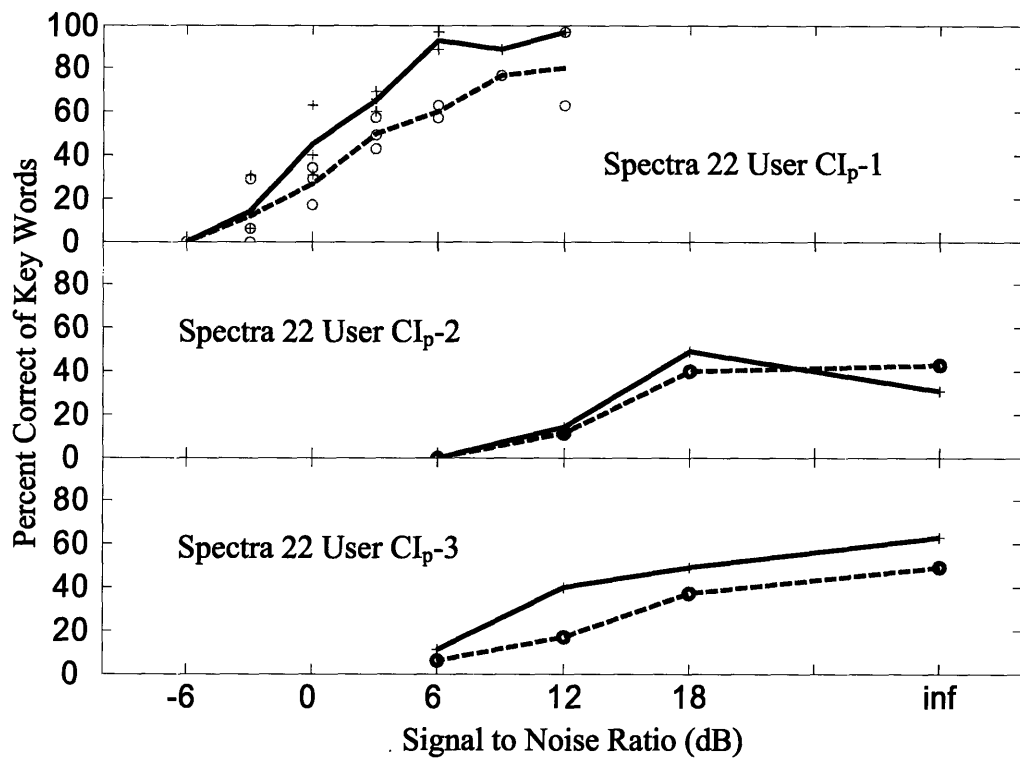


Figure 5.3: Subject results for Audallion BEAMformer. Solid and dashed lines represent speech reception performance in sum and BEAMformer modes, respectively.

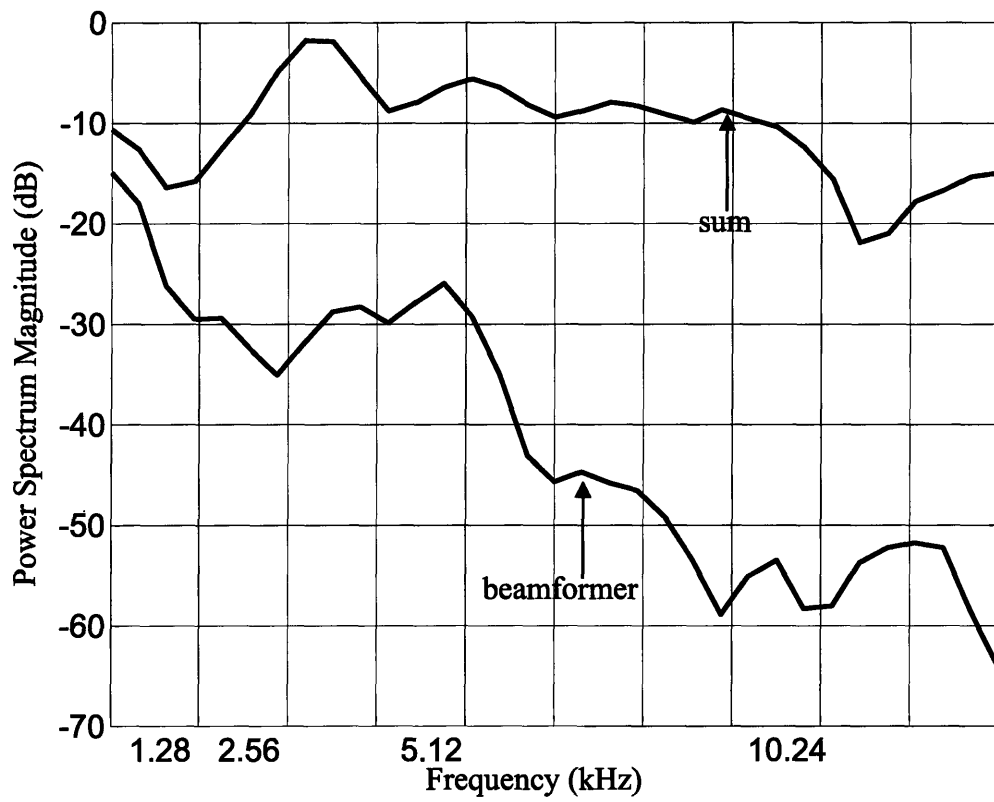


Figure 5.4: Physical response of our implementation of binaural algorithm.

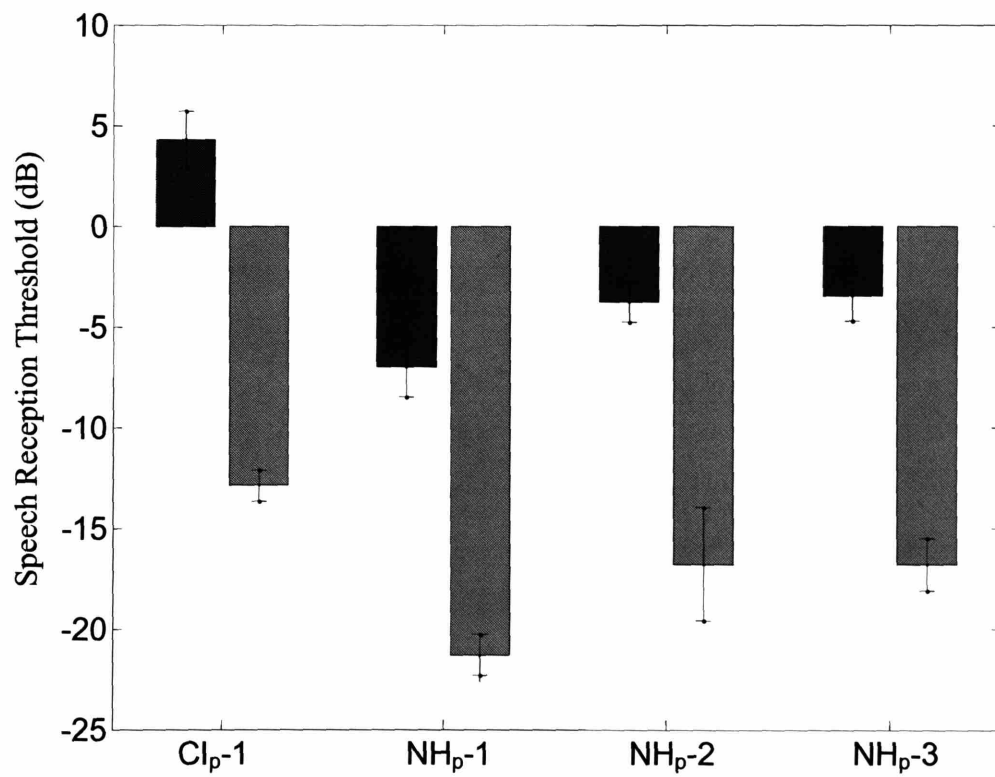


Figure 5.5: Speech reception threshold without (darker bars) and with (lighter bars) the binaural noise reduction algorithm for one CI user and three normal-hearing subjects.

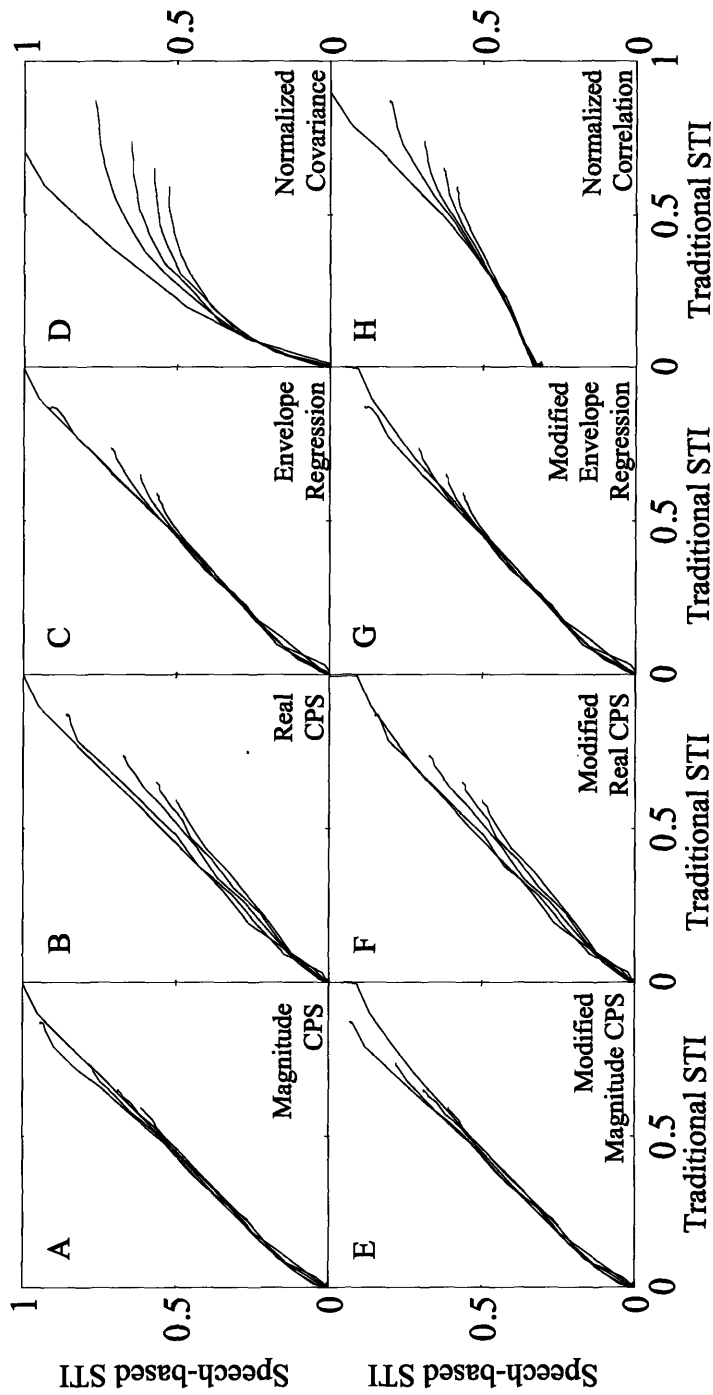


Figure 5.6: Comparison of speech-based STI methods to the traditional STI. Each plot shows the relationship between one speech-based method and the traditional STI. Each curve corresponds to the 45-dB range of SNR values for one level of reverberation. More reverberant conditions terminate at lower values of the traditional STI.

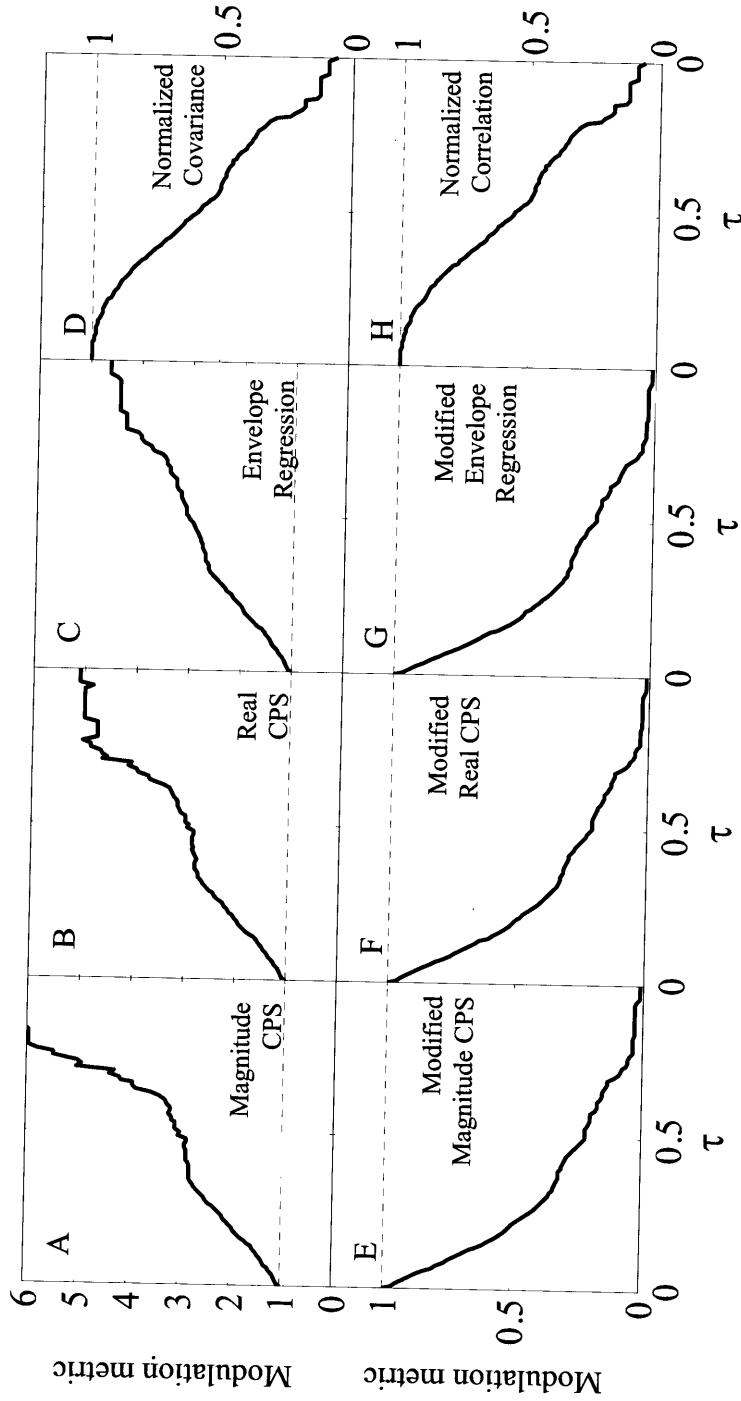


Figure 5.7: Intermediate modulation metrics of speech-based STI methods for envelope thresholding as a function of threshold, τ . For the CPS methods, the intermediate metrics are the MTFs from Eqs. 2.7 and 2.8 averaged over modulation frequency. For the envelope regression methods, the intermediate metric is M (Eq. 3.2). For the normalized covariance and normalized correlation methods, the intermediate metrics are r (Eq. 2.11) and ρ (Eq. 3.5), respectively. All results are for the octave-band centered at 1 kHz. The dotted line indicates unity, the maximum valid value for all metrics.

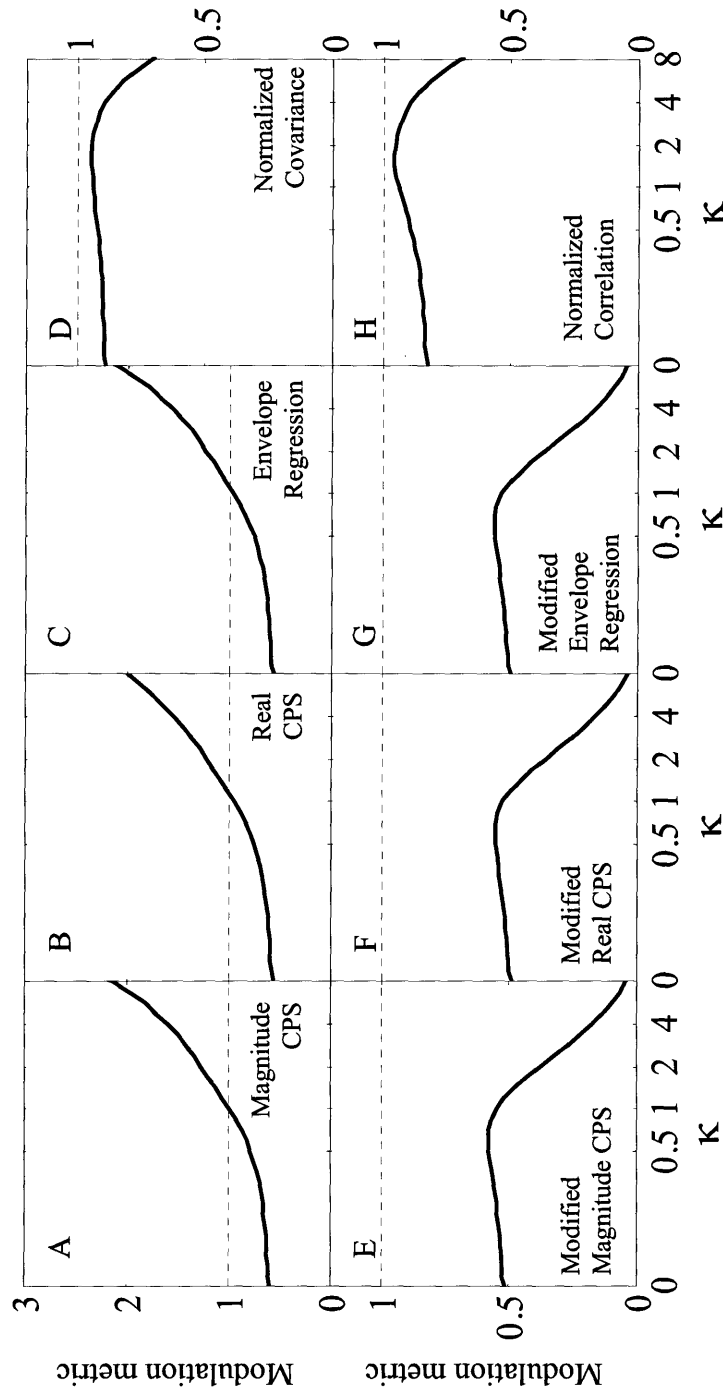


Figure 5.8: Intermediate modulation metrics of speech-based STI methods (as in Figure 5) for spectral subtraction as a function of control parameter, κ . All results are for the octave band centered at 1 kHz. The dotted line indicates unity, the maximum valid value for all metrics.

Chapter 6

Experiment 1: Acoustic Degradation

Acoustic degradations such as additive noise and reverberation decrease the intelligibility of speech. The STI predicts the intelligibility of acoustically degraded speech for normal hearing and hearing-impaired listeners. The experiments presented in this chapter are designed to assess the ability of the various STI and the proposed NCM methods to predict the effects of acoustic degradation on speech reception for CI-processed speech. Subjects included NH-CI₈ and actual CI users. Stimuli consisted of sentences, with multiple noise types, noise levels, and reverberation levels. Results show that objective intelligibility scores for both NH-CI₈ and actual CI users follow different trends than for normal-hearing (*not* listening to a vocoder simulation of CI sound-processing) subjects. All three intelligibility metrics investigated produce reasonable and comparable predictions. Possibilities for improvements upon the different metrics are developed in the discussion.

6.1 Introduction

A fundamental problem in hearing research is to understand how noise and reverberation affect the intelligibility of speech. Additive noise and reverberation degrade speech reception for both normal-hearing listeners and CI users. However, speech reception degrades more rapidly in the presence of background noise for CI users. It has been shown that CI users require 5 to 13 dB gain in SNR (using speech-shaped noise) in order to achieve comparable speech reception to normal-hearing listeners (Hochberg et al., 1992; Fu et al., 1998). Nelson et al. (2003) found that the SNR required by CI listeners was at least 25 dB greater than normal-hearing listeners when the noise source was amplitude modulated. Clearly, speech reception for CI users is more sensitive to the effects of additive noise. In addition, other differences exist between normal-hearing listeners and CI users. For example, Nelson et al. (2003) showed that normal-hearing listeners exhibit significant release from masking for modulated noise sources compared to unmodulated sources, while CI users receive very little release from masking and actually show negative effects of modulated noise for maskers at syllabic modulation rates (2-4 Hz). Qin and Oxenham (2003) illustrated—using noise vocoder simulations—that the intelligibility of CI-processed speech degrades more rapidly for modulated noise sources than for unmodulated sources when 8 or fewer channels are used in the simulation.

Much research has focused on attempts to quantify the effects of acoustic degradations on speech reception. For normal-hearing listeners, the STI is well correlated with speech reception for additive noise, reverberation, and their combination (see Section 2.2). In addition, STI has been modified and evaluated for use with hearing-impaired listeners (Humes et al., 1986, Ludvigsen, 1987, Payton et al., 1994). However, few studies have addressed the effects of modulated noise sources on STI predictions (Payton et al., 2002). Nor have previous studies attempted to predict the effects of these degradations on CI-processed speech.

A basic assumption of STI is that modulations arise from the desired source and that both additive noise and reverberation act to reduce the level of modulations in the received signal. Therefore, traditional STI methods treat the preservation of modulations in the received signal as having positive implications for intelligibility. A problem occurs

when the noise source itself is modulated. For example, consider the case of a single competing talker. At a sufficiently low SNR, the competing talker will reduce speech reception. As an extreme example, if the SNR was less than -60 dB then the target speaker probably could not be heard at all. However, the long-term modulation spectrum of the competing talker interference will be approximately the same as the desired talker; thus, the resulting STI would remain high (since the modulations appear to be transferred). One simple solution that addresses this problem is to require the modulation transfer function to be phase-locked. That is, the modulations in the output must occur at the same time as in the clean envelope signal. All of the candidate metrics are phase-locked methods. This issue has not been addressed in previous evaluations of STI, which tend to use unmodulated noise sources.

The experiment described in this chapter is designed to evaluate the ability of the candidate intelligibility metrics to predict the intelligibility of CI-processed speech when a signal is acoustically degraded. Previous studies have shown the STI to be well correlated to speech reception for normal-hearing subjects for additive stationary noise and reverberation; the experiments described in this section will extend STI theory in two dimensions. First, modulated as well as unmodulated noise sources will be considered. Second, the different intelligibility metrics will be applied to speech reception results for CI users and for NH-CI₈ listeners.

6.2 Conditions

The basic problem addressed in this chapter is illustrated in Figure 6.1. Clean speech is acoustically degraded and delivered to either a CI subject or to a normal hearing subject listening to the 8-channel vocoder simulation of CI sound-processing. The clean and

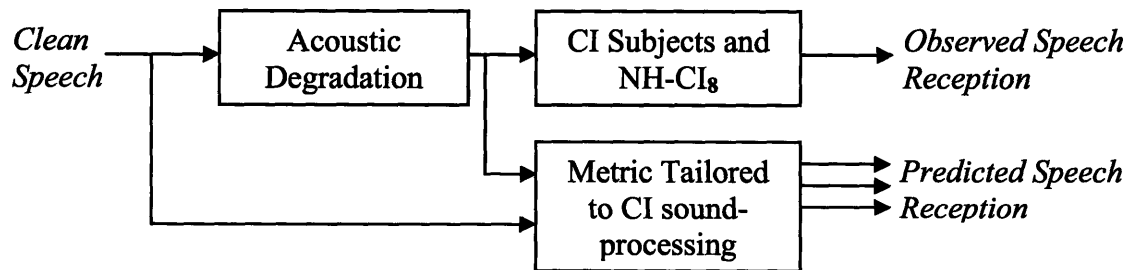


Figure 6.1: Block diagram of the experimental procedure for acoustic conditions.

degraded signals are used to calculate the intelligibility metric and the resulting predicted speech reception.

16 experimental conditions were chosen to answer the following questions:

- 1) Are the candidate metrics accurate predictors of speech reception for additive noise and reverberation for NH-CI_s and for actual CI users?
- 2) Do the candidate metrics quantify the speech reception effects of modulation in the noise source?

The first question addresses the applicability of the metrics to CI simulations and CI processing. The application of STI to quantify the intelligibility of CI-processed speech is novel; accordingly, a variety of experimental conditions need to be chosen to investigate this issue. Included in the test conditions should be a number of conditions for which it has already been shown that the STI is a good predictor of intelligibility for normal-hearing listeners. Specifically, combinations of speech-shaped noise and reverberation will be tested.

The second question concerns the application of STI to noise sources with inherent modulation. As discussed in the previous sections, traditional STI methods do not account for the impact of modulated noise sources on STI predictions. The test conditions chosen evaluate the capability of the STI methods in the context of inherent modulations in the noise source. Specifically, a time-reversed speaker and multi-talker babble will be used at three different SNRs to observe the effects. These noise types were chosen since they have different levels of modulation, yet none have any linguistic information that could confuse the listener.

The 16 conditions selected are based on 3 SNRs, 3 noise types, and 3 levels of reverberation. Table 6.1 summarizes these conditions. The three noise types are speech-shaped noise, multi-talker babble, and time-reversed speech. For normal-hearing listeners, the signal to noise ratios tested were -3 , 0 , and $+3$ dB. The reverberation times (T_{60}) tested were 0 , 0.15 (mild), and 1.2 (high) seconds. The experiment was divided into three trials that were tested on three separate days. Each trial consisted of the 16 conditions each tested using one complete list from the CUNY database. The six divisions (columns) of the conditions in Table 6.1 were used to partially counterbalance

the conditions across subjects. Within each subject, the SNR and reverberation levels were partially counterbalanced across trials. Details of the experimental methods are given in Chapter 4.

| | Noise Condition | | | | | |
|----------|-----------------|---------------------|-------|---------------------|----------------------|-------|
| | Quiet | Speech-Shaped Noise | | Multi-Talker Babble | Time-Reversed Speech | |
| Anechoic | ∞ | +3, 0 and -3 dB | | +3, 0 and -3 dB | +3, 0 and -3 dB | |
| Mild | ∞ | | +3 dB | | | +3 dB |
| High | ∞ | | +3 dB | | | +3 dB |

Table 6.1: Summary of experimental conditions for acoustic degradation. The conditions are separated into six columns corresponding to the six condition groups used to counterbalance the material as described in the text.

6.3 Results of Listening Experiment

6.3.1 NH-CI₈ Subjects

Six normal-hearing listeners participated in this experiment. Figure 6.2 illustrates the subject scores for each condition averaged across the NH-CI₈ subjects and trials. The data is divided into two subplots for ease of comparison. (The anechoic speech-shaped noise and time-reversed speech conditions at 3 dB SNR appear in both subplots.) Subplot A emphasizes the effect of reverberation and includes the quiet, speech-shaped noise, and time-reversed speech conditions tested at each reverberation level. Subplot B emphasizes the effect of SNR and includes each noise type at each SNR in an anechoic room.

An initial repeated measures analysis of variance (RMANOVA_1)¹² was performed using trials as the repetition variable. The dependent variable was the speech reception score transformed to RAU, and subject and condition were main factors. Subject was a significant factor ($p < 0.001$). Of the six subjects tested, NH-4 had a relatively high average score (10 RAU above mean for all subjects) and NH-6 had a relatively low average score (more than 10 RAU below mean for all subjects). The

¹² All variance and post-hoc measures are calculated in Matlab[®] in accordance with Winer et al. (1991).

interaction between subject and condition was not significant ($p > 0.1$); thus, the trends measured for the different conditions are consistent across subjects.

A second repeated measures analyses of variance (RMANOVA_2) was performed on the conditions represented in Figure 6.2A. This set of conditions represents a balanced set to consider the effect of reverberation. The dependent variable was the speech reception score transformed to RAU, and subject, noise type, and reverberation level were main factors. Both noise type and reverberation levels were statistically significant factors ($p < 0.001$). The interaction between noise type and reverberation level is also significant ($p < 0.001$).

The *post hoc* analysis of noise type and reverberation level was implemented according to Tukey's honestly significantly different (HSD) test ($\alpha = 0.05$). The three noise types in this group are quiet, speech-shaped noise, and time-reversed speech; each noise type was significantly different from the others. Quiet, of course, was the easiest condition and time-reversed speech the most difficult at the SNR of 3 dB. Each reverberation level was also significantly different from the others when averaged across noise types. However, when analyzed within noise type, the anechoic and mildly reverberant conditions were found to be significantly different only for speech-shaped noise. For all noise types, the highly reverberant condition resulted in significantly lower speech reception scores.

A third repeated measures analyses of variance (RMANOVA_3) was performed on the conditions represented in Figure 6.2B. This set of conditions represents a balanced set to consider the effect of SNR. The dependent variable was the speech reception score transformed to RAU, and subject, noise type, and SNR were main factors. As expected, the effect of SNR was statistically significant ($p < 0.001$) with higher speech reception scores associated with a higher SNR. An initial surprise was that the effect of noise type was not significant ($p > 0.1$). However, the interaction between noise type and SNR was significant. The analysis of this interaction yields insight into noise type trends as explained below.

The *post hoc* analysis of noise type and SNR was implemented according to Tukey's HSD ($\alpha = 0.05$). The analysis reveals strong trends between noise type and SNR. The general trend exhibited at relatively high SNRs is for speech reception scores

to be higher for unmodulated noise. At the highest SNR tested (3 dB), scores were significantly higher for speech-shaped noise and babble compared to time-reversed speech. At 0 dB SNR, scores were significantly higher for speech-shaped noise compared to time-reversed speech (the score for the babble conditions fell between the two but was not significantly different from either). This trend reversed at the lowest SNR tested; the speech reception score for time-reversed speech was significantly higher than the other two conditions. The underlying reason behind this interesting interaction between noise type and SNR will be discussed in Section 6.5.

6.3.2 CI Subjects

Three CI subjects participated in this experiment. The CI subjects were tested using a similar set of conditions as those summarized in Table 6.1. However, the SNR of each condition was shifted by a certain amount, Δ , in order to compensate for individual performance differences. The process for determining Δ for each subject is given in Section 4.4.2. Table 4.1 summarizes the Δ values found for each subject. Figure 6.3 illustrates the subjects' speech reception scores for each condition averaged across subjects and trials.

The analysis of variance performed was identical to those described in the previous section except using the CI data. The results found were similar to the NH-CI₈ results. First, as with the NH-CI₈ data, RMANOVA_1 implemented with the CI data indicates that both subject and condition were significant. The variance among subject scores was greater for the CI subjects. The average scores for the three subjects are 63.0, 44.9, and 26.2 RAU (respectively for CI-1, CI-2, and CI-3). In contrast to the NH-CI₈ data, the interaction between subject and condition was moderately significant ($p = 0.031$) for the CI data. Subsequent analysis illustrates this interaction reflects different performance trends in reverberation. Thus, care must be taken to understand different trends exhibit by individual subjects. To this end, Figure 6.4 illustrates speech reception scores for individual CI users.

As with the NH-CI₈ data, RMANOVA_2 calculated using CI data shows that both noise type and reverberation levels were statistically significant ($p < 0.001$). The noise type trends were the same as for the NH-CI₈ data with speech reception scores highest in

quiet, followed by speech-shaped noise, and then time-reversed speech. The reverberation trends were comparable to the NH-CI₈ data with speech reception scores significantly different for the reverberant conditions and ranked, as expected, with the anechoic case highest and highly reverberant conditions lowest. However, in contrast to the NH-CI₈ data, the interaction between noise type and reverberation level was not significant ($p > 0.1$). The difference found for the NH-CI₈ data was attributed to different performance trends in terms of the magnitude of the reverberation effect across noise type. *Post hoc* analysis indicates that this difference between reverberation trends in quiet compared to reverberation trends in noise was smaller for the CI data. In particular, the effect of high levels of reverberation on speech reception scores was comparable in quiet and in noise for the CI users.

In addition, RMANOVA_2 for the CI data clarifies the subject by condition interaction shown in RMANOVA_1. In particular, RMANOVA_2 shows that the interaction between subject and reverberation was significant ($p < 0.001$) while the interaction between subject and noise type was not significant ($p > 0.1$). Subject performance in mild reverberation varied among subjects from being approximately equal to the corresponding anechoic condition to being significantly lower than anechoic. The largest drop in performance attributed to mild reverberation was for subject CI-3 in speech-shaped noise who performed 25 RAU lower in mild reverberation compared to the anechoic condition. The detriment in speech reception scores due to high levels of reverberation compared to anechoic was always significant; however, the magnitude of the detriment ranged from approximately 30 to 70 RAU.

As with the NH-CI₈ data, RMANOVA_3 calculated using CI data shows that the effect of SNR is statistically significant ($p < 0.001$). As expected, higher speech reception scores occur for higher SNRs. Unlike the NH-CI₈ data, the impact of noise type was significant with speech reception scores for both speech-shaped noise and babble significantly higher than for time-reversed speech (but not from each other). Furthermore, the interaction between noise type and level is moderately significant ($p = 0.016$). The interaction between noise type and SNR was comparable to the NH-CI₈ data with higher speech reception scores in speech-shaped noise and babble at relatively high SNRs. As with the NH-CI₈ data, the trend reverses at lower SNRs

yielding lower speech reception scores for speech-shaped noise. A minor difference exists between CI and NH-CI₈ data that was confirmed by the *post-hoc* analysis: for the comparison of noise types in the -3 dB SNR conditions, scores for the babble condition were significantly higher than the speech-shaped noise condition. The time-reversed speech results were not significantly different from either the speech-shaped noise or babble conditions.

Taken together, the main differences between CI users and NH-CI₈ correspond to inter-subject variability with respect to reverberation. The trend that subjects performed better in unmodulated noise for a relatively high SNR was true for both groups of subjects. The trend that subjects performed better in modulated noise for a relatively low SNR was true for both groups of subjects; however, the trend was not as stark for the CI users in that scores were highest for the babble condition, which is less modulated than time-reversed speech.

6.4 Results of Intelligibility Predictions

6.4.1 NH-CI₈ Subjects

The procedure for calculating particular metrics from the clean and degraded speech waveforms is detailed in Section 4.5. As discussed in Section 5.3, we have selected the envelope-regression STI method, the modified envelope-regression STI method, and the NCM method for further evaluation. The metrics are calculated for the conditions tested and then a psychometric function is fit to the mapping between metric value and the mean reception scores (see Section 4.6). The resulting psychometric function thus yields a predicted score (in RAU) for a given metric value. Figures 6.5, 6.6 and 6.7 compare the observed scores for NH-CI₈ to the predicted scores for the candidate methods averaged over trials and NH-CI₈ subjects.

Two measures are given for assessing the predictions made by the different intelligibility metrics: 1) the model error defined as the standard deviation between predicted and observed scores and 2) the correlation coefficient between predicted and observed scores. All three intelligibility metrics have comparable performance in fitting the acoustic degradation data: the model errors differ by up to 0.5 RAU and the correlation coefficient by up to 0.01.

One minor trend that is observed for all three metrics is that performance on highly reverberant conditions is consistently over-predicted when noise is present. In contrast, performance is under-predicted for the highly reverberant condition in quiet. We discuss possible modifications that may produce more accurate predictions in quiet in Section 6.5.

A second, more significant, trend of interest is the prediction of the effect of noise type. None of the three metrics accurately predict the trends. With respect to noise type, both the original and modified envelope-regression STI methods consistently order the speech reception predictions with speech-shaped noise the lowest, time-reversed speech slightly higher, and babble approximately 10 RAU higher. The NCM method consistently orders the predictions with time-reversed speech the lowest, followed by speech-shaped noise, and then babble. These predictions do not correspond to the observed trends (c.f. Section 6.3.1). We discuss the need for modifications and give general suggestions in Section 6.5.

6.4.2 CI Subjects

The psychometric function was fitted for the NH-CI₈ data based on the mean subject scores. However, for actual CI users, we expect a wider variance in observed scores. It is possible that a particular subject may not be able to score 100% in quiet. To compensate for this potential difference, the psychometric function was fit to each subject and R_{max} of Equation 4.10 was allowed to vary. The added degrees of freedom in the model were taken into account in the calculation of the model error by lowering the corresponding degrees of freedom (N in equation 4.13). All three intelligibility metrics have comparable performance in fitting the acoustic degradation data: the model errors differ by up to 1.5 RAU and the correlation coefficient by up to 0.03. Figures 6.8, 6.9 and 6.10 illustrate the comparison between observed scores for the CI users and predicted scores for the respective methods.

Analysis of the predictions yield similar, yet less pronounced, findings compared to the NH-CI₈ analysis. First, all three metrics tend to over-predict performance on highly reverberant conditions when noise is present. This trend is not as stark as in the NH-CI₈. Consider, for example, in Figure 6.10 the NCM predictions underestimate

speech reception for highly reverberant conditions in the -10 to 10 RAU region. However, the general trend is still for overestimation of the highly reverberant conditions.

Second, as with the NH-CI₈ analysis, the effect of noise-type is not well predicted. The metric predictions follow the same trend as with the NH-CI₈ data—indeed, it is the same metric except for consideration of the CI users SNR shift (Δ)—and consequently, do not capture the appropriate ranking of speech reception with respect to noise type. Potential modifications for the metric to better capture the effect of noise type are developed in Section 6.5.

6.5 Discussion

All three metrics examined in this chapter produce reasonable predictions for the conditions tested. However, the metrics could be improved upon in a number of directions. In this discussion, we outline methods for improving the metrics by explicitly considering noise modulation and reverberation.

6.5.1 Noise Source Modulation

The metric predictions do not capture the trends associated with noise source modulation. Our results with CI users and NH-CI₈ generally confirm the finding that modulated noise is a more effective masker than speech-shaped noise for CI-processed speech (Qin and Oxenham, 2003; Nelson et al., 2003). This trend does not hold for relatively low SNRs—we hypothesize that both NH-CI₈ subjects as well as CI users were able to listen within temporal gaps of the time-reversed speech. However, the general trend of modulated noise being a more effective masker was found for mid to high regions of the speech reception range. Given this result, we analyze the proposed metrics to determine if a simple modification could capture this trend.

Since all of the candidate metrics take into account the phase of the degraded envelopes, we expected a given SNR to correspond to a particular metric value regardless of noise type. However, we noted in Section 6.4 that each method produced different values depending on noise type. For example, for the conditions tested the envelope-regression STI methods inaccurately predicted the lowest scores for the speech-shaped noise conditions compared to the other noise sources. This ranking of metric values

associated with noise type is not currently understood. Furthermore, while each method produces a ranking based on noise type, none of these rankings captures the complex interaction between noise type and SNR seen in the speech reception scores.

All of the candidate metrics are related to phase-locked MTFs. One starting point for improving the metric design would be consider the ramifications of using a non-phase locked method. The traditional STI method is related to a non-phase-locked MTF as described by

$$MTF(f) = \alpha \left(\frac{S_{yy}(f)}{S_{xx}(f)} \right)^{1/2}. \quad (6.1)$$

We chose not to develop metrics based on this form because preliminary investigations showed that predictions were inaccurate for nonlinear operations. Furthermore, this non-phase-locked form actually produces higher values for modulated noise since the noise source contributes to the overall modulation levels of the degraded envelopes. We are not suggesting that the non-phase-locked MTF might be used on its own to produce a superior metric; what we are suggesting is that a calibration term could be based on the non-phase-locked MTF.

The non-phase-locked MTF might be used to quantify the level of modulation in the noise source. To test this idea, we calculated the MTF as in Equation 6.1 for the speech degraded by speech-shaped noise, babble, and time-reversed speech at 0 dB SNR. We used the one-third octave binning procedure discussed in Chapter 5 to produce an average value and then averaged these values across frequency bands. The resulting value for speech-shaped noise, babble, and time-reversed speech at 0 dB were 0.59, 0.70, and 0.92. Thus, the higher the level of modulation, the closer this quantity is to 1.

The important conclusion from this analysis is that the non-phase-locked MTF produces distinct results dependent on the degree of noise source modulation. It should then be theoretically possible to use this result to modify the various metrics to produce the needed distinction between noise types. We leave determination of the exact manner of the transformation for future investigation.

Any function that forms a similar distinction between noise source modulation levels could be used to modify the metric predictions. The function could be based on

clean and degraded envelopes, or on the noise source itself when known. Many possibilities exist; if a suitable metric for quantifying the level of modulation in the noise (or similarity of the noise to the desired speech) is determined, then it could be incorporated into either the STI or NCM methods to account for the intelligibility differences between different noise types.

6.5.2 Effect of Reverberation

In this section, we consider minor modifications for optimizing the various methods with respect to reverberation. The candidate metrics fairly predict the effect of reverberation on speech reception. A few minor trends were pointed out in Section 6.4. Despite the fact that the reverberation trends are minor, we develop two different approaches that can be used to compensate for reverberation trends. The first approach is based on the effect of lag in the autocorrelation function used in the various metrics and should compensate for the low speech reception prediction in quiet. The second approach is adjusting the range of modulation frequencies used in the metric calculations.

To understand the justification for the first approach, it is insightful to analyze the reverberant impulse response, and the envelope of this impulse response, given in Figure 6.11. In the impulse response, the impulse corresponding to the direct wave propagation occurs before 5 ms; however, it is clear from the envelope of the transfer function that the energy resulting from room reverberation is sustained over 100 ms and has a peak near 60 ms. Consequently, significant speech energy—as well as information—of the desired speech signal may be delayed relative to the metric’s reference signal.

The question then is, “do the intelligibility metrics do an adequate job of characterizing this prolonged dissipation of acoustic energy?” To answer this question, we analyze the NCM method in terms of shifting the envelope signals. The normalized correlation can be expressed in terms of the autocorrelation function at zero-lag as

$$\rho^2 = \frac{\phi_{xy}^2}{\phi_x \phi_y} = \frac{R_{xy}^2[0]}{R_x[0]R_y[0]}. \quad (6.2)$$

Evaluating the cross-correlation at zero-lag implies that $x(t)$ and $y(t)$ are temporally aligned. However, the primary effect of reverberation is to retard the dissipation of

acoustic energy; as such, it may be more accurate to consider shifting $y(t)$ relative to $x(t)$.

To accomplish this comparison, we suggest calculating the normalized correlation based on the maximum value of the cross-correlation rather than the zero-lag value. That is, normalized correlation could be calculated as

$$\rho^2 = \frac{\max(R_{xy}^2[k])}{R_x[0]R_y[0]}, \quad (6.3)$$

where $R_{xy}[k] = E\{x[n] \cdot y[n-k]\}$. Note that the maximum value of the auto-correlation function is necessarily the zero-lag value, so the denominator terms need not be redefined. Figure 6.12 illustrates the importance of redefining the normalized correlation to take into account the effect of reverberation. In Figure 6.12A, a clean speech envelope and a corresponding reverberant envelope are plotted. It is clear that the envelope energy in the reverberant envelope decays more slowly after a peak than the anechoic envelope. Figure 6.12B illustrates the cross-correlation function as a function of lag for values between -100 and 100 ms. It is clear that the maximum value of the cross-correlation function does not occur at 0 lag, but near -40 ms.

Examining Figure 6.12B, we find that $\max(R_{xy}^2[k])$ is approximately 20% greater than $R_{xy}^2[0]$. Thus, using $\max(R_{xy}^2[k])$ should result in significantly larger values of the NCM metric for the reverberant condition. On the other hand, the change is not expected to be significant for additive noise. Redefining the normalized correlation as in Equation 6.3 allows the model to account for temporally aligning the clean and degraded envelopes to compensate for the retardation of the acoustic energy dissipation and produce more accurate predictions.

The envelope regression STI method can be amended in a similar manner. The intermediate modulation metric, M , of Eq. 3.2 can be expressed in terms of the covariance function as

$$M = \alpha \frac{\lambda_{xy}}{\lambda_x} = \alpha \frac{C_{xy}[0]}{C_x[0]}, \quad (6.4)$$

and can be redefined to account for the effect of reverberation as

$$M = \alpha \frac{\lambda_{xy}}{\lambda_x} = \alpha \frac{\max(C_{xy}[k])}{C_x[0]}, \quad (6.5)$$

where $C_{xy}[k] = E\{(x[n] - \mu_x) \cdot (y[n-k] - \mu_y)\}$. This proposed amendment should result in higher values of the metric predictions compared to values without the amendment; therefore, it would only be helpful for the highly reverberant conditions in quiet that are underestimated.

A second, simpler, modification that would affect the reverberation results is to simply change the maximum modulation frequency considered in the metric analysis. The effect of additive noise on the modulation transfer function is approximately constant across modulation frequency (see Figures 4.1 and 4.2), while the effect of reverberation is time varying. In general, increasing the maximum modulation frequency would decrease metric values for reverberation (i.e. since higher modulation frequencies generally have lower MTF values) but should theoretically not impact the results for additive noise (i.e. averaging constant values). Therefore, we suggest investigating the maximum modulation frequency included in the metric analysis as a free parameter to better fit the reverberation results.

6.6 Conclusions

The main conclusions of this chapter are:

- (1) The listening experiment confirmed that observed speech reception is lower for modulated noise than for unmodulated noise for CI-processed speech with the exception of relatively low SNRs where the subjects apparently benefit from temporal gaps in modulated noise sources.
- (2) The original and modified envelope regression STI methods and the NCM method all produce reasonable predictions for the wide range of acoustic conditions tested for both NH-CI₈ and actual CI users.

- (3) All three methods may be improved by explicitly accounting for the effects of noise source modulation and for the temporal shift in the reverberant envelope.

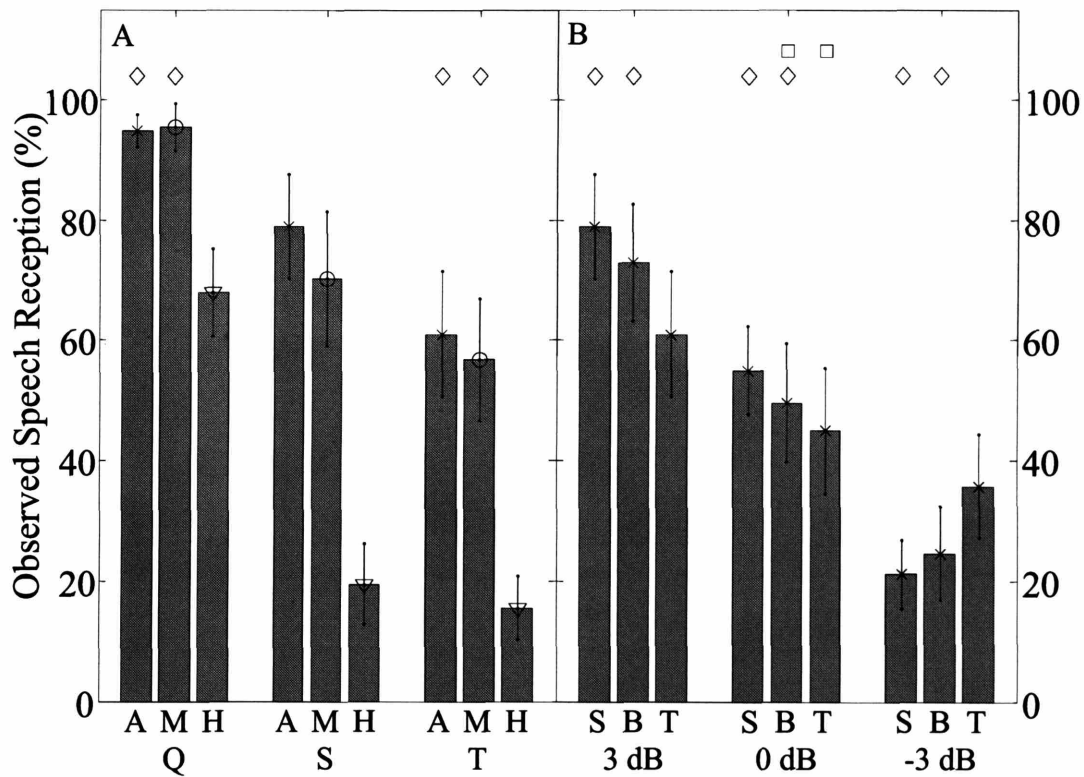


Figure 6.2: NH-CI₈ scores for acoustic degradation conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$). The figure is divided into two sub-plots to emphasize the effects of A) reverberation and B) SNR. Abbreviations: quiet (Q), speech-shaped noise (S), multi-talker babble (B), time-reversed speech (T), anechoic (A), mild (M), and high (H).

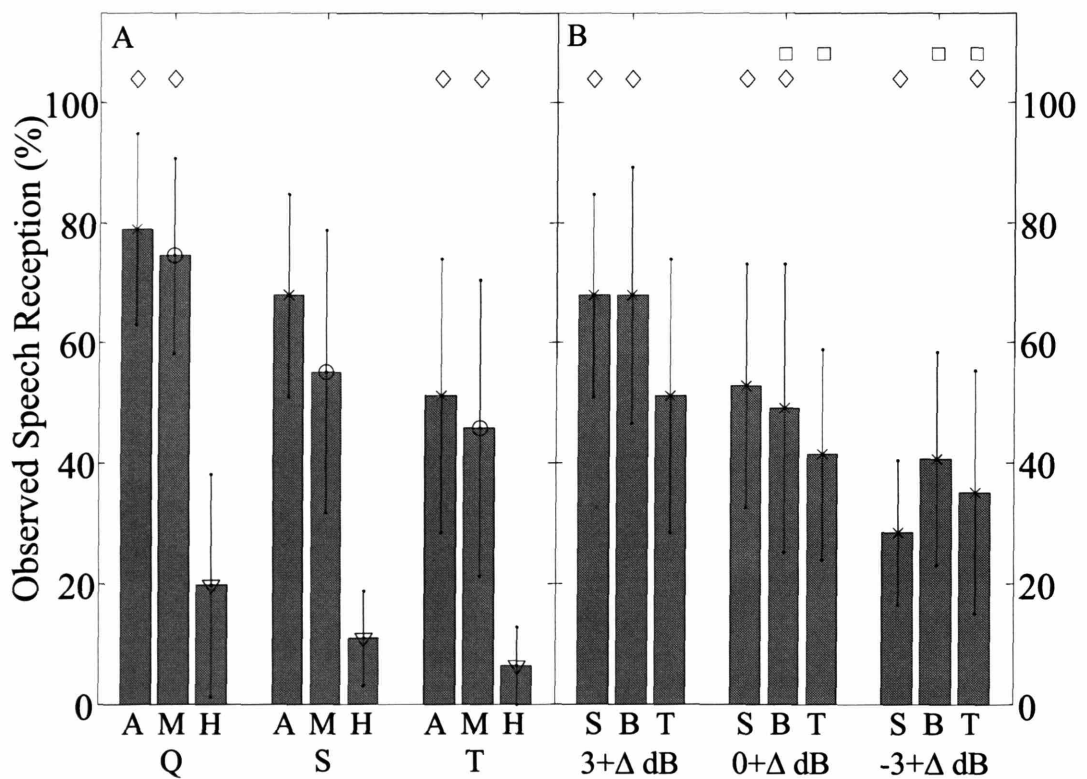


Figure 6.3: Speech reception scores for 3 CI users tested on the acoustic degradation conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$). The figure is divided into two sub-plots to emphasize the effects of A) reverberation and B) SNR. Abbreviations: quiet (Q), speech-shaped noise (S), multi-talker babble (B), time-reversed speech (T), anechoic (A), mild (M), and high (H).

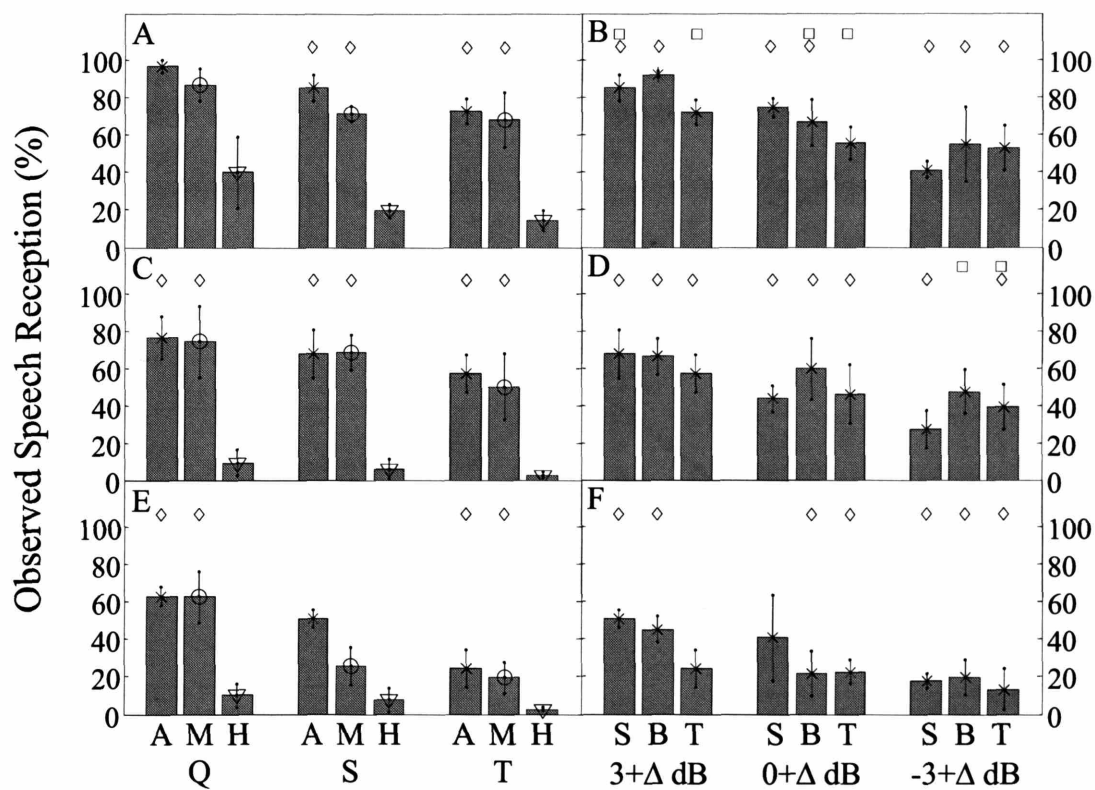


Figure 6.4: Individual speech reception scores for 3 CI users tested on the acoustic degradation conditions. The bars represent the mean scores averaged across trials for each subjects. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$). The figure is divided into two sub-plots to emphasize the effects of A,C,E) reverberation and B,D,F) SNR. Abbreviations: quiet (Q), speech-shaped noise (S), multi-talker babble (B), time-reversed speech (T), anechoic (A), mild (M), and high (H).

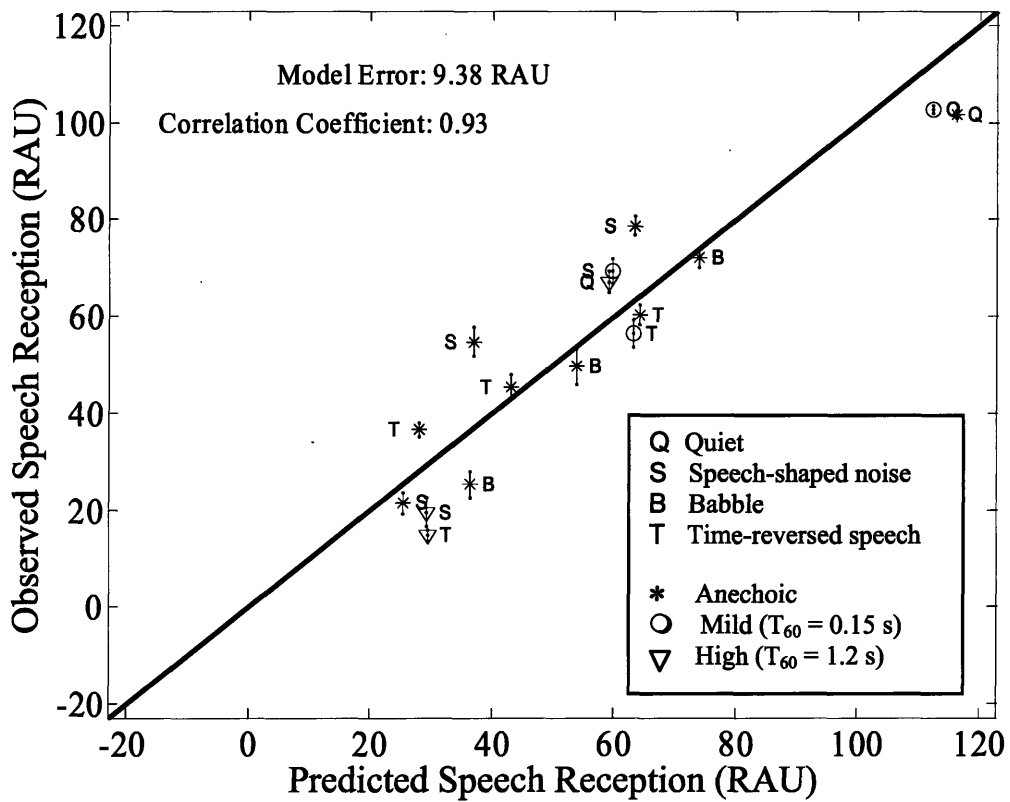


Figure 6.5: Comparison of observed scores for NH-Cl₈ and predicted scores from the envelope-regression STI method. The error bars represent the standard error of the mean.

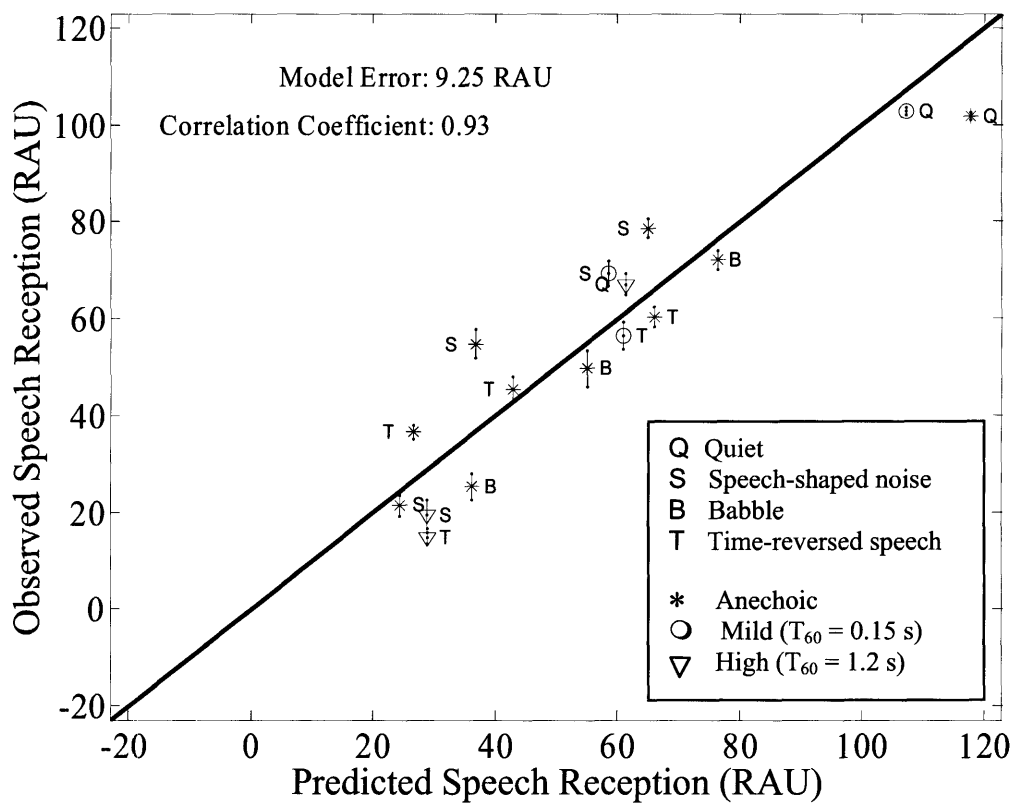


Figure 6.6: Comparison of observed scores for NH-Cl₈ and predicted scores from the modified envelope-regression STI method. The error bars represent the standard error of the mean.

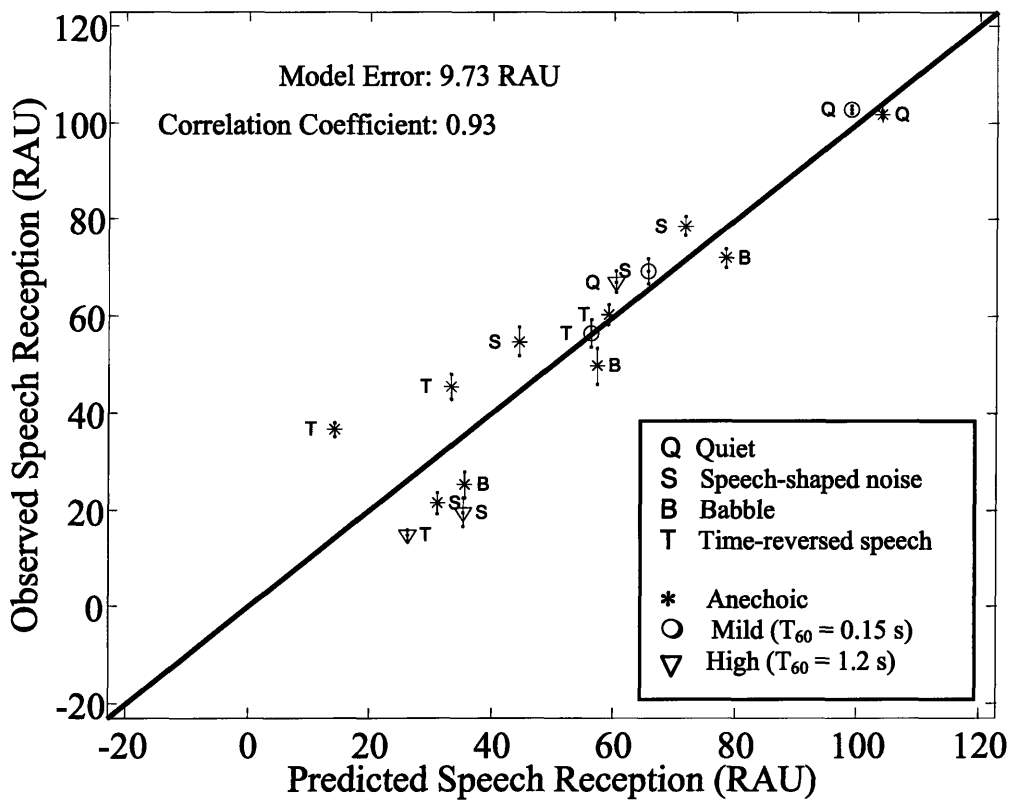


Figure 6.7: Comparison of observed scores for NH-Cl₈ and predicted scores from the NCM method. The error bars represent the standard error of the mean.

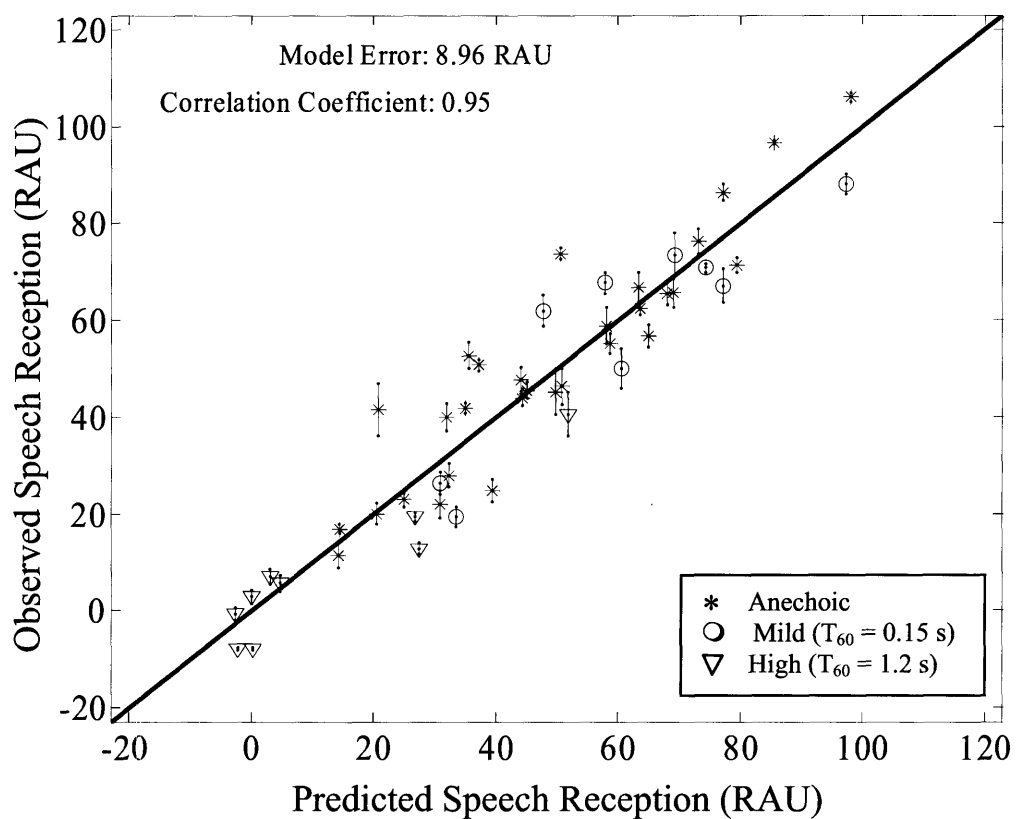


Figure 6.8: Comparison of observed scores for CI users and predicted scores from the envelope-regression STI method. The psychometric curve fitting was performed on individual subjects to account for large inter-subject differences. The error bars represent the standard error of the mean.

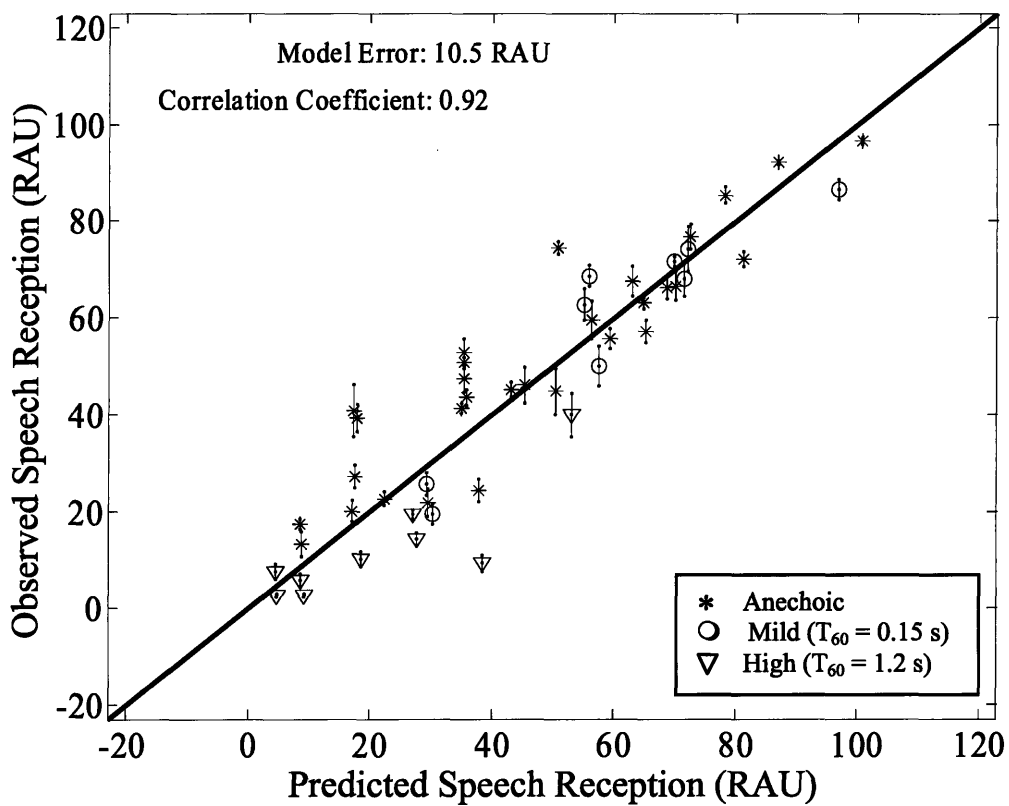


Figure 6.9: Comparison of observed scores for CI users and predicted scores from the modified envelope-regression STI method. The psychometric curve fitting was performed on individual subjects to account for large inter-subject differences. The error bars represent the standard error of the mean.

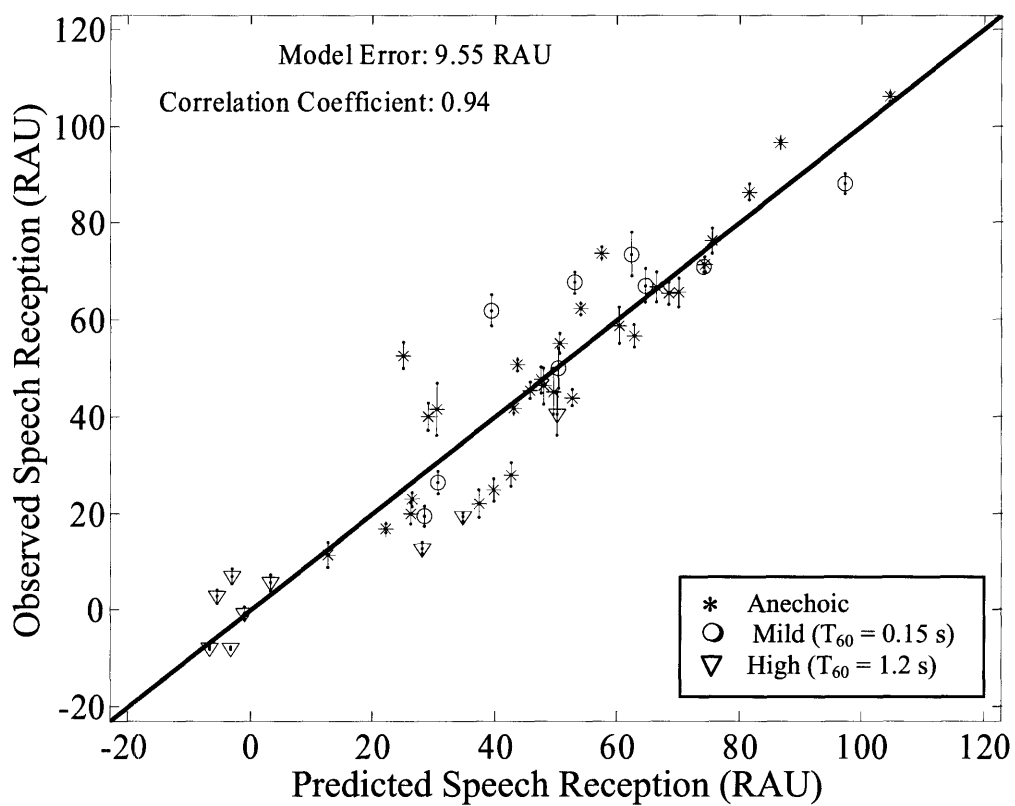


Figure 6.10: Comparison of observed scores for CI users and predicted scores from the NCM method. The psychometric curve fitting was performed on individual subjects to account for large inter-subject differences. The error bars represent the standard error of the mean.

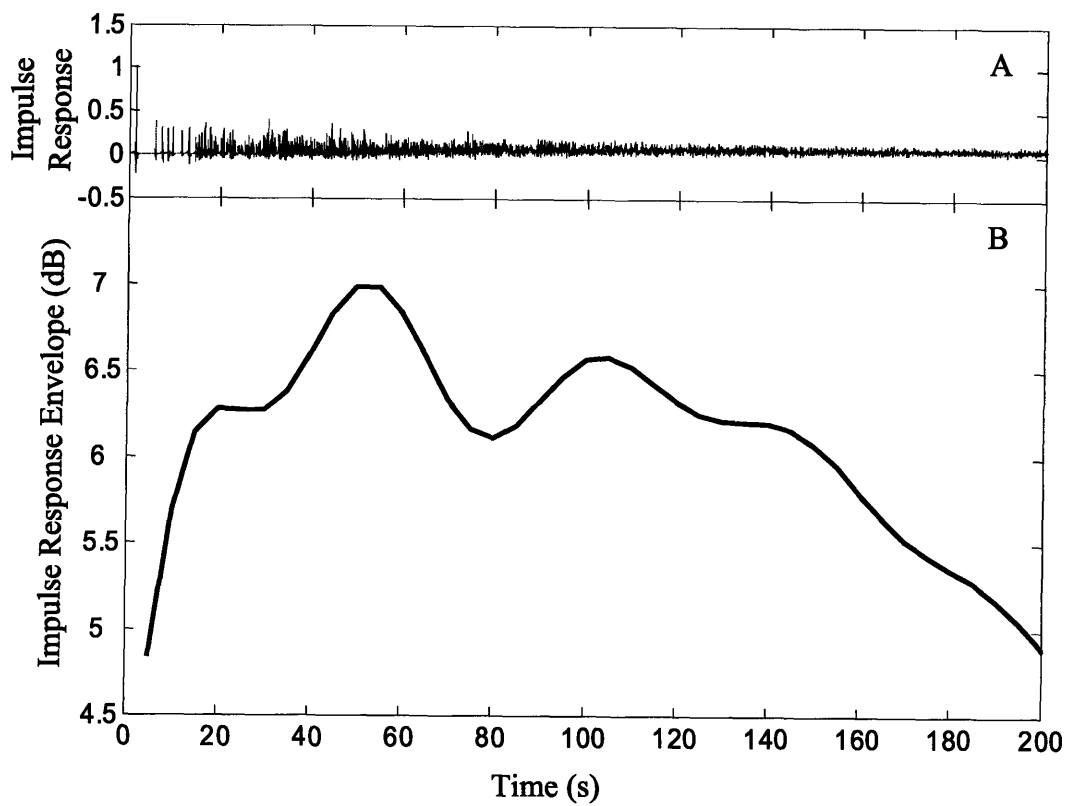


Figure 6.11: A) Reverberant impulse response, $T_{60} = 1.2$ seconds. B) Envelope of reverberant impulse response.

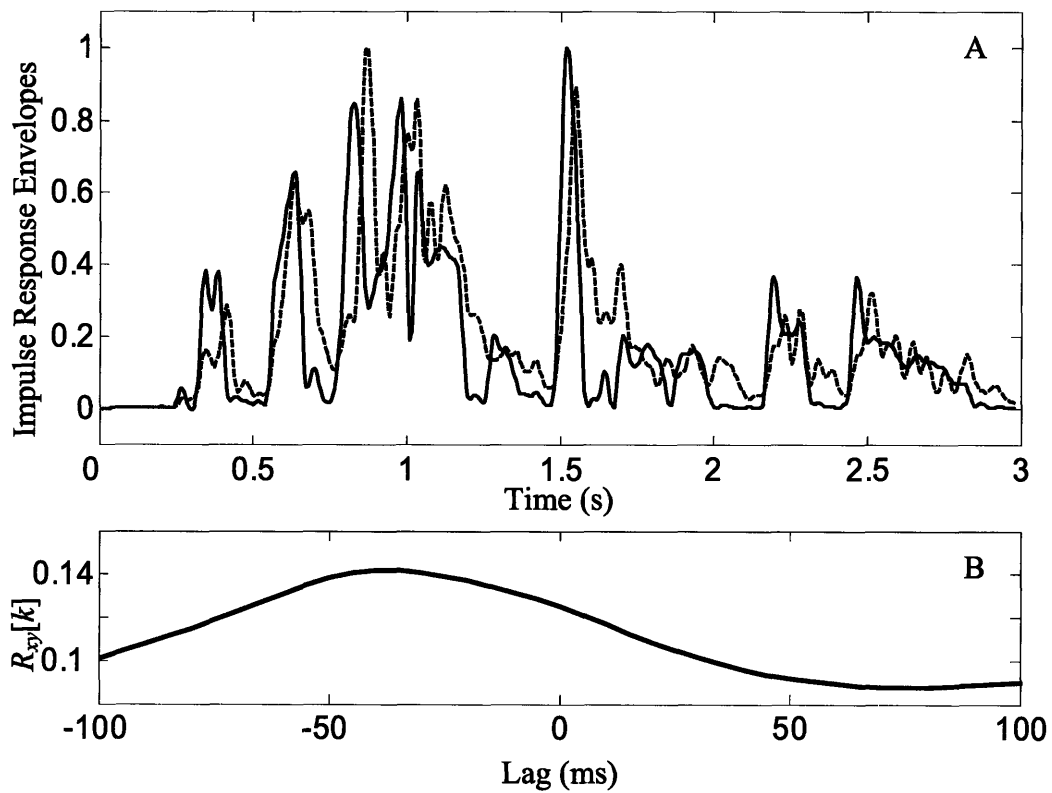


Figure 6.12: Effect of reverberation on a speech envelope signal. A) Impulse response envelopes of anechoic (solid line) and reverberant (dashed line) rooms. B) The autocorrelation function between the two envelopes of A as a function of lag.

Chapter 7

Experiment 2: N-of-M Processing

The experiment described in this chapter considers the ability of the NCM and STI variations to predict the effect of N-of-M strategies on CI-processed speech. Both clean and acoustically degraded speech is investigated for normal-hearing subjects listening to a 20-channel noise-vocoder simulation of CI sound-processing that includes an N-of-M algorithm. The values of N used in the N-of-M algorithm were 20, 9, 6, and 3. Subjects were tested for each value of N in quiet and using speech-shaped noise, multi-talker babble, and time-reversed speech as an interference at 0 dB SNR. Observed speech reception scores decreased monotonically with decreasing N for each condition. The unmodified STI method does not produce reasonable predictions for these conditions; however, the modified STI method, as well as the NCM method, produce reasonable predictions. Improvement upon the intelligibility models based on incorporating frequency band redundancy is discussed.

7.1 Introduction

N-of-M processing refers to a signal processing technique widely used in Nucleus[®] CI systems manufactured by Cochlear Corporation. As described in Section 2.1.1, the N-of-M strategy analyzes the envelope information of M channels and selects N channels for stimulation. The rationale for adopting the N-of-M strategy is that the subset of envelope signals with the highest energy will convey the essential speech information. By only coding a subset of the channels during any stimulation cycle, the algorithm allows the use of a higher pulse rate.

In the study presented in this chapter, we investigate the effect of coding only a subset of the envelopes on speech reception. The conditions are designed to evaluate the ability of performance metrics to predict speech reception of acoustically degraded speech when subjected to CI sound-processing strategies that includes N-of-M processing. We are interested in how performance changes for different noise types as a function of the number of channels coded, N , in the N-of-M operation. To investigate this effect, an N-of-M operation is included in the noise-vocoder simulation of CI sound-processing. By using noise vocoder simulations of CI sound-processing, we avoid issues concerning the stimulation rate of the electrodes. In other words, we desire to investigate the effects of the N-of-M processing independently from the effects of electrode stimulation rate.

We incorporate the N-of-M operation into the intelligibility metric calculation (see Section 3.2) and assess the predictive power of each candidate metric. By incorporating the N-of-M operation, the metric calculation is further tailored to specific CI sound-processing strategies. This additional tailoring allows the metrics to be used in conjunction with a larger set of CI sound-processing strategies. Furthermore, by analyzing the physical effect of N-of-M processing on the speech envelopes, rather than simply the speech reception consequences, researchers will better understand how the loss of envelope information affects speech reception. In this way, the performance metric framework can be useful for developing optimal N-of-M strategies.

7.2 Conditions

The N-of-M problem considered in this chapter is illustrated in Figure 7.1. Clean speech is acoustically degraded and then delivered to normal-hearing subjects listening to a 20-

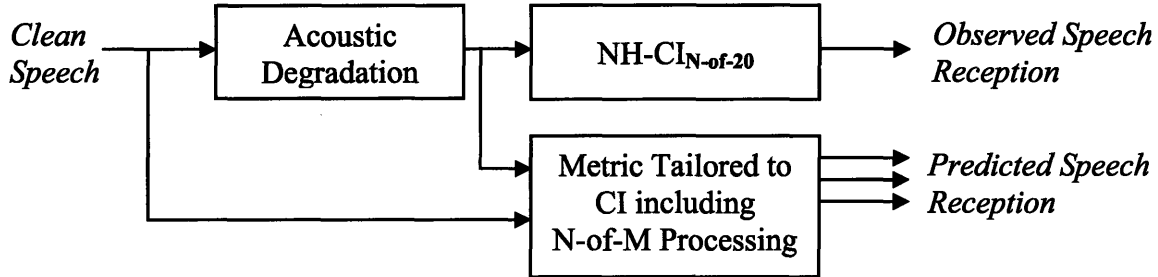


Figure 7.1: Block diagram of the experimental procedure for N-of-M conditions.

channel vocoder simulation of CI sound-processing that includes a simulation of the N-of-M operation. CI users were not tested for the experiment described in this chapter since that would require additional hardware to control the subjects' CI sound processors. The clean and degraded signals are used to calculate the various intelligibility metrics and the corresponding predicted speech reception scores.

16 experimental conditions were chosen to answer the following questions:

- 1) What is the effect of N-of-M processing on the intelligibility of CI-processed speech? In particular, is the intelligibility of speech processed by the N-of-M algorithm affected by the degree of modulation in the noise source?
- 2) Do any of the candidate metrics predict these effects?

The 16 conditions selected were based on quiet, 3 noise types, and 4 values of N in the N-of-M algorithm. Table 7.1 summarizes these conditions.

| Quiet (Q) | Speech-Shaped Noise (0 dB SNR) (SSN, S) | Multi-Talker Babble (0 dB SNR) (B) | Time-Reversed Speech (0 dB SNR) (TRS, T) |
|--------------------------|---|--|--|
| 3, 6, 9, and 20 of 20 | 3, 6, 9, and 20 of 20 | 3, 6, 9, and 20 of 20 | 3, 6, 9, and 20 of 20 |

Table 7.1: Summary of experimental conditions for N-of-M processing. Abbreviations in parenthesis are used to denote conditions in the figures presented in this chapter.

The experiment was divided into three trials that were tested on three separate days. Each trial consisted of the 16 conditions summarized in Table 7.1. Each condition was tested using one complete list from the CUNY database. The four divisions (columns) of the conditions found in Table 7.1 were used to partially counterbalance the conditions across a 4-subject set and the value of N tested was partially counterbalanced within each subject across trials (with a fourth ‘null’ trial serving only for counterbalancing purposes). Two groups of four subjects participated for a total of 8 normal-hearing subjects.

7.3 Results of the Listening Experiment

Figure 7.2 illustrates the speech reception scores for each condition averaged across subjects and trials. Figure 7.3 represents the same data, but grouped differently to emphasize the effect of noise type.

An initial repeated measures analysis of variance (RMANOVA_1)¹³ was performed using trials as the repetition variable. The dependent variable was the speech reception scores transformed to RAU, and subject and condition were main factors. Subject was a significant factor. The range of subject scores was 14.5 RAU. The lowest average score was 50.5 RAU and the highest was 65.0 RAU; the mean score across conditions and subjects was 58.6 RAU. However, the interaction between subject and condition was not significant ($p > 0.1$). Thus, the trends observed for the different conditions were consistent across subjects.

A second repeated measures analyses of variance (RMANOVA_2) was performed using the speech reception scores transformed to RAU as the dependent variable, and subject, noise type, and numbers of active channels (N , in the N-of-M algorithm) as main factors. Noise type, N , and interaction between them were significant ($p < 0.001$); this interaction was similar to the noise type and SNR interaction revealed in Experiment 1 and is analyzed further below.

Post hoc comparisons were made according to Tukey’s HSD ($\alpha = 0.05$). The first set of *post hoc* comparisons compared scores for different N averaged across noise type.

¹³ All variance and post-hoc measures are calculated in Matlab[®] in accordance with Winer et al. (1991).

The general trend was, as expected: lower scores occur for lower values of N . One minor exception was that the $N = 6$ and $N = 9$ conditions did not produce significantly different scores for time-reversed speech. A more interesting exception was that in quiet, scores did not significantly differ for the $N = 6, 9, \text{ or } 20$. Aside from those exceptions, scores followed the trend that more channels coded corresponded to higher speech reception.

The second set of *post hoc* comparisons compared scores for different noise types averaged across N . Speech reception scores in quiet were, as expected, significantly higher than scores in noise. Amongst the noise conditions, scores for the speech-shaped noise and the time-reversed speech conditions did not significantly differ; however, both produced higher scores than the multi-talker babble. The similarity in average scores for the least modulated noise source (speech-shaped noise) and the most modulated noise source (time-reversed speech) can be understood by considering *post hoc* comparisons for a given N . For $N = 20$, scores were highest for speech-shaped noise, second highest for babble, and lowest for time-reversed speech (with all comparisons significant). For $N = 9$, scores for speech-shaped noise and babble were not significantly different, but both were significantly higher than for time-reversed speech. For $N = 6$, scores for speech-shaped noise and time-reversed speech were not significantly different, but both were significantly higher than for babble. For $N = 3$, scores for speech-shaped noise and babble were not significantly different, but both were significantly *lower* than for time-reversed speech. Thus, the general trend was for the speech reception to be higher for the unmodulated noise source for large values of N . For smaller values of N , speech reception was highest for the time-reversed speech condition. These trends illustrated an interaction between noise type and N similar to that seen for noise type and SNR in Chapter 6.

7.4 Results of Intelligibility Predictions

The procedure for calculating particular metrics from the clean and degraded speech waveforms is detailed in Section 4.5. As described in Section 5.2, it is possible for the original envelope regression method to fail by producing invalid values of the intermediate metric. In particular, when the modulation metric (Eq. 3.2) is outside the range between 0 and 1, then the apparent SNR calculated (Eq. 3.1) is a complex—in the

mathematical sense—number and cannot be interpreted in the existing STI framework. In this chapter, we avoid this problem by clipping the modulation metric (Eq. 3.2) to values between 0 and 1.

The metrics were calculated for the conditions tested and then a psychometric function was fit to the mapping between metric value and the mean scores. The resulting psychometric function thus yields a predicted score (in RAU) for a given metric value. Figures 7.4, 7.5 and 7.6 illustrate the comparison between observed scores for NH-CI_{No20} and predicted scores for the respective methods.

As was done in Chapter 6, two measures are given to assess the predictions made by the different intelligibility metrics: 1) the model error defined as the standard deviation between predicted and observed scores and 2) the correlation coefficient between predicted and observed scores. The envelope regression STI method without the proposed modification fails to produce reasonable predictions of speech reception for the N-of-M conditions. This failure is quantified by the high model error (23.9 RAU) and the low correlation coefficient (0.67). In contrast, both the modified envelope regression STI method and the NCM method produce reasonable results. Their respective model errors are 10.0 RAU and 11.3 RAU; while their respective correlation coefficients are 0.95 and 0.93.

Both the modified envelope regression STI and the NCM methods produce reasonable predictions of scores, but certain inaccuracies need to be highlighted. Scores for the different noise-types exhibited an interaction trend with noise-type and N . Specifically, for higher values of N , scores were lower for the more modulated noise sources. This trend reverses for lower values of N , where scores are higher for time-reversed speech conditions. In contrast, the metrics generally produced consistent rankings in terms of predictions. For example, the NCM method always predicts the time-reversed speech conditions to have the lowest scores (that prediction is only true for the $N = 9$ and 20 conditions). For the modified envelope regression method, predicted scores were generally highest for time-reversed speech and lowest for multi-talker. The inability of the metrics to capture the interaction between noise type and N is similar to the problem of capturing the effect of noise type and SNR addressed in Chapter 6.

Second, both the modified envelope regression STI and the NCM methods fail to predict the effect of N-of-M processing on scores in quiet. The predictions are too low when compared to observed scores. For example, the 3-of-20 condition in quiet had a mean observed score of 86%; however, the corresponding predictions for the STI and NCM methods were 58% and 42% respectively. This prediction error was the largest for any condition in this data set. Analysis of this failure with suggestions for improvements is found in Section 7.6.

7.5 Frequency-Band Analysis

The conditions chosen for Experiment 1 presented in the previous chapter were chosen to investigate a wide range of acoustic degradations. A central issue was how the transmission index is calculated based on the clean and degraded envelopes. The conditions were chosen such that there was little variation across frequency bands. In particular, all noise sources had the same long-term spectra as the clean speech signal, and the reverberant impulse responses were designed to have the same T_{60} independent of frequency. This design placed the focus on how the degradations affected the envelope independent of the frequency band of interest.

In contrast, the N-of-M processing strategy might affect the TI values differently depending on the frequency band of interest. The N-of-M strategy chooses the N highest energy bands during a particular cycle. The N-of-M strategy will follow certain trends. For example, when a vowel is present, the low-frequency bands will generally be selected. By analyzing the TI values across frequency we may gain insight into the behavior of the N-of-M algorithm.

Figure 7.7 illustrates the TI values (for the NCM method) as a function of band number for the 6-of-20 condition in quiet. The TI values are close to one for the first five frequency bands. This is because when a vowel is present, the low frequencies dominate, and the first five frequency bands are almost always selected amongst the 6 chosen bands. On the other hand, during consonants that have little low frequency energy, those bands are not selected; however, the effect on the envelope is small. In a similar manner, certain consonants have a predominantly high-frequency spectrum. When those consonants occur, the highest frequency bands are faithfully represented. Consequently,

the high-frequency bands have fairly high TI values. The middle frequency bands between 7 and 15 have the lowest TI values in quiet. These low TI values result from not selecting these bands despite having significant energy present because the energy in either the lowest or highest frequency bands is comparably higher.

Figure 7.8 illustrates the TI values (for the NCM method) as a function of channel number for the 6-of-20 condition in speech-shaped noise. The trend for the lowest frequency bands is similar to the trend in quiet: the TI values are fairly high since those bands are chosen when the vowel sounds dominate. However, the highest frequency channels are much lower in noise than in quiet. In the quiet case, when a consonant is present with predominantly high frequency energy, the N-of-M strategy selects the high-frequency bands; however, when noise is present, the noise contributes significant energy to low-frequency bands causing the N-of-M strategy to select low-frequency bands. Consequently, the TI values for the high-frequency bands in noise are low compared to the values in quiet.

7.6 Discussion

The unmodified envelope regression STI did not produce reasonable predictions of speech reception for N-of-M processing. The modified envelope regression STI and the NCM methods did produce reasonable predictions. In this discussion we first develop a method for improving the model predictions in quiet and then consider ways for using the metric for optimizing the N-of-M procedure. It should be noted before continuing that the effect of noise source modulation was very strong for the N-of-M conditions. However, the trends are similar to results presented in Chapter 6 so we refer the reader to Section 6.5.1 for the relevant discussion.

7.6.1 Results in Quiet can be Improved by Considering Mutual Information Model

A significant disparity between speech reception prediction and objective score occurred for the 3-of-20 condition in quiet. We hypothesize that incorporating mutual dependence of adjacent frequency bands into the intelligibility models would reduce this disparity.

Grant and Braida (1991) suggested that adjacent frequency bands in articulation index (AI) analysis would be more correlated than non-adjacent frequency bands. We

hypothesize that this effect would be more pronounced for narrow frequency bands. In other words, narrow adjacent frequency bands are more likely to contain redundant information. For the 3-of-20 condition in quiet, the algorithm must select 3 frequency bands every 4 ms and set the other bands to zero. However, it is possible that a significant portion of the envelope energy set to zero carries redundant information. The intelligibility models should be framed to account for the possibility of redundant information—or frequency band correlation—and be reevaluated to see if predictive performance is improved upon.

Steeneken and Houtgast (1999) developed an STI model that incorporates mutual dependence of adjacent frequency bands. The revised model was found to produce more accurate results for the conditions they considered and the results were included in the revised IEC standard (IEC, 1998). The revised method incorporates mutual dependence by introducing redundancy factors, γ_j , into the psychoacoustic STI weighting function,

$$STI = \sum_i^N w_i TI_i - \sum_j^{N-1} \gamma_j \sqrt{TI_j \cdot TI_{j+1}}, \quad (7.1)$$

with the constraint that

$$\sum_i^N w_i - \sum_j^{N-1} \gamma_j = 1. \quad (7.2)$$

They found that this revised model improved the data fit for a set of acoustic conditions. This revised form could be used in conjunction with any of the modified methods we have developed and used for analysis with the N-of-M operation.

We also propose a second revision that accounts for redundant information. It should be noted that the revision proposed by Steeneken and Houtgast is based on the calculated TI values for each frequency band. Consider the case of additive speech-shaped noise. If the speech-shaped noise truly has the same long-term spectrum as the desired speech signal, then theoretically, the TI values will be the same for each frequency band. For this case, the revision would not alter the STI calculated. However, a subject might perform better than predicted by capitalizing on *short-term redundant information*. At a given moment in time, one particular frequency band may be more

clear and convey similar information as an adjacent band, while at the next moment in time the roles are reversed. We suggest that redundant information should be accounted for by taking into account short-term comparisons of frequency bands.

One procedure for doing this would be to calculate the intermediate metrics on a short-term scale and then average across bands. For example, in the NCM method we could generalize the TI calculation as

$$TI_{ij} = \frac{E[x_i(t)y_j(t)]}{E[x_i(t)]E[y_j(t)]}, \quad (7.3)$$

that could then be calculated in a short-time manner (e.g. every 30 ms). Adjacent frequency bands could then be averaged in manner similar to the Steeneken and Houtgast revision and then averaged across time. The key difference in this revision is that analysis is performed first on short-time segments allowing redundant information to be analyzed with finer temporal resolution. The above revision based on short-time analysis provides one example of how we might quantify adjacent channel correlation and then average; similar methods could also be developed with different functions for quantifying the redundancy.

7.6.2 Using Intelligibility Models for Optimizing N-of-M Processing

The results of the frequency-band analysis presented in Section 7.5 facilitate analysis of N-of-M processing and might be used for optimization. The results presented in Figures 7.6 and 7.7 clearly illustrate that the N-of-M operation does not affect all frequency bands equivalently. For both the quiet and additive noise condition, the low-frequency bands always produced significantly higher TI values. In fact, for the 6-of-20 quiet condition, the lowest five frequency bands all had TI values greater than 0.96. We might ask if it is possible to alter the N-of-M strategy to improve higher-frequency performance without significantly reducing low-frequency performance.

One possible alteration would be to restrict the N-of-M algorithm such that *adjacent bands would not be selected*. This proposal assumes the above argument that adjacent bands will carry redundant information. It would be straightforward to evaluate this proposal using $NH-CI_{N\text{-of-M}}$. The analysis of the TI values across frequency bands

could be used as a guide towards modifying the N-of-M algorithm. For example, the across-band correlation suggested in Equation 7.3 could serve as a guide as to the degree of adjacent frequency band redundancy. If two adjacent bands have a low level of redundant information, then they would be excluded from the rule that adjacent bands not be selected.

Another possibility would be to pre-emphasize the spectrum in a manner that would theoretically produce the highest overall metric value before applying the N-of-M algorithm. This approach could be used to shift more of the N-of-M decisions to the higher frequency components. Again, the intelligibility models could be used to determine a range of possible pre-emphasis filters and then subject testing could be used to determine optimum settings.

7.7 Conclusions

The main conclusions of this chapter are:

- (1) Speech reception was generally higher for larger values of N . The two exceptions seen were in quiet where $N = 6, 9$, and 20 were not significantly different, and in time-reversed speech where $N = 6$ and 9 were not significantly different.
- (2) Speech reception for the N-of-M processing conditions exhibited an interaction between noise type and N similar to the interaction between noise type and SNR seen in Chapter 6.
- (3) The original speech-based STI methods do not produce reasonable predictions for N-of-M processing.
- (4) The modified STI and the NCM methods produce reasonable predictions for N-of-M processing but fail to capture the interaction between noise type and N .
- (5) We propose that the intelligibility models would produce better predictions, especially for N-of-M in quiet conditions, by incorporating redundant information.

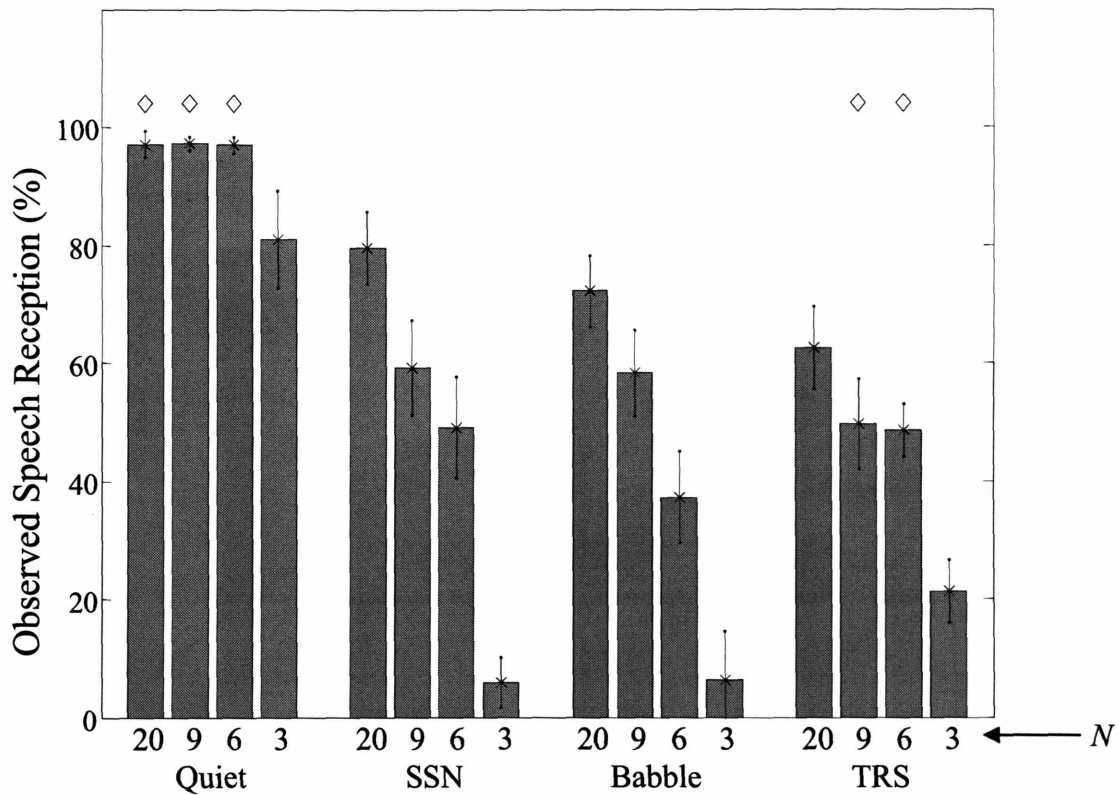


Figure 7.2: NH- CI_{Sim} scores for N-of-M processing conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$).

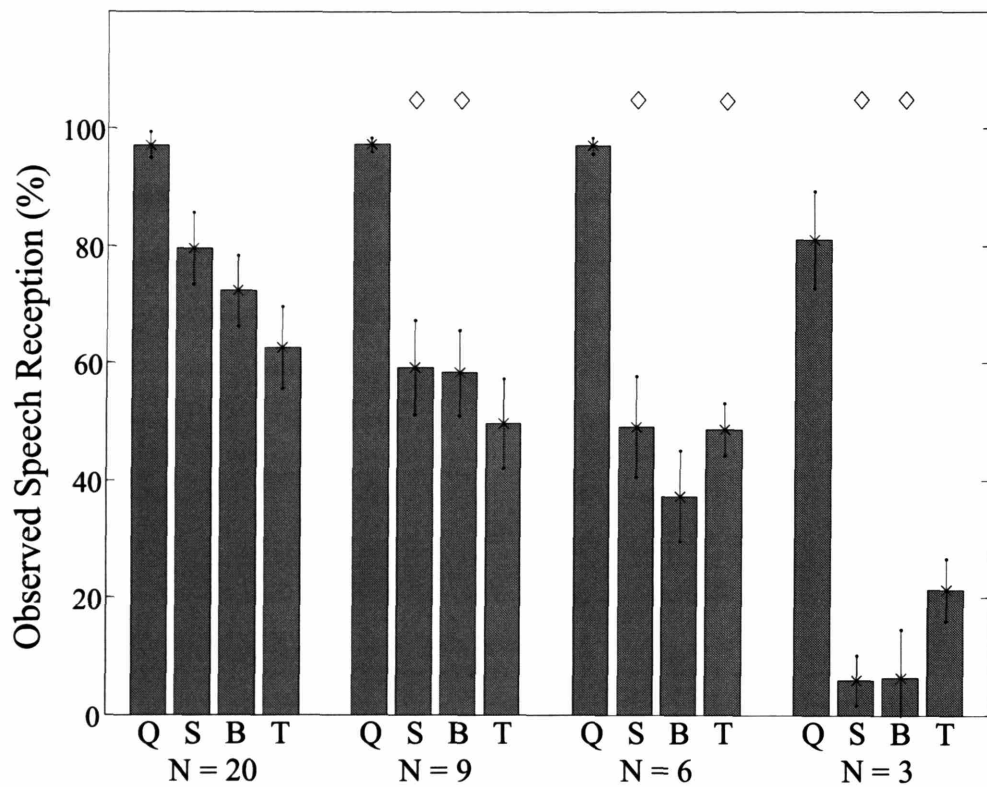


Figure 7.3: Same data as Figure 7.2 but arranged to emphasize the effect of noise type. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$). Abbreviations: quiet (Q), speech-shaped noise (S), multi-talker babble (B), and time-reversed speech (T).

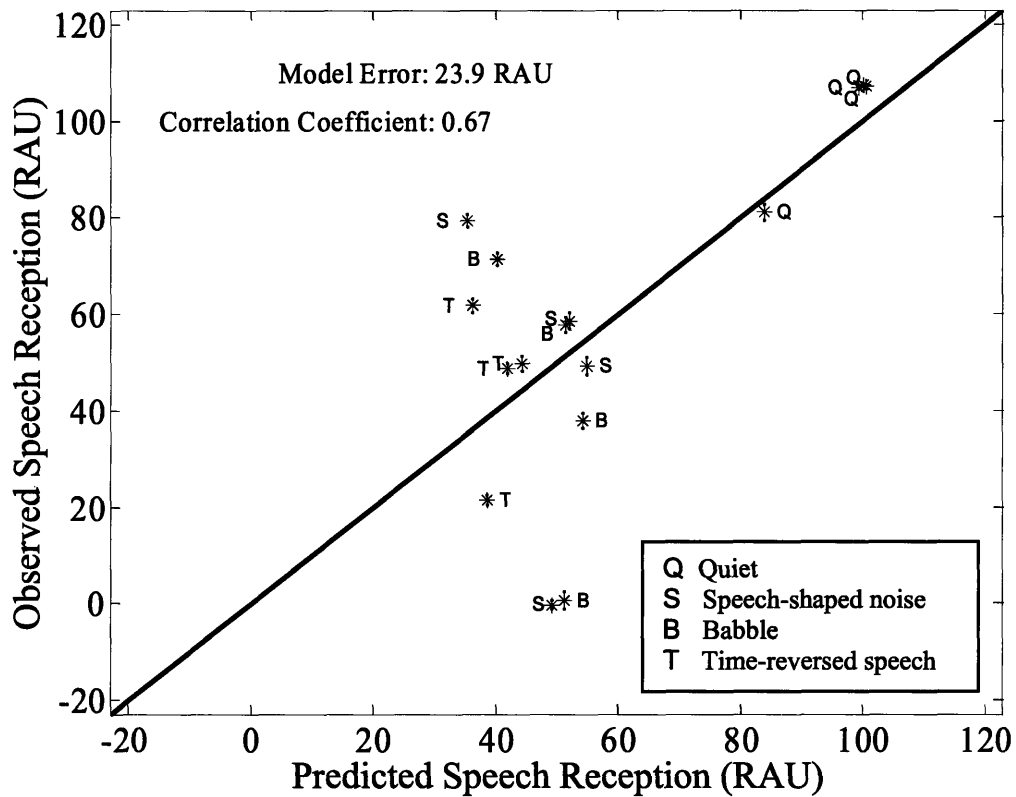


Figure 7.4: Comparison of observed scores for $NH-CI_{Sim}$ and predicted scores from the envelope-regression STI method. The error bars represent \pm one standard error of the mean.

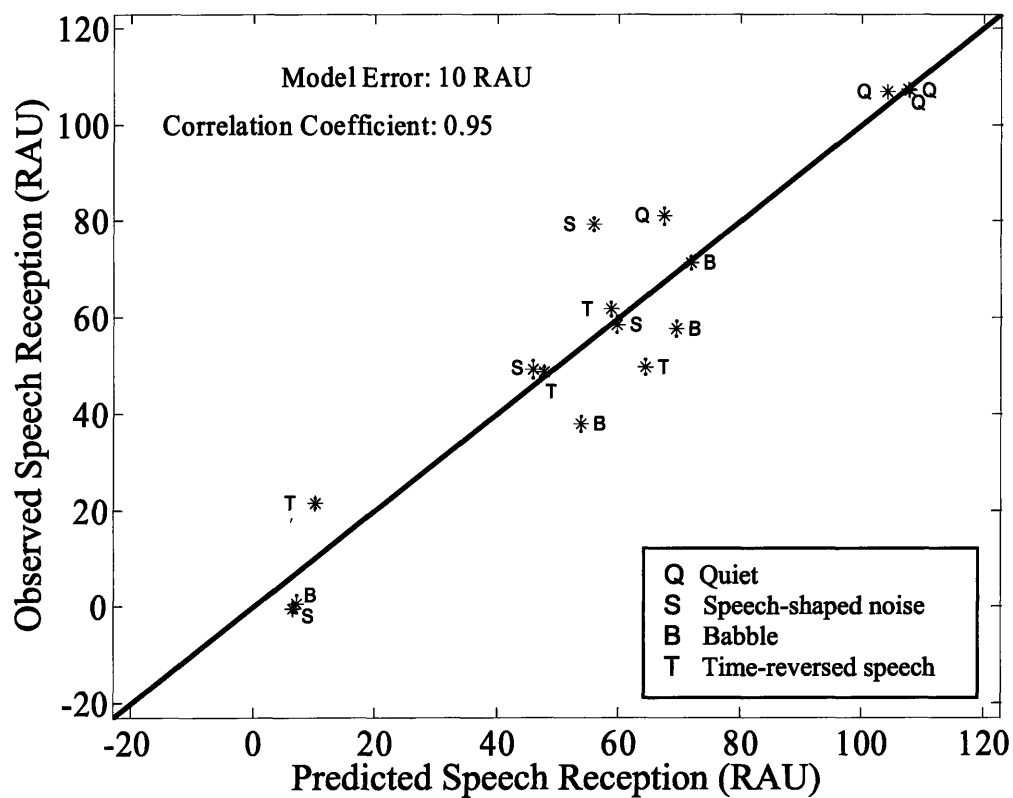


Figure 7.5: Comparison of observed scores for NH-CI_{Sim} and predicted scores from the modified envelope-regression STI method. The error bars represent \pm one standard error of the mean.

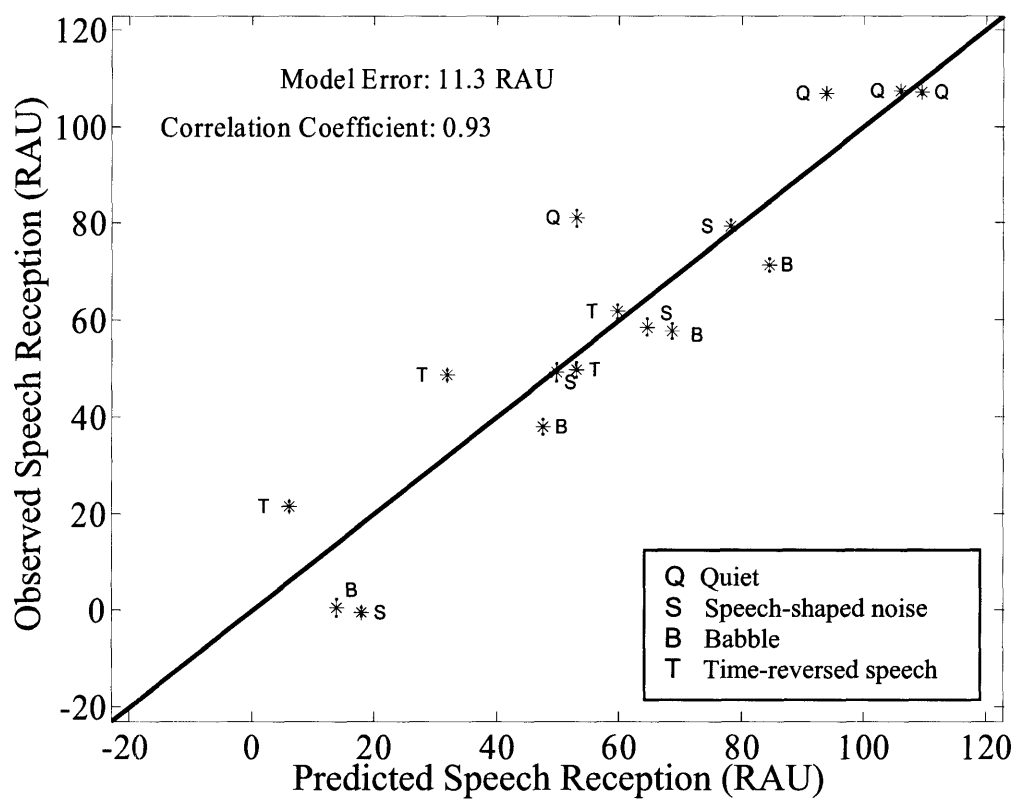


Figure 7.6: Comparison of observed scores for NH-CI_{Sim} and predicted scores from the NCM method. The error bars represent \pm one standard error of the mean.

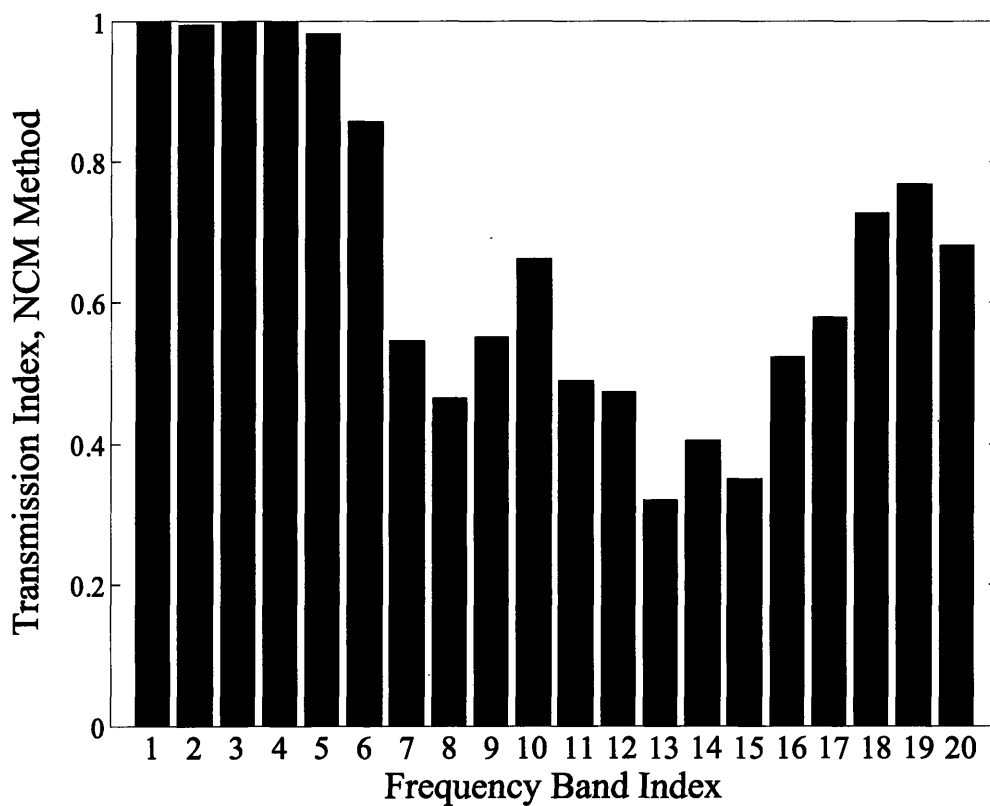


Figure 7.7: TI values for the NCM method (Eq. 3.5) calculated as intermediate variables for the 6-of-20 condition in quiet. The TI values are calculated as intermediate metrics in the NCM calculation and are based on the same clean and degraded material as the NCM data presented in Figure 7.6.

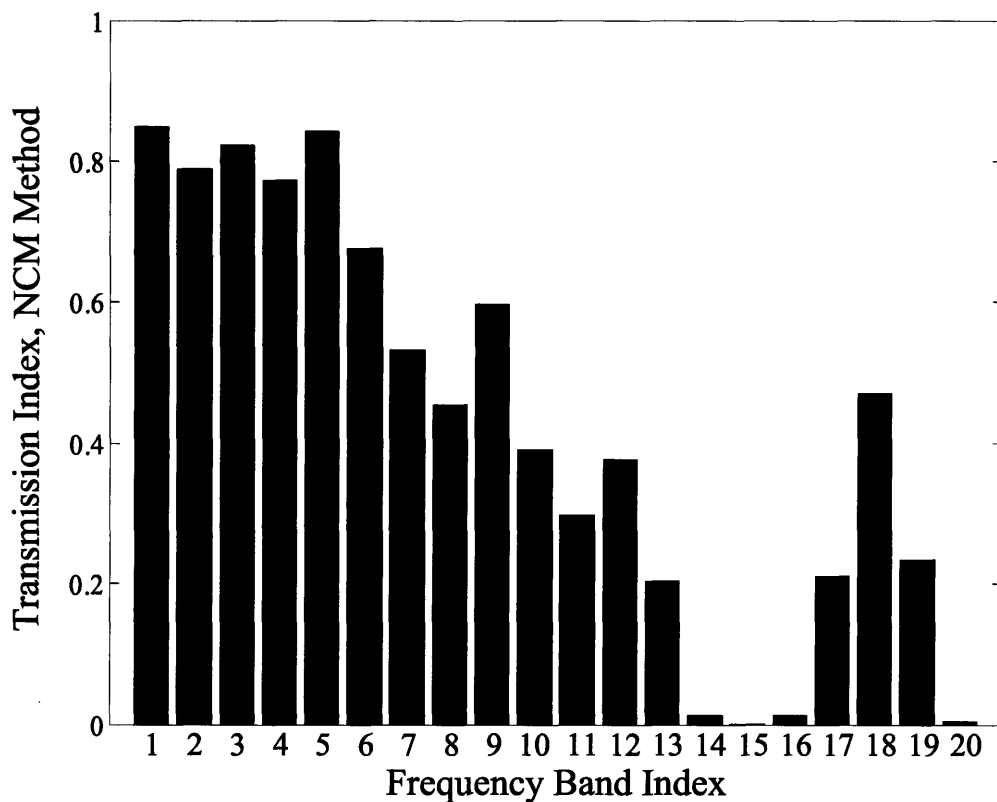


Figure 7.8: TI values for the NCM method calculated as intermediate variables for the 6-of-20 condition in speech-shaped noise (0 dB). The TI values are calculated as intermediate metrics in the NCM calculation and are based on the same clean and degraded material as the NCM data presented in Figure 7.6.

Chapter 8

Experiment 3: Spectral Subtraction

Spectral subtraction is a noise reduction algorithm that has been studied for normal-hearing listeners and for CI users. Previous studies suggest that spectral subtraction does not improve speech reception in noise for normal-hearing subjects. In contrast, there is evidence that spectral subtraction does improve speech reception in noise for CI users. The experiment presented in this chapter is designed to evaluate the effects of spectral subtraction on speech reception in noise for CI-processed speech. A generalized form of spectral subtraction is investigated to allow for control of the level of noise removal. Subjects include NH-CI₈, NH-CI₂₀ and actual CI users. The results clearly indicate that spectral subtraction improves speech reception in noise for all subjects tested. Further, the STI and the NCM are investigated as predictors of intelligibility for the processed speech. The unmodified STI method does not produce reasonable predictions for these conditions; however, the modified STI and NCM methods do produce reasonable predictions. The NCM, in particular, produces accurate predictions. Use of the metrics to determine optimal values of a control parameter is discussed. An explanation of why spectral subtraction improves speech reception for CI users is also discussed.

8.1 Introduction

Spectral subtraction is a single-microphone noise-reduction strategy (reviewed in Section 2.3.1). The primary application for spectral subtraction is the suppression of stationary noise in a degraded signal. A generalized form of spectral subtraction containing two control parameters was introduced in Equation 2.14. For the experiment presented in this chapter, we set the control parameter α equal to one and optimize the algorithm for the control parameter κ . Thus, the corresponding frequency domain equation of interest is

$$|P(F, n)| = |D(F, n) - \kappa \hat{N}(F)|, \quad (8.1)$$

where $P(F, n)$ is the estimated speech spectrum of the n^{th} segment, $D(F, n)$ is the degraded speech spectrum, and $\hat{N}(F)$ is the estimated noise spectrum. The phase information is retained such that the phase of the output signal is the same as the input (degraded speech) signal. The parameter, κ , allows the strength of the noise suppression to be controlled.

Investigation of spectral subtraction will prove insightful into a number of areas in our research. Our research interests focus on noise reduction strategies, intelligibility metrics, and how those two areas interact with CI sound-processing. It has been clearly shown in the past that spectral subtraction does not improve speech reception in noise for normal-hearing listeners (Lim and Oppenheim, 1979); in contrast, mounting evidence suggests that spectral subtraction does improve speech reception in noise for CI-processed speech (Weiss, 1993, Hochberg et al., 1992). Further, investigations of the STI indicate that STI predicts that spectral subtraction should improve speech reception in noise (Ludvigsen et al., 1990, 1993). Ludvigsen argued that this represented a failure of STI since he found no intelligibility gains in normal-hearing subjects.

Previous studies have not investigated if STI could serve as an accurate predictor of speech reception for CI users. In Chapter 6, it was mentioned that CI users are more sensitive to the effects of noise and to noise source modulations. Another difference exists that suggests that STI may actually be a better model for CI users than normal-hearing listeners. STI predicts that the spectral subtraction noise reduction algorithm (see Chapter 5) should improve intelligibility.

Our contention is that the STI is better suited to predict intelligibility gains for CI users than for normal-hearing listeners. As such, the STI and the NCM need to be investigated with respect to spectral subtraction and CI sound-processing to evaluate if the predicted gains are quantitatively accurate.

8.2 Conditions

The problem addressed in this chapter is illustrated in Figure 8.1. Clean speech is first

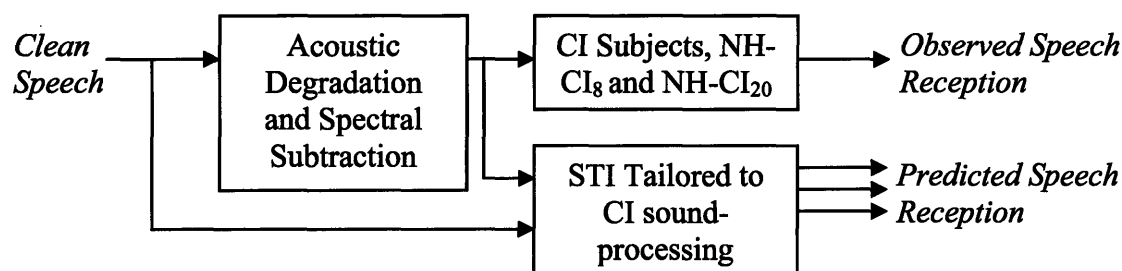


Figure 8.1: Block diagram of the experimental procedure for spectral subtraction conditions.

acoustically degraded and then processed through the spectral subtraction algorithm. The resulting signal is delivered to either a CI subject or a normal hearing subject listening to the vocoder simulation of CI sound-processing. The clean and degraded signals are used to calculate the various intelligibility metrics and the corresponding predicted speech reception.

16 conditions were selected to answer the following questions:

- 1) Does spectral subtraction improve speech reception in noise for CI-processed speech?
- 2) Are speech-reception gains from the spectral subtraction algorithm dependent on the number of channels in the CI processor?
- 3) What is the optimal value for the control parameter κ ?
- 4) Do any of the candidate metrics predict the effects of spectral subtraction on the intelligibility of speech in noise?

Eight normal-hearing listeners were tested on speech processed with NH-CI₈ and NH-CI₂₀. Clean speech was degraded by additive speech-shaped noise at 0 dB for the 8-channel condition and -3 dB for the 20-channel condition. Eight values of κ were selected from 0 (no processing) to 8. The conditions are summarized in Table 8.1.

The experiment was divided into three trials that were tested on three separate days. Each trial consisted of the 16 conditions each tested using one complete list from the CUNY database. The conditions were partially counterbalanced across subjects for 4 groups (2

subjects in each group) corresponding to the two columns in Table 8.1 divided between $\kappa = 1.26$ and 1.59. The conditions within these groups were counterbalanced across trials (with a 4th ‘null’ trial serving only for counterbalancing purposes). Details of the experimental methods are given in Chapter 4.

| NH-CI ₈ (SNR = 0) | NH-CI ₂₀ (SNR = -3) |
|---------------------------------|-----------------------------------|
| $\kappa = 0$ | $\kappa = 0$ |
| $\kappa = 0.5$ | $\kappa = 0.5$ |
| $\kappa = 1$ | $\kappa = 1$ |
| $\kappa = 1.26$ | $\kappa = 1.26$ |
| $\kappa = 1.59$ | $\kappa = 1.59$ |
| $\kappa = 2$ | $\kappa = 2$ |
| $\kappa = 4$ | $\kappa = 4$ |
| $\kappa = 8$ | $\kappa = 8$ |

Table 8.1: Summary of experimental conditions for spectral subtraction.

8.3 Results of the Listening Experiment

8.3.1 NH-CI₈ and NH-CI₂₀ Subjects

The subjects’ responses were scored as percentage of words correct for each trial. Figure 8.2 illustrates the subject scores for each condition averaged across subjects and trials. The data was divided into two groups corresponding to NH-CI₈ and NH-CI₂₀ results. Speech reception as a function of κ was similar for the 8 and 20-channel simulations: both monotonically increase to 1.59 and then monotonically decrease with the exception $\kappa = 1.26$ for the 8-channel simulation.

An initial repeated measures analysis of variance (RMANOVA_1)¹⁴ was performed using trials as the repetition variable. The dependent variable was the speech reception score transformed to RAU and subject and condition were main factors. Subject is a significant factor. The lowest average subject score was 57.1 RAU, and the

¹⁴ All variance and post-hoc measures are calculated in Matlab[®] in accordance with Winer et al. (1991).

highest was 68.7 RAU. The interaction between subject and condition was not significant ($p > 0.1$). Thus, the trends observed for the different conditions were consistent across subjects.

A second repeated measures analyses of variance (RMANOVA_2) was performed using the speech reception score transformed to RAU as the dependent variable, and subject, number of channels in the NH-CI_{Sim}, and control parameter value, κ , as main factors. RMANOVA_2 indicates that the effect of the number of channels in the NH-CI_{Sim} was moderately significant ($p = 0.014$). It was also found that κ was significant. The interaction between subject and κ was moderately significant ($p = 0.029$). The interaction between the number of channels in the simulation and κ was significant ($p < 0.001$).

The *post hoc* analysis of κ values was implemented according to Tukey's HSD ($\alpha = 0.05$). For both NH-CI₈ and NH-CI₂₀, the values of κ between 0.5 and 2 all significantly improved speech-reception scores compared to no processing ($\kappa = 0$). For NH-CI₈, scores were highest for $\kappa = 1.59$; however, scores for κ values of 1 and 2 were not significantly lower. For NH-CI₂₀, scores were highest for $\kappa = 1.59$; however, scores for κ values between 1 and 2 were not significantly lower. In general, an optimal parameter range of κ values between 1 and 2 exists; however, the variance in scores was too great to indicate an exact optimal value within this range. For both NH-CI₈ and NH-CI₂₀, the κ values of 4 and 8 were found to significantly decrease scores compared to the optimal parameter range.

8.3.2 CI Subjects

Three CI subjects—one Clarion (8 channel) and two Nucleus (22 channels)—participated in this experiment. The CI subjects were tested using a similar set of conditions as those summarized in Table 8.1. However, the SNR of each condition was shifted by a certain amount, Δ , in order to compensate for individual performance differences. The process for determining Δ for each subject is given in Section 4.4.2. Table 4.1 summarizes the Δ values found for each subject. Figure 8.3 illustrates the subject's scores for each condition averaged across subjects and trials.

A repeated measures analysis of variance was performed similar to RMANOVA_1 but using the CI data. This analysis indicates that subject and parameter value, κ , were both significant. The average scores for the three subjects are 60.6, 49.2, and 56.5 RAU (respectively for CI-4, CI-5, and CI-6). The interaction between subject and κ was not found to be significant. Figure 8.4 illustrates individual scores for the three CI users tested.

The *post hoc* analysis of κ values was implemented according to Tukey's HSD ($\alpha = 0.05$). The results were very similar to the NH-CI_{sim} results. Speech reception scores were significantly higher than no processing for κ values between 0.5 and 2. The highest average speech reception score occurred for $\kappa = 2$; however, κ values between 0.5 and 2 did not produce significantly different results. Thus, an optimal parameter range was determined to be between 0.5 and 2.

8.4 Results of the Intelligibility Predictions

8.4.1 NH-CI₈ and NH-CI₂₀ Subjects

The procedure for calculating particular metrics from the clean and degraded speech waveforms is detailed in Section 4.5. As discussed in Section 5.3, we have selected the envelope-regression STI method, the modified envelope-regression STI method, and the NCM method for further investigation. As described in Section 5.2, it is possible for the original envelope regression method to fail by producing invalid values of the intermediate metric. In particular, when the modulation metric (Eq. 3.2) is outside the range between 0 and 1, then the apparent SNR (Eq. 3.1) is a complex—in the mathematical sense—number and cannot be interpreted in the existing STI framework. In this chapter, we avoid this problem by clipping the modulation metric (Eq. 3.2) to values between 0 and 1.

The metrics are calculated for the conditions tested and then a psychometric function is fit to the mapping between metric value and the mean reception scores. The resulting psychometric function thus yields a predicted score (in RAU) for a given metric value. Figures 8.5, 8.6 and 8.7 illustrate the comparison between observed scores for NH-CI₈ and predicted scores for the respective methods.

Two measures are given for assessing the predictions made by the different intelligibility metrics: 1) the model error defined as the standard deviation between predicted and observed scores and 2) the correlation coefficient between predicted and observed scores. The unmodified envelope regression STI method poorly predicts speech reception scores for these spectral subtraction conditions as seen by the high model error, 29.3 RAU, and the low correlation coefficient, 0.04. The modified envelope regression STI method produces reasonable predictions as seen by the low model error (10.2 RAU) and high correlation coefficient (0.92). The NCM method produces the most accurate predictions as quantified by the lowest model error (5.04 RAU) and highest correlation coefficient (0.99).

While both the modified envelope regression STI and the NCM methods produce reasonable speech reception predictions, certain trends need to be highlighted. As mentioned above, the mean scores for the four κ values ranging from 1 to 2 are not significantly different. The interpretation of the psychometric curve can be facilitated by dividing the conditions into five groups: no processing ($\kappa = 0$), mild processing with $\kappa = 0.5$, the optimal performance range ($1 \leq \kappa \leq 2$), a moderately high processing with $\kappa = 4$, and high processing ($\kappa = 8$).

Considering these groupings, we see that the modified envelope regression STI method (Figure 8.6) fails to capture key trends. In particular, the speech reception predictions are approximately the same for the optimal performance range, the $\kappa = 0.5$ range, and the no processing range. In other words, the modified envelope regression fails to allow the optimal parameter range to be predicted. This failure is paramount since our interest in developing an intelligibility metric for noise reduction operations is motivated by our desire to use the metric to optimize performance.

The NCM method, in contrast, does predict the optimal parameter range. In Figure 8.7, the data points corresponding to this optimal range are tightly clustered near 80 RAU for both observed and predicted scores. Thus, there is a clear distinction between the optimal range and the other conditions.

A smaller trend within the NCM predictions should be mentioned. It should be noted that the predicted speech reception for the no processing and $\kappa = 4$ region are approximately the same, even though speech reception is on the average 20 RAU less for

the $\kappa = 4$ region. In other words, in terms of the trade off between noise removal and introduction of signal distortion, there is a slight bias to underestimate the detriment of signal distortion to speech reception.

8.4.2 CI Subjects

The psychometric function was fitted for the NH-CI₈ data based on the mean subject scores. However, for actual CI users, we expect a wider variance in observed scores. It is possible that a particular subject may not be able to score 100% in quiet. To compensate for this potential difference, the psychometric function is fit to each subject and allowing R_{max} of Equation 4.10 to vary. The added degrees of freedom in the model are taken into account in the calculation of the model error and the correlation coefficient.

Figures 8.8, 8.9 and 8.10 illustrate the comparison between observed and predicted scores for the three candidate methods. The accuracy of the speech reception predictions is comparable to the NH-CI_{Sim} results: the unmodified envelope regression STI method produces grossly inaccurate predictions while the other two methods produce reasonable predictions. We use a similar classification of ranges used in the previous section; however, one exception is that the $\kappa = 0.5$ condition is grouped with the optimal range since the associated speech reception scores were not significantly different. The groupings are: no processing ($\kappa = 0$), the optimal performance range ($0.5 \leq \kappa \leq 2$), a moderately high processing with $\kappa = 4$, and high processing ($\kappa = 8$). The results are comparable to the NH-CI_{Sim} case with only the NCM method accurately predicting the optimal range of parameter values.

8.5 Frequency-Band Analysis

The analysis of TI values across frequency bands may prove insightful for the spectral subtraction algorithm. As with the N-of-M strategy, spectral subtraction is a nonlinear operation that may have effects that vary across frequency bands. Investigating how the TI values differ across frequency bands may provide insight as to how well the spectral subtraction algorithm performs in different frequency regions.

The TI values calculated using the NCM method for the 20-channel condition with $\kappa = 0$ and 1.26 are presented in Figure 8.11. The TI values for $\kappa = 1.26$ are

generally greater than for the no processing condition. The highest four frequency bands are exceptions. There is only a small difference in TI values for the fourth highest frequency band. The TI values are actually lower for the processed condition in the three highest frequency bands. Further analysis of these bands suggests that other values of κ yield higher TI values. Figure 8.12 illustrates the TI values for all κ values for the three highest bands. We see that for frequency bands 18 and 19, the TI values are highest for $\kappa = 2$. Thus, insofar as the overall metric is indicative of intelligibility, we might select κ corresponding to the highest TI value for each frequency band in order to maximize speech reception.

8.6 Discussion

The unmodified envelope regression STI did not produce reasonable predictions of speech reception for spectral subtraction, while both the modified envelope regression STI and the NCM methods did. However, only the NCM method accurately predicted the range of optimal parameter settings. In this discussion we first consider a possible explanation for why spectral subtraction improves speech reception for CI users but not for normal-hearing listeners, and then discuss possibilities for using the NCM metric to optimize spectral subtraction.

8.6.1 CI Specificity of the Results

As mentioned in Section 8.1, spectral subtraction does not improve speech reception for normal-hearing listeners. In contrast, other studies have shown that spectral subtraction does improve speech reception for CI users. Our study clearly shows that spectral subtraction improves the intelligibility of CI-processed speech. We hypothesize that this discrepancy is because the algorithm operates using spectral information that the CI user does not have access to, but that normal-hearing listeners do.

The process of coding speech information for CI stimulation reduces the information present in the signal. One fundamental way that the signal information is reduced is that the spectral resolution is limited. The bandpass filters used in the CI sound-processing strategy (see Figure 2.2) limit the spectral resolution of CI-processed speech. Thus, if speech and noise exist in the same band, then they will be combined in

the envelope signal that is used to modulate the electric stimulation. The bandwidth of a normal-hearing auditory system is much narrower than the corresponding CI bandpass filters. Furthermore, the CI bandpass filters are non-overlapping in contrast to overlapping filters in the normal auditory system. The normal auditory system is therefore privy to higher resolution when analyzing the input signal.

This access to higher resolution frequency information is precisely why we contend that the STI and NCM models are better suited for predicting the intelligibility of CI-processed speech than for unprocessed speech. These models are based on non-overlapping frequency bands that can be tailored to fit the bandpass filters used in CI sound-processing. To have a more accurate model for normal-hearing listeners, the front-end of the model would have to include overlapping filters with higher resolution. Rules might be specified for such a model prescribing how overlapping filters are combined to determine an overall metric value. In any case, for the normal hearing model to be accurate, it would have to predict no speech reception gains for spectral subtraction.

8.6.2 Optimizing Spectral Subtraction using the NCM

The NCM method accurately predicts the speech reception trends for spectral subtraction for both the 8 and 20 channel processing conditions. This method clearly isolated a range of κ values corresponding to optimal performance. The NCM value as a function of κ is illustrated in Figure 8.13. The NCM method predicts a global maximum near $\kappa = 1.7$. Unfortunately, the variance of the mean intelligibility scores was too high to determine if this global maximum corresponds to a speech reception maximum.

Nevertheless, the NCM method can be used as a tool to isolate an optimal range of parameters. Another parameter worth investigating is the window length used to parse the signal since this parameter determines the frequency resolution implemented in the algorithm. Figure 8.14 illustrates the NCM score as a function of window length. The NCM value increases as the window size increases, reaching a maximum at 51 ms, and then decreases. The decrease in NCM value for windows greater than 51 ms can be attributed to smearing information across phoneme boundaries. A comparable study to the one presented in this chapter could be formulated to investigate if the NCM predictions of Figure 8.14 correspond to speech reception.

Further evaluation of the NCM method is required to assess its ability as a tool for algorithm optimization. We have shown that the NCM successfully predicts an optimal range of parameters for the spectral subtraction parameter κ . However, we cannot blindly assume that this ability will carry over to other parameters or to other algorithms. We suggest that the NCM method be used as a guide in selecting parameter values; but at the same time, testing a range of parameters in order to verify the predictions. In this manner, the predictive power of the metric can be evaluated for other parameters and other algorithms.

8.7 Conclusions

The main conclusions of this chapter are:

- (1) Spectral subtraction improves the intelligibility of CI-processed speech in the presence of stationary background noise.
- (2) Speech reception gains are seen for both 8 and 20-channel CI sound-processing strategies for a range of optimal parameters.
- (3) The original speech-based STI methods do not produce reasonable speech-reception predictions for N-of-M processing.
- (4) The modified speech-based STI method produces reasonable predictions; however, it does not isolate an optimal range of κ values for spectral subtraction.
- (5) The NCM method produces reasonable predictions and also isolates an optimal range of κ values for spectral subtraction.
- (6) We suggest using the NCM method as a guide for selecting a range of optimal parameter values for different parameters and different algorithms so long as the NCM predictions are verified in the process.

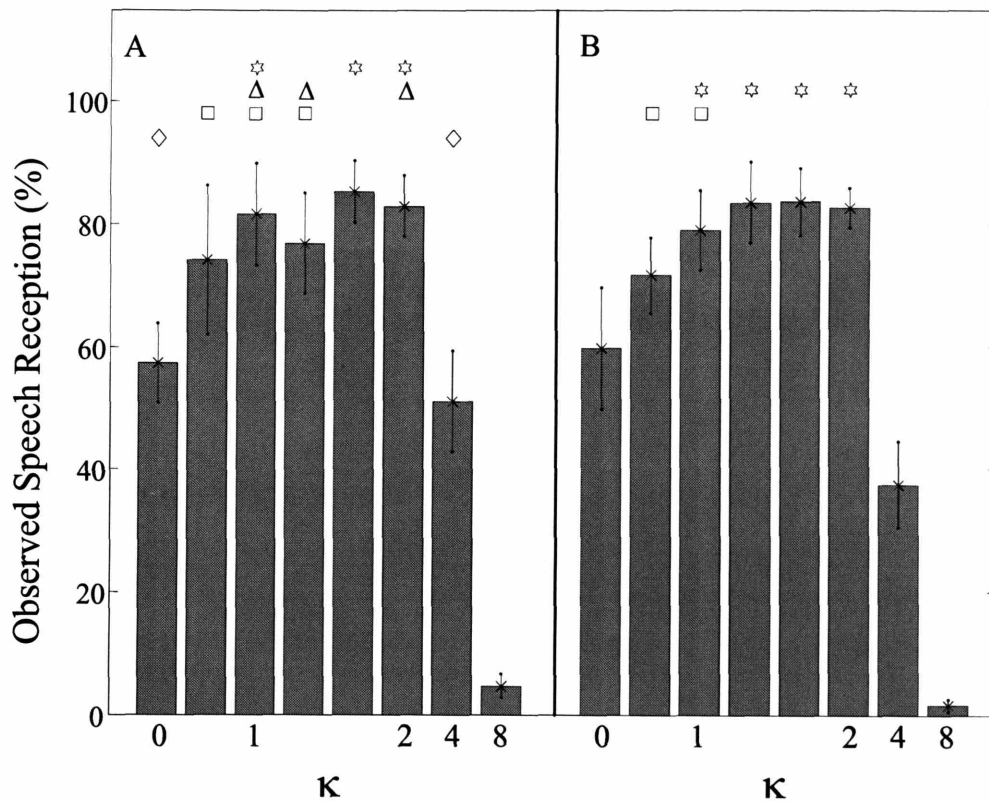


Figure 8.2: NH-Cl₈ and NH-Cl₂₀ scores for spectral subtraction conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$). The two subplots represent results from A) NH-Cl₈ and B) NH-Cl₂₀.

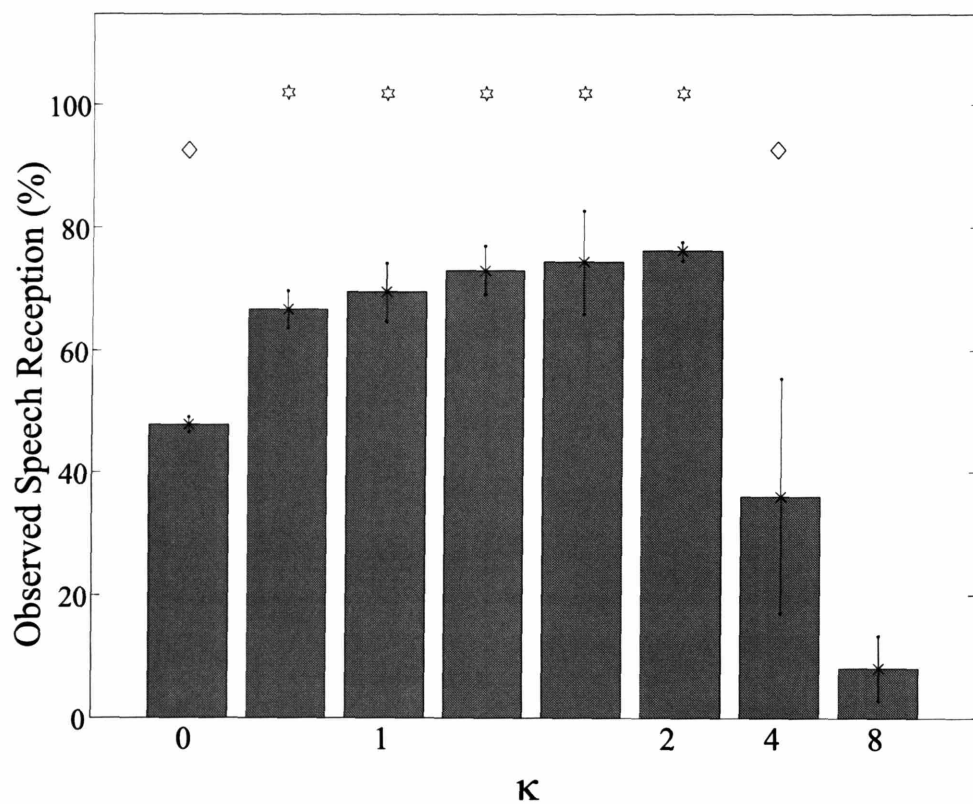


Figure 8.3: Speech reception scores for CI users tested on spectral subtraction conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$).

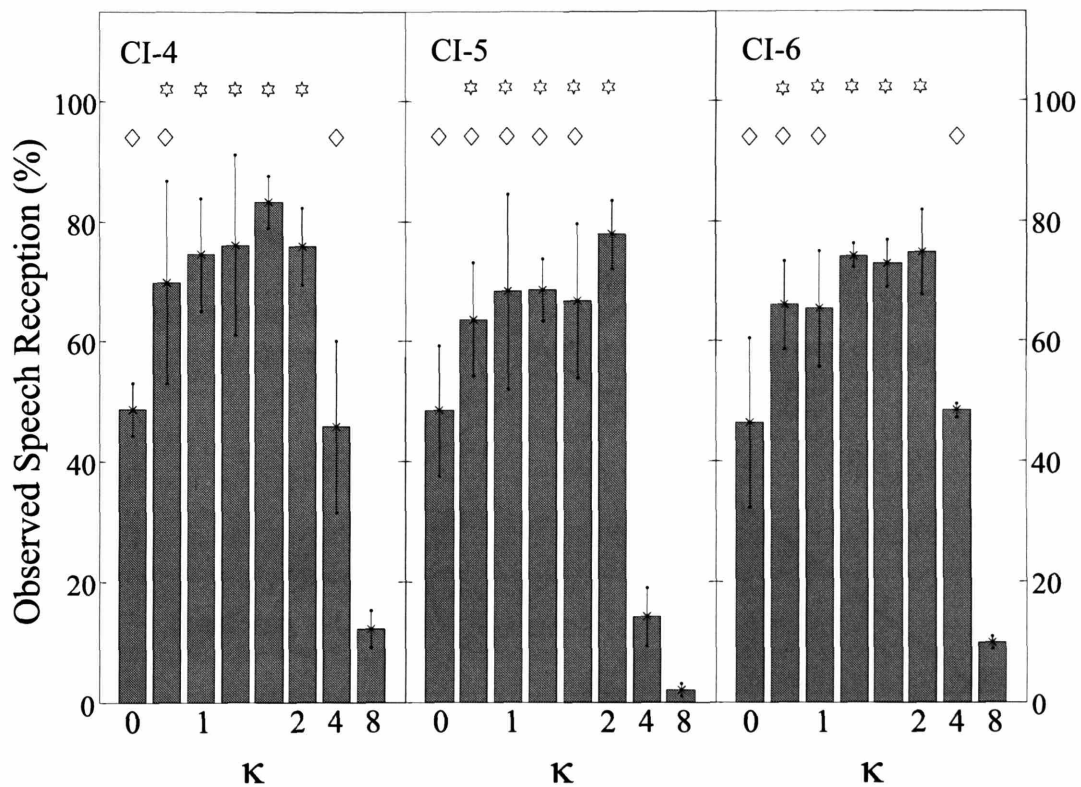


Figure 8.4: Individual speech reception scores for CI users tested on spectral subtraction conditions. The bars represent the mean scores averaged across trials for each subject. The error bars represent \pm one standard deviation of the mean. For each set of bars, conditions with the same symbols above the bars were not significantly different according to a *post hoc* Tukey HSD test ($p > 0.05$).

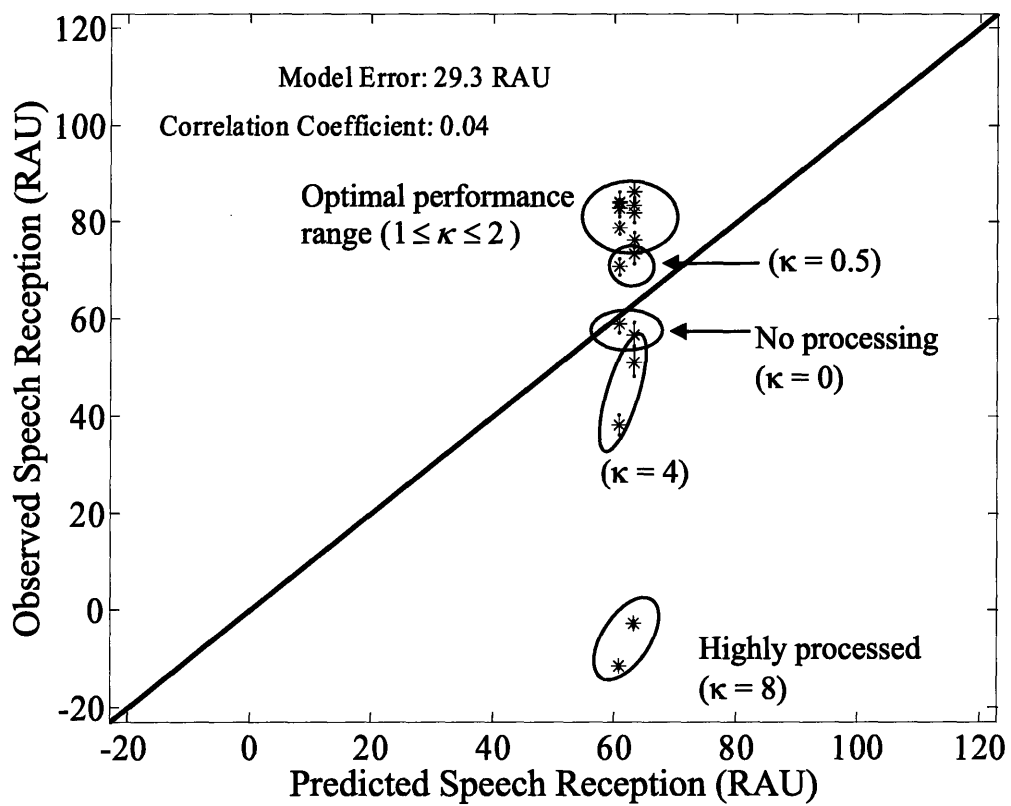


Figure 8.5: Comparison of observed scores for NH-CI_{Sim} and predicted scores from the envelope-regression STI method. The error bars represent \pm one standard error of the mean.

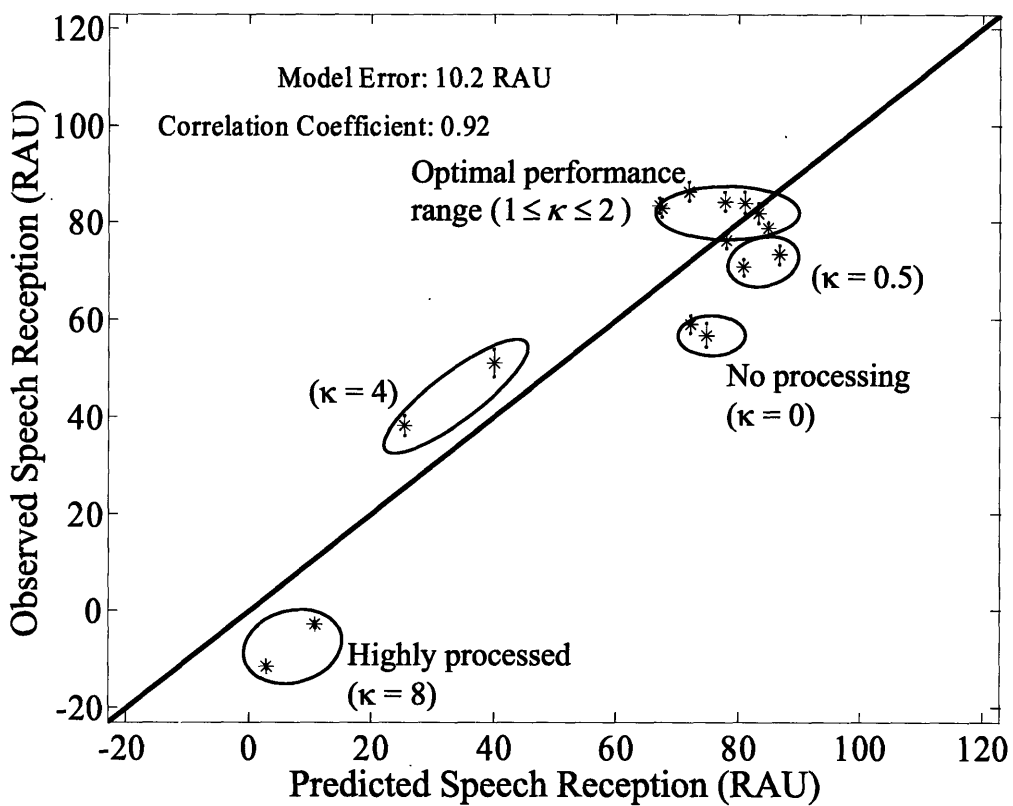


Figure 8.6: Comparison of observed scores for NH-CI_{sim} and predicted scores from the modified envelope-regression STI method. The error bars represent \pm one standard error of the mean.

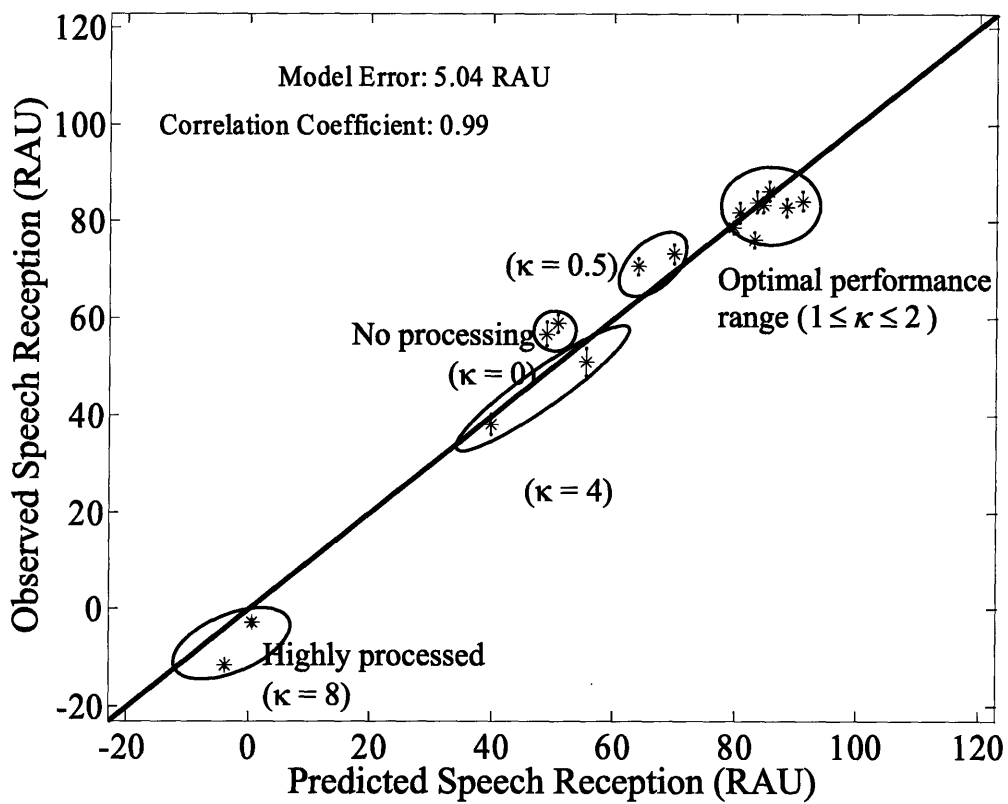


Figure 8.7: Comparison of observed scores for NH-Cl_{Sim} and predicted scores from the NCM method. The error bars represent \pm one standard error of the mean.

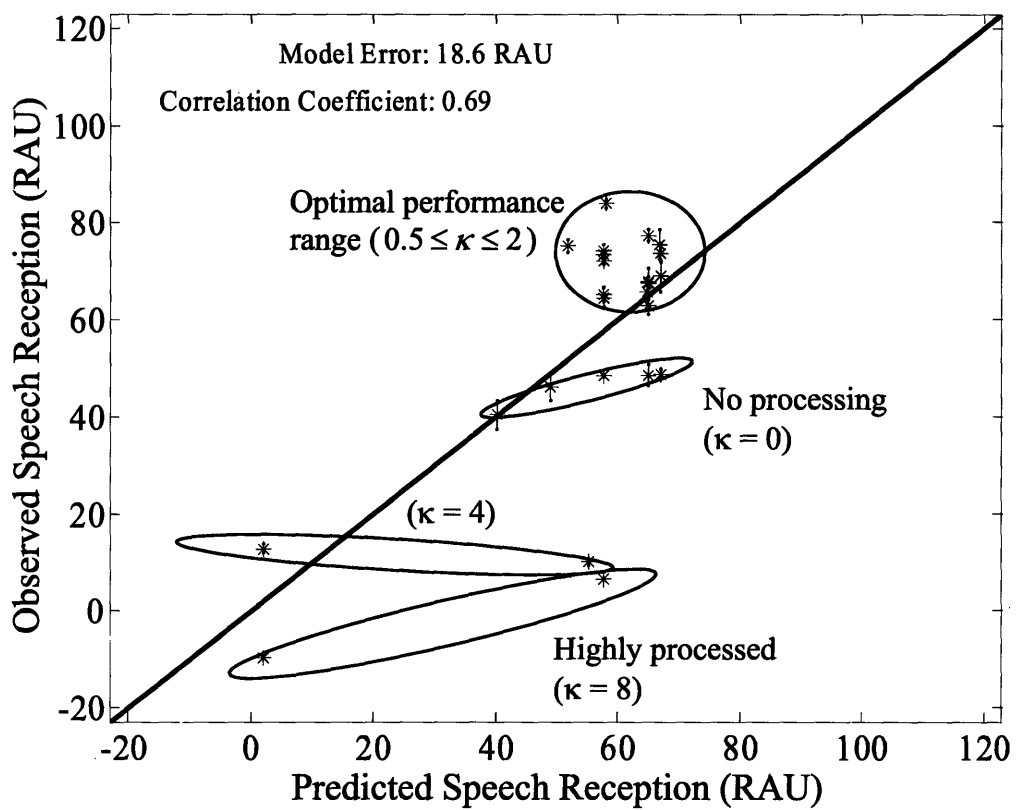


Figure 8.8: Comparison of observed scores for CI users and predicted scores from the envelope-regression STI method. The error bars represent \pm one standard error of the mean.

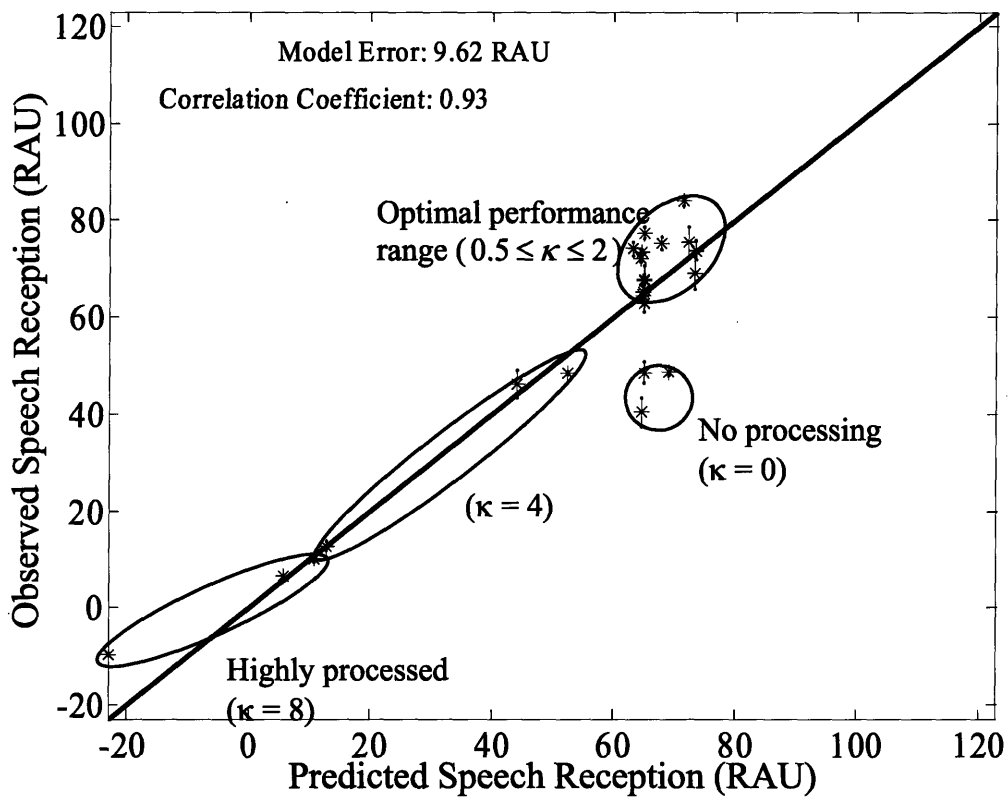


Figure 8.9: Comparison of observed scores for CI users and predicted scores from the modified envelope-regression STI method. The error bars represent \pm one standard error of the mean.

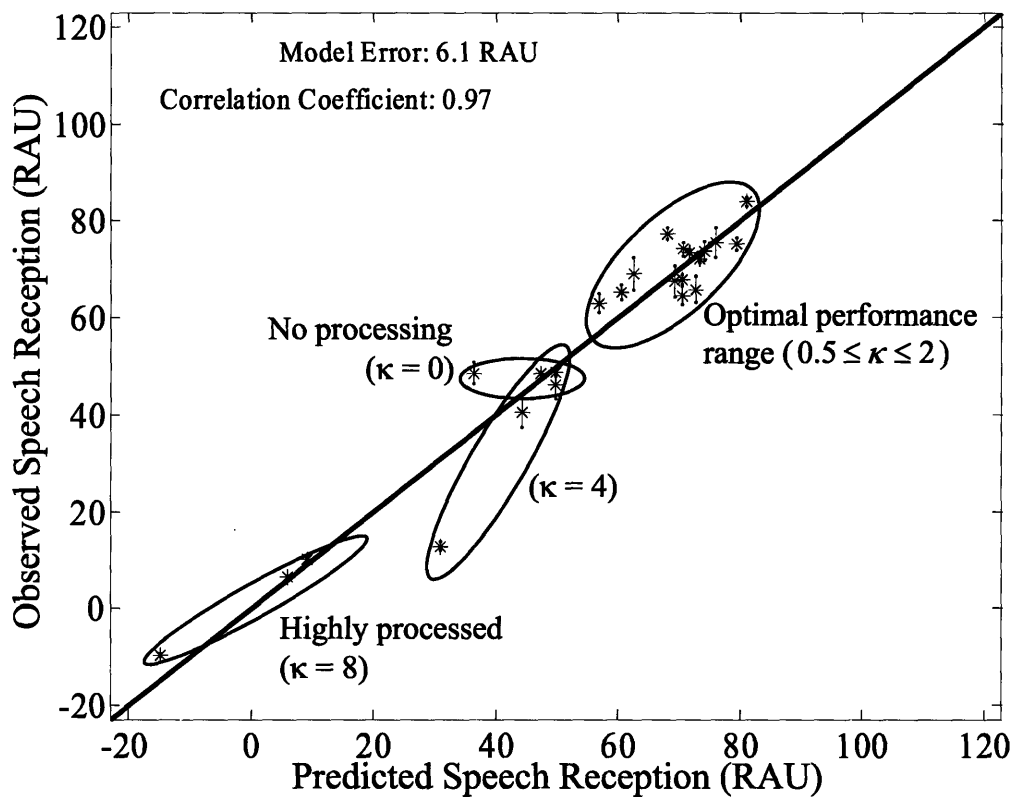


Figure 8.10: Comparison of observed scores for CI users and predicted scores from NCM method. The error bars represent \pm one standard error of the mean.

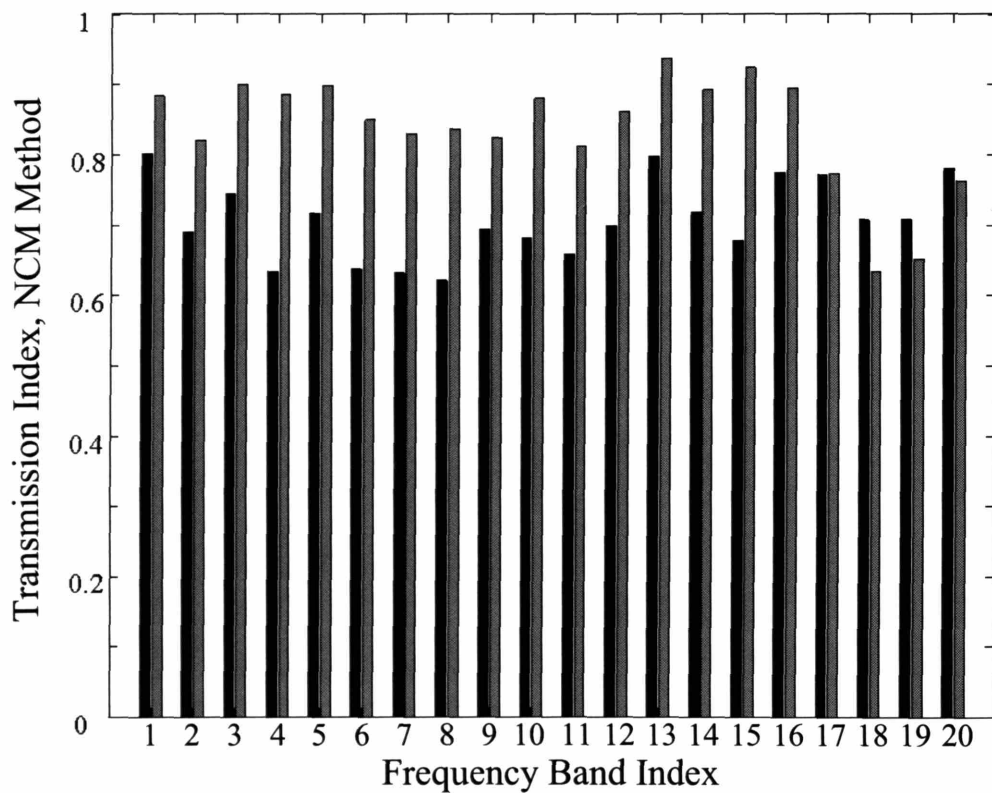


Figure 8.11: TI values for the NCM method (Equation 3.5). Each pair of bars corresponds to TI values for a given frequency band for the 20-channel system. The left and right bars in each pair correspond to TI values with the spectral subtraction algorithm off and on ($\kappa = 1.26$), respectively. The TI values are calculated as intermediate metrics in the NCM calculation and are based on the same clean and degraded material as the NCM data presented in Figure 8.10

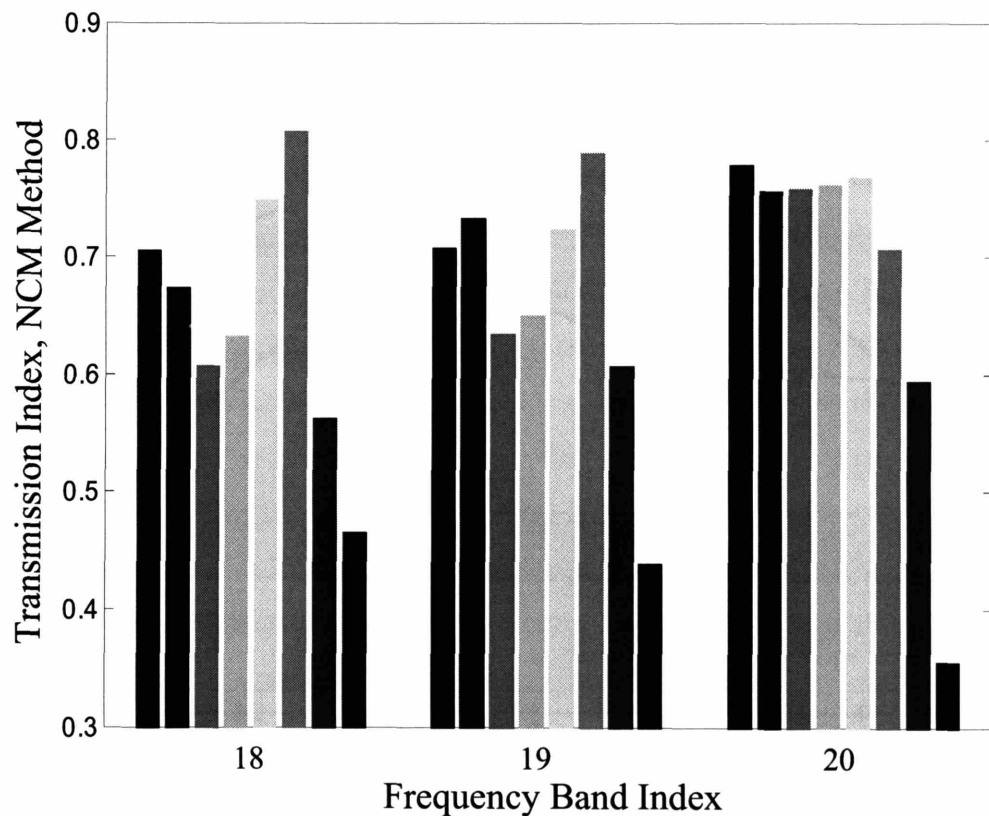


Figure 8.12: TI values for the NCM method (Eq. 3.5). Each set of eight bars correspond to TI values for one of the three highest frequency bands of the 20-channel system. Each bar within a set corresponds to a different value of the control parameter, κ , (see Table 8.1) ranging from 0 to 8 (ordered left to right). The TI values are calculated as intermediate metrics in the NCM calculation and are based on the same clean and degraded material as the NCM data presented in Figure 8.10.

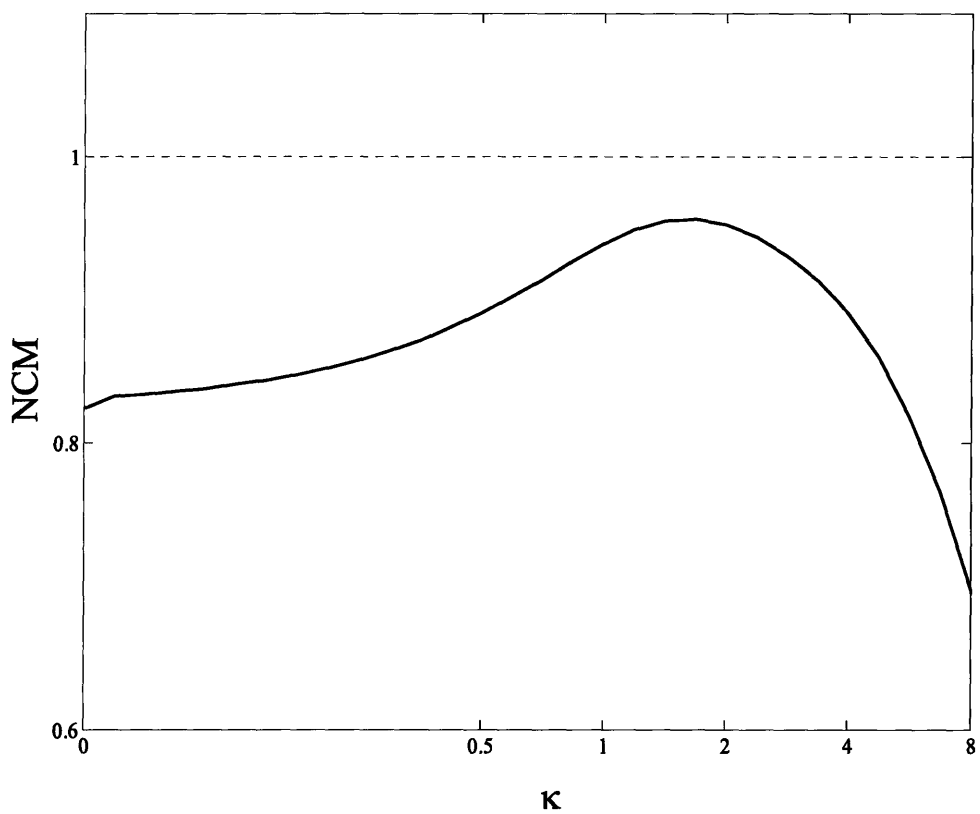


Figure 8.13: NCM as a function of κ for the 8-channel analysis in speech-shaped noise (0 dB).

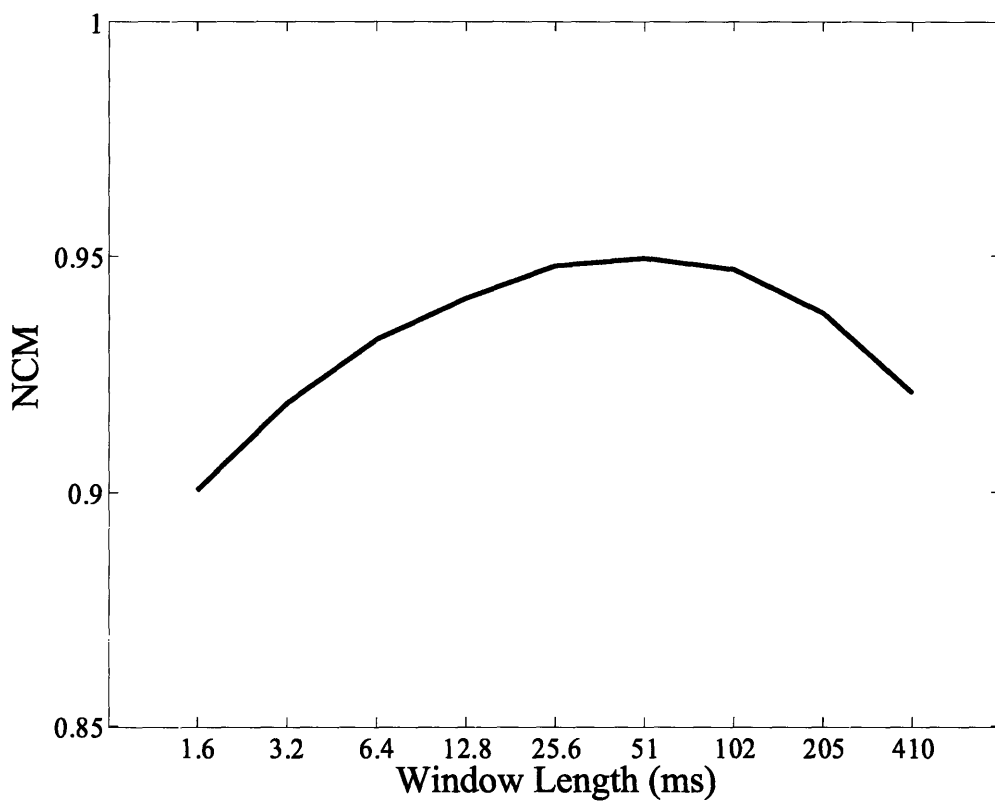


Figure 8.14: NCM as a function of window length for the spectral subtraction algorithm for the 8-channel analysis calculated for speech-shaped noise (0 dB) and using $\kappa = 1$.

Chapter 9

Experiment 4: Binaural Noise Reduction

Binaural noise reduction algorithms are based on comparisons of inter-microphone level and phase differences. The experiment presented in this chapter is designed to evaluate the effects of binaural noise reduction on speech reception in noise for CI-processed speech. Subjects include NH-CI₈, NH-CI₂₀ and actual CI users. The results clearly indicate that binaural noise reduction improves speech reception in noise for CI-processed speech for both 8 and 20-channel processors. Further, the STI variations and the NCM are investigated as predictors of intelligibility for the processed speech. The *modified* STI method does not produce reasonable predictions for these conditions; however, the *unmodified* STI method and the NCM methods do produce reasonable predictions. Failure of the modified envelope regression STI method is discussed. Use of the NCM to determine optimal values of a control parameter is also discussed.

9.1 Introduction

Despite the success of CI sound-processing strategies in quiet (Loizou, 1998), speech reception by CI users is still badly degraded by background noise. Comparisons of speech reception by cochlear implantees to that by normal hearing (NH) listeners show that implantees require anywhere from 5 to 13 dB higher speech-to-noise ratio (SNR) to achieve performance comparable to that of normal-hearing listeners when listening in stationary noise (Hochberg et al., 1992; Fu et al., 1998). Nelson et al. (2003) found that the SNR required by CI listeners was at least 25 dB greater than normal-hearing listeners when the noise was modulated.

Several factors associated with cochlear implantation and profound hearing impairment contribute to reduced speech reception in noise. Among these factors are reduced spectral resolution resulting from the limited number and location of implanted electrodes (Hannekom and Shannon, 1998; Henry and Turner, 2003), reduced temporal resolution associated with the carrier that modulates the electric pulse train (Muchnik et al., 1994), and a dynamic range that is less than 20 dB, compared with the normal hearing range of 100 dB (Zeng et al., 2002). In addition, for the vast majority of implantees whose CI systems use only one microphone, there are none of the benefits that can be gained by exploiting interaural differences, which for normal-hearing listeners yield substantial binaural advantages in speech reception (Zurek, 1993).

Given these limitations of both the CI user's impaired auditory system and the implant itself, solutions have been explored along two avenues. In the first, the processing performed by the implant processor is manipulated to find the best feature-extraction processing strategies and parameters for use in quiet and in noise (Holden et al., 1995; Fu, Shannon, and Wang, 1998). This approach essentially aims to optimize performance for the CI user within the bounds of the limited information available. This optimization procedure cannot overcome the fundamental limitations imposed by the auditory impairment, nor can it replace the loss of binaural hearing. In other words, there is no implant processing that could be termed noise reduction, *per se*.

The second approach to improving speech reception in noise attempts to reduce the noise at the input to the implant. In this chapter we focus on binaural noise reduction algorithms (reviewed in Section 2.3.2). For this class of algorithms, two microphone

signals—one over each ear—are processed to improve the overall signal to noise ratio. The first stage of the processing is to determine the inter-microphone phase and level differences as a function of frequency. Different inter-microphone phase and level differences correspond to different spatial locations. The desired signal is generally assumed to be arriving from straight ahead of the listener, such that the inter-microphone differences for the desired signal are expected to be zero. Frequency components that have inter-microphone differences not corresponding to the direction of the desired signal are suppressed. In this manner, the overall SNR of the signal is improved. Details of our implementation of the binaural noise reduction algorithm are described in Section 4.3.5. An evaluation of a commercial device and a preliminary evaluation of the algorithm we developed are given in Section 5.1.

9.2 Conditions

The problem addressed in this chapter is illustrated in Figure 9.1. The clean speech is

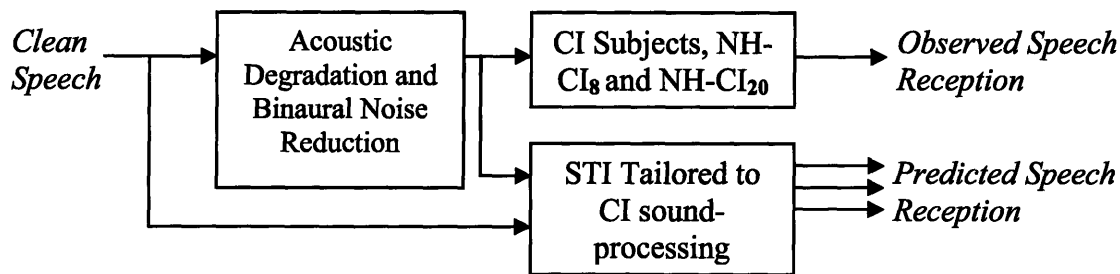


Figure 9.1:Block diagram of the experimental procedure for binaural processing conditions.

acoustically degraded and for half the conditions is then processed through the binaural noise reduction algorithm. The resulting signal is delivered either to a CI subject or a normal hearing subject listening to a vocoder simulation of CI sound-processing. The clean and degraded signals are used to calculate the various intelligibility metrics and the corresponding predicted speech reception.

16 conditions were selected to answer the following questions:

- 1) Does the binaural noise reduction algorithm improve speech reception in noise?
 - a. How does reverberation affect performance?
 - b. How does noise source modulation affect performance?

- 2) Are speech-reception gains from the binaural noise reduction algorithm dependent on the number of channels in the CI processor?
- 3) Do any of the candidate intelligibility metrics predict the effects of binaural noise reduction on the intelligibility of speech in noise?

We chose to investigate 8 and 20-channel CI processors. This decision was based on the fact that our subject pool contains primarily Clarion (8 channels) and Nucleus (22 channels) CI users. Clean speech was degraded by additive noise at -3 dB for the 8-channel condition and -6 dB for the 20-channel condition. We investigated speech-shaped noise as well as a single time-reversed talker. Both anechoic and mildly reverberant rooms were considered. The conditions are summarized in Table 9.1.

| NH-CI ₈ (-3 dB SNR) | | NH-CI ₂₀ (-6 dB SNR) | |
|--------------------------------|--------------|---------------------------------|--------------|
| Algorithm Off | Algorithm On | Algorithm Off | Algorithm On |
| SSN, A | SSN, A | SSN, A | SSN, A |
| SSN, M | SSN, M | SSN, M | SSN, M |
| TRS, A | TRS, A | TRS, A | TRS, A |
| TRS, M | TRS, M | TRS, M | TRS, M |

Table 9.1: Summary of experimental conditions for binaural noise reduction conditions. Abbreviations: anechoic (A) and mild (M) reverberation.

The experiment was divided into three trials that were tested on three separate days. Eight normal-hearing and 3 cochlear-implant subjects participated as subjects. Each trial consisted of the 16 conditions, each tested using one complete list from the CUNY database. The four divisions (columns) of the conditions found in Table 9.1 were used to partially counterbalance the conditions across subjects and the SNR or reverberation levels were partially counterbalanced within each subject across trials. Details of the experimental methods are given in Chapter 4.

9.3 Results of the Listening Experiment

9.3.1 NH-CI₈ and NH-CI₂₀ Subjects

The subjects' responses were scored as percentage of words correct for each trial. Figure 9.2 illustrates the subject scores for each condition averaged across subjects and trials. The data is divided into two groups corresponding to NH-CI₈ and NH-CI₂₀ results.

An initial repeated measures analysis of variance (RMANOVA_1)¹⁵ was performed using trials as the repetition variable. The dependent variable was the speech reception score transformed to RAU, and subject and condition were main factors. Subject is a significant factor. The lowest average subject score was 52.8 RAU, and the highest was 71.1 RAU. The interaction between subject and condition was not significant ($p > 0.05$). Thus, the trends observed for the different conditions were consistent across subjects.

A second repeated measures analyses of variance (RMANOVA_2) was performed using the speech reception score transformed to RAU as the dependent variable and subject, number of channels in the NH-CI_{Sim}, noise type, reverberation level, and algorithm (on vs. off) as main factors. The number of channels in the NH-CI_{Sim} was statistically significant ($p = 0.002$). Noise type is significant ($p < 0.001$) with higher scores for the time-reversed speech condition. Reverberation and algorithm function were significant ($p < 0.001$) with higher scores in the anechoic and algorithm on conditions, respectively. All second order interactions between noise type, reverberation, and algorithm function were significant, as was the interaction between algorithm function and number of channels in the CI simulation. Speech reception scores were lower in reverberation for both speech-shaped noise and time-reversed speech; however, the drop in performance was greater for the speech-shaped noise conditions. No higher order interactions were significant.

A few *post hoc* comparisons are made to emphasize the average speech reception gains that the binaural algorithm yields in different conditions. These results are illustrated in Figure 9.3. The binaural algorithm improved speech reception scores for all conditions tested and the average speech reception gain comparing the algorithm on

¹⁵ All variance and post-hoc measures are calculated in Matlab[®] in accordance with Winer et al. (1991).

versus off was 36.1 RAU for NH-CI₈ and 44.3 RAU for NH-CI₂₀. The higher average gain experienced for the 20-channel conditions contributed to the interaction between algorithm and number of channels in the CI simulation noted above. Considering the effect of reverberation, the average speech reception gain for anechoic conditions was 42.6 RAU for NH-CI₈ and 52.9 RAU for NH-CI₂₀; in comparison, the average speech reception gain for mild reverberation conditions was 29.6 RAU for NH-CI₈ and 35.8 RAU for NH-CI₂₀. Thus, average speech reception gains were more than 10 RAU smaller in mild reverberation compared to anechoic. Considering the effect of noise type, the average speech reception gain for speech-shaped noise conditions was 32.3 RAU for NH-CI₈ and 41.4 RAU for NH-CI₂₀; in comparison, the average speech reception gain for time-reversed speech was 39.8 RAU for NH-CI₈ and 47.3 RAU for NH-CI₂₀. Thus, the algorithm provides slightly more benefit for the time-reversed speech conditions. These comparisons of speech reception gains for reverberation and noise type for a given NH-CI_{Sim} were significant ($\alpha = 0.05$).

9.3.2 CI Subjects

Three CI subjects (CI-7, CI-8, CI-9) participated in this experiment. The CI subjects were tested using a similar set of conditions as those summarized in Table 9.1. However, the SNR of each condition was shifted by a certain amount, Δ , in order to compensate for individual performance differences. The process for determining Δ for each subject is given in Section 4.4.2. Table 4.1 summarizes the Δ values found for each subject. Figure 9.4 illustrates the subject's scores for each condition. The scores reported are mean values across subjects and trials.

A repeated measures analysis of variance was performed similar to RMANOVA_1 but using the CI data. This analysis indicates that subject and conditions are both significant. The average scores for the three subjects are 7.4, 30.9, and 48.5 RAU (respectively for CI-7, CI-8, and CI-9). The interaction between subject and condition was found to be significant ($p < 0.01$). Subsequent analysis illustrates this interaction primarily reflects different performance trends in reverberation. Thus, care must be taken to understand different trends exhibit by individual subjects. To this end, Figure 9.5 illustrates scores for individual CI users.

A repeated measures analysis of variance was performed similar to RMANOVA_2 but using the CI data. The key differences in the findings were that the interaction between algorithm function and noise type is not significant ($p > 0.1$) for CI users and that the interaction between noise type and reverberation is not significant ($p > 0.1$). All other main and interaction effects were comparable to NH-CI_{Sim} results.

In addition, RMANOVA_2 for the CI data clarifies the subject by condition interaction showed in RMANOVA_1. In particular, the interaction between subject and reverberation (0.023) and between subject and noise type ($p = 0.046$) were both moderately significant. Interaction between subject and algorithm performance was not significant ($p > 0.1$). Subject performance in mild reverberation varied from being approximately equal to the corresponding anechoic condition to being significantly lower than anechoic. The largest drop in performance attributed to mild reverberation was for subject CI-9 in speech-shaped noise who performed 25 RAU lower in mild reverberation compared to the anechoic condition.

A few *post hoc* comparisons are made to emphasize the effect of the binaural noise reduction algorithm. These results are illustrated in Figure 9.6. The overall average speech reception gain comparing the algorithm on versus off was 32.3 RAU. Considering the effect of reverberation, the average speech reception gain for anechoic conditions was 42.1 RAU; in comparison, the average speech reception gain for mild reverberation conditions was 22.5 RAU. Considering the effect of noise type, the average speech reception gain for speech-shaped noise conditions was 36.4 RAU; in comparison, the average speech reception gain for mild reverberation conditions was 28.1 RAU. The comparisons of speech reception gains for reverberation is significant ($\alpha = 0.05$); but the comparison across noise types is not significant ($\alpha = 0.05$).

9.4 Results of the Intelligibility Predictions

9.4.1 NH-CI₈ and NH-CI₂₀ Subjects

The procedure for calculating particular metrics from the clean and degraded (or processed) speech waveforms is detailed in Section 4.5. As described in Section 5.2, it is possible for the original envelope-regression metric to fail by producing invalid values of

intermediate metrics. In particular, when the modulation metric (Eq. 3.2) is outside the range between 0 and 1, then the apparent SNR calculated as calculated in Eq. 3.1 is a complex—in the mathematical sense—number and cannot be interpreted in the existing STI framework. In this chapter, we avoid this problem by clipping the modulation metric (Eq. 3.2) to values between 0 and 1.

The metrics are calculated for the conditions tested and then a psychometric function is fit to the mapping between metric value and the mean reception scores. The resulting psychometric function thus yields a predicted score (in RAU) for a given metric value. Figures 9.7, 9.8, and 9.9 illustrate the comparison between observed scores for NH-CI₈ and predicted scores for the respective methods.

Two measures are given for assessing the predictions made by the different intelligibility metrics: 1) the model error defined as the standard deviation between predicted and observed scores and 2) the correlation coefficient between predicted and observed scores. Surprisingly, the unmodified envelope regression STI method predicts speech reception for these binaural noise-reduction processing conditions quite well as evidenced by its low model error, 5.25 RAU, and its high correlation coefficient, 0.97. In contrast, the *modified* envelope regression STI method produces poor predictions as evidenced by its low correlation coefficient, 0.75. The fact that these results are contrary to our expectations (we expected the modified method to perform better since the operation is nonlinear) is discussed in Section 9.6. The NCM method produces reasonable predictions as seen by its low model error, 6.92 RAU, and high correlation coefficient, 0.96.

Both the unmodified envelope regression STI and the NCM methods produce reasonable speech reception predictions. Both generate reasonable predictions as to the gain provided by the binaural noise reduction algorithm in a variety of conditions. A minor trend exists for both metrics in that they tend to overestimate performance for the reverberant conditions.

9.4.2 CI Subjects

The psychometric function was fitted for the NH-CI₈ data based on the mean subject scores. However, for actual CI users, we expect a wider variance in observed scores. It

is possible that a particular subject may not be able to score 100% in quiet. To compensate for this potential difference, the psychometric function is fit to each subject and allowing R_{max} of Equation 4.10 to vary. The added degrees of freedom in the model are taken into account in the calculation of the model error and the correlation coefficient.

Figures 9.10, 9.11, and 9.12 illustrate the comparison between observed and predicted scores for the respective methods. The accuracy of the speech reception predictions is comparable to the NH-CI_{Sim} results: the modified envelope regression STI method produces grossly inaccurate predictions while the other two methods produce reasonable predictions. The overall model accuracy is lower when comparing the fitting of the CI data to the NH-CI_{Sim} data for each metric.

9.5 Frequency-band analysis

Similar to the spectral subtraction analysis given in Section 9.13, the TI values can be compared with and without processing to illustrate the frequency band specificity of the binaural noise reduction. Figure 9.8 illustrates the TI value for the NCM method for the 20-channel speech-shaped noise condition with the algorithm off versus with the algorithm on.

Before comparing the algorithm on versus off conditions, note that the TI values are not constant for the speech-shaped noise condition with the algorithm off. This result is somewhat surprising since the noise source is speech-shaped so we might expect the TI values to be relatively constant. However, this variation in TI values can be explained by considering the effect of summing the left and right microphone signals on the SNR. The phase difference between the left and right ear can be approximated as (Blauert, 1996):

$$\Omega = \left(\frac{d \sin(\theta)}{c} \right) 2\pi f, \quad (9.1)$$

where d is the diameter of the head, θ is the angle of incidence, c is the speed of sound, and f is the frequency of the sound component. For our simulation, the diameter of the head and the angle of incidence were specified as 0.24 m and 60 degrees. Using $c = 340$ ms, we find that the noise source will be perfectly out of phase at 820 Hz. Thus the listener will receive substantial benefit near 820 Hz simply because the noise source

will destructively interfere. In contrast, noise components at 1640 Hz will be in phase and constructively interfere. It can be noted in Figure 9.7 that the largest TI values for the unprocessed condition occur near 820 Hz (band 4) and the lowest near 1640 Hz (band 8) supporting this argument.

Comparing TI values for algorithm on versus off clearly shows that processing results in higher values regardless of frequency band. The improvement is greatest for the middle frequency bands (bands 5 through 9 corresponding to 950 to 2323 Hz) because of the effect described in the previous paragraph.

Analysis of TI values can be used to specify gain parameters as a function of frequency. For example, the gain parameters α and β of Equation 5.3 could be specified as functions of frequency and selected to optimize individual TI values. Insofar as the TI values are indicative of the intelligibility contribution of a given band, this procedure will optimize overall performance.

In addition, the insight from the TI analysis that the noise source constructive interference can be significant has practical consequences for the binaural noise reduction algorithm. The binaural noise reduction algorithm operates by applying frequency dependent gain control on the sum of the left and right microphone signals (as illustrated in Figure 2.8). However, the gain control could readily be applied to either the left or right microphone signal independently. Therefore, if an algorithm was developed that could determine the better ear, the gain control could be applied to just that microphone signal. Such an algorithm may not be too difficult to develop since the binaural algorithm itself calculates inter-microphone phase and intensity differences that would be useful for determining the angle of incoming sounds.

In summary, the analysis of TI values for the different frequency bands can often illuminate algorithm function in ways that the overall NCM (or STI) score cannot. The TI analysis illuminates the effect of noise source interference patterns. The TI analysis can also be used for selecting frequency specific parameters.

9.6 Discussion

Our results clearly indicate that binaural noise reduction improves speech reception in noise for a wide range of conditions including different levels of noise source modulation

and reverberation levels. The original envelope regression STI method accurately predicts the speech reception results for these conditions. In contrast, the modified envelope regression method produces poor predictions. This success of the original method and failure of the modified method is in contrast to the results from the N-of-M and spectral subtraction experiments. This unexpected reversal can be explained by considering the degree of the processing implemented (i.e. the strength of the gain function applied) for the binaural algorithm and by considering the nature of the proposed modification (i.e. the new scaling factor of Eq. 3.4).

Noise reduction processing can generally be conceived as a trade-off between noise removal and signal distortion. As the processing becomes more rigorous, more noise is removed at the expense of introducing distortions in the desired signal (in this discussion we define noise as the competing acoustic sound and distortion as detrimental artifacts arising from the processing). For example, in the spectral subtraction algorithm as the control parameter κ is increased, the amount of noise present monotonically decreases; however, the distortions introduced for $\kappa > 4$ substantially reduce the intelligibility of the signal. The binaural algorithm contains comparable control parameters (α and β of Equation 5.3). Unlike in the evaluation of spectral subtraction in Chapter 8, we did not vary the control parameters in the binaural algorithm. Instead, we set the control parameters and tested a variety of conditions. The success of the original envelope regression STI method occurs in large part because the parameters chosen do not result in a processed signal that is excessively distorted. For example, the output signal does not contain the “rippled” sound often associated with heavily processed signals. We hypothesize that the original envelope regression STI would not effectively characterize signals that were distorted when processed with high values of the control parameters.

In fact, the proposed normalization term introduced in Section 4.3 and repeated here:

$$\beta = \frac{\mu_x}{\mu_x + \mu_z}, \quad (9.1)$$

was introduced primarily because nonlinear operations may lower the clean envelope energy such that the original normalization term repeated here:

$$\alpha = \frac{\mu_x}{\mu_y}, \quad (9.2)$$

increases too rapidly. However, for the case at hand, the energy of the processed envelopes is not greatly reduced compared to the clean envelopes. This can be seen by comparing the clean and degraded envelope signals. Figure 9.14 illustrates the clean and degraded envelope signals for the speech-shaped noise (-6 dB SNR) condition with the binaural algorithm on for the lowest and highest frequency band of the 8-band processor. For the lowest band (Figure 9.14A), the processing does a remarkable job of extracting the speech envelope of the desired signal without introducing distortion. For that case, the clean and degraded envelopes are almost equal and, consequently, $\mu_y \approx \mu_x$. Furthermore, for that band, $\mu_z \ll \mu_x$; consequently, $\beta \approx \alpha \approx 1$. In other words, the processing does an excellent job of extracting the speech component of the signal and both metrics are able to quantify this result.

In contrast, for the high-frequency band, the binaural algorithm over-processes the signal resulting in a distorted and suppressed clean envelope as seen in Figure 9.14B. The normalization terms were calculated for the high band envelopes and found to be: $\alpha = 4.8$ and $\beta = 0.55$. Thus, the original normalization term is effectively giving the degraded envelope a 4.8 scale factor. This was precisely the reason that β was introduced, because the α scale factor would increase as the mean of the degraded envelope decreases. The modified normalization term is effectively giving the degraded envelope a 0.55 scale factor.

The effect of the disparity between scale factors in the high band is that the original method produces a higher TI value than the modified method. For the N-of-M and spectral subtraction operations, we found that the predictions from the original method were too high. That result does not occur here perhaps in part because the distortions in the signal are not excessive compared to the N-of-M and spectral subtraction conditions. Nonetheless, the bottom line is that the modified scaling produces inaccurate predictions. Yet, the original method produces inaccurate predictions for the

N-of-M and spectral subtraction conditions. Thus, only the NCM method has proven to be a reliable predictor of performance for all conditions tested.

9.7 Conclusions

The main conclusions of this chapter are:

- (1) The binaural noise reduction algorithm improves speech reception for CI-processed speech for a variety of conditions including different noise types and mild levels of reverberation.
- (2) The modified envelope regression STI method fails to produce reasonable speech reception predictions for the binaural noise reduction conditions considered.
- (3) Both the unmodified envelope regression STI and NCM methods produce reasonable speech reception predictions.

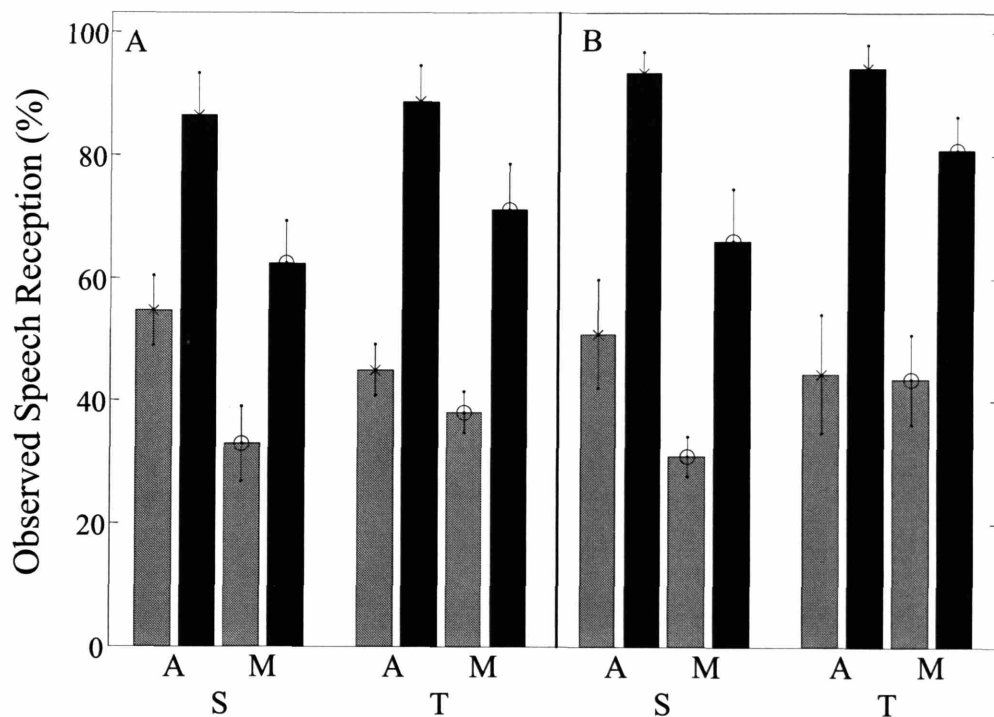


Figure 9.2: Speech reception scores for NH-CI_{Sim} tested on the binaural noise reduction conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. The darker shaded bars correspond to conditions with the binaural algorithm on. Speech reception with the binaural algorithm on was significantly higher than speech reception with the algorithm off for each condition tested according to a *post hoc* Tukey HSD test ($p < 0.05$). The two subplots represent results from A) NH-CI₈ and B) NH-CI₂₀. Abbreviations: speech-shaped noise (S), time-reversed speech (T), anechoic (A), and mild (M).

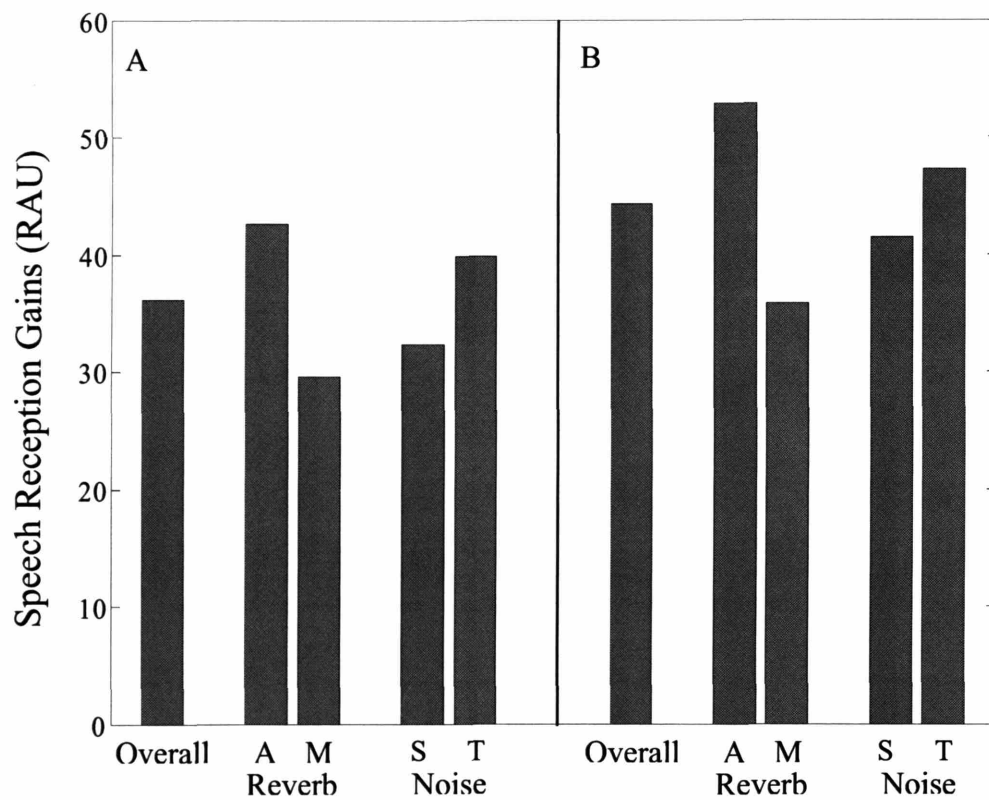


Figure 9.3: Average speech reception gains resulting from the binaural algorithm. The two subplots represent results from A) NH-CI₈ and B) NH-CI₂₀. Abbreviations: speech-shaped noise (S), time-reversed speech (T), anechoic (A), and mild (M).

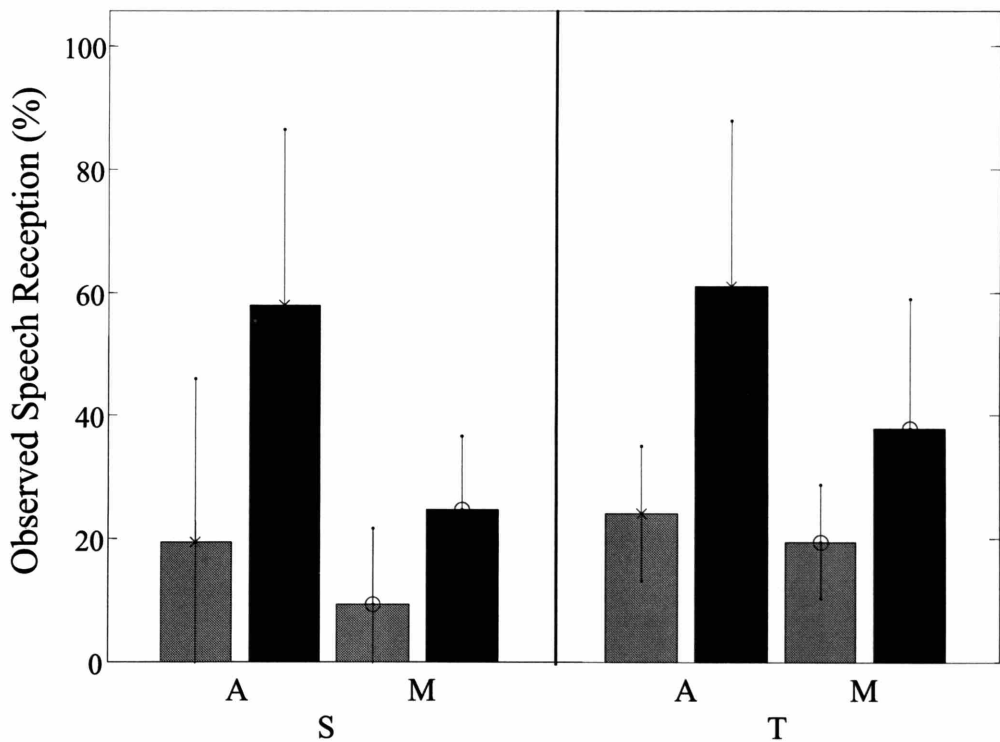


Figure 9.4: Speech reception scores for CI users tested on the binaural noise reduction conditions. The bars represent the mean scores averaged across trials and subjects. The error bars represent \pm one standard deviation of the mean. The darker shaded bars correspond to conditions with the binaural algorithm on. Abbreviations: speech-shaped noise (S), time-reversed speech (T), anechoic (A), and mild (M). Speech reception with the binaural algorithm on was significantly higher than speech reception with the algorithm off for each condition tested according to a *post hoc* Tukey HSD test ($p < 0.05$).

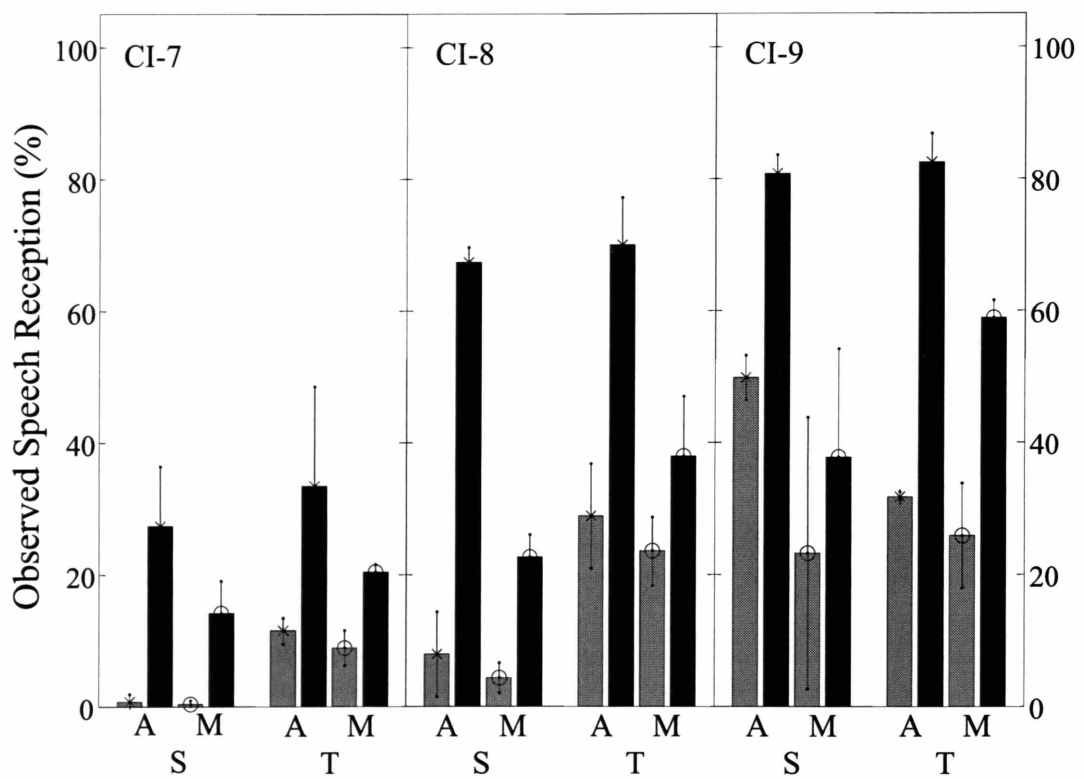


Figure 9.5: Individual speech reception scores for CI users tested on the binaural noise reduction conditions. The bars represent the mean scores averaged across trials for each subjects. The error bars represent \pm one standard deviation of the mean. The darker shaded bars correspond to conditions with the binaural algorithm on. Abbreviations: speech-shaped noise (S), time-reversed speech (T), anechoic (A), and mild (M). Speech reception with the binaural algorithm on was significantly higher than speech reception with the algorithm off for each condition tested according to a *post hoc* Tukey HSD test ($p < 0.05$).

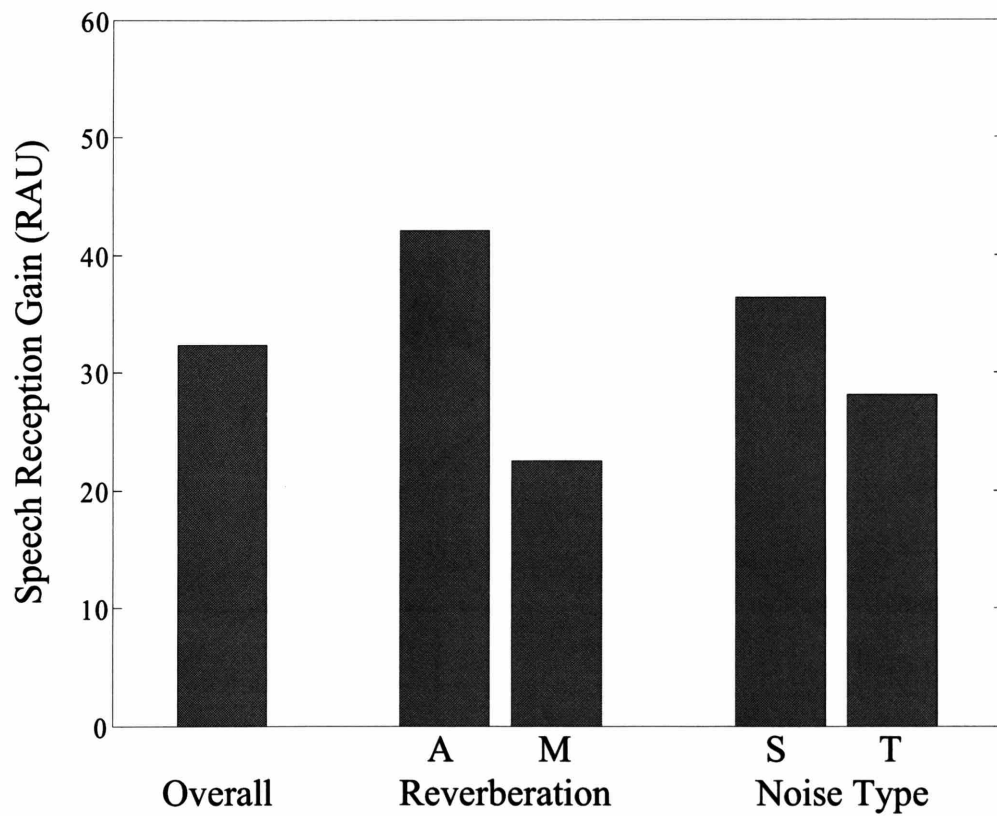


Figure 9.6: Average speech reception gains resulting from the binaural algorithm for CI users. Abbreviations: speech-shaped noise (S), time-reversed speech (T), anechoic (A), and mild (M).

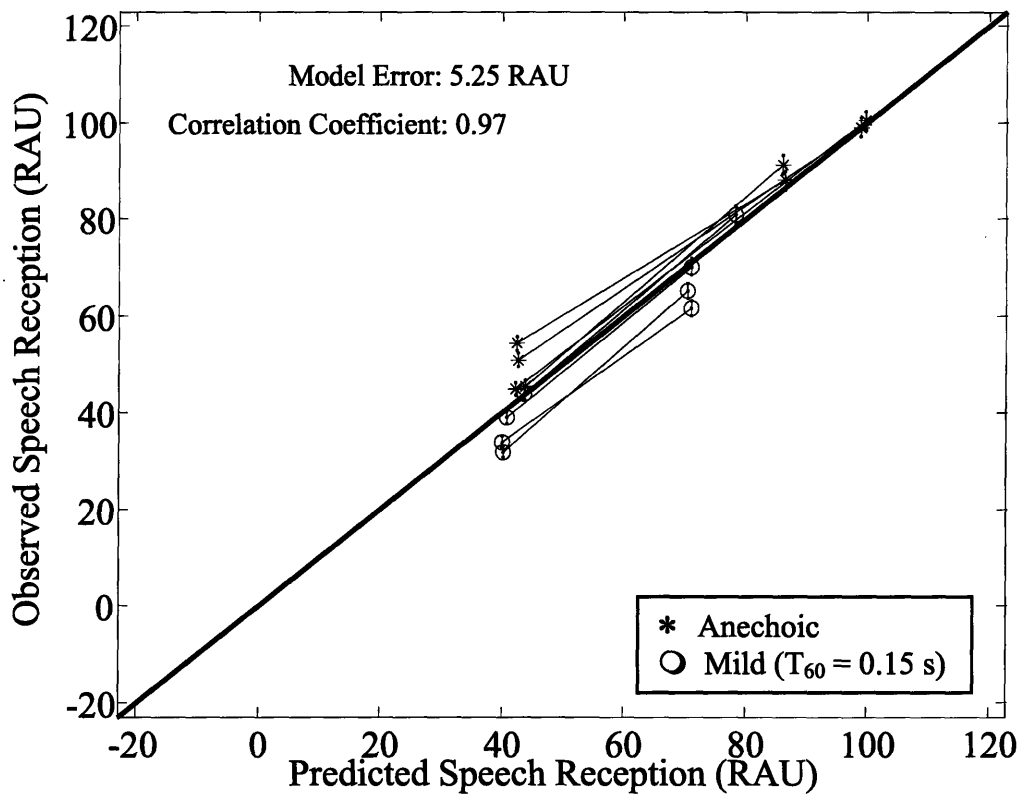


Figure 9.7: Comparison of observed scores for $NH-CI_{Sim}$ and predicted scores from the envelope-regression STI method. The error bars represent \pm one standard error of the mean. The dashed lines connect conditions corresponding to algorithm on and off for a particular acoustic degradation with the algorithm on condition always having the higher observed speech reception.

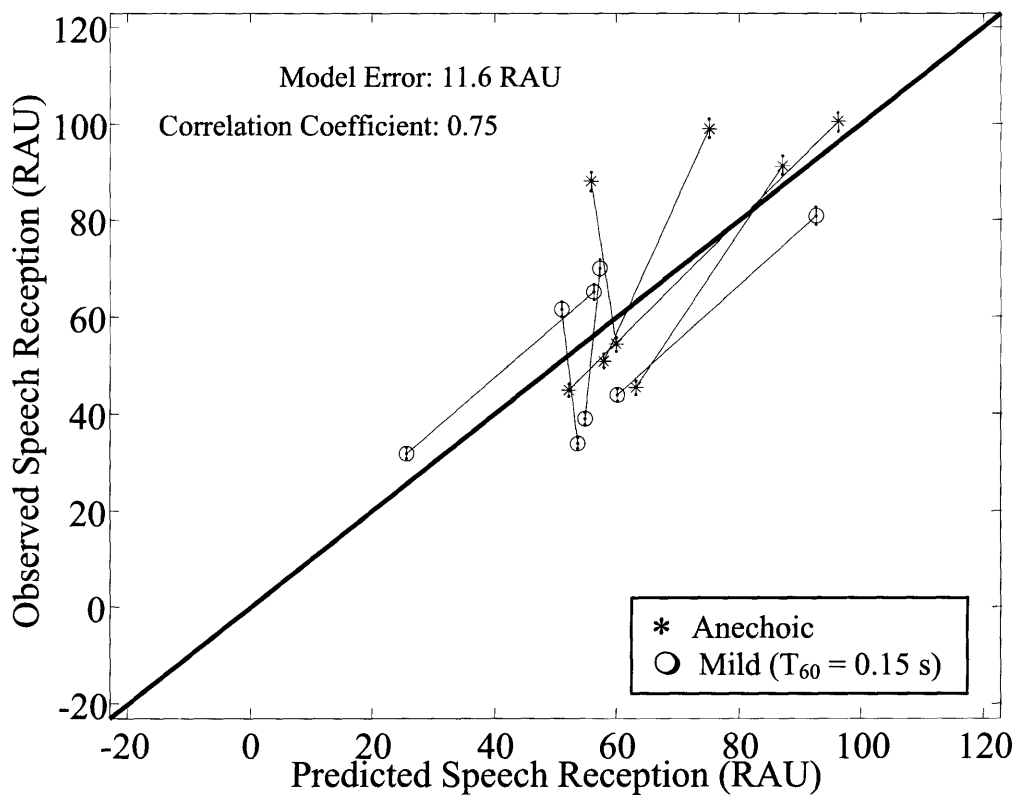


Figure 9.8: Comparison of observed scores for $NH-CI_{Sim}$ and predicted scores from the modified envelope-regression STI method. The error bars represent \pm one standard error of the mean. The dashed lines connect conditions corresponding to algorithm on and off for a particular acoustic degradation with the algorithm on condition always having the higher observed speech reception.

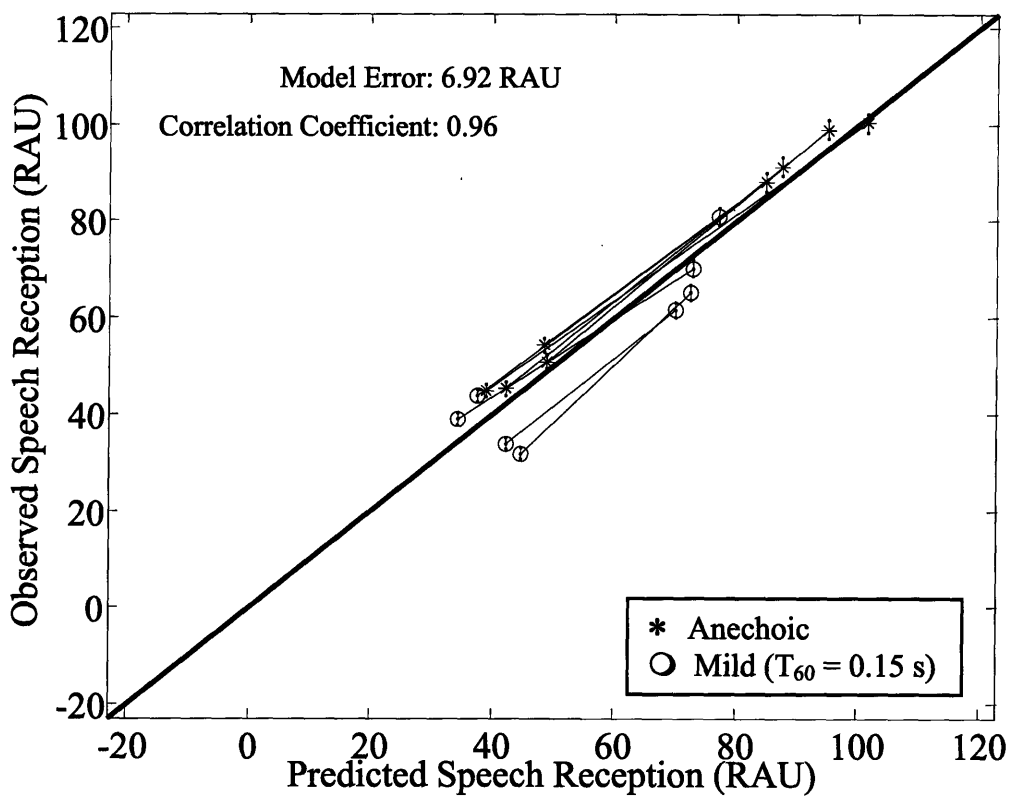


Figure 9.9: Comparison of observed scores for $NH-CI_{Sim}$ and predicted scores from the NCM method. The error bars represent \pm one standard error of the mean. The dashed lines connect conditions corresponding to algorithm on and off for a particular acoustic degradation with the algorithm on condition always having the higher observed speech reception.

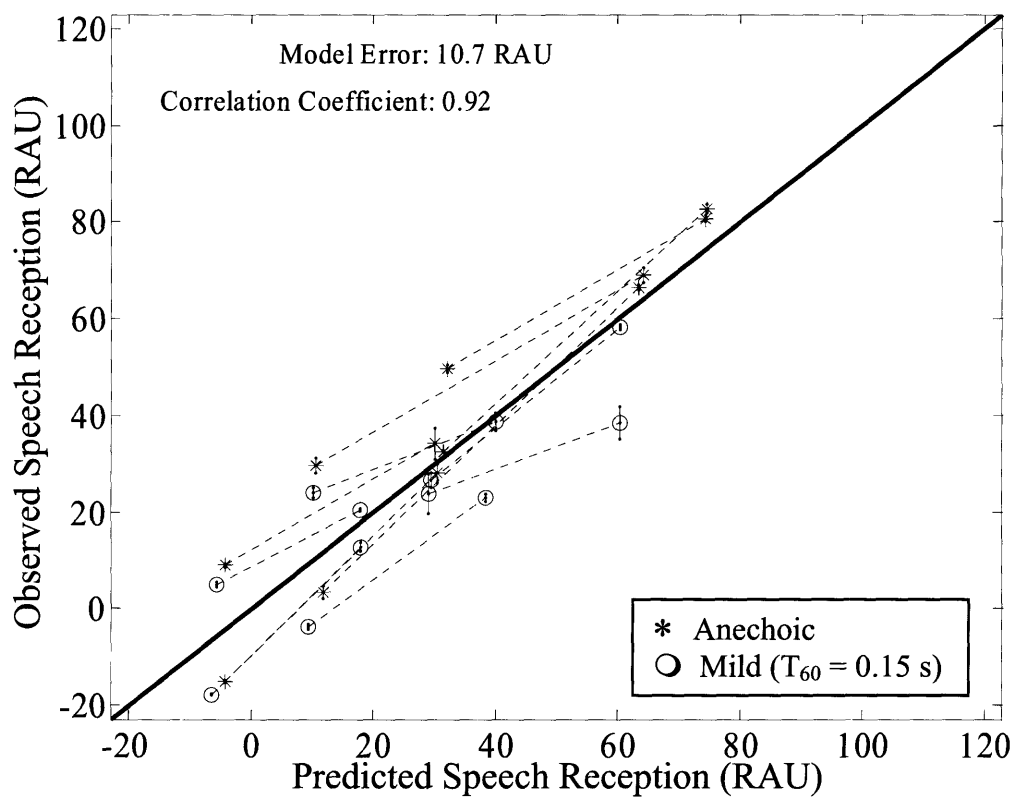


Figure 9.10: Comparison of observed scores for CI users and predicted scores from the envelope-regression STI method. The error bars represent \pm one standard error of the mean.

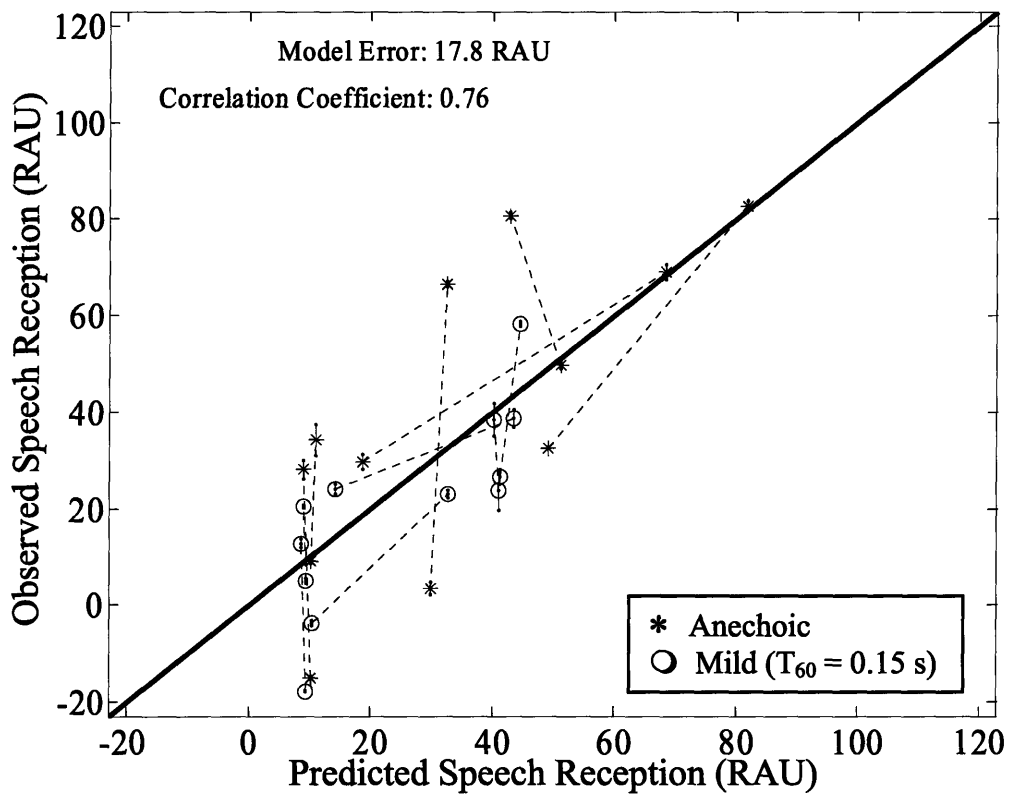


Figure 9.11: Comparison of observed scores for CI users and predicted scores from the modified envelope-regression STI method. The error bars represent \pm one standard error of the mean.

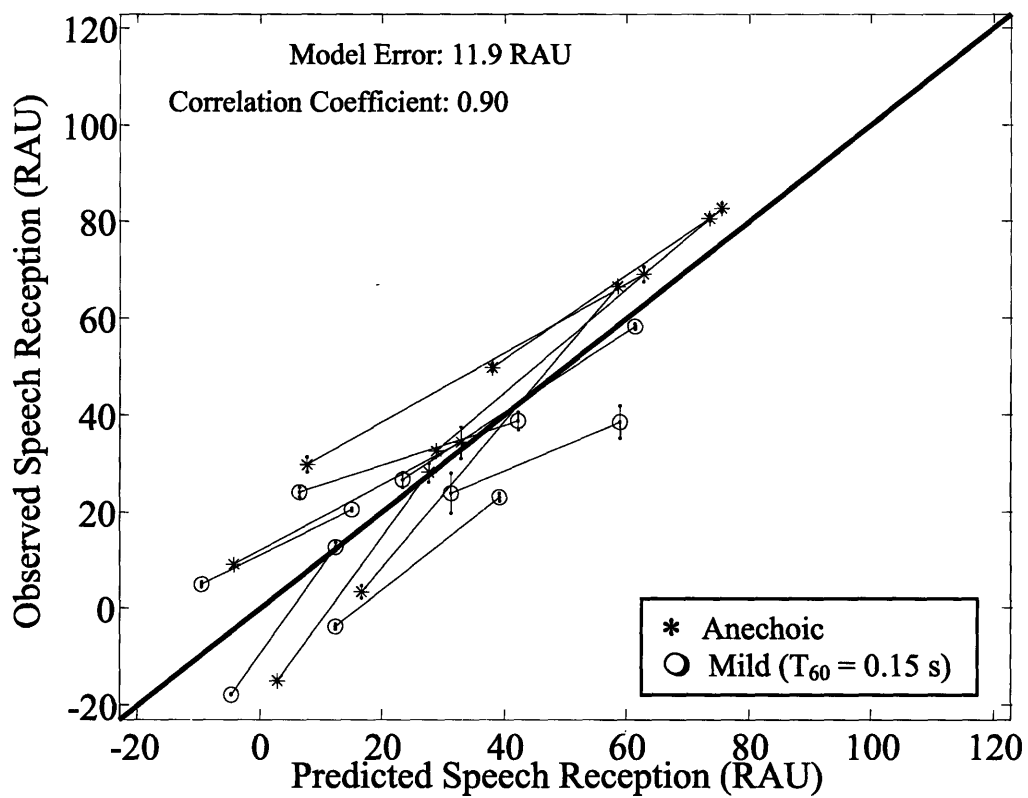


Figure 9.12: Comparison of observed scores for CI users and predicted scores from the NCM method. The error bars represent \pm one standard error of the mean.

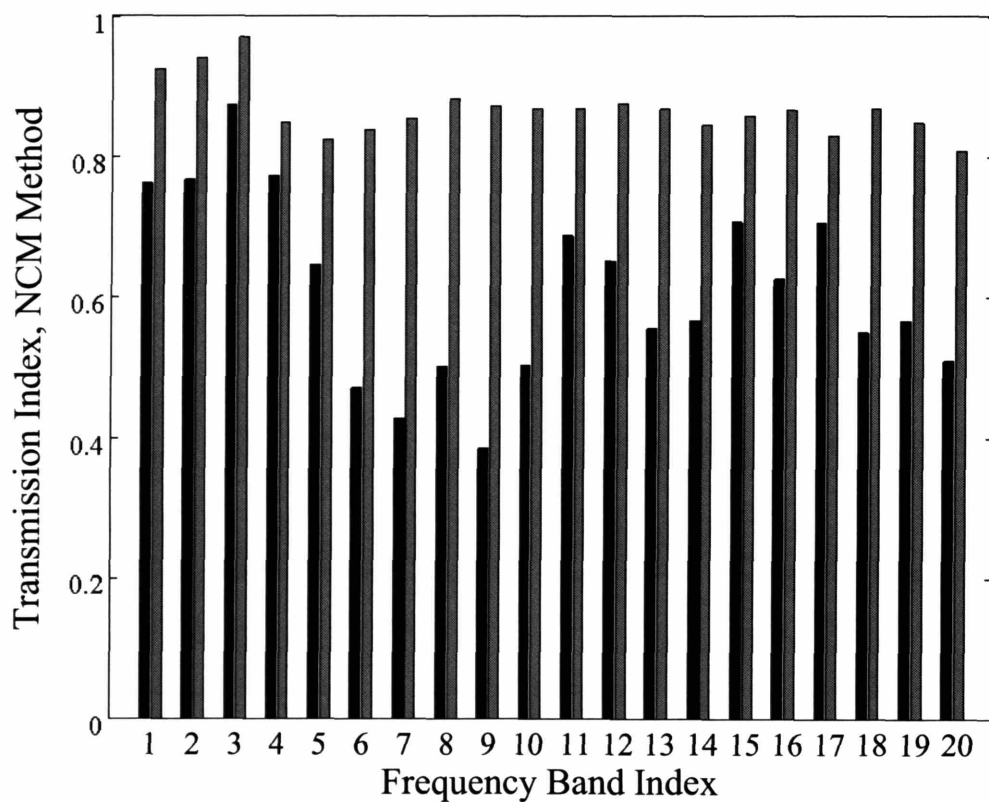


Figure 9.13: Analysis of TI values for the NCM method. Each set of two bars correspond to TI values for a given frequency band for the twenty channel analysis. The left and right bars correspond to TI values with the binaural algorithm off and on, respectively. The TI values are calculated as intermediate metrics in the NCM calculation and are based on the same clean and degraded material as the NCM data presented in Figure 9.12.

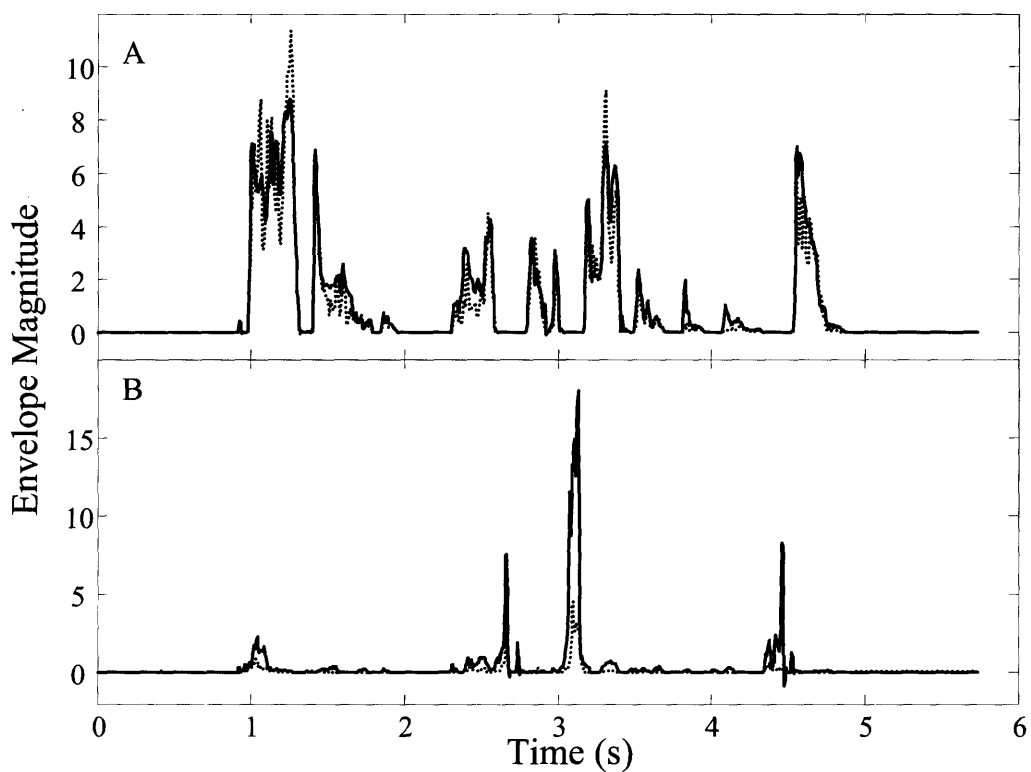


Figure 9.14: A clean speech is degraded with speech-shaped noise at -6 dB SNR and the binaural algorithm is applied. Envelope signals are determined for clean (solid line) and degraded (dotted line) for an 8-channel vocoder. Envelopes are plotted for a A) low-frequency band and a B) high-frequency band.

Chapter 10

Discussion

In this chapter we discuss the overall performance of the intelligibility metrics developed and assessed in this thesis. We first discuss the general success and failure of the candidate metrics for each of the four main experiments. Only the NCM was found to be an accurate predictor of speech reception for all experiments. In addition, we evaluate the candidate metrics across conditions illustrating the overall success of the NCM method. We conclude this chapter by summarizing suggested future work associated with the speech reception models and with the noise reduction algorithms.

10.1 Summary of Intelligibility Predictions

Our stated goal for this thesis was to identify an intelligibility metric that is an accurate predictor of speech reception for CI-processed speech for a wide range of conditions including nonlinear operations. This goal is motivated by the desire to have a single metric that can be used to optimize noise reduction algorithms specifically for CI users. The advantages of determining a relevant physical performance metric for CI users is that it makes algorithm evaluation efficient, consistent, and subject independent.

Our pursuit of such a metric began by investigating the STI. We noted that the STI might serve as an excellent candidate for assessing speech reception in CI users since the mechanics of STI calculation are quite similar to the mechanics of CI sound-processing. Both are dependent on the envelope signals in a number of frequency bands spanning the relevant spectrum for speech. We introduced a procedure in Section 3.2 allowing the STI to be tailored to a particular CI processing strategy.

However, this procedure for tailoring STI to a particular CI processing strategy does not address the failure of STI for nonlinear operations, which is fundamentally rooted in the underlying STI calculation. Our preliminary work discussed in Section 5.2 describes how STI calculation results in invalid intermediate metrics that cannot be logically interpreted in the STI framework. As such, we were required to introduce modifications in order to apply STI to nonlinear operations.

These modifications resulted in five novel metrics introduced in Section 3.1. Of these metrics, three were based on a novel normalization term and result in very similar STI predictions. Of these three metrics based on a new normalization term, the modified envelope regression method was selected for detailed evaluation in Chs. 6-9. Similarly, previously-proposed metrics resulted in similar predictions and the unmodified envelope regression method was selected for detailed evaluation in Chs. 6-9. Also introduced in Section 3.1 is the normalized correlation metric (NCM). We suggest that NCM be considered a novel metric independent of the STI. The commonalities between STI and NCM are that both metrics are based on clean and degraded envelope signals in a number of frequency bands spanning the relevant spectrum for speech. However, the procedures for calculating a single TI value based on the clean and degraded envelopes are

fundamentally different between the two metrics. As such, the NCM metric predictions—especially for nonlinear operations—are unique.

The gross accuracy of the metrics analyzed in Chs. 6 through 9 is summarized in Table 10.1. By gross accuracy we simply mean whether or not the metric produces reasonable predictions for the conditions tested. For example, the unmodified envelope regression STI predicts that N-of-M processing should increase the intelligibility of speech in noise; this prediction is in the opposition direction from the observed trend and is therefore labeled as a failure. Similarly, the modified envelope regression STI method does not predict an increase in speech reception resulting from the application of the binaural noise reduction; this prediction is contrary to the significant observed improvement in speech reception.

| | ER STI | Modified ER STI | NCM |
|-----------|--------|-----------------|-----|
| Acoustic: | ✓ | ✓ | ✓ |
| N-of-M: | × | ✓ | ✓ |
| Spectral: | × | ✓ | ✓ |
| Binaural: | ✓ | × | ✓ |

Table 10.1: General success (✓) and failure (×) of the investigated metrics for each experiment.

Only the NCM is successful in providing reasonable predictions of speech reception for all four experiments. As such, it deserves further consideration. In Section 10.2 we analyze the performance of the NCM across conditions tested in this experiment. In Section 10.3 we discuss future directions for developing and analyzing the NCM.

10.2 Evaluation of the Performance Metrics across Experiments

As mentioned in the previous section, only the NCM method produces reasonable speech reception predictions for all conditions tested. In evaluating the speech reception predictions for the various conditions, a distinct psychometric function was fitted to the data for each experiment. An important question to answer is how well the different intelligibility metrics predict performance across all conditions tested. Towards answering this question, we fitted single psychometric functions for all of the NH-Cl₈ and NH-Cl₂₀ conditions. That is, we fitted a single psychometric function for the NH-Cl₈ data and a second function for the NH-Cl₂₀ data.

Figures 10.1, 10.2, and 10.3 illustrate the speech reception predictions for the three candidate metrics for the NH-CI₈ conditions. This set of conditions includes acoustic degradation, spectral subtraction, and binaural noise reduction conditions.

In Figure 10.1 we see that the unmodified envelope regression STI method has the highest model error, 16.4 RAU, and lowest correlation coefficient between predicted and observed scores, 0.71. These indicators primarily reflect the inability of this method to capture trends for the spectral subtraction conditions. The highly processed condition for spectral subtraction is particularly poorly predicted.

In Figure 10.2 we see that the modified envelope regression STI method performs slightly better with a model error of 15.4 RAU and a correlation coefficient of 0.80. Certain conditions are labeled on the figure to emphasize that this method has difficulty fitting the acoustic degradation data and the noise reduction data simultaneously. In particular, only three conditions from the acoustic degradation data are under-predicted (they fall above the psychometric function).

In Figure 10.3 we see that the NCM method has the best performance with a model error of 10.4 RAU and a correlation coefficient of 0.90. Certain conditions are labeled on the figure to emphasize that this method has difficulty fitting the highly processed conditions of spectral subtraction. This trend was observed in Chapter 8 and basically implies that the NCM method is biased towards underestimating the negative impact of signal distortion on speech reception.

Figures 10.4, 10.5, and 10.6 illustrate the speech reception predictions for the three candidate metrics for the NH-CI₂₀ conditions. This set of conditions includes N-of-M, spectral subtraction, and binaural noise reduction conditions.

In Figure 10.4 we see that the unmodified envelope regression STI method has the highest model error, 24.7 RAU, and lowest correlation coefficient between predicted and observed scores, 0.45. These indicators reflect the inability of this method to capture trends for the N-of-M and spectral subtraction conditions. The highly processed conditions ($\kappa = 4$ and $\kappa = 8$) for spectral subtraction are particularly poorly predicted. Certain conditions are labeled on the figure to emphasize these failings.

In Figure 10.5 we see that the modified envelope regression STI method performs considerably better with a model error of 16.1 RAU and a correlation coefficient of 0.83.

Certain conditions are labeled on the figure to emphasize that this method has difficulty fitting the N-of-M data and the noise reduction data simultaneously. In particular, a number of the N-of-M conditions are overestimated (fall beneath the psychometric function) while a number of the noise reduction processing conditions are underestimated.

In Figure 10.6 we see that the NCM method has the best performance with a model error of 12.3 RAU and a correlation coefficient of 0.89. Again we see that this method has difficulty fitting the highly processed conditions of spectral subtraction. Another trend is that the N-of-M conditions in quiet are underestimated; we hypothesized in Chapter 7 that this effect could be compensated for by incorporating redundant information into the model. Certain conditions are labeled on the figure to emphasize these failings.

In summary, the NCM method produces reasonable speech reception predictions even when fitting the data across experiments. The other two metrics have worse performance when fitting across experiments, which is not surprising given that both of the envelope regression STI methods exhibited extremely poor predictions for at least one set of experimental conditions.

10.3 Future Work

The work described in this thesis integrates three fields of speech and hearing sciences: cochlear implant speech reception, intelligibility metrics, and noise reduction algorithms. In the course of this work, we successfully developed and evaluated a predictor of speech reception for CI-processed speech for a wide range of conditions. Spectral subtraction and binaural noise reduction algorithms were developed and evaluated. The evaluations clearly indicated that these algorithms improve speech reception in noise for CI-processed speech. In this final section of the discussion, we develop possible ramifications of these successful results.

10.3.1 Intelligibility Metrics

The NCM was developed and assessed in this thesis specifically for CI-processed speech. The thesis commenced by considering the STI framework for developing intelligibility

metrics for CI-processed speech. However, the STI methods needed to be modified to produce reasonable predictions for nonlinear operations. The NCM was developed while attempting to modify STI for nonlinear operations. The resulting NCM method has proven to be a fair predictor of performance for a wide range of conditions for both noise vocoder simulations of CI-processed speech, as well as for actual CI user performance. In addition, computation of the NCM is considerably less complex when compared to that of the STI. Because of its success across all four experiments of this thesis, and because the metric can be more efficiently calculated than the STI methods, in the following we focus solely on the NCM. Parallels on the following topics could be developed for STI if desired.

First, the accuracy of the NCM method in predicting the intelligibility of CI-processed speech could be improved by a number of possible modifications. Three potential modifications were suggested concerning the effects of noise source modulation (Section 6.5.1), reverberation (Section 6.5.2), and adjacent frequency band redundancy (Section 7.6.1). The modifications suggested related to reverberation were explicit and require only minor modifications to existing software developed for this thesis. The modified metrics could then be calculated and the predictions compared to observed scores as before. The suggested modifications concerning noise source modulation and frequency band redundancy were more open-ended. These modifications would require analysis, further development and extensive evaluation.

Second, we specifically developed intelligibility metrics for CI-processed speech. We did so by specifying the bandpass filtering and envelope extraction strategies used in the metric to match the CI processing. Since we specifically tailor the metric to particular CI processing parameters, we require separate psychometric function fittings for each CI processor. Figure 10.7 shows the different psychometric functions calculated for the NH-CI₈ and NH-CI₂₀ data in Figures 10.3 and 10.6. For a given NCM value, predicted speech reception is higher for the NH-CI₂₀ conditions. The reason behind this result is that speech reception is generally higher for subjects listening to noise vocoder simulations of cochlear implant processing when the simulation has more channels. The NCM does not currently allow for direct comparison of the NH-CI₈ and the NH-CI₂₀ conditions using the same psychometric function. To achieve such a direct comparison,

factors accounting for higher speech reception with more channels would need to be included. In particular, the metric could account for the increased redundancy of information across channels as the number of channels increases.

10.3.2 Noise reduction algorithms

We have shown that both spectral subtraction and binaural noise reduction can improve speech reception in noise for CI-processed speech. It was argued that these algorithms are particularly effective for CI users since they capitalize on high resolution processing and binaural information before the signal is transmitted to the CI sound-processing strategy. The success of these algorithms motivates the development of a body-worn noise reduction accessory for CI sound processors.

A number of issues must be addressed before developing such an accessory. A primary issue for development of a spectral subtraction based noise reduction accessory will be how to estimate the noise spectrum level. For the purpose of this thesis, we assumed that the noise spectral estimate was known. A real-time accessory, however, would have to perform running estimates of the noise spectrum.

For binaural noise reduction based accessory, the effect of reverberation needs to be given more attention. The reverberation levels considered in Chapter 9 were relatively mild; consequently, further evaluation in strong reverberation is needed.

10.4 Final Conclusions

We have successfully developed and assessed performance metrics specifically designed for CI-processed speech. Conditions tested included acoustic degradation, N-of-M processing, spectral subtraction, and binaural noise reduction. The NCM method proved to be an accurate predictor of speech reception for both noise vocoder simulations of CI sound-processing as well as actual CI user performance. The other methods did not successfully predict intelligibility for all conditions tested. In the process of evaluating these performance metrics, we have shown that both spectral subtraction and binaural noise reduction improve the intelligibility of CI-processed speech.

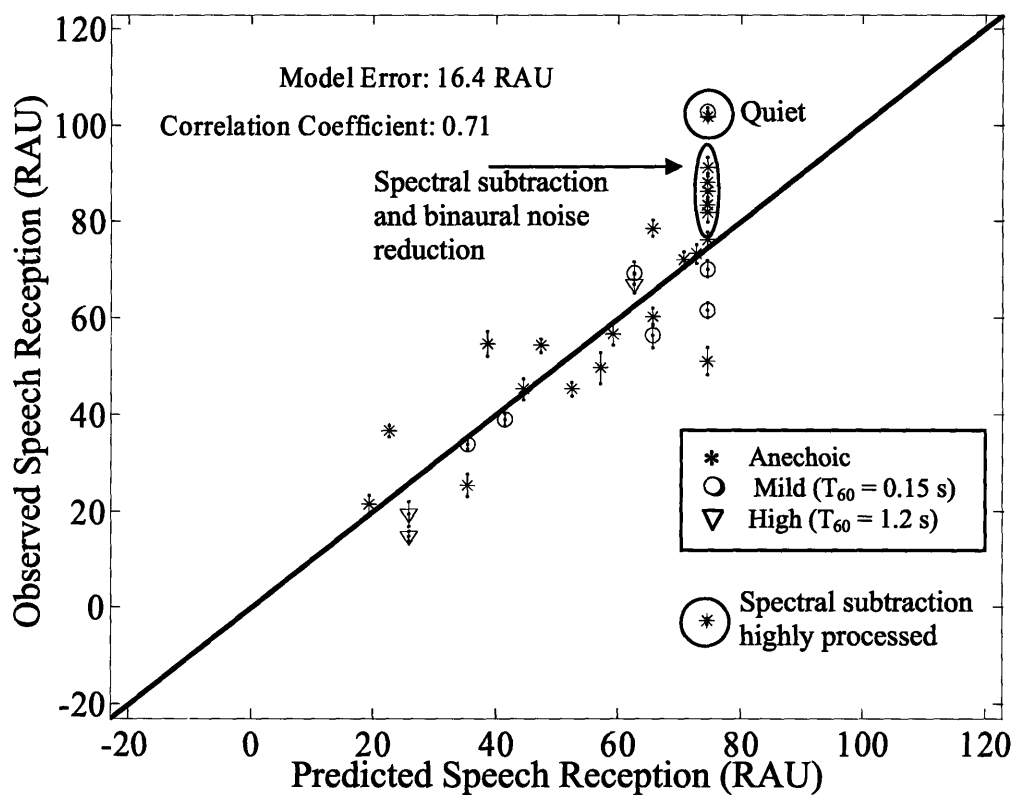


Figure 10.1: Comparison of observed scores for NH-Cl₈ and predicted scores from the envelope regression STI method. A single psychometric function is fitted to the data pooled across experiments. Conditions include acoustic degradation, spectral subtraction and binaural noise reduction.

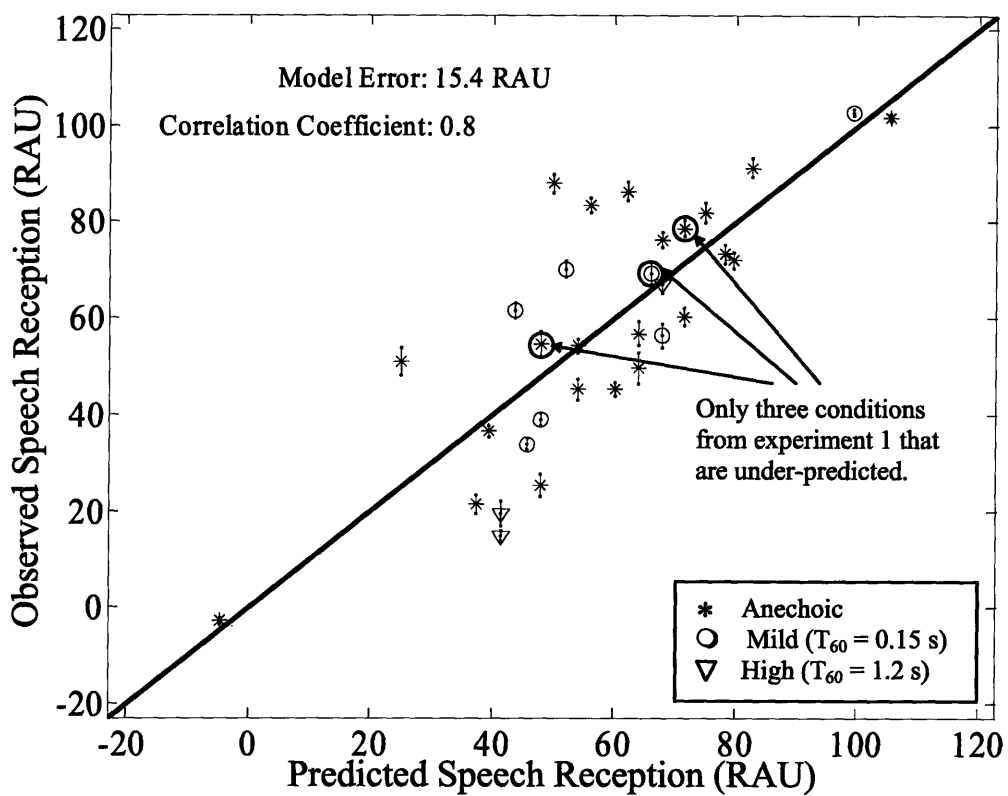


Figure 10.2: Comparison of observed scores for NH-Cl₈ and predicted scores from the modified envelope regression STI method. A single psychometric function is fitted to the data pooled across experiments. Conditions include acoustic degradation, spectral subtraction and binaural noise reduction.

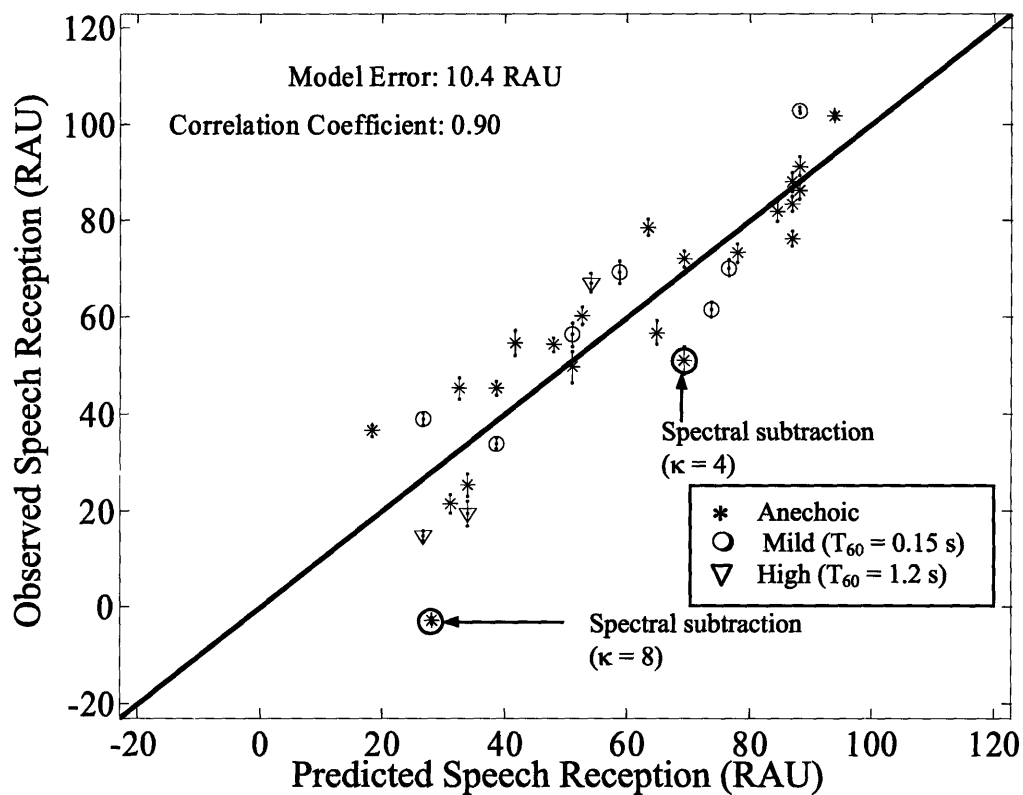


Figure 10.3: Comparison of observed scores for NH-Cl₈ and predicted scores from the NCM method. A single psychometric function is fitted to the data pooled across experiments. Conditions include acoustic degradation, spectral subtraction and binaural noise reduction.

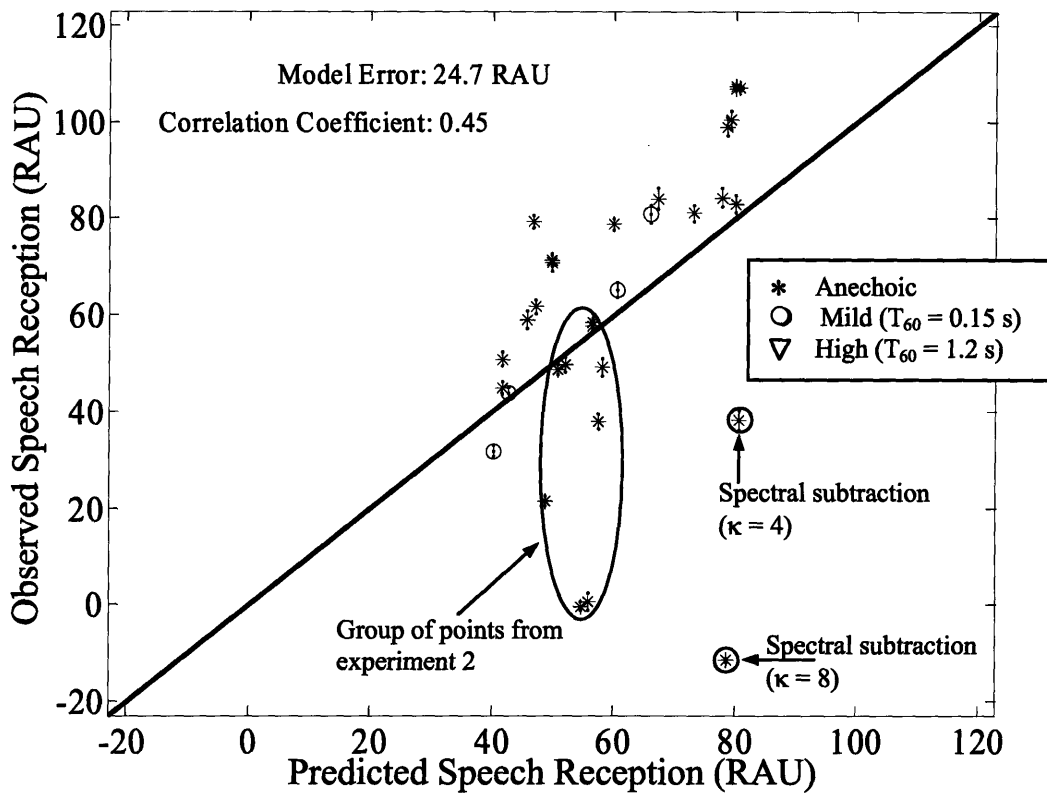


Figure 10.4: Comparison of observed scores for NH-Cl₂₀ and predicted scores from the envelope regression STI method. A single psychometric function is fitted to the data pooled across experiments. Conditions include N-of-M processing, spectral subtraction and binaural noise reduction.

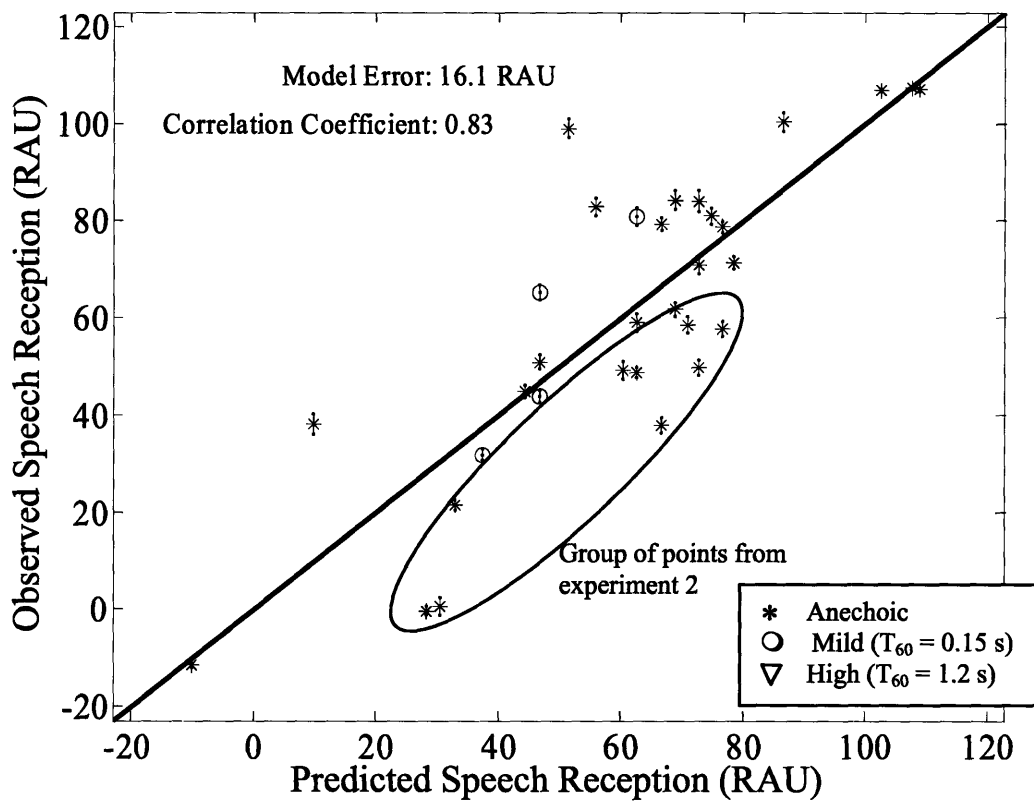


Figure 10.5: Comparison of observed scores for NH-Cl₂₀ and predicted scores from the modified envelope regression STI method. A single psychometric function is fitted to the data pooled across experiments. Conditions include N-of-M processing, spectral subtraction and binaural noise reduction.

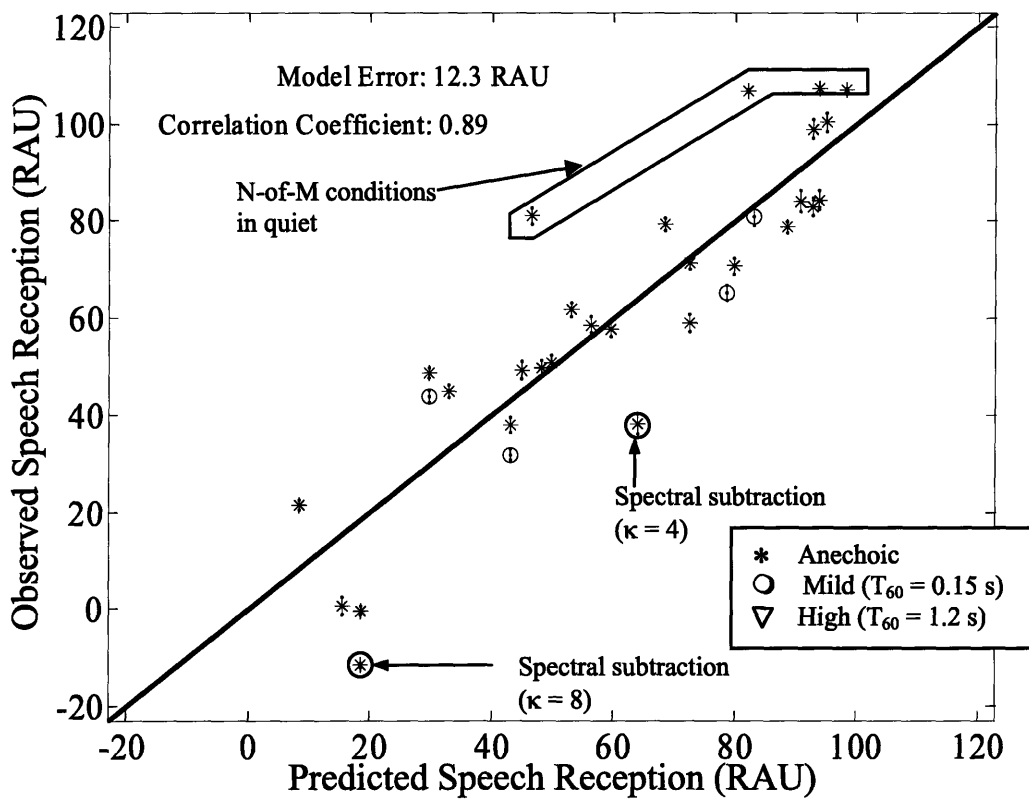


Figure 10.6: Comparison of observed scores for NH-Cl₂₀ and predicted scores from the NCM method. A single psychometric function is fitted to the data pooled across experiments. Conditions include N-of-M processing, spectral subtraction and binaural noise reduction.

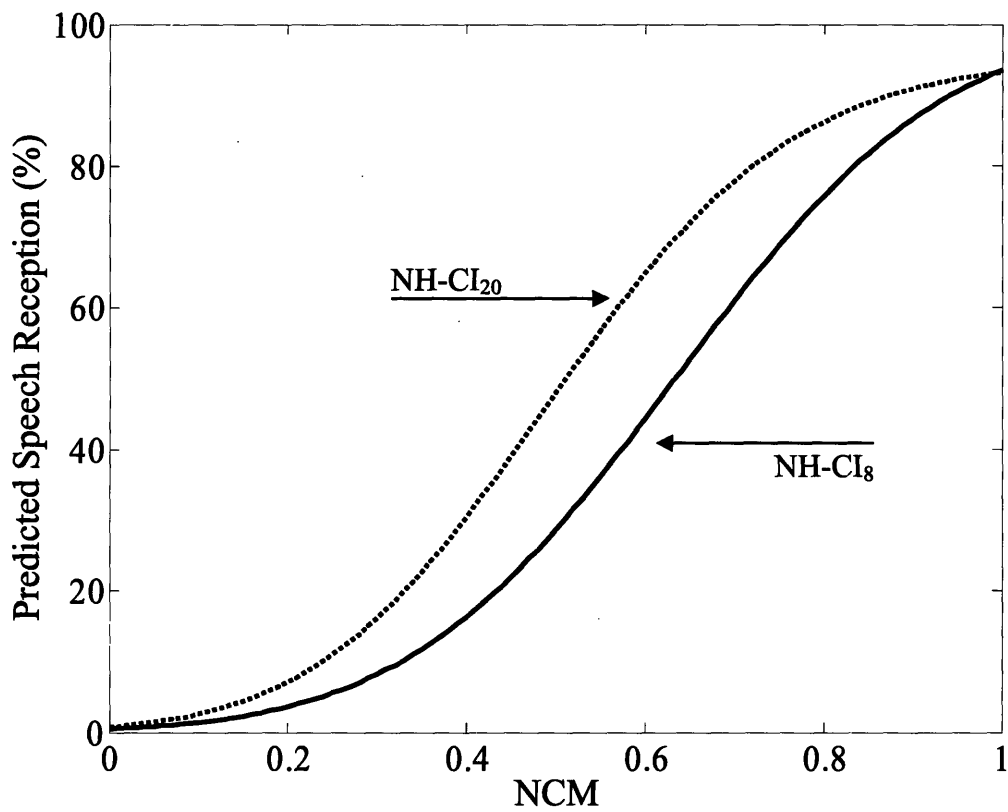


Figure 10.7: Comparison of psychometric functions for fitting NH-Cl₈ and NH-Cl₂₀ scores from the NCM method. A single psychometric function is fitted to the data pooled across experiments.

Appendix A

Selection of Candidate Metrics

Appendix A includes the correlation analysis that justifies the selection of three candidate metrics for detailed consideration in Chapters 6 through 9 of this thesis.

In this appendix we present correlation analysis between the various performance metrics considered in this thesis. This analysis helps justify the selection of three candidate metrics from among the nine metrics introduced in Chapters 2 and 3. Each table gives the correlation coefficients between pairs of metrics that were calculated on the clean and processed speech signals associated with each experiment. The details of the calculations are as given in Section 4.5. The abbreviations used in the tables are as follows: envelope regression STI (ER_α), real cross-power spectrum STI ($RCPS_\alpha$), magnitude cross-power spectrum STI ($MCPS_\alpha$), the associated (respectively) modified STIs (ER_β , $RCPS_\beta$, $MCPS_\beta$), the normalized correlation metric (NCM), the normalized covariance STI (NCov), and the normalized correlation STI (NCor).

Table A.1 gives the correlation analysis results for the acoustic degradations (Experiment 1). Note that the metrics are all well correlated with each other for these acoustic degradation conditions except for the normalized covariance STI. This further justifies the exclusion of the normalized covariance method from further consideration.

| | ER_α | $RCPS_\alpha$ | $MCPS_\alpha$ | ER_β | $RCPS_\beta$ | $MCPS_\beta$ | NCM | NCov | NCor |
|---------------|-------------|---------------|---------------|-------------|--------------|--------------|-------------|------|------|
| ER_α | 1.00 | | | | | | | | |
| $RCPS_\alpha$ | 1.00 | 1.00 | | | | | | | |
| $MCPS_\alpha$ | 0.98 | 0.98 | 1.00 | | | | | | |
| ER_β | 0.99 | 0.99 | 0.97 | 1.00 | | | | | |
| $RCPS_\beta$ | 0.99 | 0.99 | 0.97 | 1.00 | 1.00 | | | | |
| $MCPS_\beta$ | 0.96 | 0.97 | 0.99 | 0.98 | 0.97 | 1.00 | | | |
| NCM | 0.95 | 0.96 | 0.92 | 0.94 | 0.95 | 0.90 | 1.00 | | |
| NCov | 0.67 | 0.69 | 0.58 | 0.67 | 0.69 | 0.57 | 0.83 | 1.00 | |
| NCor | 0.99 | 0.99 | 0.95 | 0.98 | 0.98 | 0.94 | 0.97 | 0.77 | 1.00 |

Table A.1: Summary of the correlation analysis between intelligibility metrics for the acoustic degradation conditions. Metrics within a column emphasized by bold-face type are suggested as a single group.

Table A.2 summarizes the correlation analysis for the N-of-M processing conditions (Experiment 2). A key result is that the real and magnitude CPS methods are well correlated to the envelope regression method when using the same normalization procedure. That is, all of the α methods and β methods perform in a similar manner. In addition, the NCM, normalized covariance STI, and normalized correlation STI methods are well correlated justifying their grouping for these conditions. Note that the α and β

methods are not well correlated indicating the effect of the modification for these nonlinear conditions.

| | ER_α | $RCPS_\alpha$ | $MCPS_\alpha$ | ER_β | $RCPS_\beta$ | $MCPS_\beta$ | NCM | NCov | NCor |
|---------------|-------------|---------------|---------------|-------------|--------------|--------------|-------------|------|------|
| ER_α | 1.00 | | | | | | | | |
| $RCPS_\alpha$ | 0.99 | 1.00 | | | | | | | |
| $MCPS_\alpha$ | 0.98 | 0.98 | 1.00 | | | | | | |
| ER_β | 0.78 | 0.83 | 0.73 | 1.00 | | | | | |
| $RCPS_\beta$ | 0.77 | 0.81 | 0.72 | 1.00 | 1.00 | | | | |
| $MCPS_\beta$ | 0.74 | 0.79 | 0.70 | 0.99 | 1.00 | 1.00 | | | |
| NCM | 0.66 | 0.70 | 0.56 | 0.95 | 0.95 | 0.93 | 1.00 | | |
| NCov | 0.67 | 0.71 | 0.56 | 0.92 | 0.92 | 0.89 | 0.98 | 1.00 | |
| NCor | 0.70 | 0.74 | 0.61 | 0.98 | 0.98 | 0.96 | 1.00 | 0.97 | 1.00 |

Table A.2: Summary of the correlation analysis between intelligibility metrics for the N-of-M conditions. Metrics within a column emphasized by bold-face type are suggested as a single group.

Table A.3 summarizes the correlation analysis for the spectral subtraction conditions (Experiment 3). A key result is that the real and magnitude CPS methods are well correlated to the envelope regression method when using the same normalization procedure. That is, all of the α methods and β methods perform in a similar manner. In addition, the NCM, normalized covariance STI, and normalized correlation STI methods are well correlated justifying their grouping for these conditions. Note that the α and β methods are not well correlated indicating the effect of the modification for these nonlinear conditions.

| | ER_α | $RCPS_\alpha$ | $MCPS_\alpha$ | ER_β | $RCPS_\beta$ | $MCPS_\beta$ | NCM | NCov | NCor |
|---------------|-------------|---------------|---------------|-------------|--------------|--------------|-------------|------|------|
| ER_α | 1.00 | | | | | | | | |
| $RCPS_\alpha$ | 0.98 | 1.00 | | | | | | | |
| $MCPS_\alpha$ | 1.00 | 1.00 | 1.00 | | | | | | |
| ER_β | -0.45 | -0.31 | -0.39 | 1.00 | | | | | |
| $RCPS_\beta$ | -0.47 | -0.33 | -0.41 | 1.00 | 1.00 | | | | |
| $MCPS_\beta$ | -0.47 | -0.34 | -0.41 | 1.00 | 1.00 | 1.00 | | | |
| NCM | -0.01 | 0.15 | 0.06 | 0.87 | 0.85 | 0.85 | 1.00 | | |
| NCov | -0.13 | 0.03 | -0.06 | 0.92 | 0.91 | 0.91 | 0.93 | 1.00 | |
| NCor | 0.11 | 0.27 | 0.18 | 0.82 | 0.81 | 0.81 | 0.97 | 0.94 | 1.00 |

Table A.3: Summary of the correlation analysis between intelligibility metrics for the spectral subtraction conditions. Metrics within a column emphasized by bold-face type are suggested as a single group.

Table A.4 summarizes the correlation analysis for the binaural noise reduction conditions. A key result is that the real and magnitude CPS methods are well correlated to the envelope regression method when using the same normalization procedure. That is, all of the α methods and β methods perform in a similar manner. In addition, the NCM, normalized covariance STI, and normalized correlation STI methods are well correlated justifying their grouping for these conditions. Note that the α and β methods are not well correlated indicating the effect of the modification for these nonlinear conditions.

| | ER_α | $RCPS_\alpha$ | $MCPS_\alpha$ | ER_β | $RCPS_\beta$ | $MCPS_\beta$ | NCM | NCov | Ncor |
|---------------|-------------|---------------|---------------|-------------|--------------|--------------|-------------|------|------|
| ER_α | 1.00 | | | | | | | | |
| $RCPS_\alpha$ | 1.00 | 1.00 | | | | | | | |
| $MCPS_\alpha$ | 0.99 | 0.99 | 1.00 | | | | | | |
| ER_β | 0.55 | 0.55 | 0.54 | 1.00 | | | | | |
| $RCPS_\beta$ | 0.57 | 0.57 | 0.56 | 1.00 | 1.00 | | | | |
| $MCPS_\beta$ | 0.37 | 0.38 | 0.41 | 0.91 | 0.91 | 1.00 | | | |
| NCM | 0.98 | 0.98 | 0.95 | 0.58 | 0.60 | 0.34 | 1.00 | | |
| NCov | 0.80 | 0.80 | 0.72 | 0.62 | 0.63 | 0.29 | 0.89 | 1.00 | |
| NCor | 0.97 | 0.97 | 0.95 | 0.64 | 0.66 | 0.42 | 0.99 | 0.89 | 1.00 |

Table A.4: Summary of the correlation analysis between intelligibility metrics for the binaural noise reduction conditions. Metrics within a column emphasized by bold-face type are suggested as a single group

Appendix B

Theoretical Derivations Concerning the STI and NCM Methods

B.1 Stochastic Reformulation of the Envelope Regression Method¹⁶

The following is a stochastic reformulation of the envelope regression method that facilitates comparison with other methods. The reformulation begins with the assumption that the linear regression of the sampled degraded envelope, $y[n]$, onto the sampled clean envelope, $x[n]$, is performed using a minimum mean-square-error criterion (Ross, 1998). In this case, the optimal fit is

$$y_{MMSE}[n] = \mu_y + \frac{\lambda_{xy}}{\lambda_x}(x[n] - \mu_x), \quad (\text{B.1})$$

where λ_{xy} and λ_x are defined in Eqs. 2.12 and 2.13. Thus, the slope (A) and the y-intercept (B) calculated using a minimum mean-square-error criterion are

$$A = \frac{\lambda_{xy}}{\lambda_x} \quad (\text{B.2})$$

and

$$B = \mu_y - \frac{\lambda_{xy}}{\lambda_x} \mu_x. \quad (\text{B.3})$$

Substituting Eqs. B.2 and B.3 into 2.9 and rearranging allows the apparent SNR to be expressed as

$$aSNR = 10 \log_{10} \left(\frac{M}{1 - M} \right), \quad (\text{B.4})$$

where M is a modulation metric defined as

$$M = \frac{\mu_x \lambda_{xy}}{\mu_y \lambda_x}. \quad (\text{B.5})$$

¹⁶ Appendix B.1 and B.2 are reproduced from Goldsworthy and Greenberg, 2004: Appendix A. Changes were made to equation numbers to be internally consistent with this thesis.

B.2 Normalized Correlation Expressed as an Energy-Weighted MTF¹⁷

The normalized correlation is defined as

$$\rho = \frac{\phi_{xy}}{\phi_x^{1/2} \phi_y^{1/2}}. \quad (\text{B.6})$$

Using the relationship between the cross-correlation function, $R_{xy}[k]$, and the cross-power spectrum, $S_{xy}(f)$ (Papoulis, 1984), together with the observation that ϕ_{xy} equals the cross-correlation function computed at zero lag, yields

$$\phi_{xy} = R_{xy}[0] = \int_{f=-1/2}^{1/2} S_{xy}(f) df, \quad (\text{B.7})$$

where $\phi_{xy} \triangleq E\{x[n]y[n]\}$ and $R_{xy}[k] \triangleq E\{x[n]y[n-k]\}$. The normalized correlation can then be expressed as

$$\rho = \frac{\int_{f=-1/2}^{1/2} S_{xy}(f) df}{\phi_x^{1/2} \phi_y^{1/2}}. \quad (\text{B.8})$$

Bringing the denominator inside the integral and multiplying numerator and denominator by the same terms yields

$$\rho = \int_{f=-1/2}^{1/2} \left(\frac{\phi_x}{\phi_y} \right)^{1/2} \left[\frac{S_{xy}(f)}{S_{xx}(f)} \right] \left[\frac{S_{xx}(f)}{\phi_x} \right] df. \quad (\text{B.9})$$

Defining a new MTF,

$$MTF_{\rho}(f) \triangleq \left(\frac{\phi_x}{\phi_y} \right)^{1/2} \frac{S_{xy}(f)}{S_{xx}(f)}, \quad (\text{B.10})$$

and a weighting function,

$$W(f) \triangleq \frac{S_{xx}(f)}{\phi_x}, \quad (\text{B.11})$$

¹⁷ Appendix B.1 and B.2 are reproduced from Goldsworthy and Greenberg, 2004: Appendix A. Changes were made to equation numbers to be internally consistent with this thesis.

allows describing ρ as an energy-weighted average of this new MTF, that is,

$$\rho = \int_{f=-1/2}^{1/2} MTF_{\rho}(f) \cdot W(f) df. \quad (\text{B.12})$$

The weighting function, $W(f)$, is the ratio of the power of the clean envelope at each modulation frequency to the total power in the clean envelope.

The MTF defined in Eq. B.12 is similar in form to the MTFs defined for the cross-power spectrum methods. All three MTFs are based on the normalized ratio of the cross-power spectrum between clean and degraded envelopes to the power spectrum of the clean envelope. The main differences are the factor used for normalization ($\sqrt{\phi_x / \phi_y}$ rather than $\alpha = \mu_x / \mu_y$) and the fact that in Eq. B.12 the MTF is complex-valued. However, since $S_{xx}(f)$ is real and symmetric, and $S_{xy}(f)$ is complex-conjugate symmetric, the integral over equal ranges of positive and negative frequencies will be real-valued.

B.3 Modulation Metric (M) Expressed as an Energy-Weighted MTF

A similar derivation as given in Section B.2 exists for relating the modulation metric, M (Eq. B.5), to an energy-weighted MTF. The derivation is similar in form to that given in Section B.2, but is more complex since M is based on covariance terms rather than correlation. These covariance variables can be expressed as

$$\lambda_{xy} = C_{xy}[0] = R_{xy}[0] - \mu_x \mu_y = \int_{f=-1/2}^{1/2} S_{xy}(f) df - \mu_x \mu_y \quad (\text{B.13})$$

and

$$\lambda_x = C_{xx}[0] = R_{xx}[0] - \mu_x^2 = \int_{f=-1/2}^{1/2} S_{xx}(f) df - \mu_x^2. \quad (\text{B.14})$$

The next step is to bring the mean terms ($\mu_x\mu_y$ and μ_x^2) into the integrand. It should be noted that multiple ways exist of completing this step yielding different interpretations of the variables. One useful method is to express the mean terms as

$$\mu_x\mu_y = \int_f \mu_x\mu_y\delta(f) \quad \text{and} \quad \mu_x^2 = \int_f \mu_x^2\delta(f)df, \quad (\text{B.15})$$

where $\delta(f)$ is the Dirac delta function. The covariance variables can be expressed as

$$\lambda_{xy} = \int_f S_{xy}(f)df - \int_f \mu_x\mu_y\delta(f)df = \int_f [S_{xy}(f) - \mu_x\mu_y\delta(f)]df \quad (\text{B.16})$$

and

$$\lambda_x = \int_f S_{xx}(f)df - \int_f \mu_x^2\delta(f)df = \int_f [S_{xx}(f) - \mu_x^2\delta(f)]df. \quad (\text{B.17})$$

Then defining

$$\tilde{S}_{xy}(f) = S_{xy}(f) - \mu_x\mu_y\delta(f) \quad (\text{B.18})$$

and

$$\tilde{S}_{xx}(f) = S_{xx}(f) - \mu_x^2\delta(f), \quad (\text{B.19})$$

the modulation metric, M (Eq. B.5), can then be expressed as

$$M = \frac{\mu_x}{\mu_y} \frac{\int_f \tilde{S}_{xy}(f)df}{\int_{f'} \tilde{S}_{xx}(f')df'}. \quad (\text{B.20})$$

The prime notation is used to distinguish the denominators variables of integration to make the following equations accurate. Multiplying inside the numerator's integrand by $\tilde{S}_{xx}(f)/\tilde{S}_{xx}(f)$ and rearranging terms allows M to be expressed as

$$M = \int_f \left[\frac{\mu_x}{\mu_y} \frac{\tilde{S}_{xy}(f)}{\tilde{S}_{xx}(f)} \right] \left[\frac{\tilde{S}_{xx}(f)}{\int_{f'} \tilde{S}_{xx}(f')df'} \right] df. \quad (\text{B.21})$$

The second term in brackets ([-]) is a weighting function based on the energy in the clean envelope signal. A modified phase-locked MTF is defined as

$$M\tilde{T}F(f) = \frac{\mu_x \tilde{S}_{xy}(f)}{\mu_y \tilde{S}_{xx}(f)}, \quad (\text{B.22})$$

and a corresponding weighting function as

$$\tilde{W}(f) \triangleq \frac{\tilde{S}_{xx}(f)}{\int_f \tilde{S}_{xx}(f') df'}, \quad (\text{B.23})$$

thus allowing the modulation metric to be defined as an energy-weighted average of this modified phase-locked MTF,

$$M = \int_f M\tilde{T}F(f) \cdot \tilde{W}(f). \quad (\text{B.24})$$

It should be noted that the MTF defined in Eq. B.22 is based on power spectral densities (Eqs. B.18 and B.19) that are modified to account for the envelope means. However, this modification only affects the DC frequency value, and consequently, the MTF of Eq. B.22 can be written as:

$$M\tilde{T}F(f) = \begin{cases} \frac{\mu_x S_{xy}(f) - \mu_x \mu_y}{\mu_y S_{xx}(f) - \mu_x^2} & f = 0 \\ \frac{\mu_x S_{xy}(f)}{\mu_y S_{xx}(f)} & \text{otherwise.} \end{cases} \quad (\text{B.25})$$

It should also be noted that the phase-locked MTF is usually defined as the real part of the ratio of the cross-spectral density to the auto spectral density. For the above discussion, taking the real part was unnecessary since the cross-spectral density is complex-conjugate symmetric and since the integration occurs over a symmetric range of positive and negative frequencies.

B.4 Relation of Energy-Weighted to One-Third Octave Averaging

The derivations given in Sections B.2 and B.3 illustrate that the intermediate metrics of the normalized correlation and envelope regression methods can be expressed as energy-weighted averages of alternate MTFs. In this section, energy-weighted and traditional one-third octave averaging are compared.

The comparison is facilitated by considering the one-third octave procedure in terms of a weighting function expressed in terms of frequency. The one-third octave procedure is based on averaging a number of frequency values that are logarithmically spaced with equal contribution. As such, the frequency contribution of a bin centered at $2f$ has the same contribution to the resulting apparent SNR as a bin centered at f despite being twice the size. In general, specifying a discrete set of logarithmically spaced bins having equal contribution is comparable to using a weighting function that is inversely proportional to frequency,

$$W(f) = \frac{1/f_i}{\sum_i (1/f_i)}, \quad (\text{B.26})$$

where f_i is the center frequency of the i^{th} bin and the denominator insures that the weighting function sums to 1.

The energy-weighted weighting functions are calculated from clean speech signals. We calculate the function here using the concatenation of 4 lists of sentences from the IEEE database. The intensity envelope signal is calculated for an octave-band centered at 1 kHz using square-law rectification and using a 50 Hz lowpass filter. The envelope signals are down-sampled to 200 Hz and the power spectra are calculated using a 4096-point FFT. The energy-weighted weighting function of Eq. B.11 is calculated from the resulting power-spectra normalized by the total energy of the signal.

Figure B.1 illustrates the resulting weighting functions for the one-third octave and the energy-weighted procedures. The one-third octave procedure is generated using Eq. B.26 with a maximum modulation frequency of 20 Hz. The two weighting functions are similar in that both place emphasis on the low modulation frequencies. A primary difference is that the energy-weighted function shows a more constant contribution for modulation frequencies between 0 and 4 Hz. Further, the weighting function for the one-

third octave method exists only when the corresponding bin is included in the summation. Typically, bins with center frequencies ranging from 0.6 to 12.7 Hz are included. The bin centered at 12.7 would have a maximum edge at 16 Hz, thus frequency components for that case would be zero above 16 Hz.

Figure B.2 illustrates the cumulative summations of the respective weighting functions. The cumulative summations are useful for comparing the two methods. It is clear from Figure B.2 that the energy-weighted method places more emphasis on lower modulation frequencies. In fact, 90% of the cumulative weight occurs between 0 and 6 Hz for the energy-weighted method. Also note that the energy-weighted method has nearly 100% of its cumulative weight for frequencies less than 20 Hz despite containing energy up to 50 Hz. In other words, the envelope signal energy is very low for frequencies above 20 Hz, thus the contribution of higher frequencies to the energy-weighted method is negligible.

The energy-weighted and one-third octave weighting functions are similar as seen in Figures B.1 and B.2; however, the functions are slightly different. An examination of the results given in this thesis (Appendix A) indicates that this difference does not result in substantially different STI values. As such, we propose using the energy-weighted (i.e. envelope regression) methods since they are much more efficient to compute.

B.5 Relation of Apparent SNR to True SNR

Traditional Method

For the traditional STI methods, the MTF for additive stationary noise can be expressed as

$$MTF = \frac{\mu_x / \mu_n}{\mu_x / \mu_n + 1} \quad (\text{B.27})$$

(Houtgast and Steeneken, 1973) where μ_n is the mean intensity of the noise envelope. Plugging this MTF into Eq. 2.2 allows the apparent SNR for additive stationary noise to be expressed as

$$SNR = 10 \log_{10} \left(\frac{\mu_x}{\mu_n} \right). \quad (\text{B.28})$$

Thus, for the case of additive stationary noise, the apparent SNR is equal to the true SNR. However, this result depends on using intensity envelopes when calculating the MTF. If magnitude envelopes were used instead, the apparent and true SNRs would not be equivalent.

Normalized Correlation Method

The normalized correlation method is an extension of the method proposed by Holube and Kollmeier (1996). The basis of the normalized covariance method is its simple relation to the SNR. To see this, consider a signal, $x(t)$, degraded by noise, $n(t)$,

$$y(t) = x(t) + n(t). \quad (\text{B.29})$$

The normalized covariance between two zero-mean signals can be written in terms of statistical expectations as

$$r^2 = \frac{E^2[x(t)y(t)]}{E[x^2(t)]E[y^2(t)]}. \quad (\text{B.30})$$

Assuming that $x(t)$ and $n(t)$ are uncorrelated then the expectations are given as

$$E[x(t)y(t)] = \sigma_x^2, \quad E[x^2(t)] = \sigma_x^2, \quad \text{and} \quad E[y^2(t)] = \sigma_x^2 + \sigma_n^2, \quad (\text{B.31})$$

where σ_x^2 and σ_n^2 are the variances of $x(t)$ and $n(t)$ respectively. Thus, the normalized covariance can be written as

$$r^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}. \quad (\text{B.32})$$

Consequently,

$$\frac{r^2}{1-r^2} = \frac{\sigma_x^2}{\sigma_n^2} = \text{SNR}. \quad (\text{B.33})$$

Thus, Holube and colleagues developed a quick and reliable method for estimating SNR from zero-mean signals. The above derivation requires that the signals are zero-mean and does not hold for envelope signals. Thus, in contrast to Eq. B.28 which is based on

envelope signals, the SNR calculated using Eq. B.33 does not produce SNR values that correspond to the apparent SNR values of STI theory. The normalized covariance STI method tested in this thesis is based on envelope signals. Thus, the above derivation does not hold and the consequent metric produces results that are different from those calculated using the more traditional methods.

B.6 Effect of Using ρ^2 Directly as the Transmission Index

In Section 5.3 we justify the selection of three candidate metrics for consideration based on correlation analysis. One of the metrics chosen, the NCM, excludes certain calculation procedures normally taken in STI methods. The procedures excluded for the NCM method are a transformation from ρ to an apparent SNR (Eq. 2.10, replacing r with ρ), a clipping to ± 15 dB (Eq. 2.3), and a linear scaling (Eq. 2.5) to produce values between 0 and 1. One rationale for the exclusion of these procedures is that while reasonable for STI where the apparent and true SNRs are theoretically equal (see Section B.5), they are not useful for the normalized correlation approach since a simple relationship does not exist between apparent and true SNR when calculated on envelope signals (Section B.5). Thus, the transformation embodied by Equations 2.2 through 2.5 may be little more than an added inconvenience.

However, it is important to comprehend the analytical effect that these procedures have on the resulting TI value. Figure B.3 illustrates TI values for the normalized correlation STI method and the NCM method as a function of ρ^2 . The effect of excluding the procedures is to skew the weight of particular TI values. TI values ranging from 0.05 to 0.5 are decreased while values between 0.5 and 0.95 are increased. However, the maximum change resulting from the exclusion of the procedures is approximately 0.1 and occurs near values of 0.2 and 0.8. The transformation is one-to-one and monotonically increasing (except for a small range of values less than 0.04 and greater than 0.96 where the function is flat). The monotonicity of the transformation is relevant since it implies that the ordering of TI values with and without the transformation will be the same (thus furthering the argument that the transformation is irrelevant).

However, the role of the transformation may be important when combining the resulting TI values across frequency bands. Including the transformation will produce slightly different overall NCM values. However, in light of the fact that it is not known if these slight differences will in any way increase the accuracy of the model, and considering that there is no theoretical justification of including the transformation, we chose to exclude the transformation.

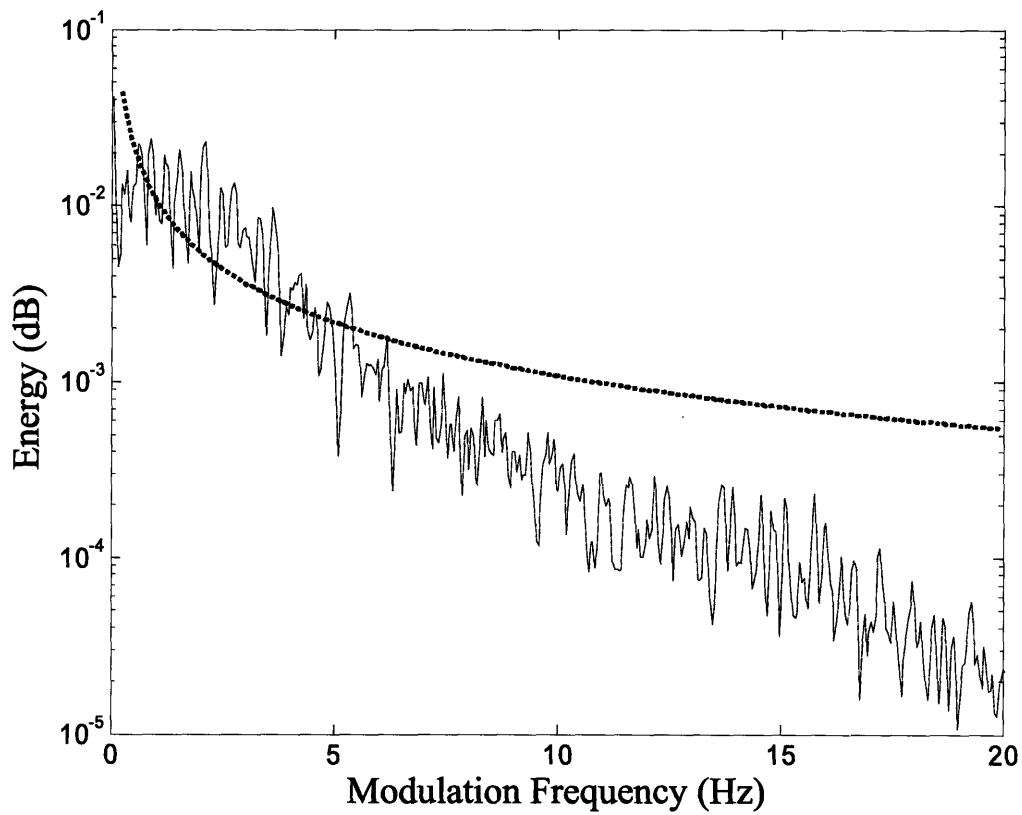


Figure B.1: Energy of speech envelopes as a function of modulation frequency. Energy is normalized such that the cumulative sum is one. The solid line represents actual energy calculated with clean speech signals. The dotted line represents the energy of a signal that would have equal energy per $1/3^{\text{rd}}$ octave bins.

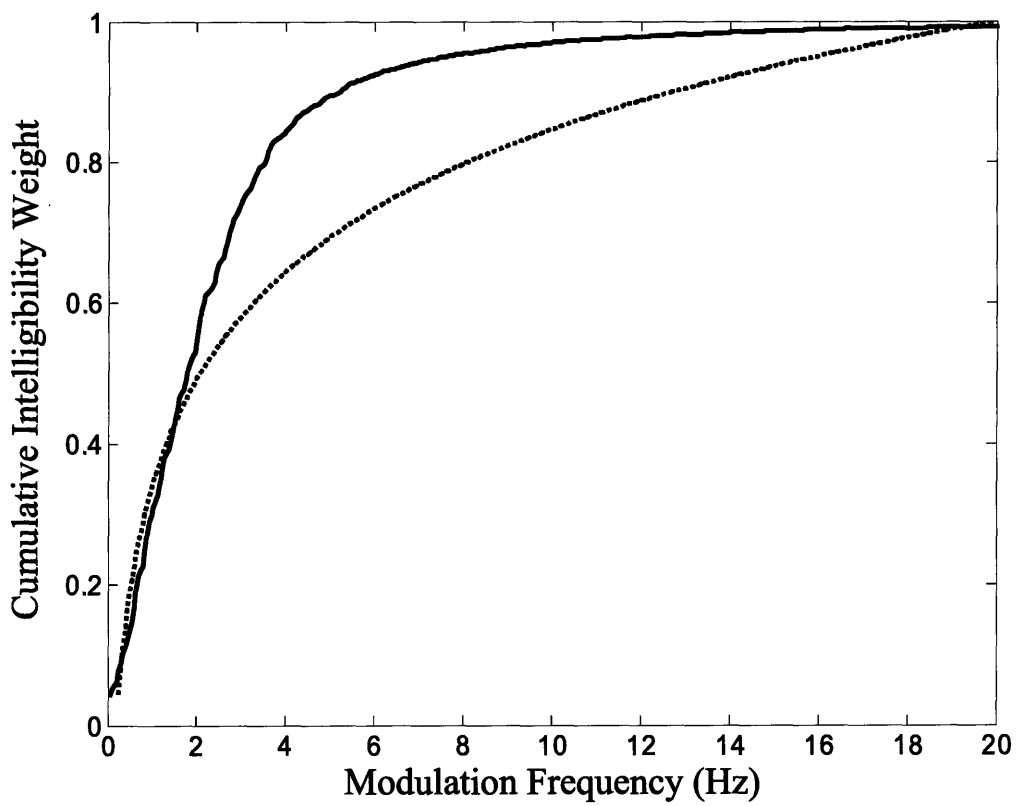


Figure B.2: Same as Figure B.1 except represented as cumulative energy.

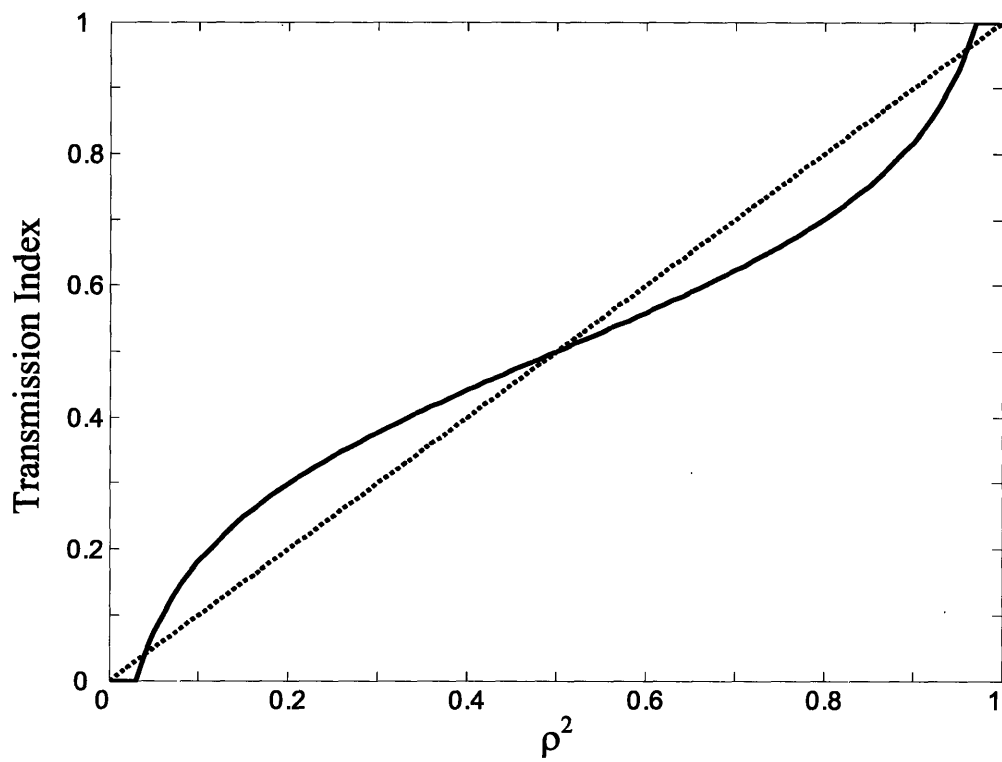


Figure B.3: The solid line illustrates the TI values calculated using the normalized-correlation STI method which *includes* the apparent SNR transformation. The dotted line illustrates the TI values calculated using the NCM method (i.e. $TI = \rho^2$) which *excludes* the apparent SNR transformation..

Appendix C

Repeated Measures Analysis of Variance

Tables

In this appendix we give the repeated measures analysis of variance tables associated with each main experiment and subject group.

C.1 Experiment 1: Acoustic Degradation

C.1.1 NH-Cl₈ RMANOVAs

| Source | SS | d.f. | MS | F | p |
|------------------|--------|------|------------|-------|-----------------|
| Subject | 11196 | 5 | 2239 | 19.7 | 0 ¹⁸ |
| Condition | 214420 | 15 | 14295 | 125.9 | 0 |
| S x C | 7619 | 75 | 102 | 0.9 | 0.707 |
| Error | 21800 | 192 | 114 | | |
| Total | 255035 | 287 | | | |

Table C.1: RMANOVA_1 for NH-Cl₈ as described in Section 6.3.

| Source | SS | d.f. | MS | F | p |
|-------------------|--------|------|------------|-------|-------|
| Subject | 6140 | 5 | 1228 | 11.9 | 0 |
| Noise Type | 68042 | 2 | 34021 | 330.6 | 0 |
| Reverb | 77618 | 2 | 38809 | 377.1 | 0 |
| S x NT | 411 | 10 | 41 | 0.4 | 0.944 |
| S x R | 1538 | 10 | 154 | 1.5 | 0.151 |
| NT x R | 3086 | 4 | 772 | 7.5 | 0 |
| S x NT x R | 2694 | 20 | 135 | 1.3 | 0.189 |
| Error | 13809 | 108 | 103 | | |
| Total | 170645 | 161 | | | |

Table C.2: RMANOVA_2 for NH-Cl₈ as described in Section 6.3.

| Source | SS | d.f. | MS | F | p |
|-------------------|-------|------|------------|-------|-------|
| Subject | 7008 | 5 | 1402 | 11.9 | 0 |
| Noise Type | 536 | 2 | 268 | 2.3 | 0.108 |
| Level | 52602 | 2 | 26301 | 222.8 | 0 |
| S x NT | 662 | 10 | 66 | 0.6 | 0.842 |
| S x L | 1171 | 10 | 117 | 1.0 | 0.455 |
| NT x L | 6125 | 4 | 1531 | 13.0 | 0 |
| S x NT x L | 2367 | 20 | 118 | 1.0 | 0.465 |
| Error | 15114 | 108 | 118 | | |
| Total | 83218 | 161 | | | |

Table C.3: RMANOVA_3 for NH-Cl₈ as described in Section 6.3.

¹⁸ p values listed as 0 indicate that $p < 0.0001$.

C.1.2 CI Subjects RMANOVAs

| Source | SS | d.f. | MS | F | p |
|------------------|--------|------|------------|-------|-------|
| Subject | 32487 | 2 | 16243 | 128.8 | 0 |
| Condition | 77715 | 15 | 5181 | 41.1 | 0 |
| S x C | 6348 | 30 | 212 | 1.7 | 0.031 |
| Error | 12110 | 96 | 126 | | |
| Total | 128660 | 143 | | | |

Table C.4: RMANOVA_1 for CI users as described in Section 6.3.

| Source | SS | d.f. | MS | F | p |
|-------------------|-------|------|------------|-------|-------|
| Subject | 17543 | 2 | 8771 | 66.2 | 0 |
| Noise Type | 9875 | 2 | 4937 | 37.3 | 0 |
| Reverb | 57365 | 2 | 28682 | 216.5 | 0 |
| S x NT | 544 | 4 | 136 | 1.0 | 0.402 |
| S x R | 2647 | 4 | 662 | 5.0 | 0.002 |
| NT x R | 691 | 4 | 173 | 1.3 | 0.280 |
| S x NT x R | 959 | 8 | 120 | 0.9 | 0.519 |
| Error | 8113 | 54 | 132 | | |
| Total | 96778 | 80 | | | |

Table C.5: RMANOVA_2 for CI users as described in Section 6.3.

| S | SS | d.f. | MS | F | P |
|-------------------|-------|------|------------|-----|-------|
| Subject | 20439 | 2 | 10219 | 92 | 0 |
| Noise Type | 1656 | 2 | 828 | 7 | 0.002 |
| Level | 10932 | 2 | 5466 | 49 | 0 |
| S x NT | 1074 | 4 | 269 | 2 | 0.063 |
| S x L | 363 | 4 | 91 | 1 | 0.527 |
| NT x L | 1516 | 4 | 379 | 3 | 0.016 |
| S x NT x L | 793 | 8 | 99 | 0.9 | 0.538 |
| Error | 6872 | 62 | 111 | | |
| Total | 42852 | 80 | | | |

Table C.6: RMANOVA_3 for CI users as described in Section 6.3.

C.2 Experiment 2: N-of-M Processing

C.2.1 NH-Cl₈ RMANOVAs

| Source | SS | d.f. | MS | F | p |
|------------------|--------|------|------------|-------|-------|
| Subject | 7044 | 7 | 1006 | 7.7 | 0 |
| Condition | 463729 | 15 | 30915 | 236.5 | 0 |
| S x C | 10265 | 105 | 98 | 0.7 | 0.956 |
| Error | 33457 | 256 | 131 | | |
| Total | 514495 | 383 | | | |

Table C.7: RMANOVA_1 for NH-Cl₂₀ as described in Section 7.3.

| Source | SS | d.f. | MS | F | p |
|--------------------------|--------|------|------------|-------|-------|
| Subject | 7044 | 7 | 1006 | 7.7 | 0 |
| Noise Type | 250685 | 3 | 83562 | 639.4 | 0 |
| N (# of Channels) | 180235 | 3 | 60078 | 459.7 | 0 |
| S x NT | 1810 | 21 | 86 | 0.7 | 0.870 |
| S x N | 3920 | 21 | 187 | 1.4 | 0.105 |
| NT x N | 32810 | 9 | 3646 | 27.9 | 0 |
| S x NT x N | 4535 | 72 | 72 | 0.6 | 0.997 |
| Error | 37993 | 256 | 131 | | |
| Total | 514495 | 383 | | | |

Table C.8: RMANOVA_2 for NH-Cl₂₀ as described in Section 7.3.

C.2.2 CI Subjects RMANOVAs

No CI Subjects for N-of-M Experiment

C.3 Experiment 3: Spectral Subtraction

C.3.1 NH-Cl_{Sim} RMANOVAs

| Source | SS | d.f. | MS | F | p |
|------------------|--------|------|------------|-------|-------|
| Subject | 7349 | 7 | 1050 | 8.9 | 0 |
| Condition | 363162 | 15 | 24211 | 206.2 | 0 |
| S x C | 13527 | 105 | 129 | 1.1 | 0.277 |
| Error | 30060 | 256 | 117 | | |
| Total | 414097 | 383 | | | |

Table C.9: RMANOVA_1 for NH-Cl_{Sim} as described in Section 8.3.

| Source | SS | d.f. | MS | F | p |
|-------------------------------------|--------|------|-------|-------|-------|
| Subject | 7349 | 7 | 1050 | 8.9 | 0 |
| CI_{Sim} | 726 | 1 | 726 | 6.2 | 0.014 |
| κ | 358915 | 7 | 51274 | 436.7 | 0 |
| S x CI | 912 | 7 | 130 | 1.1 | 0.358 |
| S x κ | 8495 | 49 | 173 | 1.5 | 0.029 |
| CI x κ | 3521 | 7 | 503 | 4.3 | 0 |
| S x CI x κ | 4120 | 49 | 84 | 0.7 | 0.920 |
| Error | 34180 | 256 | 117 | | |
| Total | 414097 | 383 | | | |

Table C.10: RMANOVA_2 for NH-CI_{Sim} as described in Section 8.3.

C.3.2 CI User RMANOVAs

| Source | SS | d.f. | MS | F | p |
|--------------------------------|-------|------|------|------|-------|
| Subject | 1617 | 2 | 808 | 7.6 | 0.001 |
| κ | 41932 | 7 | 5990 | 56.3 | 0 |
| S x κ | 2372 | 14 | 169 | 1.6 | 0.116 |
| Error | 5103 | 48 | 106 | | |
| Total | 51024 | 71 | | | |

Table C.11: RMANOVA_1 for CI users as described in Section 8.3.

C.4 Experiment 4: Binaural Noise Reduction

C.4.1 NH-CI_{Sim} RMANOVAs

| Source | SS | d.f. | MS | F | p |
|------------------|--------|------|-------|-------|-------|
| Subject | 9834 | 7 | 1405 | 18.1 | 0 |
| Condition | 199387 | 15 | 13292 | 171.1 | 0 |
| S x C | 10141 | 105 | 97 | 1.2 | 0.086 |
| Error | 19890 | 256 | 78 | | |
| Total | 239252 | 383 | | | |

Table C.12: RMANOVA_1 for NH-CI_{Sim} as described in Section 9.3.

| Source | SS | d.f. | MS | F | p |
|-------------------------|-----------|-------------|-----------|----------|----------|
| Subject | 9833.5 | 7 | 1405 | 18.1 | 0 |
| CI_{Sim} | 789 | 1 | 789 | 10.2 | 0.002 |
| Noise Type | 1861 | 1 | 1861 | 23.9 | 0 |
| Reverb | 37539 | 1 | 37539 | 483.1 | 0 |
| Binaural | 146176 | 1 | 146176 | 1881.4 | 0 |
| S x CI | 318 | 7 | 45 | 0.6 | 0.769 |
| S x NT | 538 | 7 | 77 | 1.0 | 0.439 |
| S x R | 1180 | 7 | 169 | 2.2 | 0.037 |
| S x B | 1073 | 7 | 153 | 2.0 | 0.059 |
| CI x NT | 266 | 1 | 266 | 3.4 | 0.066 |
| CI x R | 38 | 1 | 38 | 0.5 | 0.482 |
| CI x B | 1587 | 1 | 1587 | 20.4 | 0 |
| NT x R | 3883 | 1 | 3883 | 50.0 | 0 |
| NT x B | 1308 | 1 | 1308 | 16.8 | 0 |
| R x B | 5443 | 1 | 5443 | 70.1 | 0 |
| Error | 27420 | 256 | 78 | | |
| Total | 239252 | 383 | | | |

Table C.13: RMANOVA_2 for NH-CI_{Sim} as described in Section 9.3. Values are calculated including all higher order interactions between variables; however, all higher order interactions were found to be not significant and are excluded from this summary.

C.4.2 CI Subjects RMANOVAs

| Source | SS | d.f. | MS | F | p |
|-----------|-------|------|-------|-------|-------|
| Subject | 20355 | 2 | 10177 | 103.2 | 0 |
| Condition | 30012 | 7 | 4287 | 43.5 | 0 |
| S x C | 4273 | 14 | 305 | 3.1 | 0.002 |
| Error | 4734 | 48 | 99 | | |
| Total | 59374 | 71 | | | |

Table C.14: RMANOVA_1 for CI users as described in Section 9.3.

| Source | SS | d.f. | MS | F | p |
|----------------|-------|------|-------|-------|--------|
| Subject | 20355 | 2 | 10177 | 103.2 | 0 |
| Noise Type | 2674 | 1 | 2674 | 27.1 | 0 |
| Reverb | 6209 | 1 | 6209 | 63.0 | 0 |
| Binaural | 18747 | 1 | 18747 | 190.1 | 0 |
| S x NT | 796 | 2 | 398 | 4.0 | 0.0457 |
| S x R | 991 | 2 | 495 | 5.0 | 0.0226 |
| S x B | 273 | 2 | 136 | 1.4 | 0.3343 |
| NT x R | 326 | 1 | 326 | 3.3 | 0.1078 |
| NT x B | 314 | 1 | 314 | 3.2 | 0.1142 |
| R x B | 1731 | 1 | 1731 | 17.5 | 0.0004 |
| S x NT x R | 405 | 2 | 202 | 2.1 | 0.140 |
| S x NT x B | 1402 | 2 | 701 | 7.1 | 0.002 |
| S x R x B | 344 | 2 | 172 | 1.7 | 0.186 |
| N x R x B | 12 | 1 | 12 | 0.1 | 0.734 |
| S x NT x R x B | 63 | 2 | 32 | 0.3 | 0.727 |
| Error | 6960 | 57 | 122 | | |
| Total | 59374 | 71 | | | |

Table C.15: RMANOVA_2 for CI users as described in Section 9.3.

References

- Advance Bionics (1996). *Clarion multi-strategy implant system*. Device fitting manual, Version 2.0.
- Allen, J.B. and Berkley, D.A. (1979). "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* 65, 943-950.
- Bernstein, L.R. and Trahiotis, C. (1996). "The normalized correlation: Accounting for binaural detection across center frequency," *J. Acoust. Soc. Am.* 100(6), 3774-84.
- Blauert, J. (1996). *Spatial Hearing: The psychophysics of human sound localization*. MIT Press, Cambridge, MA.
- Boll, S. (1979). "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech and Sig. Proc.* ASSP-27(2), 61-68.
- Boothroyd, A. (1995). "A wearable tactile intonation display for the deaf," *IEEE Trans. Acoust., Speech and Sig. Proc.* 33(1), 111-117.
- Cochlear Corporation (1996). *Fundamentals of programming*. Technical reference manual.
- Dorman, M., Loizou, and Rainey, D. (1997a). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.* 102(4), 2403-2411.
- Dorman, M., Loizou, and Rainey, D. (1997b). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," *J. Acoust. Soc. Am.* 102(5), 2993-2996.
- Dorman, M., Loizou, P., and Fitzke, J. (1998a). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," *Ear and Hearing* 19, 481-484.
- Dorman, M., Loizou, P., Fitzke, J., and Tu, Z. (1998b). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," *J. Acoust. Soc. Am.* 104(6), 3583-3585.
- Drullman, R., Festen, J.M., and Plomp, R. (1994a). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* 95(2), 1053-1064.
- Drullman, R., Festen, J.M., and Plomp, R. (1994b). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* 95(5), 2670-2680.
- Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* 97(1), 585-592.
- Eddington, D.K. and Pierschalla, M.L. (1994). "Cochlear implants: Restoring hearing to the deaf," *On the brain*. The Harvard Mahoney Neuroscience Institute letter. 3(4).

- French, N.R. and Steinberg, J.C. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* 19, 90-119.
- Fu, Q.J., Shannon, R.V., and Wang, X.S. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," *J. Acoust. Soc. Am.* 104, 3586-3596.
- Fu, Q.J. and Shannon, R.V. (2000). "Effects of dynamic range and amplitude mapping on phoneme recognition in nucleus-22 cochlear implant users," *Ear and Hearing* 21(3), 227-235.
- Glasberg, B. and Moore, B. (1990). "Derivation of auditory filter shapes from notched noise data," *Hearing Research* 47, 103-138.
- Goldsworthy, R. L. and Greenberg, J.E. (1999). "Evaluation of the Audallion BEAMformer" Biomedical Technology Seminar, Health Sciences and Technology Forum, Boston, MA.
- Goldsworthy, R.L. and Greenberg, J.E. (2000). "Algorithms used for Noise Reduction," American Speech-Language-Hearing Association Audiology Conference, San Francisco.
- Goldsworthy, R.L. and Greenberg, J.E. (2001). "Using STI as a Performance Metric for Cochlear-implant users," Conference on Implantable Auditory Prosthesis, Asilomar, CA (August).
- Goldsworthy, R.L. and Greenberg, J.E. (2003). "Predicting the Intelligibility of Cochlear Implant Speech Processing," Conference on Implantable Auditory Prosthesis, Asilomar, CA (August).
- Goldsworthy, R.L. and Greenberg, J.E. (2004). "Analysis of speech-based Speech Transmission Index methods with implications for nonlinear operations," *J. Acoust. Soc. Am.* 116(6), 3679-3689.
- Grant, K.W. and Braida, L.D. (1991). "Evaluating the articulation index for auditory-visual input," *J. Acoust. Soc. Am.* 89(6), 2952-2960.
- Hamacher, V., Doering, W.H., Mauer, G., Fleishmann, H., Hennecke, J. (1997). "Evaluation of noise reduction systems for cochlear-implant users in different acoustic environment," *Am. J. of Otology* 18, S46-S49.
- Hanekom, J.J., Shannon, R.V. (1998). "Gap detection as a measure of electrode interaction in cochlear implants," *J. Acoust. Soc. Am.* 104(4), 2372-2384.
- Henry, B.A. and Turner, C.W. (2003). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," *J. Acoust. Soc. Am.* 113(5), 2861-2873.
- Hochberg, I. Boothroyd, A., Weiss, M. and Hellman, S. (1992). "Effects of noise and noise suppression on speech perception by cochlear-implant users," *Ear and Hearing* 13(4), 263-271.
- Hohmann, V. and Kollmeier, B. (1995). "The effect of multichannel dynamic compression on speech intelligibility," *J. Acoust. Soc. Am.* 97, 1191-1195.
- Holden, L.K., Skinner, M.W., and Holden, T.A. (1995). "Comparison of the normal and noise-suppression settings on the Spectra 22 speech processor of the Nucleus 22-Channel Cochlear Implant System," *Am. J. Audiology* 4(3): 55-58.
- Holube, I. and Kollmeier, K. (1996). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," *J. Acoust. Soc. Am.* 100(3), 1703-15.

- Houtgast T. and Steeneken, H.J.M. (1971). "Evaluation of speech transmission channels by using artificial signals," *Acustica* 25, 355-367.
- Houtgast T. and Steeneken, H.J.M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* 77(3), 1069-77.
- Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (1986). "Application of the articulation index and the speech transmission index to the recognition of speech by normal hearing and hearing-impaired listeners," *J. Speech Hear. Res.* 29, 447-462.
- IEC (1998). *Sound System Equipment – Part 16: Objective rating of speech intelligibility by Speech Transmission Index; 2nd Ed*, International Standard No. 60268-16.
- IEEE (1969). "IEEE recommended practice for speech quality measurements," IEEE, NY.
- Kalikow, D.N., Stevens, K.N., and Elliott, L.L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* 61, 1337-1351.
- Koch, R. (1992). "Gehörgerechte Schallanalyse zur Vorhersage und Verbesserung der Sprachverständlichkeit," ("Auditory sound analysis for the prediction and improvement of speech intelligibility"), Dissertation, Universität Göttingen.
- Kollmeier, B. and Koch, R. (1994). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," *J. Acoust. Soc. Am.* 95(3), 1593-1602.
- Kollmeier, B. Pessig, J., and Hohmann, V. (1993). "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," *Journal of rehabilitation research and development*, 30(1), 82-94.
- Kryter, K.D. (1962a). "Methods for the calculation and use of the articulation index," *J. Acoust. Soc. Am.* 34, 1689-1697.
- Kryter, K.D. (1962b). "Validation of the articulation index," *J. Acoust. Soc. Am.* 34, 1698-1706.
- Lim, J.S. and Oppenheim, A.V. (1979). "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE* 67(12), 1586-1604.
- Lindemann, W. (1986). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation for lateralization for stationary signals," *J. Acoust. Soc. Am.* 80, 1608-1622.
- Lockwood, M.E., Jones, D.L., Bilger, R.C., Lansing, C.R., O'Brien, W.D., Wheeler, B.C., and Feng, A.S. (2004). "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms," *J. Acoust. Soc. Am.* 115(1), 379-391.
- Loizou, P. (1998). "Mimicking the human ear," *IEEE Signal Proc. Mag.* 15(5), 101-130.
- Loizou, P., Dorman, M., and Tu, Z. (1999). "On the number of channels needed to understand speech," *J. Acoust. Soc. Am.* 106(4): 2097-2103, 1999.
- Ludvigsen, C. (1987). "Prediction of speech intelligibility for normal hearing and cochlear hearing-impaired listeners," *J. Acoust. Soc. Am.* 82, 1162-1171.
- Ludvigsen, C., Elberling, C., Keidser, G. and Poulsen, T. (1990). "Prediction of intelligibility of nonlinearly processed speech," *Acta Otolaryngol. Suppl.* 469, 190-195.

- Ludvigsen, C., Elberling, C., and Keidser, G. (1993). "Evaluation of a noise reduction method – Comparison of measured scores and scores predicted from STI," *Scand. Audiol. Suppl.* 38, 50-55.
- Margo, V., Schweitzer, C., and Feinman, G. (1997). "Comparisons of Spectra 22 performance in noise with and without an additional noise reduction preprocessor," *Seminars in Hearing*, 18 (4), 405-415.
- Muchnik, C., Taitelbaum, R., Tene, S., and Hildesheimer, M. (1994). "Auditory temporal resolution and open speech recognition in cochlear implant recipients," *Scandinavian Audiology* 23(2), 105-109.
- National Institute on Deafness and Other Communication Disorders (2004). Statistics about Hearing Disorders, Ear Infections, and Deafness. Reference to website: www.nidcd.nih.gov/health/statistics/hearing.asp. Update referenced: June 18th, 2004.
- Nelson, P.B., Jin, S.H., Carney, A.E., and Nelson, D.A. (2003). "Understanding speech in modulated interference: Cochlear-implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* 113(2), 961-968.
- Osberger, M. and Fisher, L. (1990). "SAS-CIS preference study in postlingually deafened adults implanted with the clarion cochlear implant," *Annals of oto., rhino. and laryng.* 108(4), 74-79.
- Papoulis, A. and Pillai, S.U. (2002). *Probability, Random Variables and Stochastic Processes*. McGraw Hill Publishers, 4th Edition.
- Pavlovic, C.V. (1987). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," *J. Acoust. Soc. Am.* 82(2), 413-422.
- Payton, K.L., Uchanski, R.M. and Braida, L.D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* 95(3), 1581-1592.
- Payton, K.L. and Braida, L.D. (1999). "A method to determine the speech transmission index from speech waveforms," *J. Acoust. Soc. Am.* 106, 3637-3648.
- Payton, K.L., Braida, L.D., Chen, S, Rosengard, P., and Goldsworthy, R. (2002). "Computing the STI using speech as a probe stimulus," *Past, Present and future of the speech transmission index*. (TNO Human Factors, Soesterberg, The Netherlands), pp. 125-138.
- Qin, M. and Oxenham, A. (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* 114(1), 446-454.
- Ross, S. (1998), *A first course in probability. 5th edition*, (Prentice Hall, New Jersey, USA), pp. 350-354.
- Schweitzer, H.C., Terry, A.M., Grim, M.A (1996). "Three experimental measures of a digital beamforming signal processing algorithm," *J. Am. Acad. Audiol.* 7, 230-239.
- Shannon, R.V., Zang, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues," *Science* 270, 303-304.
- Steeneken, H.J.M. and Houtgast T. (1980). "A physical method for measuring speech transmission quality," *J. Acoust. Soc. Am.* 67(1), 318-326.
- Steeneken, H.J.M. and Houtgast T. (1982). "Some applications of the speech transmission index (STI) in auditoria," *Acustica* 51, 229-234.

- Steeneken, H.J.M. and Houtgast T. (1999). "Mutual dependence of the octave-band weights in predicting speech intelligibility," *Speech Communication* 28, 109-123.
- Studebaker, G.A. (1985). "A "Rationalized" arcsine transform," *J. Speech and Hear. Res.* 28, 455-462.
- Van Buuren, R.A., Festen, J.M., and Houtgast, T. (1998). "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," *J. Acoust. Soc. Am.* 105, 2903-2913.
- Van Hoesel, R. J., and Clark, G. M. (1995). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *J. Acoust. Soc. Am.* 97, 2498-2503.
- Weiss, M.R. (1993). "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," *J. Rehab. Res.* 30(1), 117-128.
- Winer, B.J., Brown, D.R., and Michels, K.M. (1991). "Statistical principles in experimental design," 3rd edition, McGraw-Hill.
- Wittkopp, T., Albani, S., Hohmann, V., Pessig, J., Woods, W., and Kollmeier, B. (1997). "Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction," *Acustica*, 83, 684-699.
- Yariv, E. and Van Trees, H. (1995). "A signal subspace approach for speech enhancement," *IEEE Trans. on Speech and Audio Proc.*, 3(4), 251-266.
- Zeng FG, Grant G, Niparko J, Galvin J, Shannon R, Opie J, Segel P. (2002). "Speech dynamic range and its effect on cochlear implant performance," *J. Acoust. Soc. Am.* 111(1), 377-386.
- Zurek, P.M. (1993). "Binaural advantages and directional effects in speech intelligibility," In *Acoustical Factors Affecting Hearing Aid Performance II*, edited by G.A. Studebaker and I. Hochberg (Allyn and Bacon, Boston).