

ABSTRACT

Title of Dissertation: THE ROLE OF FREQUENCY, TIMING AND LEVEL DISTORTION ON BINAURAL PROCESSING IN SIMULATIONS OF COCHLEAR IMPLANT USERS WITH SINGLE-SIDED DEAFNESS.

Jessica Marie Wess
Doctor of Philosophy, 2017

Dissertation directed by: Adjunct Professor Joshua G. W. Bernstein
Neuroscience and Cognitive Science Program
University of Maryland – College Park
and
Audiology & Speech Pathology Center
Walter Reed National Military Medical Center
Bethesda, Maryland

Professor Sandra Gordon-Salant
Department of Hearing and Speech Sciences
University of Maryland College Park

Cochlear implants are a promising new treatment option for single-sided deafness.

Cochlear implants for single-sided deafness have been shown to improve speech perception in noise and aid in sound localization. However, this intervention is not as

good as acoustic hearing and listeners' exhibit large amounts of variability in hearing outcomes. These limitations may be caused by certain distortions inherent in the processing of the sound signals by the cochlear implant. This dissertation examined the role that three key cochlear implant distortions might play in limiting speech perception in noise for listeners with single-sided deafness. The first distortion examined was the frequency mismatch between the cochlear implant and the acoustic ear. The next distortion examined was the effect of timing differences between the cochlear implant and the normal hearing ear. Finally, the effect of compression on hearing speech in spatial noise was investigated. These limitations and distortions could limit binaural processing ability in those with single-sided deafness who receive a cochlear implant. The goal of this dissertation was to examine the role of cochlear-implant distortions on binaural hearing using simulations of cochlear implant processing presented to normal-hearing listeners. Normal-hearing listeners were presented with vocoder simulations of cochlear-implant processing to one ear, and unprocessed signals to the other ear. These simulations were used to examine the ability to understand binaural speech signals in noisy environments and to examine auditory object formation in simulated free-field environments. These data provided insight into how CI distortions and mapping strategies can limit binaural benefits for those with single-sided deafness. Knowledge of these limitations could lead to better programming strategies to improve binaural hearing and quality of life for those with single-sided deafness who receive a cochlear implant.

THE ROLE OF FREQUENCY, TIMING AND LEVEL DISTORTION ON BINAURAL
PROCESSING IN SIMULATIONS OF COCHLEAR IMPLANT USERS WITH
SINGLE-SIDED DEAFNESS.

By

Jessica Marie Wess

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:

Joshua Bernstein, Ph.D., Co-Chair
Sandra Gordon-Salant, Ph.D., Co-Chair
Douglas Brungart, Ph.D.
Kenneth Grant, Ph.D.
Matthew Goupell, Ph.D.
Jonathan Simon, Ph.D.

© Copyright by
Jessica Marie Wess
2017

Dedication

This dissertation is dedicated to my parents, Rose and John, for their unconditional love support and encouragement.

Acknowledgements

First and foremost I'd like to thank my adviser Josh Bernstein. Josh took me in as a wayward graduate student, as I was looking for greener scientific pastures. I have learned so much in the last four years at Walter Reed and working with such an amazing scientist has been a real honor. Josh has been extremely patient with me and my colloquial personality. Josh's door was always open whenever I needed help with anything and I'm really grateful for all his guidance, support and encouragement.

I would like to extend my sincere gratitude to my Co-Adviser Dr. Sandra Gordon-Salant, for her patience and the time she has spent helping me become a better writer and scientist.

Many thanks to and my dissertation committee members: Dr. Matthew Goupell, Dr. Douglas Brungart, Dr. Ken Grant and Dr. Jonathan Simon for their time, input and guidance.

Finally I would like to thank my husband Paul, for all the love and encouragement, for being my best friend and for providing interesting and helpful scientific discussions. Also for help with the occasional data analysis or plot generation. ☺

This research has been supported by a grant from the Defense Medical Research and Development Program (DM130007; PI: Joshua Bernstein)

Table of Contents:

Abstract	
Dedication	ii
Acknowledgements.....	iii
Table of Contents	iv
List of Tables	xi
List of Figures	xii
List of Abbreviations	xiv
Chapter 1. Introduction to binaural hearing and cochlear implants for single-sided deafness.....	1
General Introduction	1
Dissertation Aims.....	3
Binaural hearing is critical for speech perception in noisy environments	8
The role of binaural fusion and auditory grouping in spatial hearing.....	3
Single-sided deafness and treatment options	14
Possible sources of distortion in CI users with SSD.....	24
Spectral mismatches and their effects on binaural hearing.....	24
Temporal disparities between cochlear implants and normal hearing ears	29
Loudness growth, compression and their effects on binaural hearing.....	32
Chapter 2. The effect of interaural mismatches on contralateral unmasking in vocoder simulations of single-sided deafness.....	37

Introduction.....	37
Experiment 2.1. The role of spectral mismatches on contralateral unmasking in simulations of CI users with SSD.	45
Experimental question and hypothesis.....	45
Methods.....	45
Participants.....	45
Approach.....	45
Stimuli	46
Procedure	48
Results.....	49
Summary.....	52
Experiment 2.2. The role of temporal mismatches on contralateral unmasking in simulations of CI users with SSD	53
Experimental question and hypothesis	53
Methods.....	53
Participants.....	53
Stimuli	54
Procedure	55
Results.....	55
Summary.....	56

Experiment 2.3. The role of spectral mismatches and vocoder channel resolution on contralateral unmasking in simulations of CI users with SSD.	56
Experimental question and hypothesis.....	56
Methods.....	57
Participants.....	57
Stimuli.....	57
Procedure	58
Results.....	60
Summary.....	63
Experiment 2.4. The role of spectral and temporal mismatches on contralateral unmasking in simulations of CI users with SSD	63
Experimental question and hypothesis.....	63
Methods.....	63
Participants.....	63
Stimuli.....	64
Procedure	64
Results.....	65
Summary.....	67
Discussion.....	68
Impacts of a spectral mismatch.....	69
Impacts of spectral resolution	71
Effects of temporal mismatch	73

Effects of combined spectral and temporal mismatch	75
Implications for SSD-CI listeners.....	76
Study Limitations.....	78
Conclusions.....	80
Chapter 3. Effect of compression and expansion on binaural hearing in simulations of CI users with SSD.....	82
Introduction.....	82
Experiment 3.1. The effect of compression and expansion on squelch in simulations of cochlear implants for SSD listeners.....	88
Experimental question	89
Hypothesis.....	89
Methods.....	90
Approach.....	91
Participants.....	93
Stimuli.....	94
Generation of HRTFs.....	94
Noise Vocoding	95
Loudness manipulations.....	96
Procedure	97
Results.....	99
Summary.....	103

Experiment 3.2. The effect of compression and expansion on head-shadow benefit in simulations of cochlear implants for SSD listeners.	104
Experimental question	104
Hypothesis.....	104
Methods.....	107
Approach.....	107
Participants.....	108
Stimuli.....	108
Procedure	109
Results.....	110
Discussion	115
The effect of compression and expansion on squelch	118
The effect of compression and expansion on head-shadow benefit	120
Implications for CI listeners.....	122
Study Limitations.....	122
Conclusions.....	125
Chapter 4. The role of spectral mismatch on perceived binaural fusion in vocoder stimulations of cochlear implant listening.	128
Introduction.....	128
Experiment 4.1. Numerosity judgments of binaural fusion.....	136
Study objectives	136
Experimental questions.....	136

Hypothesis.....	136
Methods.....	137
Participants.....	137
Stimuli.....	137
Procedure	137
Noise Vocoding	144
Results.....	145
Interim Discussion 4.1	150
Experiment 4.2. Discrimination, spectral mismatch and binaural fusion.	152
Study objectives	152
Experimental question	152
Hypothesis.....	152
Methods.....	152
Participants.....	152
Stimuli.....	153
Procedure	155
Apparatus	155
Results.....	156
Discussion.....	157
Impacts of spectral mismatch.....	159
Disruption of temporal processing.....	162
Implications for SSD-CI listeners.....	163

Study limitations165

Conclusions.....168

Chapter 5. Summary of dissertation and general discussion170

General discussion175

References.....182

List of Tables

Table I.	Hypothesis table for effect of compression and expansion on contralateral unmasking	88
Table II.	Statistical post hoc results for squelch experiment	100
Table III.	Hypothesis table for effect of compression and expansion on head-shadow benefit.....	103
Table IV.	Statistical post hoc results for head-shadow experiment.....	109
Table V.	Experimental parameters fusion diotic conditions vs foils for Experiment 4.1A	135
Table VI.	Analysis and synthesis channel allocation table standard vs place-matched	138
Table VII.	Experimental parameters monaural and bilateral vocoded and unprocessed conditions for experiment 4.1A.....	139

List of Figures

Figure 1.1.	Release from informational masking in CI users with SSD.....	20
Figure 1.2.	Performance variability among CI users with SSD.....	21
Figure 1.3.	Release from masking in vocoder simulations of CI users with SSD.....	24
Figure 2.1.	The effect of spectral shift on contralateral unmasking	49
Figure 2.2.	The effect of spectral shift on contralateral unmasking as a function of TMR	50
Figure 2.3.	The effect of temporal shift on contralateral unmasking	54
Figure 2.4.	Analysis and synthesis band edges for +4 ERB spectral shift	57
Figure 2.5.	The effect of spectral shift and spectral resolution on contralateral unmasking	59
Figure 2.6.	The effect of spectral shift and temporal resolution on contralateral unmasking as a function of TMR.....	60
Figure 2.7.	The effect of spectral shift and temporal resolution on contralateral unmasking	64
Figure 2.8.	The effect of spectral shift and temporal resolution on contralateral unmasking as a function of TMR.....	65
Figure 3.1.	Prediction for squelch experiment.....	90
Figure 3.2.	Spatial configuration for HRTF contralateral unmasking experiment.....	92
Figure 3.3.	HRTF acquisition schematic	92
Figure 3.4.	Compression and Expansion Input/Output function	94

Figure 3.5.	Contralateral unmasking monaural vs linear bilateral.....	96
Figure 3.6.	The effect of compression and expansion on contralateral unmasking.....	98
Figure 3.7.	Contralateral unmasking compression parameter as a function of TMR....	99
Figure 3.8.	Prediction for head-shadow experiment.....	99
Figure 3.9.	Spatial configuration for HRTF head-shadow benefit experiment	104
Figure 3.10.	Head-shadow benefit monaural vs linear bilateral	106
Figure 3.11.	The effect of compression and expansion on head-shadow benefit.....	107
Figure 3.12.	Head-shadow compression parameter as a function of TMR	110
Figure 3.13.	Loudness growth in cochlear implants and normal hearing.....	124
Figure 4.1.	Data from experiment 4.1A. Sets A, B, C and D	140
Figure 4.2.	Data from experiment 4.1B. Numerosity judgments for all vocoded and unprocessed conditions.....	142
Figure 4.3.	Schematic of two example trials from experiment 4.2.....	149
Figure 4.4.	Data from experiment 4.1B. Numerosity judgments for all vocoded and unprocessed conditions.....	150

List of Abbreviations

- Alternative forced choice – AFC
- Analysis of variance –ANOVA
- Auditory scene analysis – ASA
- Automatic gain control – AGC
- Behind the ear – BTE
- Bilateral cochlear implant – BICI
- Bone-anchored hearing aids – BAHA
- Cochlear implant – CI
- Computerized tomography – CT
- Continued interleaved sampling – CIS
- Contralateral routing of signal (hearing aid) – CROS
- Coordinate response measure – CRM
- Decibels – dB
- Dynamic range – DR
- Electrical auditory brainstem response – eABR
- Equivalent rectangular bandwidth – ERB
- Fast fourier transform – FFT
- Graphical user interface – GUI
- Head related transfer functions – HRTFs
- Hearing impaired – HI
- Hearing loss – HL
- Hertz – Hz
- In the ear – ITE
- Interaural level differences – ILDs
- Interaural timing differences – ITDs
- Just noticeable differences – JND
- Knowles Electronic Manikin for Acoustic Research – KEMAR
- Normal hearing – NH
- Root mean squared – RMS
- Signal to noise ratios – SNRs
- Single-sided deafness – SSD
- Spatial release from masking – SRM
- Sound pressure level – SPL
- Superior olivary complex – SOC
- Target-to-masker ratios – TMR

Chapter 1: Introduction to binaural hearing and cochlear implants for single-sided deafness

General Introduction

The human auditory system allows normal hearing (NH) individuals to detect speech and other signals embedded in dynamic noisy environments. Having two ears with NH (binaural hearing) is immensely important for hearing in these situations. Normal human interaction requires being able to hear a particular talker of interest when multiple people are talking in the background. For example, the ability to understand and communicate with someone across the table in a crowded cafeteria, is facilitated by binaural hearing. Individuals who have only one functional ear are at a severe disadvantage in terms of normal social interaction in such situations.

Cochlear implants (CIs) are the world's first widely successful neuroprosthetic devices. CIs can restore partial hearing in completely deaf individuals to a point where they can carry on verbal conversations completely normally. Traditionally CIs have only been implanted in profoundly hearing impaired individuals, but more recent outcomes indicate that they may also be useful for individuals with deafness in one ear—referred to as single-sided deafness (SSD) — to partially restore binaural hearing. More specifically, for listeners with SSD, CIs have been shown to restore basic spatial hearing functions, including facilitation of speech understanding in spatial noise and improved sound localization ability. CIs are not as good as acoustic ears and there exists considerable

variability in outcomes. Despite the benefits CIs can provide, especially for speech understanding in quiet environments, the auditory signals provided by CIs are crude relative to a NH ear. Several alterations occur to a sound signal after being transduced by a CI that may limit the benefit of CIs for SSD individuals who still have a functional acoustic ear. Many of these alterations can be grouped into frequency, timing and level distortions. These distortions likely cause very different neural representations of two identical acoustic signals presented to a CI and NH ear.

The overall goal of this dissertation was to investigate CI distortions and measure the effect on listeners' ability to segregate voices in competing talker environments. Knowledge of the impact of these distortions will allow clinicians to make better programming and mapping choices for SSD-CI listeners, which could potentially improve their binaural hearing outcomes. Experiments utilized simulated CI speech processing (vocoded) presented to NH listeners. This was accomplished by presenting unprocessed sound to one ear and vocoded sound (sounds designed to mimic the auditory stimulus provided to SSD patients by the CI) to the other ear. The aims of this dissertation investigated how vocoder-simulated CI distortions might affect speech perception in competing-talker environments when the signals presented to the ears are controlled independently (Chapter 2), under more realistic simulated spatial configurations (Chapter 3) and whether the effects of interaural mismatch can be attributed to the perceived fusion of speech signals across the ears (Chapter 4). Exploration of the effect of CI distortions on binaural hearing will help determine: (i) which distortions are particularly problematic and (ii) possible mapping and programming techniques that could improve binaural hearing for SSD-CI users.

Dissertation Aims

The overall goal of this dissertation was to examine how distortions associated with CI processing could potentially limit hearing speech in multiple concurrent talker environments for those with SSD. Specifically, frequency, timing and level CI distortions were examined. The technique used to accomplish this goal was to incorporate distortions in these three dimensions into vocoder simulations presented to NH listeners. This overarching goal was approached through a series of three specific Aims.

Aim 1 (Chapter 2): Measure the extent to which negative impacts of interaural mismatch in frequency and timing on contralateral unmasking in vocoder simulations of CI listeners with SSD.

In this dissertation, contralateral unmasking is operationally defined as the improvement in speech perception associated with adding interfering voices to an ear contralateral to the target speech. CI users with SSD receive contralateral unmasking but there exists large variability across users and they do not appear to benefit as much as NH participants listening to a vocoder. The contralateral unmasking metric is a proxy for how well listeners are able to combine information across the ears to facilitate hearing speech in background noise. Frequency mismatches between place of stimulation of the implant array and place of excitation in the NH ear could reduce the ability to benefit from binaural cues. Latency dissimilarities between the CI processor and NH ear are also likely to be

present and could cause disruptions in binaural hearing. Chapter 2 consisted of 4 experiments examining the effects of spectral and temporal mismatch on speech understanding in the presence of interfering talkers using a contralateral unmasking paradigm that focused on the processes of binaural integration in a speech task. Experiment 2.1 examined the effect of a spectral shift on contralateral unmasking and found a strong dependence of frequency match on performance in the task (i.e., less spectral shift = more contralateral unmasking). Experiment 2.2 examined temporal mismatch and its effect on contralateral unmasking, and found that physiologically plausible temporal mismatches did not greatly disrupt contralateral unmasking. Experiment 2.3 investigated the interaction between the frequency resolution of the vocoder and spectral mismatch and found that broader channel vocoding made listeners more immune to spectral shifts. Finally, Experiment 2.4 examined the potential interaction between spectral and temporal shifts and found that once a mismatch was implemented in the vocoder, the addition of a second mismatch (either temporal or spectral) did not further disrupt performance. More specifically, instead of finding an additive effect, once a mismatch was present the additional mismatch had a negligible effect on performance. Therefore, spectral mismatch effects performance more than temporal mismatch for the perceptual separation of a target from a masker background.

Aim 2 (Chapter 3): Determine how contralateral unmasking and head-shadow benefit can be affected by envelope compression and expansion in HRTF-generated virtual auditory environments.

The goal of this chapter was to examine how compression distortions are likely to manifest in a simulated free-field environment. The experiments in Chapter 3 examined binaural squelch (3.1) and head-shadow (3.2) in a more realistic auditory environment than was used in Chapter 2. Specifically, the experiments examined how compression and expansion might affect the relative interaural level differences (ILDs) and target-to-masker ratios (TMRs) in the two ears and impact speech perception in the presence of interfering talkers that are spatially separated from the target talker of interest. Normally, listeners have access to spatial cues (interaural timing differences [ITDs] and ILDs) of environmental signals arriving at the two ears, which help listeners segregate competing talkers and other background noises to aid them in streaming auditory sources of interest. The contralateral unmasking paradigm employed in Chapter 2 was an artificial situation that would never occur in the free field. In a simulated free field environment, the listener has two different TMRs in the two ears and these will likely be distorted by compression. In Chapter 3, two experiments examined spatial hearing benefit with spatial cues provided via generalized head-related transfer functions (HRTFs) that mimic the effects of head-shadow and path-length differences that are encountered for signals in the free field. Level compression and expansion were implemented in the vocoder to determine what effect amplitude manipulation had on speech perception in the presence of spatially separated interfering talkers. Envelope compression was found to have a negative effect on both squelch and head-shadow benefit. Envelope expansion had little effect on head-shadow benefit (Experiment 3.2) but increased binaural squelch, relative to the compression conditions (Experiment 3.1). It is likely that in the squelch experiment compression and expansion exerted their effects by changing the ILDs between the target and the maskers;

with the ILD between the target and maskers decreasing with compression and increasing with expansion. For the head-shadow experiment, diminished performance after compression could have been a result of change in the TMR at the vocoded ear (closest to target), which reduced audibility of the target. However, since both expansion and compression disrupted performance it is likely that envelope distortion reduced intelligibility of the target.

Aim 3 (Chapter 4): Elucidate a possible fusion mechanism for the contralateral unmasking (squench) effect in Chapters 2 and 3. More specifically, to develop and test a paradigm to measure binaural fusion in the presence of a spectral mismatch.

Chapters 2 and 3 described experiments in which the addition of the vocoder provided contralateral unmasking, but this benefit was eliminated or largely diminished after CI distortions were implemented in the vocoder. The loss of binaural squench and contralateral unmasking after vocoder distortion could be explained by a loss of binaural fusion ability. To more directly test this hypothesis, the experiments in Chapter 4 aimed to measure binaural fusion ability with and without a spectral mismatch. The spectral mismatch distortion was tested in these experiments, because that distortion profoundly disrupted contralateral unmasking in Chapter 2, compared to the other distortions tested in this dissertation. In Chapter 4, spectral mismatch was implemented not by linearly shifting the vocoder channels (as was done in Chapter 2), but rather by utilizing a more realistic mismatch based on published radiographic data averaged across CI users. The experimental approach had listeners identify the number of voices in the environment,

instead of relying on intelligibility. Binaural fusion was tested in two ways. The first experiment (4.1) involved listeners counting the number of voices they heard in a complex mixture. If an unprocessed and a vocoded version of the same voice presented to opposite ears were fused, the listener should report one voice; if not, they should report hearing two. This should occur even when the voice to be fused was accompanied by other voices in the mixture. Experiment 4.2 was a two-alternative forced choice (2AFC) task, in which the listeners had to discriminate a fusion interval from a non-fusion interval. In this experiment, the “fused” interval had the same voice in the two ears (one vocoded, one unprocessed), and the non-fused interval had two different voices in the two ears. If listeners were able to perceptually fuse the two voices in the first interval, they should have been able to more easily tell the difference between the mixture containing the same voice in the two ears and the mixture containing no common voices in the two ears. Experiment 4.1 found that people reported a number of voices that indicated they were not fusing the fusion stimulus, regardless of the vocoder condition. In Experiment 4.2, when binaural fusion was assessed via a discrimination test, listeners were generally more likely to select the correct fusion interval with a place-matched vocoder mapping than with a mismatched mapping. Taken together, these results suggested that the listeners were achieving incomplete fusion. For spectrally matched stimuli, the speech stimuli were sufficiently fused between the two ears and this was enough to detect that there was a common voice presented to the two ears (Experiment 4.2). Yet, the stimuli were unfused enough that listeners still reported a diotically presented voice as two voices when they were asked to count the number of talkers they heard in the mixture.

The remainder of this chapter will review literature relevant to the main question raised by this dissertation: How do simulated CI distortions affect listeners' ability to segregate voices in competing talker environments? First, the role of binaural hearing in speech perception in noisy environments is discussed. Second, the concepts of binaural fusion and auditory grouping are introduced. Third, an overview of SSD and its treatment options are presented. Fourth, possible sources of distortion in CI processing that can affect binaural hearing are described.

Binaural hearing is critical for improving speech perception in noisy environments

Binaural hearing provides a number of benefits for listening in complex acoustic environments. Two of the most important benefits are the ability to localize sounds and the ability to understand speech in noise. The phenomenon of being able to successfully focus attention on a particular stimulus or talker while filtering out or ignoring competing talkers has been referred to as “the cocktail-party effect” (Cherry, 1953; Bronkhorst, 2000). Binaural hearing is critical for successful hearing in these environments (Hawley, Litovsky, & Culling, 2004). Having two ears allows for computations of spatial cues to perceptually separate sound sources based on their different locations.

Perceiving a talker of interest in multiple talker environments is difficult due to auditory masking. Auditory masking occurs when the presence of one sound interferes with the perception of another. There are multiple types of masking (Gelfand, 2004), which are generally divided into two categories: energetic and informational (Kidd, Mason, & Deliwala, 1994; Leek, Brown, & Dorman, 1991; Watson, 2005). Energetic masking can

occur when the masking energy renders the signal inaudible, which tends to occur when there is a high degree of spectral and temporal overlap between targets and maskers. Energetic masking occurs from sound-wave interference in the cochlea. For NH listeners spatially separating a target from noise results in reduced energetic masking of the target. Up to a 10 decibel (dB) benefit in speech-reception thresholds (i.e., binaural unmasking) can occur in energetic masking situations when the target is spatially separated from the maskers (Best, Thompson, Mason, & Kidd, 2013; Bronkhorst, 2000). Release from energetic masking is thought to occur via a combination of two mechanisms: (i) the head-shadow effect, utilized predominantly for higher sound frequencies (van Hoesel, 2012) and (ii) binaural squelch, which requires neural computation of interaural difference cues. The head-shadow effect results in one ear having a better signal-to-noise ratio (SNR) than the other ear when the source of interest and the masking sounds are spatially separated. A target to masker ratio (TMR) is related to an SNR, but where the signal is denoted as the target and the maskers are the noise. Change in speech reception thresholds due to the head-shadow effect is approximately 6 dB in the speech frequency range (500-2000 Hertz (Hz)) and up to 15 dB at higher frequencies (Schleich, Nopp, D'Haese, 2004a). When a target and masker are spatially separated, this results in one ear having a better SNR than the other ear. Therefore, the head-shadow benefit involves attending to the ear with the more favorable SNR for the target signal of interest.

Binaural squelch requires neural computations of interaural cues to facilitate hearing in spatial noise. The shape and size of the human head creates ITDs and ILDs. ITDs are created by a sound originating from a specific location having differential arrival times at each ear, because it has to take a longer path to reach the far ear. ILDs are created

by the intensity difference that occurs when sounds are attenuated in one ear relative to the other (due to the head shadow, for example). The auditory system takes advantages of these interaural differences to improve signal detection in noise. According to the equalization-cancellation model of masking release, the binaural system can reduce the impact of masking noise by carrying out neural computations that effectively attenuate and delay the entire signal in one ear relative to the other ear (equalization). By subtracting the resulting signals between the ears (cancellation), the binaural system can effectively reduce the amount of masking experienced by the listener (Durlach, 1963). An alternative theory for masking release involves “glimpse listening,” which requires that the listener take advantage of dips in the background noise in order to better detect the target signal of interest. This is thought to occur by providing the brain with the “lost” signal components in each individual ear and then integrating this information from each ear (Cooke, 2006). In situations involving multiple competing talkers, binaural squelch can also be thought of in terms of added listening advantage obtained by perceived spatial separation between a target and masker. Binaural squelch can of course arise from actual spatial separation but it is important to note that *perceived* spatial separation is often sufficient (Freyman, Helfer, McCall, & Clifton, 1999) at least in situations with multiple simultaneous talkers that are difficult to perceptually separate based on monaural cues alone.

The other category of masking that complicates speech perception in noisy environments is informational masking. Informational masking occurs due to a difficulty in identifying an audible signal that is accompanied by other similar sounding signals (Leek et al., 1991). An example of informational masking is the difficulty encountered when trying to listen to one talker in the midst of multiple competing talkers, all of which are

audible to the listener. For example, the difficulty is even greater when the talkers are the same gender (Brungart, 2001). The problem is likely a failure of auditory stream segregation or auditory scene analysis (ASA). ASA is the process by which the auditory system can separate and segregate sounds coming from different sources and locations (Bregman, 1994). For speech stimuli, a failure of stream segregation or informational masking occurs most often when targets and maskers are perceptually and semantically similar (Ihfeldt & Shinn-Cunningham, 2008). Informational masking is much more likely to occur when targets and maskers sound alike, for example talkers with comparable voice pitch. In this situation, fewer cues are available to either stream the target and masker speech apart or to identify which of the words spoken belong to the target and which belong to the masker talker. Confusability can also be encountered by trying to follow a string of words spoken by a target talker while a masker talker is also reciting a string of words concurrently. It has been proposed that the two main mechanisms driving informational masking are stimulus uncertainty and target-masker similarity (Durlach et al., 2003). Informational masking can also occur with non-speech stimuli such as a task involving a complex mixture of tonal stimuli (Kidd, Mason, & Arbogast, 2002).

For speech stimuli, a number of cues are known to aid listeners in achieving release from informational masking, including voice pitch, relative onset timing of different speakers, and spatial separation between targets and maskers. Differences in voice pitch between talkers can aid in streaming targets of interest from a background of competing talkers. Onset timing differences can aid in release from masking due to the strong influence of timing on stream segregation (Darwin & Hukin, 1998; de Cheveigné, McAdams, & Marin, 1997). It is known that spatial cues can also provide a great deal of

release from informational masking (Arbogast, Mason, & Kidd, 2002; Freyman, Helfer, & Balakrishnan, 2005; Hall, Buss, & Grose, 2005; Kidd, Mason, Rohtla, & Deliwala, 1998). In particular, the two primary binaural cues—ILDs and ITDs—are theorized to play a role in contributing to binaural unmasking (Hawley et al., 2004; Kidd, Mason, Best, & Marrone, 2010; Middlebrooks & Green, 1991), but the extent to which each cue is involved is still a matter of debate.

Binaural localization cues have been shown to facilitate binaural unmasking. In cases of high informational masking, a squelch or unmasking benefit arises from ILD and ITD cues that allow the listener to perceive the target and maskers as arriving from different points in space (Freyman, Balakrishnan, & Helfer, 2001). ITDs are generally only useful for frequencies below 1500 Hz (except for envelope ITDs for modulated stimuli, which can be relayed at higher carrier frequencies), and ILDs are only useful for frequencies above 1500 Hz (Middlebrooks, Macpherson, & Onsan, 2000; Rayleigh, 1907; Wightman & Kistler, 1992). High- or low-pass filtering of speech allows for the “removal” of usable ILD or ITD information in a signal. Several studies have examined the role of either ILDs or ITDs in binaural unmasking using young listeners with NH (Hawley et al., 2004; Ihlefeld & Shinn-Cunningham, 2008; Kidd et al., 2010). Some research points to the dominance of ITDs as the cue most necessary for binaural unmasking (Hawley et al., 2004). Other research has found that either cue can provide sufficient binaural unmasking when controlling for head-shadow benefits (Kidd et al., 2010). Gallun et al. (2005) assessed the role of ILDs and ITDs separately in NH listeners. They employed a word-identification task using the coordinate response measure (CRM) corpus (Bolia, Nelson, Ericson, & Simpson, 2000), which has been shown to produce a great deal of informational masking

(Brungart, 2001). Gallun et al. (2005) presented the target monaurally and the maskers diotically (identical signals presented to each ear) to create a perceived spatial difference between the locations of the target and masker signals. They systemically varied the ILD and ITD components in the masker signal to examine the role of each cue in release from informational masking. They found substantial release from masking with ITDs or ILDs alone. More importantly, they found that ILDs played a role in masking release when TMRs were held constant at the “better-ear.” Therefore, this masking release cannot be explained by better-ear listening alone.

The role of binaural fusion and auditory grouping in spatial hearing

An important prerequisite to the ability of the binaural system to facilitate the perceptual separation of concurrent voices is that the listener must be able to perceptually fuse the coherent auditory information arriving at the two ears. This is referred to as binaural fusion and it allows NH listeners to perceive diotic sounds as a single centered sound.

Binaural fusion is believed to occur in the mammalian superior olivary complex (SOC) in the brainstem, with the higher-order auditory areas receiving a more complete and summed auditory object after subcortical processing (Moore, 2000). It is believed that coincidence detectors and/or interaural cross-correlation give rise to fused perception of binaural signals (Roberts, Seeman, & Golding, 2013). Shinn, Baran, Moncrieff and Musiek (2005) tested NH listeners on a variety of dichotic speech tasks and found that binaural fusion was less likely to be affected by memory or the listener’s attention than by other

speech tasks. They therefore concluded that binaural fusion likely occurs below a listener's conscious control at the subcortical level. Even when asked to switch focus to one ear or another, listeners still reported one fused stimulus, indicating they did not have conscious "control" over the percept. Further evidence of fusion occurring in the brainstem comes from electrophysiological experiments. With presentation of matched interaural input, large binaural-difference response amplitudes can be measured at the level of the brainstem. This binaural-difference response has been measured in humans, for NH, hearing-impaired (HI), and CI listeners as well as for animals (Cai et al., 2015; Goksoy, Demirtas, Yagcioglu, & Urgan, 2005; Pelizzone, Kasper, & Montandon, 1990; Riedel & Kollmeier, 2002). Additionally, the binaural-interaction component (i.e., the difference waveform between the summed monaural response and the binaural response) has been linked to perceptual fusion ability (Zhou & Durrant, 2003). Therefore, proper integration of binaural stimuli at the level of the brainstem is paramount to successful binaural fusion.

Single-sided deafness and treatment options

Due to the immense importance of binaural hearing for communication in noisy environments, individuals with only one functional ear are at a severe disadvantage. SSD—the profound loss of hearing in one ear while the other ear remains normal-hearing or near-normal hearing—is a form of hearing loss with functional limitations that has been traditionally underappreciated. It is estimated that there are nearly 60,000 new cases of SSD a year in the US (Baguley et al., 2009; Carlyon et al., 2010; Sinopoli, 2003). SSD is now known to cause many problems for those afflicted with it. Some common complaints

include social isolation, driving difficulties, problems working, embarrassment and loss of confidence (McKinney, 2002). Traditionally, SSD was not treated because it was not considered incapacitating (i.e., individuals with SSD still have a normal-hearing ear). However, a number of studies have demonstrated that SSD is, in fact, a substantial disability. For example, individuals with SSD exhibit reduced language comprehension as well as reduced oral communication abilities (Lieu, Tye-Murray, Karzon, & Piccirillo, 2010). Additionally, for children with SSD, learning and academic challenges have been widely reported, with these children being 10 times more likely to be held back in school (Bess & Tharpe, 1984; English & Church, 1999). Until recently, the only treatments available for SSD involved hearing-aid solutions that routed signals from a microphone near the deaf ear to the NH ear. The two most common solutions are bone-anchored hearing aids (BAHAs) and contralateral routing of signal (CROS) hearing aids. BAHAs are surgically implanted into the bone just behind the deaf ear, and transmit sound to the opposite (functional) cochlea through the skull through bone conduction. CROS hearing aids are removable devices that contain a receiver on the deaf side of the head and transmit sound to the functional ear via a microphone in the deaf ear. These methods have been successful in alleviating some of the adverse effects of SSD mainly by giving access to signals presented toward the deaf side. However, these devices can impair performance in cases where the unwanted noise is on the deaf side. This occurs because in these cases, the device transmits noise to the normal hearing ear, thereby offsetting the head-shadow advantage that is otherwise present (Arndt et al., 2010). Moreover, these treatments do not restore binaural hearing, and as a result, these patients still experience difficulty with sound

localization and speech understanding in noise (Grantham et al., 2012; Linstrom, Silverman, & Yu, 2009).

In the past several years, CIs have been considered as a possible new treatment option for SSD. Although CIs are not currently approved by the United States Food and Drug Administration for SSD patients, criteria for implant candidacy at individual centers and hospitals has relaxed in the last few years and a substantial number of individuals with SSD in the U.S. and in Europe have received CIs. CIs are the world's first widely successful neuroprosthetic devices. They are implanted in individuals with severe or profound hearing loss, allowing restoration of basic levels of hearing and speech understanding. Over a quarter of a million people have been implanted with CIs worldwide and that number is steadily rising (NIH Report, 2013). A CI consists of an external microphone, a speech processor, a transmitter, a receiver, a stimulator and an electrode array. A behind-the-ear microphone picks up sounds from the environment and the speech processor then filters the signal into a number of frequency bands (depending on the number of electrode channels) and extracts information about the signal envelopes (i.e., slow fluctuations in the range of 2-50 Hz) in each band. The receiver and stimulator then convert the signal envelopes into a series of "signal-shaped" pulse trains that activate the electrodes of the implant array. The activated electrodes in turn directly stimulate the neurons of the auditory nerve. This provides the brain with a signal that captures important features of the original signal in the environment. This method of delivering sound to the brain, referred to as electric hearing, lacks the temporal and spectral resolution of sounds that are received by individuals with a normal auditory system (i.e., acoustic hearing) (Rubinstein & Miller, 1999). However, the impoverished signals of a CI are still able to relay enough information

for high intelligibility of speech for many individuals (O'Donoghue, Nikolopoulos, & Archbold, 2000).

Although CIs have been widely used as a treatment for the profoundly deaf, the first use of CIs in individuals with SSD was intended as a treatment for debilitating tinnitus in the deaf ear (van de Heyning et al., 2008). CIs proved to be successful in alleviating tinnitus for many patients and also had an encouraging secondary benefit: improved sound localization and hearing in noisy competing talker environments (Vermeire & van de Heyning, 2009). CIs for SSD allow for the use of two separate auditory signals (one in the implanted ear, one in the NH ear). This is in contrast to BAHAs and CROS hearing aids, which route the signals at the deaf ear to the one working ear. The availability of two distinct auditory inputs afforded with a single CI in one ear and acoustic hearing in the NH ear offers the potential for binaural hearing advantages among those with SSD.

Unfortunately, there exist several reasons that the same cues from binaural hearing utilized by NH listeners for speech perception (as previously discussed) may not be as effective for SSD-CI users. First, with regard to binaural squelch, CI users do not have access to fine structure ITDs so they would need to rely mainly on ILD information to receive contralateral unmasking (Loizou, 2006). Therefore, CI listeners must rely on accurate ILD cues for spatial hearing. Second, with regard to binaural fusion, "fused" perception is likely to be impaired due to the presence of potential distortions (discussed below) in electric hearing that are encountered by SSD-CI listeners. Because spatial release from informational masking depends on the listener perceiving the target and masking speech as coming from different spatial locations (Freyman et al., 2001), listeners would likely not get a squelch benefit if they were unable to integrate signals across the ears to

create a single perceptual object. The prediction, therefore, is that SSD-CI listeners would receive less SRM than NH listeners because diotic signals are less likely to be perceived with a fused image. Third, CI processing comes with a severe loss of pitch information making pitch cues much less effective for release from informational masking (Freyman, Balakrishnan, & Helfer, 2008).

Fortunately, there also exists compelling evidence that SSD-CIs can aid in spatial hearing. This comes from studies that have examined performance in localizing a sound source (Arndt et al., 2010; Firszt et al., 2012; Hansen et al., 2013) and from studies that have assessed the advantages for listening to speech in noise when there is a spatial separation between the two (Bernstein, Schuchman & Rivera., 2017; Buechner et al., 2010; Firszt et al., 2012; Hansen et al., 2013). CIs primarily improve speech perception for listeners with SSD in configurations where the signal is on the deaf side, and/or the masker is on the NH side. This pattern of conditions for which a benefit is observed is consistent with the idea that the CI allows users to take advantage of head-shadow effects and a better-ear listening strategy (Bernstein et al., 2017; Arndt et al., 2010; Buechner et al., 2010; Firszt et al., 2012; Hansen et al., 2013). Having two ears allows the listener to take advantage of listening to the ear with the better SNR, regardless of which side of the head receives the better SNR (Schleich et al., 2004a). The actual benefit that SSD listeners receive is smaller than in NH individuals, on the order of 2-5 dB. This is probably because the CI signal is distorted relative to that received by the NH ear, which appears to reduce the normal head-shadow advantage.

While previous studies suggest CIs can provide a head-shadow benefit, to date there is little evidence that a CI can provide people with SSD with other speech-in-noise benefits

associated with binaural hearing, namely, binaural squelch. However, the results of a pair of recent studies suggest that SSD-CI listeners may experience a binaural-squelch benefit for speech understanding in certain situations. Bernstein, Goupell, Schuchman, Rivera, and Brungart (2016) investigated whether a CI could provide benefits to speech perception in complex auditory scenes beyond those provided by the head-shadow (better-ear) advantage. They employed a paradigm that eliminated the head-shadow advantage in order to investigate whether listeners could combine information across the two ears to improve speech reception performance via a binaural benefit. This was accomplished by using headphones to present the target talker and two interfering maskers to the one acoustic ear. They then investigated the impact on performance of also presenting the same interfering masker signals to the opposite ear via direct connection to the CI. Putatively, for NH listeners presented with signals over headphones in this manner, this results in the perception that the maskers are speaking to them from the center of the head, while the target is speaking to them from the side, thereby providing a spatial cue to perceptually separate the target signal from the maskers (Bernstein et al. 2016). Figure 1.1 shows the results for the SSD-CI listeners in this study. SSD-CI listeners received a binaural benefit in conditions involving competing talkers of the same gender as the target talker. The interpretation of this result is that the spatial information provided by the CI helped listeners to perceptually separate the competing talkers in conditions where the target and maskers were easily confused with each other (informational masking). These implant users presumably capitalized on differences in the combined target and masker signals in the two ears, allowing for improved perceptual segregation of multiple competing voices. The opposite-gender masker conditions did not result in significant binaural unmasking,

presumably because these were situations with less informational masking, thus the target and interfering speech could be segregated via monaural cues (Figure 1.1; Bernstein et al., 2016). Because the target signal was not presented to the second ear, there was no better-ear advantage provided at the CI ear using this paradigm. These results show that the contralateral unmasking paradigm is an effective way to study the role of binaural squelch (integration of information across the ears via spatial cues) for the release from informational masking. Bernstein et al. (2017) found a similar result when testing SSD-CI users in the free field with the target in front and symmetric maskers on either side, so there was no long term head-shadow advantage available to the listeners. As in the Bernstein et al. (2016) study, listeners showed a benefit from the implant with same-gender interferers but not with speech-shaped noise or opposite-gender interferers.

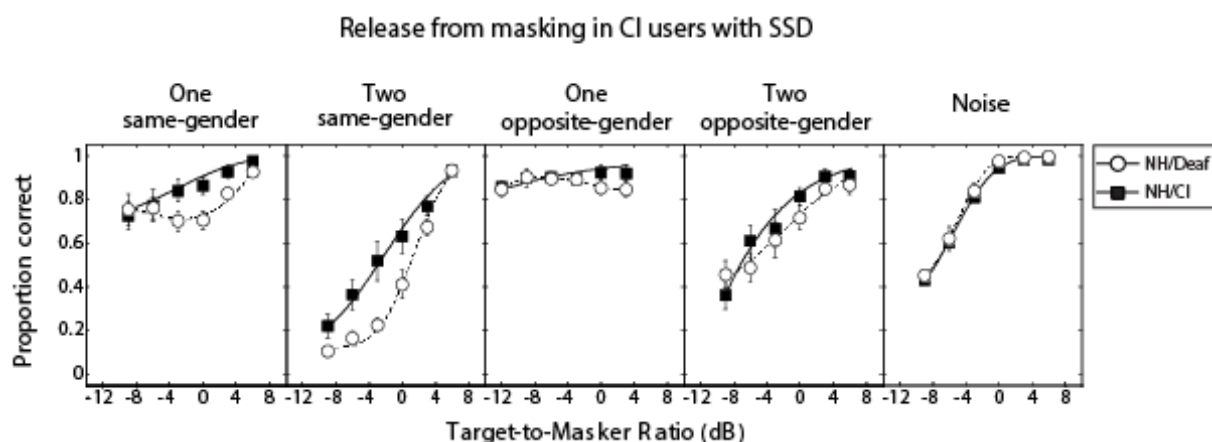


Figure 1.1. Significant improvements in performance were measured in the one and two-same gender conditions. Therefore, binaural unmasking in this population seems to occur mainly in situations with high informational masking (adapted from Bernstein et al. 2016).

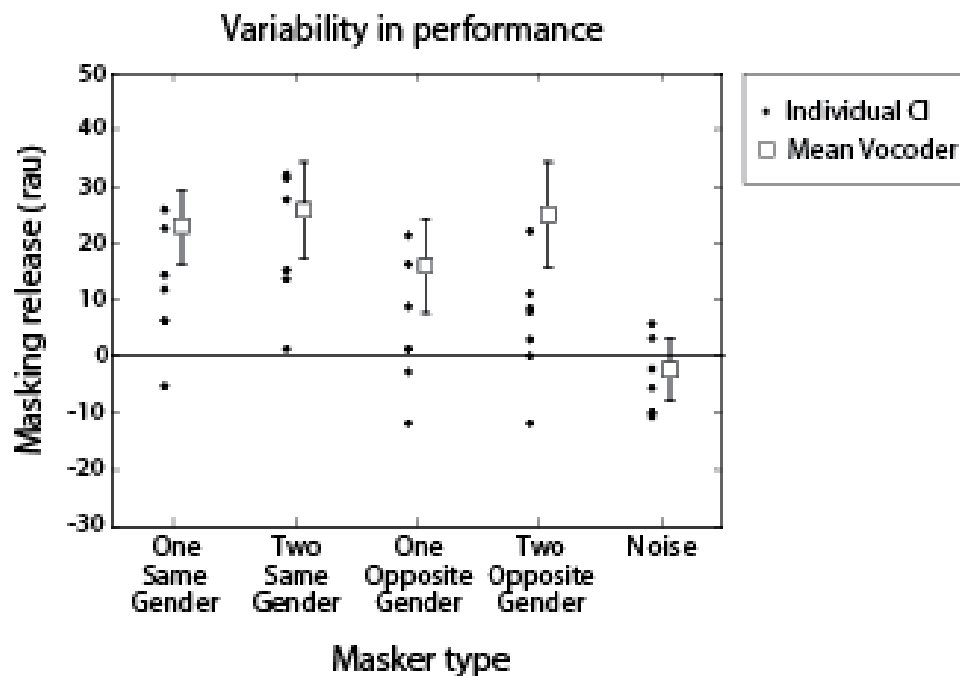


Figure 1.2. Large amounts of inter-subject variability seen among CI users with SSD. The best CI listener is at the level of the mean of the vocoder data in all speech masker situations (adapted from Bernstein et al. 2016).

Despite the advantages of CIs for SSD that have been observed, there were several indications that these listeners are not receiving the maximum benefit possible from their device. First, there was a large degree of inter-subject variability in the amount of masking release each individual patient receives with their CI (Figure 1.2) (Bernstein et al., 2016). Second, vocoder simulations of cochlear implantation for SSD presented unilaterally to NH listeners show more masking release than is observed for actual CI patients. The CI listener who had the best performance in this task was just about as good as an average NH listener who was listening to vocoder simulations (discussed in following paragraph). These results suggest that the SSD-CI listeners were not performing the task optimally. SSD-CI users could potentially receive a larger binaural benefit from their implant with performance more closely matching that of the NH vocoder listeners.

Vocoded speech presented to NH listeners is often utilized to manipulate experimental parameters involving CI processing without all of the variability inherent in actual CI users; it is a common simulation technique used in CI research (Loizou, 2006). Vocoding performs some of the same signal-processing steps that are carried out in a CI processor, including allocating the original signal into separate channels (analysis filters) within the audible speech range of 100Hz - 10,000Hz, and then extracting the envelopes from the resulting signals. These envelopes are then used to modulate an acoustic carrier signal instead of electrical pulse trains in CIs. Vocoding also permits manipulation of variables relating to CI processing while avoiding common confounds in CI data, for example, duration of CI user deafness, and differences in coding strategies and electrode configurations across different CI manufacturers. Vocoding allows for the independent manipulation of certain distortions inherent in CI processing, which allows for more careful study of each distortion.

Although useful, vocoder processing is an imperfect estimation of what CI users might hear (Freyman et al., 2008; Ihlefeld & Litovsky, 2012; Li & Loizou, 2009). Vocoder simulations can lack certain aspects that are characteristic in CI processing, such as spectral smearing, because electrical current spread is difficult to represent acoustically. Also, different coding strategies, such as continued interleaved sampling (CIS), are challenging to reproduce using a vocoder. CIS requires that the pulses sent to an electrode array are presented in non-overlapping sequences. This technique is difficult to mimic in a simulation. Nevertheless, vocoder simulations have been an invaluable tool for studying CI processing and perception. Bernstein, Iyer and Brungart (2015) and Bernstein et al. (2016) examined binaural unmasking using vocoder simulations of CI users with SSD and

the same competing talker task described above (see Figure 1.1). For the vocoder listeners, masking release was observed for all multi-talker conditions (Figure 1.3; Bernstein et al., 2016). These results are in contrast to the results from the actual CI users, because these vocoder-simulation studies show contralateral unmasking for all background masker conditions, not just for same gender interferers. While the specific reasons for the variability in performance between the CI listeners and the NH listeners who are presented with vocoded speech are unknown, there are a few possible explanations. One explanation is variation in intrinsic characteristics of individual CI listeners, which cannot easily be addressed through signal processing simulations. These include current spread (van Hoesel, 2012), lack of cortical plasticity (Litovsky et al., 2012; Maslin, Munro, & El-Dereby, 2013), spiral ganglion neural survival (Maslin et al. 2013) and duration of deafness (Blamey et al., 2012). Alternatively, the variability could be caused by certain programming characteristics of the CI, particularly distortions inherent in CI processing that could affect binaural hearing. Since these distortions can possibly be remedied by signal-processing or mapping techniques, these extrinsic characteristics will be examined in this dissertation. It is hypothesized that the actual CI users with SSD did not receive the same levels of contralateral unmasking as those in the simulation, because of distortions inherent in CI processing such as spectral, temporal and level mismatches between the CI processor and the NH ear. The next section explores how these kinds of CI distortions might affect binaural hearing.

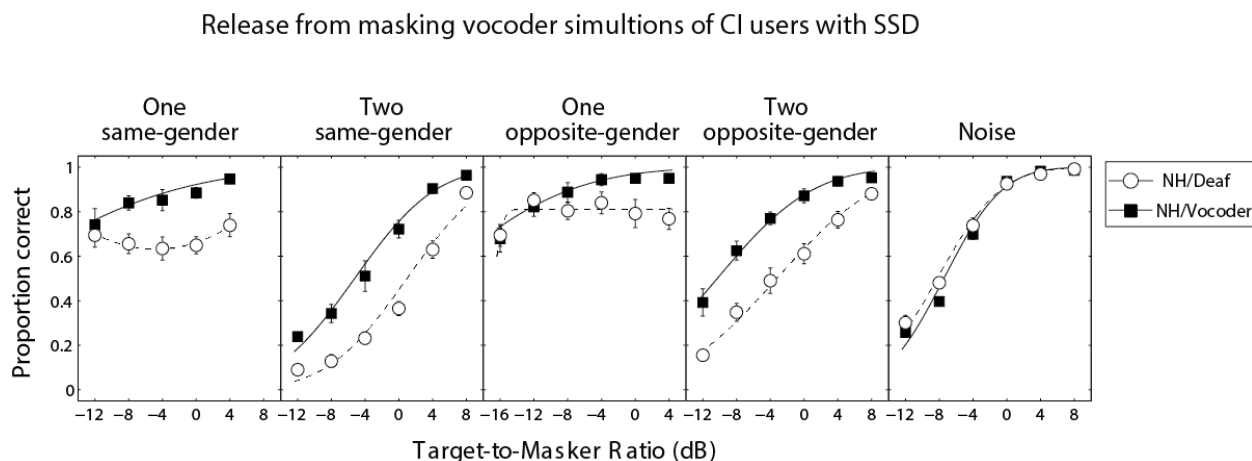


Figure 1.3. Vocoder simulations of CI users with SSD show masking release in all multi-talker conditions. No release was observed for the noise masker condition (adapted from Bernstein et al. 2016).

Possible sources of distortion in CI users with SSD

Spectral mismatches and their effect on binaural hearing.

Accurate binaural processing requires inputs that are frequency matched across the ears (Joris, Smith, & Yin, 1998). Therefore, a mismatch between the cochlear place of stimulation for the CI and acoustic ear is likely to limit binaural benefit for SSD-CI listeners. CIs are usually programmed to deliver the frequencies important for speech, between about 150 and 8000 Hz. However, the electrode array is not inserted all the way into the cochlea. As a result, this frequency mapping does not correspond to the intrinsic mapping of the basilar membrane of the inner ear. Thus, for the vast majority of CI users, there exists a large incongruity between the mapping of their CI electrode and the tonotopic

axis of their basilar membrane. For traditional CI patients with two deaf ears, this approach makes sense, because the goal of the CI is to restore as many speech cues as possible to the implanted ear. Research has shown that with months (or years) of training and experience, post-lingually deafened CI users are able to “remap” speech sounds and understand speech (Svirsky, Silveira, Neuburger, Teoh, & Suárez, 2004).

There is reason to believe, however, that this might not be the optimal approach to clinical mapping for CI users with SSD. Because SSD listeners still have one functioning ear, the main role of the CI is to assist the NH acoustic ear by providing spatial hearing benefits. For these patients, speech intelligibility via the CI alone may not be the ultimate goal. Thus, these patients might benefit from an electrode mapping that more closely matches the tonotopic organization of their NH basilar membrane, at the risk of not including some portions of the full frequency spectrum that is typically provided to CI users. This would allow for a frequency match between the implanted ear and the functioning ear, thereby potentially facilitating a larger binaural benefit. Insertion depths vary between CI recipients, either because of the properties of the electrode array or difficulties during surgery. The average insertion depth is about 20mm, but some CI users can have much shallower insertion depths (Ketten et al., 1998). A normal cochlea is about 35mm long (Fried, 1990). The Greenwood Function relates frequency selectively along the cochlea to the position of the hair cells that respond to that frequency (Fried, 1990). The Greenwood Function can be used to estimate the lowest frequencies that a CI user can hear. For example, if an electrode array was fully inserted (25mm), the lowest frequency auditory nerve fiber characteristic frequency that can be stimulated by a CI would be around 500Hz. Thus, if a place-matched mapping strategy was implemented for those with SSD-CI

listeners, they would lose some low frequency speech information from the CI signal. However, this loss of low-frequency information is much less likely to be as deleterious to SSD-CI listeners as for traditional CI listeners, because the SSD listener can rely on their NH ear for low-frequency speech cues. Because head-shadow is minimal at low frequencies (Bronkhorst & Plomp, 1988; Rayleigh, 1907), there will be very little difference in SNR between the two ears in this frequency range. Therefore, any low-frequency speech cues that are available at the CI ear will also be available at the NH ear.

Place-of-stimulation mismatches in bilateral CI (BICI) users are known to disrupt calculations of spatial cues, such as timing and level differences. Small interaural frequency offsets can cause a substantial disruption in localization of a free-field sound source (Goupell, Stoelb, Kan, & Litovsky, 2013; Kan, Stoelb, Litovsky, & Goupell, 2013; Litovsky et al., 2012). The effect of small offsets is measured by changes in just notable differences (JND) in ITD and ILD perception. Small mismatches of ± 2 electrode pairs can change perception of ITDs and ILDs in a lateralization task. However, when JNDs were estimated from the lateralization data it was found that ILDs were generally more immune to these interaural mismatches than were ITDs. Interaural mismatch led to a doubling of normalized JNDs for ITDs with only a 3mm mismatch, for ILDs the mismatch increased to 12 mm before this occurred (Kan et al., 2013). Therefore, even a small mismatch between ears would require a larger change in stimulus location to be correctly localized. Binaural fusion has also been shown to be limited in BICI when spectral mismatches are applied; listeners report unfused auditory images and often perceive multiple auditory images when there should only be one (Kan et al., 2013). Goupell et al. (2013) examined the effect of interaural frequency mismatch on binaural fusion in NH participants listening

to bilateral vocoder stimuli. They found that listeners were more likely to report multiple auditory images (i.e. stimuli were not fused) with increasing spectral mismatch between the ears. Corroborating the work from Goupell et al. (2013), work from Kan et al. (2013) performed the same interaural spectral mismatch experiment but in actual BICI listeners. They found that increasing mismatch led to perception of multiple auditory images in some of the listeners and more variability in responses across listeners. Taken together, these two studies indicate that spectral mismatch impairs fusion in both CI listeners and vocoder listeners alike. This degradation of binaural cues with frequency mismatch is likely to affect binaural squelch and subsequent contralateral unmasking in CI listeners.

SSD-CI listeners often show difficulty obtaining binaural summation after implantation, in contrast to BICI users who are able to obtain summation relatively quickly after bilateral implantation (Dunn, Tyler, Witt, Ji, & Gantz, 2012; Eapen, Buss, Adunka, Pillsbury, & Buchman, 2009). Binaural summation refers to the listening advantage obtained by having two copies of the same signal (i.e. one in each ear), the loudness of the signal is increased and it can lead to improved detection thresholds (Reynolds & Stevens, 1960). Binaural fusion is a related process but refers to a listener's ability to combine information across the ears to create the percept of a single fused sound (discussed previously for NH listeners). Aronoff, Shayman, Prasad, Suneel, and Stelmach (2015) tested binaural fusion with temporal and spectral compression in vocoder simulations of SSD-CI listening. They tested fusion by presenting vocoded stimuli in one ear and unprocessed stimuli in the other, and asked listeners if they heard one sound or two. The authors then applied various levels of spectral compression and found that more spectral mismatch resulted in less binaural fusion. Reiss et al. (2014) examined fusion using

dichotic tones by presenting dichotic tones in a five-alternative forced choice (AFC) task to CI users with residual hearing in their non-implanted ear. These listeners had moderate-to-severe hearing loss and were fitted with a hearing aid in their acoustic ear and are referred to as bimodal CI listeners. To examine fusion, the authors presented a stimulus simultaneously to the implant and the acoustic ear. The listeners were asked if they heard 1 or 2 sounds. If one sound was selected, they were asked to report which ear had the higher pitch or if they had the same pitch (indicating fusion). Many of the listeners reported fusion ranges of an octave or more (ranges much higher than measured in NH listeners). Most interestingly, the fusion ranges tended to match the pitch mismatch between mapped electrode pitches. The authors suggested that listeners might be compensating for spectral mismatch by increasing binaural fusion ranges, at least for bimodal CI users (Reiss et al., 2014). Even though bimodal CI listeners and BICI users are different than CI users with SSD, due to the availability of one relatively normal acoustic ear, the binaural cues available to SSD listeners are nevertheless limited by the poorer CI ear. Thus for SSD-CI listeners, it is likely that a spectral mismatch would reduce the ability to stream simultaneous voices and thus reduce binaural benefits for speech in noise.

The aforementioned studies examining spectral mismatch and fusion used simple tonal stimuli or stimuli presented to a single electrode. In studies examining the impact of spectral mismatch on perception of more complicated stimuli such as speech, the results are generally the same—spectral mismatch reduces binaural fusion and intelligibility of bilaterally presented stimuli. Given the widely reported effects of frequency mismatch on binaural fusion and the fact that spectral mismatch is essentially guaranteed to be present in SSD-CI listeners, spectral mismatch is a likely contributor to limitations and variability

in binaural unmasking benefits for SSD-CI listeners. Therefore, the experiments described in Chapter 2 examined the effect of a linear spectral mismatch on binaural squelch, using the contralateral masking paradigm developed by Bernstein et al. (2015; 2016) to examine the benefit of CIs for SSD in a situation where the CI does not produce a head-shadow benefit. The experiments in Chapter 4 examined the effects of a more realistic spectral shift (based on published CI insertion angle data; Landsberger, Svrakic, Roland, & Svirsky, 2015) on binaural fusion.

Temporal disparities between cochlear implants and normal hearing ears.

Many of the binaural advantages that NH listeners experience depend on the detection of temporally coherent signals across the ears. Therefore, a temporal delay between a signal received by the NH ear and the CI ear could negatively impact spatial hearing performance for SSD-CI listeners. The net temporal delay between the CI and NH ears is determined by the relative delay between electrical and acoustic processing. The temporal delay of the CI signal depends on the manufacturer of the device, the stimulation rate of the processor, and the coding strategy employed (Green, Faulkner, & Rosen, 2002). Temporal responses for NH ears depend on the mechanical properties of the traveling wave in the cochlea and the firing rates of the auditory nerve fibers to encode acoustic input. Electrically evoked auditory brainstem responses (eABRs) have been used to compare latency differences at the level of the brainstem for NH and CI users. With direct stimulation of the auditory nerve, the latency of the traveling wave in a normal cochlea is on average 4 - 8 ms slower than the rate of a CI processor, when measured at the level of

the inferior colliculus (wave V) (Dooley et al., 1993; Rasetshwane, Argenyi, Neely, Kopun, & Gorga, 2013a). This traveling wave is not replicated in a CI. However, in most cases the delay of the CI speech processor is even longer than the latency of the traveling wave and neural transduction in a NH ear. The delay of the speech processor is also not uniform across CI manufacturers. To get an idea of the estimated delay associated with the speech processor, we contacted research staff for each of the major CI manufacturers who provided estimates. Cochlear Ltd. uses a coding algorithm that induces a delay at the CI ear of about 10.5 to 12.5 ms relative to an acoustic ear (vanDijk, private communication 2015). However, Med-El uses a proprietary processing scheme that incorporates frequency-dependent group delays. Thus, their devices have a delay that more closely matches that of an acoustic ear on the order of 0.5 to 1.6 ms relative to the CI (Zirn, Arndt, Aschendorff, & Wesarg, 2015). The delay of Advanced Bionics devices falls between that of Cochlear and Med-El, with a delay of about 9 to 11 ms relative to an acoustic ear (Litvak, private communication 2016).

Differences between the latency in the acoustic ear and delay in the CI ear make it nearly impossible for a listener to obtain useful ITD information, given that the maximum delay for real-world sounds is less than 1 ms (Middlebrooks, 1999). Additionally, previous research has found that bilateral interstimulus intervals greater than 1 ms decrease binaural fusion in NH children performing binaural fusion tests (Chermak & Lee, 2005). The fusion stimuli used in Chermak and Lee (2005) were dichotic, white noise stimuli. In contrast, the binaural fusion of speech stimuli may be more resilient to interaural delays. The auditory system has the ability to suppress echoes that occur within a certain time window after the initial stimulus. This “precedence effect” results in the echo not being perceived as a

separate object. The echo threshold for speech is on the order of 30 ms (Litovsky et al. 1999; Stecker & Hafter, 2002).

Temporal disparities between a CI ear and a NH ear in people with SSD might occur not only due to differences in encoding in the CI and NH ears, but also as a result of physiological changes in the brain after deafness. Duration of unilateral deafness is known to impact cortical as well as subcortical circuits, including at the level of the brainstem (Dong, Mulders, Rodger, & Robertson, 2009). The brainstem is highly involved in ITD computations for humans and other mammals alike (Grothe, Pecka, & McAlpine, 2010). Abnormal timing delays have been measured in eABRs after unilateral deafness, suggesting the brainstem is susceptible to changes in input soon after deafness (Gordon, Valero, van Hoesel, & Papsin, 2008). These timing changes in neural circuitry can affect how well a CI ear and a NH ear can integrate temporal information and perform spatial computations, which could also impact binaural unmasking. Thus, a temporal mismatch might not only arise from processing delays in CIs, but can be inherent to the brain after deafness. Aside from changes in the brainstem that occur after deafness, this temporal disparity between a NH ear and a CI ear might be mitigated by speeding up CI processing by ~5-10 ms. This could potentially limit any binaural processing issues that could arise from altered timing between the ears for those with SSD.

The effect of timing disparities on hearing depends on the listening situation. For binaural calculations of timing differences between the ears, there is very little latitude for delays introduced by CI processing. This is because natural ITDs occur in < 1 ms. Therefore, timing differences between CI and acoustic processing would render any useable timing differences useless for spatial processing. In contrast, speech perception is

generally immune to delays up until around 40 ms (echo threshold), which affords much more leeway in terms of speech understanding, even with a large CI delay. This is an interesting contrast, since the contralateral unmasking paradigm used in this dissertation involves both spatial cues and speech perception. Experiment 2.2 in Chapter 2 will investigate the effect of interaural disparities in temporal delay on contralateral unmasking.

Loudness growth, compression and their effects on binaural hearing.

There exists a large dynamic range (DR) difference between acoustic and CI ears. This DR disparity results in very different loudness growth between the two. Because ILDs are so important for relaying binaural cues to CI listeners (Litovsky et al., 2004; van Hoesel & Tyler, 2003), differences in loudness growth between the CI and NH ears are likely to affect spatial hearing. In CIs, loudness is encoded by the amount of electrical charge delivered by the current-pulse amplitudes. When the amplitude of the current is increased, the loudness percept is also increased. The smallest possible change in charge that can be produced by the CI processor results in large increases in perceived loudness, which has the effect of reducing the available DR. It is common for CI users to have a reduced total DR of about 40dB (McDermott & Varsavsky, 2009), whereas the DR of hearing for a healthy NH ear is approximately 120 dB (Moore, 2003).

McDermott, McKay, Richardson, and Henshall (2003) describe in detail the loudness-encoding scheme of a CI processor. A signal is received at the CI microphone and converted into an electrical signal, then amplified with an automatic gain control (AGC) mechanism. The AGC circuit is similar to that used to amplify signals in a hearing aid. The

AGC limits the range of sound levels sent to the processor to include those that are above the noise floor of the processor. The AGC usually discards signals below about 25 dB sound pressure level (SPL) and maps signals in the range of 25 dB to 65 dB SPL to the listener's electrical dynamic range. Stimulus levels above about 65 dB SPL are usually "compressed" and represented at the maximum electrical stimulus level, equivalent to a 65 dB SPL signal. The internal noise of the CI in combination with the electrical response characteristics of individual neurons prohibits proper encoding below a certain threshold, usually deemed the T-level, which is the threshold for electric hearing. A comfortable C-level is then computed as a loud but bearable maximum level for the implant user. Sounds louder than the C-level are compressed to fall at or below the C-level. These measurements are made for each electrode channel. This type of loudness programming contributes to the reduced dynamic range and loudness-growth issues in CI listeners.

With respect to loudness growth, the electrode neural-interface has also been implicated as a potential source of variability in CI users. Larger than average dynamic ranges (for CI listeners) have been correlated with higher amounts of spiral ganglion cell survival (Kawano, Seldon, Pyman, & Clark, 1995). The electrode neural-interface and health of the auditory nerve determines how well a given CI listener will be able to code intensity information. The electrode-neural interface broadly refers to the physical junction between the individual electrodes on the CI array and the corresponding neurons along the basilar membrane. However, many peripheral factors contribute to the interface, such as electrode placement, scar tissue growth, bone regeneration and the number and integrity of the spiral ganglion neurons in the cochlea (DeVries, Scheperle, & Bierer, 2016). These factors can contribute to current spread and channel interactions, which can interfere with

transmission of speech information and lead to pitch perception impairments (Crew, Galvin, & Fu, 2012; Jones, Won, Drennan, & Rubinstein, 2013). Most specifically, loudness growth in CI listeners depends on the proximity of surviving spiral ganglion neurons to the location of the active electrodes on the array. Therefore, the poorer the electrode-neural interface, the more compression is needed to encode amplitude and the poorer loudness will be represented for the CI listener. A related limitation of CI amplitude processing is that there are a limited number of discriminable sound intensity steps available to the listener. Schroder, Viemeister, and Nelson (1994) estimated that the total number of discriminable intensity steps for NH listeners is about 83. In contrast, it has been estimated that the number of intensity steps for CI listeners ranges from 7 - 45 (best case) and this number is highly variable across CI listeners (Nelson, Schmitz, Donaldson, Viemeister, & Javel, 1996). The number of discriminable intensity steps is thought to be important for identifying different speech formants via differences in perceived loudness in adjacent frequency channels (Stafford, Stafford, Wells, Loizou & Keller., 2014). The ability to reliably identify formants is especially important for segregating different talkers in multi-talker environments.

In order to fit the full dynamic range of an acoustic ear into the limited dynamic range of the CI, compression algorithms are implemented in CI processing. The most common compression technique is static envelope compression. This method uses a fixed compression ratio, meaning the ratio is the same over time. Although commonly used, this strategy is not optimal and does not enable the listener to make the best use of their limited DR. In static envelope compression the ratio is fixed across channels, instead of being optimized for each individual channel. However, due to the wide use of static envelope

compression algorithms in CI processing, the effect of this type of compression on spatial hearing was examined in Chapter 3 (Experiment 3.1 and 3.2).

The presence of envelope compression in CI processing is likely to impact speech perception for SSD-CI listeners in several different ways. First, distortions occur in speech after envelope compression is applied in CI processing. Envelope compression is known to smear acoustic landmarks important not only for vowel comprehension, but for identification of word boundaries (Li & Loizou, 2009). These distortions introduced by compression could impair fusion of spatially separated maskers due to distortion of the signal envelope in the CI or vocoded ear. Second, CI envelope compression is likely to distort the ILD cues that are important for spatial hearing (Grantham et al., 2008). Because CIs do a poor job of relaying ITD information, CI listeners mainly rely on ILD cues for spatial hearing, that is in order to localize sounds (Dorman et al., 2015) and to identify differences in spatial location between concurrent sources in the environment. Finally, envelope compression is also likely to affect masked speech perception for SSD-CI listeners by changing the effective SNR (i.e., the ratio between the target and masker levels in the CI ear). In competing-talker speech tasks, performance can vary in a complex way as a function of the relative levels of the target and masker speech in each ear. Compression amplifies quieter sounds relative to louder ones; therefore, compression will have a different effect depending on the relative levels and spatial locations of the targets and maskers. Depending on the situation, compression could benefit the listener or impair performance. For example, if compression results in the talker of interest becoming louder in the acoustic ear it could improve the TMR in that ear, and consequently improve performance. In some cases, compression could make the target and masker signals more

similar to each other, and therefore reduce the perceived difference in spatial location, thereby impairing performance. The effects of amplitude compression and expansion on contralateral unmasking and head-shadow benefit were examined in the experiments described in Chapter 3 of this dissertation.

Chapter 2: The effect of interaural mismatches on contralateral unmasking in vocoder simulations of cochlear-implant listeners with single-sided deafness

The work described in this chapter is published in *Ear and Hearing*.

Wess, J.M, Brungart, D.S, Bernstein, J.G.W. (2017). The Effect of Interaural Mismatches on Contralateral Unmasking With Single-Sided Vocoders. *Ear Hear.* 38, 374-386.

Introduction

Binaural hearing provides a number of benefits for NH listeners in noisy environments (Zurek, 1993). Head-shadow effects allow listeners to obtain a substantial listening benefit simply by listening to the “better ear” where the SNR of the target is most favorable. In addition, having two ears can generate an additional “squelch” benefit by allowing the listener to take advantage of precise timing and level differences between the signals arriving at the two ears to increase the intelligibility of the target speech in the presence of spatially separated masking sounds (Drullman & Bronkhorst, 2000). As it relates to the current study, squelch or contralateral unmasking is defined as the improvement in speech understanding when the speech and noise are spatially separated and the ear with the poorer SNR is added. Squelch is particularly beneficial in situations involving multiple competing talkers, whereby interaural difference cues provide information to facilitate the perceptual separation of competing sound sources (Hawley et al., 2004). Overall, NH listeners show substantially more binaural benefit (about 3-5 dB)

for speech understanding in noise compared to BICI listeners (Aronoff, Freed, Fisher, Pal, & Soli, 2011; Culling, Jelfs, Talbert, Grange, & Backhouse, 2012).

Individuals with SSD — one NH ear and one deaf ear — are at a severe disadvantage when listening to speech in complex listening environments because they lack the benefits of binaural hearing (e.g., squelch, head-shadow) that are available to individuals with two healthy ears (Welsh, Rosen, Welsh, & Dragonette, 2004). If treated at all, typical treatments for SSD included osseointegrated or CROS hearing aids. Hearing-aid treatments have been successful in alleviating some of the adverse effects of SSD, by providing access to signals presented from the deaf side of the head by routing them to the NH side (Stewart, Clark, & Niparko, 2011). However, these treatments do not restore access to binaural cues and these patients still have trouble hearing in noisy environments and difficulty with sound localization (Grantham et al. 2008).

In the past several years, CIs have been considered as a possible new treatment option for SSD¹. Although CIs have been widely used as a treatment for the profoundly deaf, the first use of CIs in individuals with SSD was to treat debilitating tinnitus in the deaf ear (van de Heyning et al., 2008). Since then, a number of studies have found that CIs can also improve sound localization and speech perception in noise for individuals with SSD (Arndt et al., 2010; Buechner et al., 2010; Erbele, Bernstein, Schuchman, Brungart, & Rivera, 2015; Firszt et al., 2012; Hansen et al., 2013; Vermeire & Van de Heyning, 2009; Zeitler et al., 2015). In general, the benefit provided by a CI for speech perception is observed in configurations where the target signal is on the deaf side and/or the interferer is on the NH side. This is consistent with the idea that the CI allows users to take advantage of head-shadow effects and a better-ear listening strategy (Arndt et al. 2011; Buechner et

al. 2010; Firszt et al. 2012; Hansen et al. 2013; Zeitler et al. 2015), with little evidence of binaural squelch. A number of studies have measured binaural squelch for bilateral CI listeners, and have found that a second CI provides either no squelch at all (e.g., Loizou et al., 2009; Tyler, Noble, Dunn, & Witt, 2006), or very modest squelch effects on the order of 1–2 dB (e.g., Eapen et al., 2009). This likely reflects the fact that CIs do not deliver the temporal fine-structure information (van Hoesel, 2012) that allows the NH binaural system to take advantage of ITDs to increase the effectiveness of binaural hearing. Contralateral unmasking has also been demonstrated in CI users, insofar as they are able to detect tones embedded in noise that is uncorrelated across processors (Long, Eddington, Colburn, & Rabinowitz, 2003). BICI listeners have demonstrated the ability to detect changes in interaural envelopes, although just noticeable differences for CI users are much worse than for NH listeners (Goupell & Litovsky, 2015). It is unknown how well SSD CI listeners can detect changes in envelope correlation across the ears.

Bernstein et al. (2015, 2016) recently provided some evidence suggesting that SSD-CI listeners can benefit from squelch in certain situations². They employed a paradigm that did not provide any head-shadow benefit, thus ensuring that any observed advantage of the CI could be attributed to squelch. They presented a mixture containing a target talker and one or two interfering talkers to the acoustic ear, and a mixture containing a copy of the interfering talkers to the CI ear. For NH listeners presented with unprocessed signals, this paradigm results in the perception that the interferers are originating at the center of the head, while the target is speaking to them from the side, thereby providing a reliable spatial cue that can be used to perceptually segregate the target signal from the interferers (Freyman et al. 2008). Thus, it is not surprising that NH listeners obtained a substantial

benefit in this listening configuration. What is more surprising is that SSD-CI listeners (and NH listeners presented with vocoder simulations of SSD-CI listening) also received substantial benefit when target and interfering talkers were of the same gender, such that few monaural cues (i.e. voice, pitch and timbre) were available to allow the listener to perceptually separate the concurrent talkers. This effect was likely mediated by the ability of the listeners to fuse the unprocessed and CI (or vocoder)-processed interferer waveforms across the ears, to more easily perceptually separate the monaural target from the binaurally presented interferers.

Despite the evidence of squelch apparent in the average results for SSD-CI listeners, Bernstein et al. (2016) observed a large amount of intersubject variability in the magnitude of contralateral unmasking. While the specific reasons for the inter-subject variability in performance are unknown, there are several possible explanations. First, there are intrinsic characteristics of the individual listener that cannot easily be fixed or addressed through signal processing means, and that are known or suspected to influence speech perception for traditional CI listeners. These include neural survival (e.g., Maslin et al. 2013), current spread (e.g., van Hoesel 2012), duration of deafness (e.g., Blamey et al., 2013) and lack of cortical plasticity (e.g., (Litovsky et al., 2012; Maslin et al., 2013)). Second, the variability might also reflect certain extrinsic CI distortions that are potentially rectifiable via signal-processing or clinical-mapping procedures. These include frequency mismatch between the ears (brought on by electrode placement and mapping procedures), timing incongruities originating from different processing latencies between a CI and NH ear, and loudness distortions due to CI compression and reduced dynamic range. The current study focused on two of these extrinsic factors — spectral and temporal mismatch

— that have the potential to be addressed by either adjusting the frequency allocation tables or introducing a temporal delay to one ear.

Spectral mismatch — i.e., a mismatch between the cochlear places of stimulation between the CI and acoustic ears — is one of the most obvious extrinsic factors that might negatively impact contralateral unmasking for SSD-CI listeners. The average insertion depth of a CI is about 20 mm, with some CI users experiencing much shallower insertion depths (Ketten et al., 1998), whereas a normal cochlea is approximately 35 mm long (Greenwood, 1961). As a result, the CI electrodes are generally unable to stimulate the apical portions of the cochlea where the lowest frequencies (approximately 500 Hz and below) are typically processed. For the profoundly deaf, speech perception through the CI is the primary goal of cochlear implantation; therefore, CIs are often programmed to correspond to the frequencies most important for speech perception, between 150 and 8000 Hz (van Hoesel, 2012). This results in a large incongruity between the frequencies mapped to a given CI electrode and the acoustic best frequencies of the spiral ganglion neurons adjacent to the electrode (Landsberger et al., 2015; Stakhovskaya, Sridhar, Bonham, & Leake, 2007). Radiographic insertion depth data from Landsberger et al. (2015) can be used to estimate the mismatch between the cochlear place of stimulation for a given electrode and the associated place of cochlear stimulation for an acoustic stimulus at that electrode's allocated center frequency. For an average CI patient with a default frequency map, this mismatch is approximately 4-6 equivalent rectangular bandwidths (ERBs) (3.6 – 5.4 mm), depending on the manufacturer and the specific electrode within the array. However, the intersubject variability in insertion angle, and therefore the electric-acoustic mismatch, is substantial. Landsberger et al. (2015) also conducted a literature survey and

reported the across-subject mean and standard deviations of the insertion angles of the most apical electrode across a number of studies. By combining this information across studies and averaging across the three major CI manufacturers, we estimate that the range of electric-acoustic mismatch in the cochlear place of stimulation for 95% of CI users (i.e., ± 2 standard deviations from the mean) extends from -0.6 to 12 ERBs (-0.5 to 11 mm). Although there is some evidence that post-lingually deafened CI users are able adapt to the shifted stimulus to “remap” speech sounds and better understand speech (e.g., Reiss, Turner, Erenberg, & Gantz, 2007; Svirsky et al., 2004), this plasticity is likely to be incomplete, especially for individuals with an extraordinarily large mismatch, meaning that SSD-CI listeners might still benefit from an improved match between the frequency allocation of the CI and their normal acoustic ear.

A temporal delay between the NH and the CI ears could also limit contralateral unmasking. The processing delay for a CI depends on the stimulation rate of the processor as well as the coding strategy employed (Green et al., 2002). There is no uniform delay across CI manufacturers. For example, Cochlear Ltd. uses a filtering and coding strategy that causes a delay at the CI ear of about 10.5–12.5 ms relative to an acoustic ear (van Dijk, private communication 2015). In contrast, Med-El finite impulse response filters have integrating group delays with decreasing frequency that more closely match the traveling-wave latencies for an acoustic ear, resulting in frequency-dependent delays in the acoustic ear on the order of 0.5–1.6 ms relative to the CI (Zirn et al., 2015). Advanced Bionics processing latency falls somewhere in between, with 9–11 ms latency relative to an acoustic ear (Litvak, private communication 2016). Delays of this magnitude in either direction would make it very difficult to relay accurate ITD information, given that the

maximum interaural delay for real-world sources is less than 1 ms for humans (Middlebrooks, 1999), and sensitivity to ITD is known to deteriorate when the delay exceeds 300–500 μ s (Mills, 1960). However, contralateral unmasking for SSD-CI listeners is likely to depend on interaural correlation between the temporal envelopes of the stimuli processed in the two ears. Since speech envelopes are dominated by very slow 2–8 Hz temporal modulations (Elliott & Theunissen, 2009), the contralateral unmasking effect might be more resilient than ITD discrimination to interaural time delays on the order of 12 ms or less.

A third factor that could influence contralateral unmasking is spectral resolution of the CI due to physical current spread in the cochlea (Nie, Barco, & Zeng, 2006). Whereas an acoustic ear is capable of encoding 30–50 channels of spectral information (Shannon, Fu, & Galvin, 2004), a CI has only about 8 functional channels (Friesen, Shannon, Baskent, & Wang, 2001). Poor spectral resolution has been shown to limit spatial release from masking for CI users in noisy environments, such as competing talker situations (Fu & Nogaki, 2005). Spectral resolution is, for the most part, an intrinsic characteristic of the electrode-neural interface that cannot easily be overcome with signal-processing or clinical mapping solutions. Nevertheless, the degree of spectral resolution could have an impact on the extent to which spectral mismatch affects contralateral unmasking.

This study used vocoder simulations to investigate the extent to which spectral and temporal distortions may negatively impact contralateral masking release in NH listeners presented with unprocessed stimuli in one ear and vocoded stimuli in the other. Vcoders simulate certain aspects of CI processing such as filtering the original speech signal, extracting the amplitude temporal envelope and exciting a certain physical location in the

cochlea (Goupell et al., 2013). Although vocoder simulations are not considered a comprehensive CI simulation, this approach allowed us to directly control the amount of interaural spectral and temporal mismatch. This study took as a starting point the previous results from Bernstein et al. (2016) showing that (a) SSD-CI listeners experienced contralateral unmasking, and (b) that the average magnitude of contralateral unmasking for NH vocoder listeners was equal to the largest observed benefit for an individual SSD-CI listener. We expected that contralateral unmasking would be sensitive to distortions in all three dimensions of interest (spectral and temporal mismatch and spectral resolution), and that for an extreme case in each dimension, contralateral unmasking would disappear completely. Given the large amount of inter-subject variability in interaural mismatch reported in the literature – especially in the spectral dimension – the goal was to determine to what extent contralateral unmasking would be resilient to interaural mismatch.

Interaural temporal and spectral mismatch, along with the spectral resolution of the vocoder, were varied parametrically in a series of four experiments to measure how large of a mismatch could be tolerated in each dimension before contralateral unmasking was reduced or eliminated. Experiment 2.1 examined the effects of interaural spectral mismatch on contralateral unmasking. Experiment 2.2 examined the effects of an interaural temporal mismatch. Experiment 2.3 examined the interaction between frequency resolution of the vocoder and interaural spectral mismatch. Experiment 2.4 examined the interaction between interaural spectral and temporal mismatch.

Experiment 2.1: The role of spectral mismatches on contralateral unmasking in simulations of CI users with SSD

Experiment 2.1 investigated the effect of interaural spectral mismatch on contralateral unmasking. We hypothesized that spectral mismatch would reduce contralateral unmasking, with the idea that the dissimilarity between the acoustic interferers presented to one ear and the vocoded interferers presented to the other ear would negatively impact performance.

Methods.

Participants. Experiment 2.1 was a pilot study carried out at Walter Reed National Military Medical Center. Seven paid listeners (age range 18-30) participated in this experiment. All listeners had NH, defined as symmetrical thresholds equal to or better than 20 dB hearing level at octave frequencies between 125 and 8000 Hz and were free from cognitive and neurological disorders. All listeners were native English speakers.

Approach. This study employed the contralateral-unmasking paradigm of Bernstein et al. (2015, 2016) to measure the squelch benefit provided by a second (vocoded) ear in perceptually separating concurrent streams of speech. The left ear was always presented with unprocessed target and interfering speech. The right ear was

presented with either silence (in the monaural condition) or a vocoded copy of the interfering speech (in the bilateral condition).

Stimuli. The target and interfering speech were taken from the CRM speech corpus for multi-talker communication research (Bolia et al., 2000; Brungart, 2001). The CRM corpus consists of phrases of the form “Ready (call sign) go to (color) (number) now.” There were eight possible call signs (“Arrow,” “Baron,” “Charlie,” “Eagle,” “Hopper,” “Laker,” “Ringo” and “Tiger”), four possible colors (“blue,” “green,” “red” and “white”), and eight possible integer numbers (one through eight, including seven). A typical sentence would be “Ready Charlie go to white five now.” The target sentence call sign was always “Baron”, which provided the cue for the listener to identify which of the concurrent talkers was the target. The interferers used other call signs (e.g., “Arrow” or “Ringo”). Eight speakers (four females, four males) were used to record all possible combinations.

Noise vocoding was used to extract speech envelopes in a number of frequency channels and use the envelopes to excite specified regions of the cochlea (via synthesis filters). The algorithm was similar to that described by Hopkins and Moore (2009) and Bernstein et al. (2015, 2016), except that the signals were further manipulated to produce spectral mismatches between the unprocessed and vocoded ears. First, stimuli were passed through a bank of linear phase finite-impulse response “analysis” filters with bandwidths proportional to the equivalent rectangular bandwidth (ERB) of a NH auditory filter (Glasberg & Moore, 1986). This particular algorithm and the definition of channels in terms of ERBs were employed to match the processing employed by Bernstein et al., (2015, 2016), although ERBs can easily be translated into millimeter equivalent distances along

the basilar membrane (0.9mm/ERB, Moore, 1986). The order of the filters were varied so that the filter skirts had similar slopes. However, the filters were engineered to be quite steep, the shape of the filter was not symmetrical. The low-end edge of the filter had ~60 dB per octave roll-off while the top edge rolled off at about 80 dB. These steep filters reduced any potential channel interactions. Delays introduced by the filtering process were offset by removing the appropriate number of samples (half the filter length) from the beginning of the output signal, so that the output signal was time-aligned with the input signal. The envelope of the signal in each channel was extracted via a Hilbert transform. Each envelope was multiplied by a white noise carrier, with the resulting signal then passed through a bandpass “synthesis” filter with cutoff frequencies selected to stimulate a specified region of the cochlea. The level of the resulting signal in each channel was adjusted to be equal to the root-mean-squared (RMS) level of the input signal for that channel, and the delays associated with the filtering process were removed. Finally, the signals were summed across channels to create the noise-vocoded signal.

Interaural spectral mismatch was introduced through the use of synthesis filters that did not match the analysis filters used to extract the envelope, thereby stimulating a different cochlear place than would be stimulated by an unprocessed acoustic signal. This was done, rather than shift the frequencies of the analysis filters, to simulate the large range of possible electrode positions within the cochlea across a population of CI listeners (Landsberger et al., 2015), and to determine at what point an interaural spectral mismatch would harm performance. The synthesis-filter cutoff frequencies were shifted upward or downward relative to the analysis filters by 1, 2, 4 or 7 ERBs (equivalent to 0.9, 1.8, 3.6 or 6.3 mm). A 6-channel vocoder was used in this experiment, with each channel 4 ERBs (3.6

mm) wide. The frequency range of the analysis filters was 100 to 2502 Hz. It was important to ensure that any effects of the spectral processing reflected the introduction of a spectral mismatch between the two ears and not the loss of audibility of a portion of the speech spectrum. Therefore, the vocoder high-frequency cutoff was set to a lower frequency than is customary to allow for the possibility of large upward spectral shifts without removing acoustic frequency content, although shifts larger than 7 ERBs were ultimately not included based on pilot results.

Procedure. To maximize the difficulty in perceptually separating concurrent talkers in the target ear, this experiment used all same-gender interferers and targets (Brungart, Simpson, Ericson, & Scott, 2001). These are the conditions that produced the most masking release for SSD-CI listeners (Bernstein et al., 2016) and for single-sided vocoder listeners (Bernstein et al., 2015, 2016). The three talkers (target and two interferers) in a given trial were always of the same gender, although the gender varied randomly from trial to trial. The three simultaneous sentences were constrained such that they were always spoken by a different talker and had a different call sign, color and number. The target speech was presented at 60 dB SPL, with the interferer level adjusted to yield the desired target-to-masker ratio (TMR). Two TMRs were tested (0 and 4 dB), because they yielded the largest amount of contralateral unmasking in the vocoder study of Bernstein et al. (2015). In the bilateral conditions, the interferers were played at the same level to both ears.

Participants were seated in a sound booth and directed their attention to a computer screen. The speech stimulus was generated by MATLAB and played via a RME Hammerfall (Haimhausen, Germany) sound card and presented over Sennheiser HD 280

headphones. The computer screen displayed an eight-column, four-row array of colored digits corresponding to the response set of the CRM. The listener used the mouse to select the colored digit corresponding to the number and color spoken by the target talker who used the call sign “Baron”. After each response, the subject received feedback, with the button associated with the correct answer flashing briefly. In order to receive a correct response, both the color and number needed to be correctly identified. Listeners were presented with 100 trials for each TMR (0 and 4 dB) in the monaural condition, and for each combination of spectral shift and TMR in the bilateral conditions, for a total of 2000 trials for each listener. Listeners were presented with blocks of 30 trials with the spectral-shift condition held fixed (or stimuli presented monaurally) for all of the trials in the block. The TMR varied randomly from trial to trial.

Results.

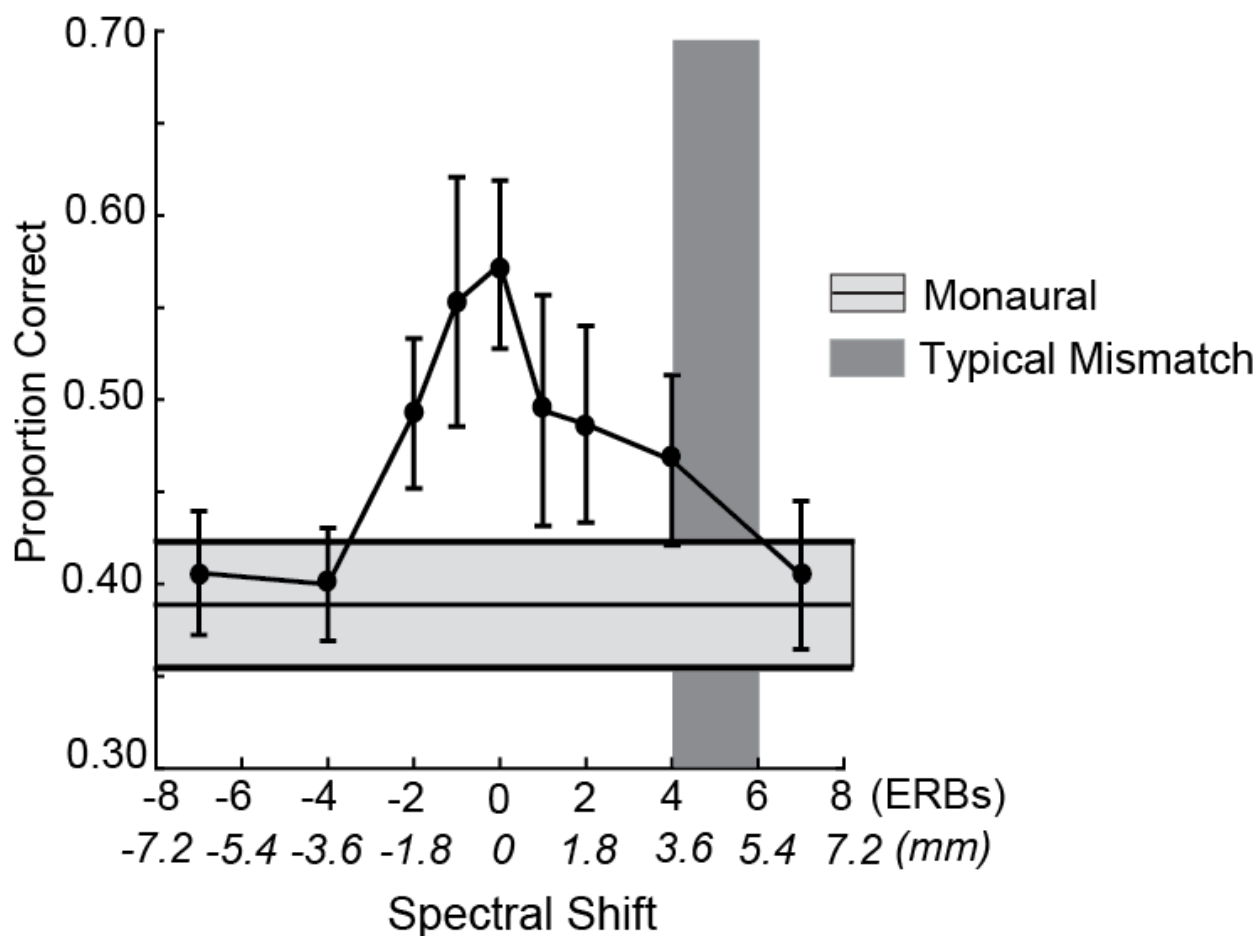


Figure 2.1. Results from experiment 2.1 plotting mean performance in correctly identifying the target number and color, with data averaged across TMR. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark gray shaded region represents the range of mismatch expected for actual SSD-CI listeners. Maximum performance was observed with no spectral shift (0 ERBs), and decreased with increasing spectral shift. Error bars represent \pm one standard error of the mean.

Figure 2.1 plots the mean proportion of trials where the color and number were both identified correctly as a function of the spectral shift. Fig. 2.1 shows the data averaged across TMR. The vertical shaded region indicates the range of expected spectral mismatch across the cochlear partition for an average CI listener (roughly 4–6 ERBs; recall, however, that the range of mismatch across individual listeners is much larger, on the order of -0.6 –12 ERBs, Landsberger et al., 2015). The horizontal shaded region indicates mean monaural performance \pm one standard error. The data in Fig. 2.1 show a clear effect of spectral mismatch on the magnitude of contralateral unmasking, with a benefit of 18 percentage points with no shift, decreasing to no benefit at all for shifts of -4 or +7 ERBs.

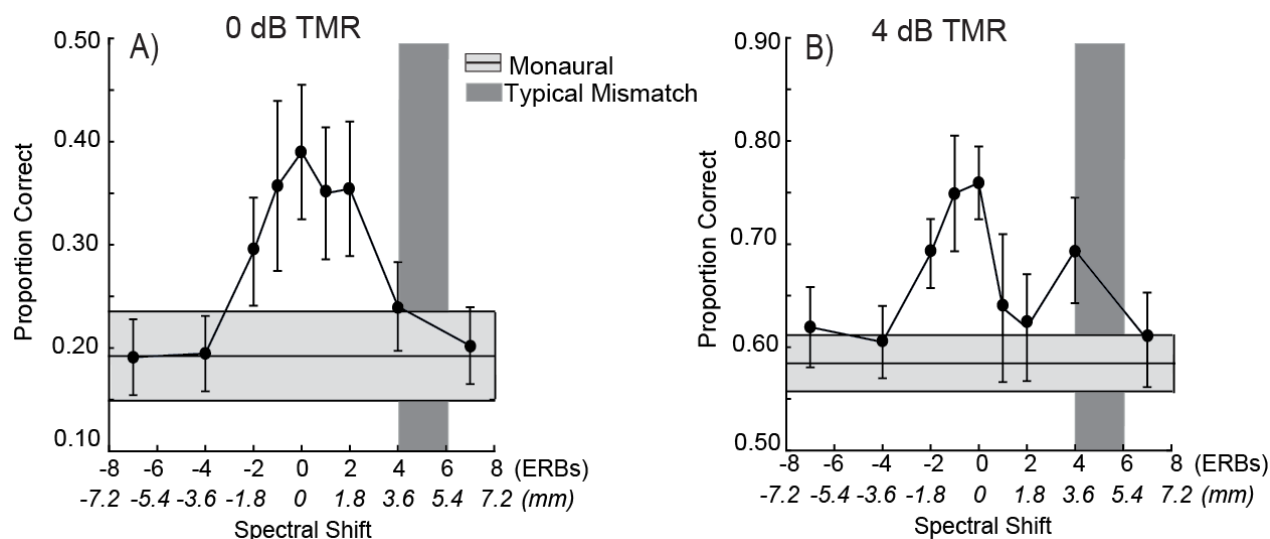


Figure 2.2. Results from experiment 2.1 plotting mean performance in correctly identifying the target number and color, with data plotted separately for the two TMRs tested: (A) 0 dB and (B) 4 dB. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark gray shaded region represents the range of mismatch expected for actual SSD CI listeners. Error bars represent \pm one standard error of the mean.

For clarity, the data have also been plotted separately for each TMR (Fig. 2.2). The data were analyzed using a repeated-measures binary-logistic regression analysis with two

within-subject factors (spectral shift and TMR). This analysis was used because the data were binary in nature (correct or not) and the analysis takes into account the likelihood that percentage-correct scores are different based on the number of trials presented. For the purposes of the statistical analysis, the monaural condition was considered as an additional spectral-shift condition. There were significant main effects of spectral mismatch [$\chi^2(9) = 2324.2, p < 0.001$] and TMR [$\chi^2(1) = 606.8, p < 0.001$] and a significant interaction between the two factors [$\chi^2(6) = 69.5, p < 0.001$]. These interactions were investigated through a series of post-hoc tests. The first set of tests sought to determine for which spectral-shift conditions contralateral unmasking was observed by comparing performance to the monaural condition, with Bonferroni corrections applied for (18) multiple comparisons. For a TMR of 0 dB, bilateral performance was significantly better than monaural performance for spectral shifts of 0, ± 1 and ± 2 ERBs ($p < 0.001$). For a TMR of 4 dB, bilateral performance was significantly better than monaural performance for spectral shifts of 0, -1, -2 and + 4 ERBs ($p < 0.001$). The second set of tests sought to determine the point at which contralateral unmasking was reduced relative to the unshifted condition. For a TMR of 0 dB, spectral shifts of -4 and ± 7 ERBs yielded reduced contralateral unmasking relative to the zero-shift condition. For a TMR of 4 dB, only the two largest negative shifts (-4 and -7 ERBs) yielded reduced contralateral unmasking.

Summary.

In summary, the results of experiment 2.1 show that significant contralateral unmasking was preserved for spectral shifts smaller than ± 2 ERBs, was significantly

reduced by spectral shifts of ± 2 -4 ERBs, and was completely eliminated by spectral shifts of ± 7 ERBs.

Experiment 2.2. The role of temporal mismatches on contralateral unmasking in simulations of CI users with SSD

Experiment 2.2 investigated the effect of interaural delays on contralateral unmasking. We hypothesized that a large interaural delay would reduce contralateral unmasking, but it was not clear to what extent the unmasking effect would be resilient to small delays.

Methods.

Participants. Eight NH paid listeners participated. Listeners were tested at the Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio. The listener panel consisted of professional listeners, in that they are paid to conduct multiple psychoacoustic experiments. All listeners were native English speakers.

Stimuli. The methods were generally the same as in experiment 2.1, except that interaural disparities were implemented in the temporal instead of the spectral dimension. Vocoder processing was carried out in 8 frequency bands covering a range of 100 to 10,000 Hz. This full-bandwidth vocoder was employed because no spectral shifts were applied.

Interaural temporal mismatches were induced by delaying the vocoded signals presented to the right ear (defined as a positive delay) or by delaying the unprocessed signals presented to the left ear (negative delay). Temporal mismatches included $\pm 100, 50, 24, 18, 12$ and six ms.

Procedure. The data reported here form a subset of the data collected from a larger experiment exploring the cues that listeners might use to perform the contralateral unmasking task. Thus, only a TMR of 0 dB was tested, and the number of trials for the reported conditions are different than in the other experiments. This incongruity occurred because this experiment was run as part of a larger unrelated experiment. Listeners were presented with 50 trials for each temporal shift condition, 200 trials for the zero-shift condition and 500 trials for the monaural condition, for a total of 2600 trials per listener. Listeners were presented with blocks of 32 or 64 trials, with all experimental parameters (i.e., temporal shift or monaural condition) varying randomly from trial to trial within each block.

Results.

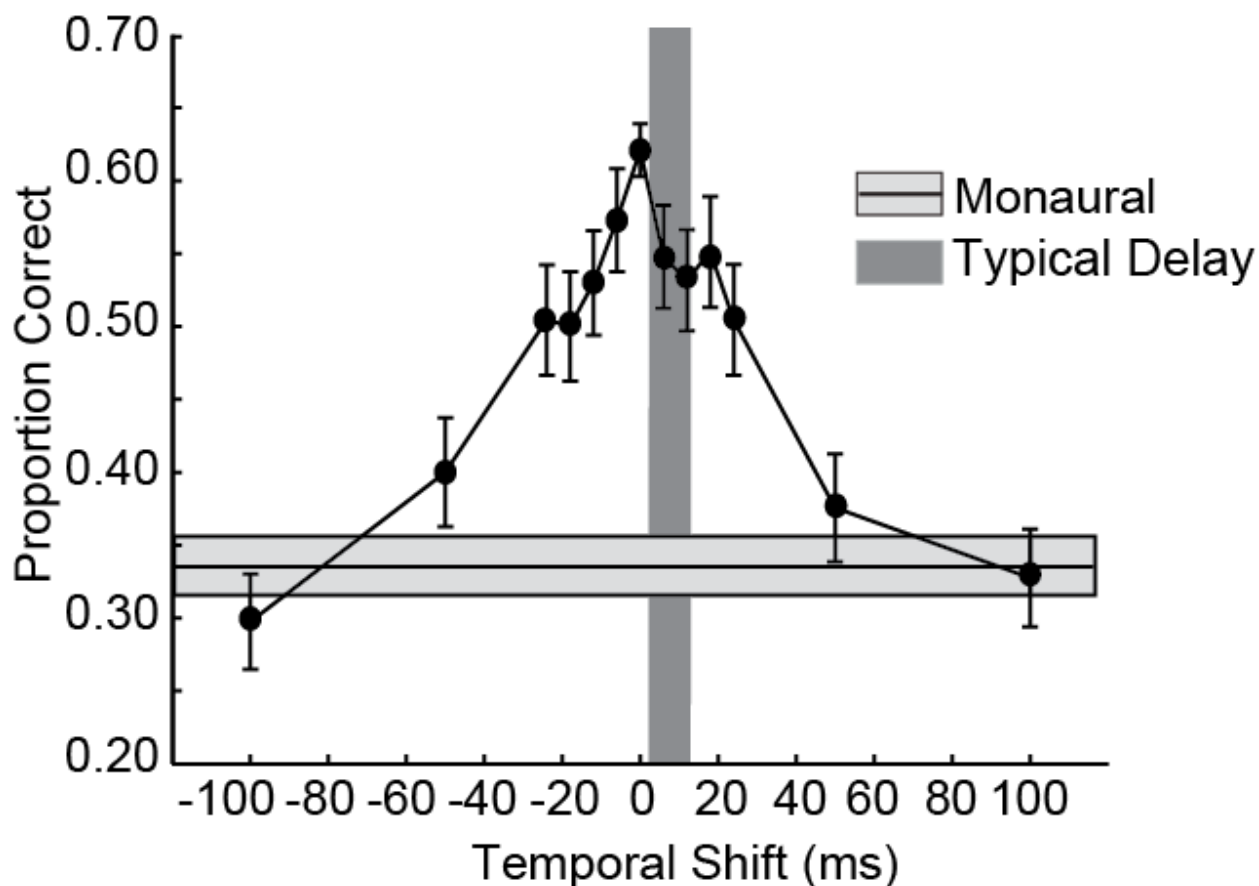


Figure 2.3. Results from Experiment 2.2 showing mean performance as a function of interaural temporal mismatch. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark grey shaded region represents the range of mismatch expected for actual SSD-CI listeners. Performance was maximum with no interaural delay (0 ms) and decreased with increasing temporal delay, although there was relatively little effect for an interaural delay of 12 ms or less (the expected range for SSD-CI listeners). Error bars represent \pm one standard error of the mean.

Figure 2.3 plots the mean proportion of keywords correctly identified as a function of temporal shift. Interaural temporal mismatch reduced contralateral unmasking, but the effect was relatively small for delays in the +0.5-12.5 ms range expected for SSD-CI

listeners. A repeated-measures binary-logistic regression analysis revealed a significant main effect of temporal shift [$\chi^2(11) = 2549.3, p < 0.001$]. Post-hoc tests were carried out to determine at what point contralateral unmasking was reduced relative to the zero-delay condition, and at which point contralateral unmasking disappeared completely with Bonferroni corrections applied for (13) multiple comparisons. Performance was significantly poorer ($p < 0.05$) than in the zero-delay condition for positive temporal shifts (i.e., vocoder leading) of 24 ms larger, and for negative shifts (i.e., vocoder lagging) of -18 ms or larger with the exception of -24 ms which was not significant. Performance was significantly better than in the monaural condition ($p < 0.05$) for all temporal shifts between -24 and +18 ms.

Summary.

In summary, temporal shifts of ± 50 – 100 ms completely eliminated contralateral unmasking and shifts smaller than ± 24 ms preserved most of the benefit. A temporal mismatch in the 0.5 – 12 ms range expected for SSD-CI listeners did not significantly reduce contralateral unmasking.

Experiment 2.3: The role of spectral mismatches and vocoder channel resolution on binaural unmasking in simulations of CI users with SSD

Experiment 2.3 examined the interaction between interaural spectral mismatch and the frequency resolution of the vocoder. Many CI users have limited spectral resolution (Loizou, 2006). We hypothesized that although reduced frequency resolution would reduce

contralateral unmasking (Bernstein et al., 2015) when no mismatch is present, it might also mitigate the negative effects of spectral mismatch on contralateral unmasking to some extent. The idea was that for a given degree of interaural mismatch, with broader vocoder channels there would be an increased likelihood of interaural correlation between the speech envelopes.

Methods.

Participants. Nine NH paid listeners participated in experiment 3, 6 of whom had also participated in experiment 2.2. All listeners who participated in both experiments completed experiment 2.2 first. Listeners were tested at the Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio.

Stimuli. The methods were generally the same as in experiment 2.1, except that the number of frequency channels in the vocoder was manipulated in addition to the introduction of interaural spectral mismatches. Four different numbers of vocoder channels were tested (3, 5, 8 and 10). Synthesis filters were shifted by 0, ± 0.5 , ± 1 , ± 2 , ± 4 and ± 7 ERBs relative to an acoustic ear. By shifting the synthesis filters, the signal excited a different cochlear place than would be excited by the unshifted vocoder signal. The frequency range of the vocoder analysis filters was 576 to 4102 Hz. The high-frequency cutoff was higher than in Experiment 2.1 because the maximum spectral shift was limited to 7 ERBs (maximum synthesis filter cutoff = 8960 Hz). The low-frequency cutoff was set higher than in Experiment 2.1 to ensure that frequency information was not removed with negative spectral shifts (minimum synthesis filter cutoff for a shift of -7 ERBs = 147 Hz).

Figure 2.4 shows examples of the analysis and synthesis filter band edges with a spectral shift of +4 ERBS, for vocoders with 3, 5, 8 and 10 channels.

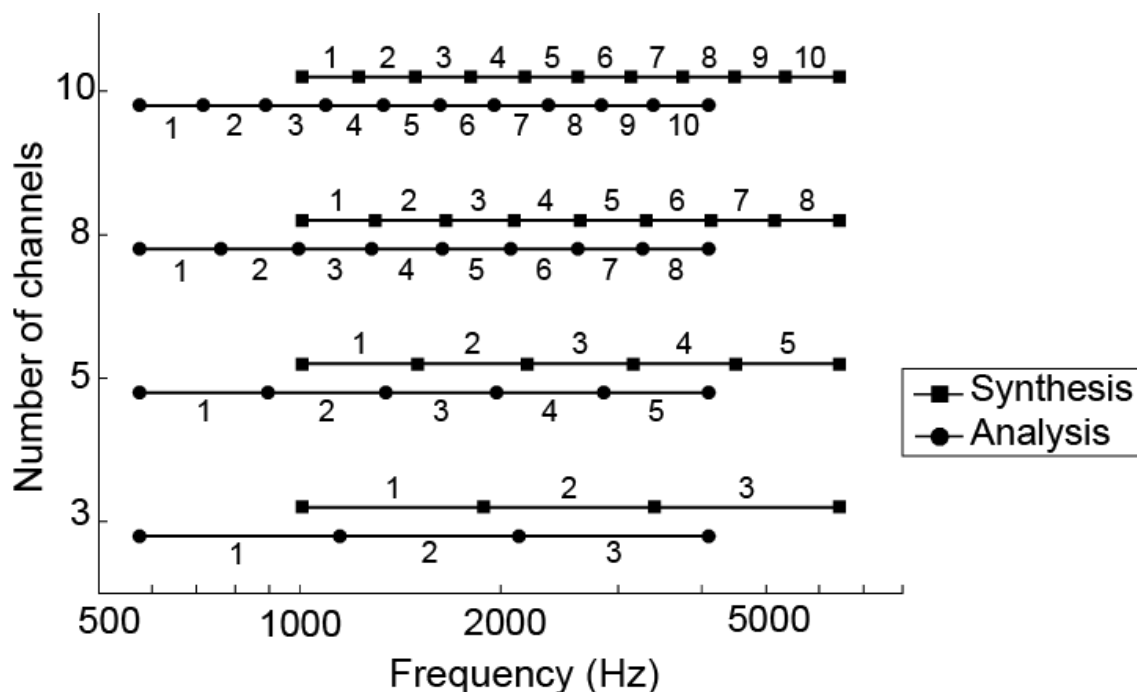


Figure 2.4. The analysis and synthesis band edges for a +4-ERB spectral mismatch for all of the vocoder-channel conditions in experiment 3. The numbers above and below each pair of lines identify the corresponding analysis and synthesis channels. For the 3-channel vocoder, the corresponding bands overlap. For the 5-channel vocoder, the bands do not overlap, but they are nearly adjacent. For the 8- and 10-channel vocoders, the analysis and synthesis bands are separated by at least one channel.

Procedure. This experiment examined an interaction between two factors and therefore included a greater number of conditions than in experiment 2.1; to compensate for this fewer trials were presented to limit the duration of the experiment. Listeners were presented with 45 trials for each combination of spectral shift, TMR (0 and 4 dB), and number of vocoder channels in the bilateral conditions. In the monaural condition, listeners were also presented with 45 trials for each TMR, but the structure of the automated program

repeated these 45 monaural trials 4 times (once for each channel-number condition), resulting in a total of 180 monaural trials for each TMR.

Thus, each listener completed a total of 4320 trials. The stimuli were presented in blocks consisting of 50 trials, including 45 bilateral trials with the number of vocoder channels held constant throughout the block and 5 monaural trials. The spectral shift condition and TMR varied randomly from trial to trial within a block.

Results.

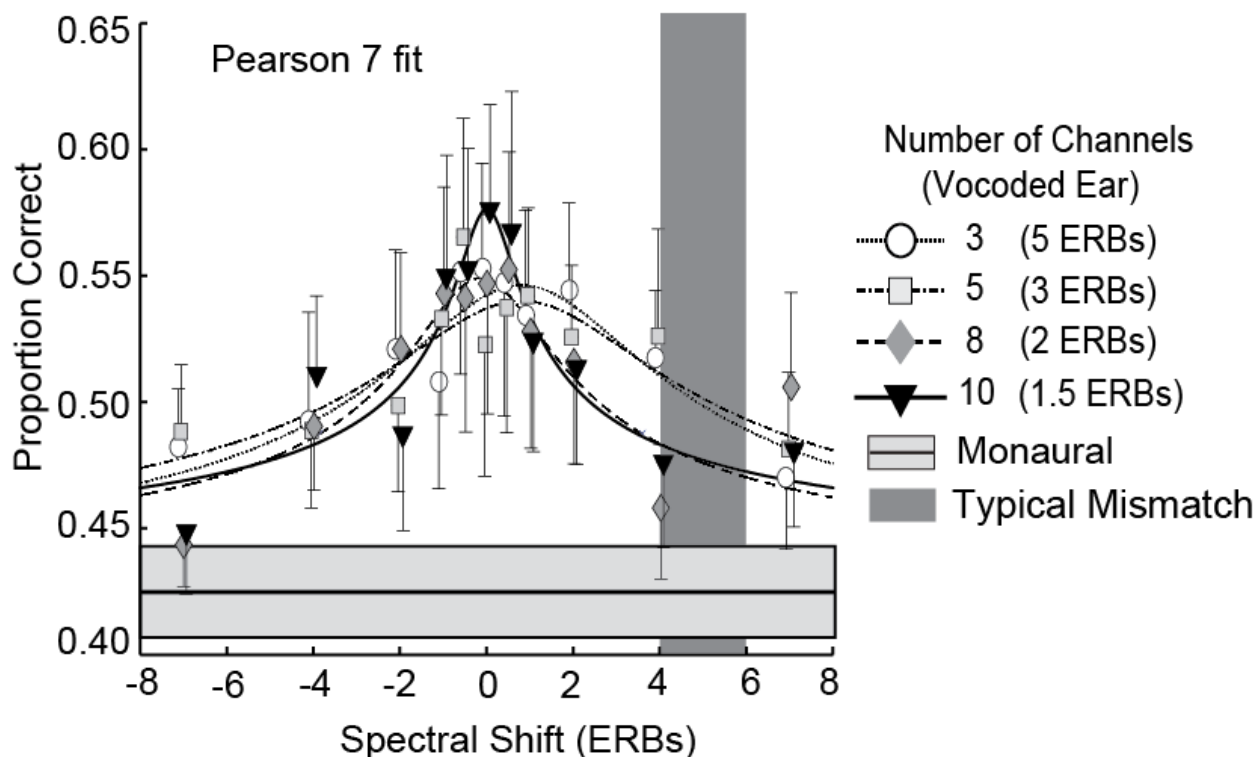


Figure 2.5. Results from Experiment 2.3 plotting mean performance in correctly identifying the target number and color as a function of spectral mismatch, with data averaged across TMR. Curves represent fits to the data for each vocoder condition (3, 5, 8 or 10 channels) using Pearson type 7 distributions with four free parameters. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark gray shaded region represents the range of mismatch expected for actual SSD CI listeners. While maximum performance was slightly better in conditions with better spectral resolution, performance dropped off substantially with relatively small spectral shifts in these conditions. Conditions with fewer vocoder channels were more immune to the effects of spectral shift. Error bars represent \pm one standard error of the mean.

Figure 2.5 plots mean performance (averaged across TMR) as a function of spectral shift for the four vocoder-channel conditions, along with curves (Pearson Type 7, four free parameters³) fitted to the data. As in Fig. 2.1, the vertical shaded region indicates the expected range of spectral mismatch across the cochlear partition for average SSD-CI

listeners (based on Landsberger et al., 2015). The horizontal shaded region indicates monaural performance. For clarity, Fig. 2.6 plots the effect of spectral resolution and spectral mismatch as a function of TMR.

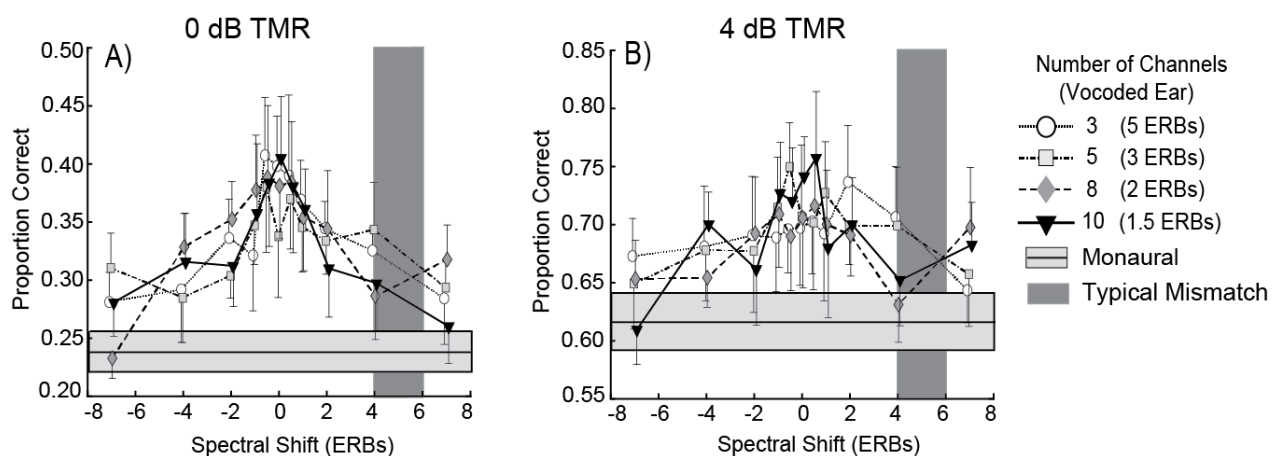


Figure 2.6. Results from experiment 2.3 plotting mean performance in correctly identifying the target number and color as a function of spectral mismatch, with data plotted separately for the two TMRs tested: (A) 0 dB and (B) 4 dB. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark grey shaded region represents the range of mismatch expected for actual SSD CI listeners. Error bars represent \pm one standard error of the mean.

A binary-logistic regression analysis revealed significant main effects of spectral mismatch [$\chi^2(9) = 50.1$ $p < 0.001$] and TMR [$\chi^2(1) = 463.4$ $p < 0.001$], but no main effect of spectral resolution ($p > 0.05$). There was a significant three-way interaction between all three factors [$\chi^2(9) = 91.6$ $p < 0.001$], and significant two-way interactions between spectral resolution and spectral shift [$\chi^2(9) = 110.0$ $p < 0.001$] and TMR and spectral shift [$\chi^2(9) = 1323.8$ $p < 0.001$]. The interaction between spectral resolution and spectral shift is visible

in Fig. 2.5, whereby conditions with fewer channels (poorer spectral resolution) were more resilient to spectral shifts. For example, for a vocoder with 3-5 channels, performance was only marginally affected by a spectral shift of 4 ERBs, whereas for a vocoder with 8-10 channels the contralateral unmasking benefit was almost completely eliminated by a 4-ERB shift.

Planned comparisons were made between performance with a spectral shift to performance in the monaural and zero-shift conditions. To reduce the number of tests, pairwise comparisons were only made between the monaural condition and the bilateral 0-ERB and +4-ERB spectral-shift conditions. These two spectral-shift conditions were selected because they represent a perfect interaural spectral match, and a mismatch that fell within the range of the shifts expected for an average SSD-CI listener. Bonferroni corrections were made for 8 multiple comparisons (4 vocoder-channel conditions x 2 TMRs).

Overall, there were many more significant effects ($p < 0.05$) for the 0-dB than for the 4-dB TMR. The 0-dB data are discussed first. With no spectral shift (0 ERBs) there was significant contralateral unmasking for vocoders with 3, 8 or 10 channels ($p < 0.05$). However, conditions with better spectral resolution (more channels) were more sensitive to spectral mismatch. A spectral shift of 4 ERBs significantly reduced performance (relative to the 0-ERB condition) for a 10-channel vocoder, but not for vocoders with 3, 5 or 8 channels. Only the 3-channel vocoder still yielded significant contralateral unmasking (relative to the monaural condition) when there was a 4-ERB spectral shift. While relatively few of the comparisons were significant for the 4-dB TMR, there was also some indication of the same basic pattern of results. Only the 10-channel condition showed reduced

performance for a 4-ERB relative to a 0-ERB shift, while only the 5-channel condition showed significant contralateral unmasking with a 4-ERB shift.

Summary.

In summary, the results of experiment 2.3 show that contralateral unmasking is greater for a vocoder with higher spectral resolution when the spectrum is matched perfectly, but that vocoders with a less spectral resolution were more robust to spectral mismatch.

Experiment 2.4: The role of spectral and temporal mismatches on contralateral unmasking in simulations of CI users with SSD

Experiment 2.4 explored the interaction between spectral and temporal mismatch in their effect on contralateral unmasking. We hypothesized that the negative impact of a mismatch in one dimension might be compounded by a mismatch in the other dimension.

Methods.

Participants. Ten NH listeners participated in this experiment, 4 of whom had participated in both experiments 2 and 3; these listeners completed experiments 2 and 3 before completing experiment 4. Listeners were tested at the Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio.

Stimuli. The methods were generally the same as in the previous experiments, except that spectral and temporal mismatches were combined. This experiment employed a 10-channel vocoder with a frequency range of 354 to 5752 Hz. The vocoder bandwidth was larger than in experiments 1 and 3 because a more narrow range of spectral shifts was tested (± 4 ERBs).

Procedure. Listeners were presented with 60 trials for each combination of spectral shift (0, ± 2 and ± 4 ERBs), temporal shift (0, ± 12 , ± 18 , ± 24 , ± 50 and ± 100 ms) and TMR (0 and 4 dB). Because the monaural conditions were coded as additional temporal-shift conditions in the experimental software, listeners completed 5 times as many trials (i.e., 300) at each TMR in the monaural conditions. Each block consisted of 48 trials with a fixed temporal shift (or monaural presentation), while the TMR and spectral shift (where applicable) varied randomly from trial to trial within the block.

Results.

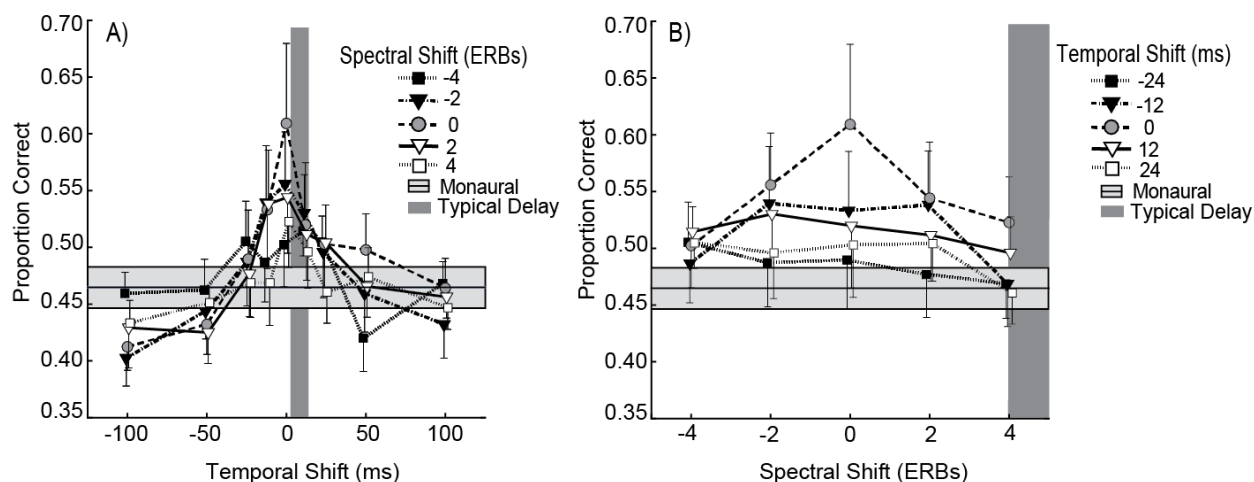


Figure 2.7. Results from experiment 2.4 plotting mean performance in correctly identifying the target number and color, averaged across TMR. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark grey shaded region represents the range of spectral or temporal mismatch expected for actual SSD CI listeners. (A) Data plotted as a function of temporal shift. (B) The same data plotted as function of spectral shift, with the ± 50 and ± 100 ms temporal-shift conditions excluded for visual clarity. Error bars represent \pm one standard error of the mean.

Figure 2.7A plots the mean performance as a function of temporal shift, with individual curves representing the different spectral-shift conditions. Figure 2.7B plots the same data as a function of spectral shift, with individual curves representing the different temporal-shift conditions tested. The ± 50 and ± 100 -ms conditions were excluded from Fig. 2.5B for visual clarity. A binary-logistic regression analysis revealed significant main effects of TMR [$\chi^2(1) = 637.2$ $p < 0.001$], temporal shift [$\chi^2(8) = 81.8$ $p < 0.001$] and spectral shift [$\chi^2(4) = 17.0$ $p < 0.05$], significant two-way interactions between spectral shift and temporal shift [$\chi^2(12) = 5.46$ $E13$ $p < 0.001$], TMR and temporal shift [$\chi^2(8) = 53.2$

$p < 0.001$], and a significant three-way interaction between all three factors [$\chi^2(11) = 63320198.5$, $p < 0.001$]. The two-way interaction between TMR and spectral shift was not significant ($p > 0.05$). The interaction between the effects of spectral and temporal mismatch is visible in Fig. 2.7, whereby the effect of a mismatch in one dimension became more muted (i.e., the individual curves in Figs. 2.7A and B became flatter) when there was also a mismatch in the other dimension. For completeness, Fig. 2.8 plots the effect of temporal and spectral mismatch as a function of TMR.

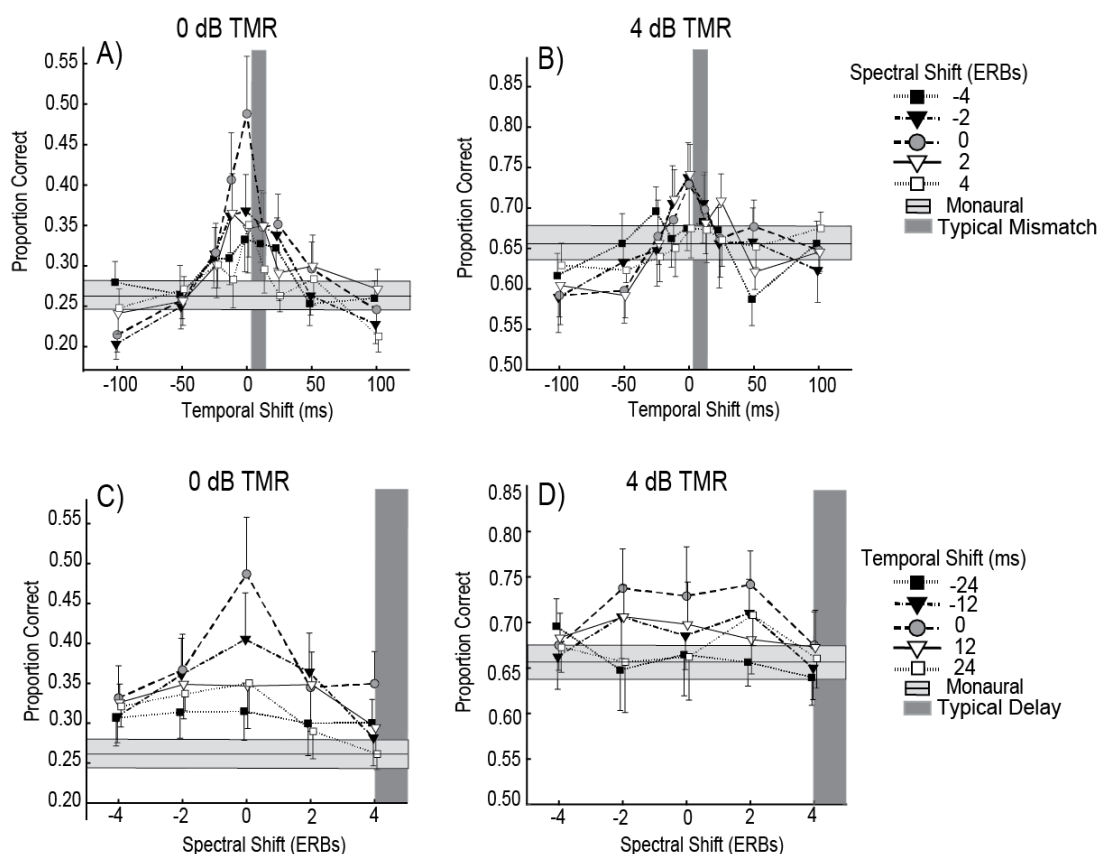


Figure 2.8. Results from experiment 2.4 plotting mean performance in correctly identifying the target number and color, with data plotted separately for the two TMRs tested. Mean monaural performance is depicted by the horizontal line, with the horizontal light grey box representing \pm one standard error of the mean. The vertical dark gray shaded region represents the range of spectral or temporal mismatch expected for actual SSD CI listeners. The scale for the ordinate changes in each panel. Top row: data plotted as a function of temporal shift for TMRs of (A) 0 dB and (B) 4 dB. Bottom row: data plotted as a function of spectral shift, with the ± 50 and ± 100 ms conditions excluded for clarity, for TMRs of (C) 0 dB and (D) 4 dB. Error bars represent \pm one standard error of the mean.

Post-hoc tests further evaluated the interaction between spectral and temporal mismatch. To limit the number of planned comparisons, pairwise comparisons were made only for conditions involving spectrally or temporally matched stimuli (0 ERBs or 0 ms) and spectral or temporal mismatches in the range expected for an average SSD-CI listeners (+4 ERBs or +12 ms). Bonferroni corrections were made for (2) multiple comparisons: +4 ERBs and +12 ms (2 TMRs). Significant differences were only observed for the 0-dB TMR. When there was no spectral mismatch (0 ERBs), a 12-ms temporal mismatch significantly reduced performance ($p < 0.05$). But there was no significant effect of a 12-ms temporal mismatch when there was also a 4-ERB spectral mismatch ($p > 0.05$). When there was no temporal mismatch, a 4-ERB spectral mismatch just failed to significantly reduce performance relative to the 0-ERB condition ($p > 0.05$). There was no effect of a 4-ERB mismatch when there was also a 12-ms temporal mismatch ($p > 0.05$).

Summary.

In summary, these results show that while mismatches in either dimension (temporal or spectral) can reduce contralateral unmasking, these two types of mismatch do not interact in an additive fashion. Instead, the results show that once masking release was diminished by a shift in one dimension, an additional shift in the other dimension had a relatively small effect.

Discussion

The results of the four experiments in this study demonstrate that interaural spectral and temporal mismatch introduced into vocoder processing can reduce or eliminate the contralateral unmasking effects experienced by NH listeners presented with unprocessed sounds in one ear and vocoded sounds in the other ear. The competing-talker paradigm produced a situation where it would have been very difficult to perceptually separate the target talker of interest from the concurrent same-gender interfering talkers based on monaural cues alone. The addition of a second copy of the interfering voices contralaterally, either via vocoder processing to a second NH ear (Bernstein et al., 2015, 2016) or via direct connection to a CI (Bernstein et al., 2016) has been previously shown to provide the listener with sufficient interaural cues to facilitate the perceptual separation of concurrent voices and improve performance in the speech-identification task. The current study replicated this finding for NH listeners presented with vocoded stimuli, and extended it by providing information about the degree of interaural spectral and temporal alignment required to facilitate the effect.

Across the four experiments, the greatest amount of contralateral unmasking was achieved in the “ideal” condition whereby a high-resolution vocoded signal was exactly matched both spectrally and temporally to the unprocessed ear. Performance in this ideal situation was found to improve performance by as much as 20 percentage points relative to the monaural condition. This result suggests that SSD-CI listeners should perform best in this type of task if signal-processing and clinical frequency mapping procedures could be established to achieve interaural alignment in both the spectral and temporal dimensions. If perfect spectral and temporal matches cannot be obtained, the “optimal” level of

performance would be difficult to achieve. In cases where there is a spectral mismatch of more than 2 ERBs or a temporal mismatch more than 12 ms or more, there was a significant reduction in contralateral unmasking performance relative to the ideally matched condition. Given the likely difficulty in achieving these tight tolerances with current technology, it will be necessary for clinicians to tolerate some level of interaural mismatch, as well as limited spectral resolution, when fitting and counseling SSD-CI patients. The implications of distortions in each of these dimensions are discussed in the following.

Impacts of a spectral mismatch.

When considered relative to the range of mismatch expected for an average SSD-CI listener, spectral mismatch had the greatest negative impact on contralateral unmasking of the three distortions examined in this study. Contralateral unmasking was maximal for a vocoder with no spectral shift, decreasing to approximately half this maximum value for a shift of ± 2 ERBs, and to nearly zero (i.e., no contralateral unmasking) for a shift of ± 4 ERBs or more, in the range expected for an average CI user (Landsberger et al., 2015).

Reduced contralateral unmasking with spectral mismatch is consistent with other examples whereby interaural frequency match affects binaural processing for bilateral CI or bilateral vocoder listeners. Small mismatches (on the order of 3 mm) disrupt interaural ILD and ITD discrimination performance for bilateral CI users (Goupell et al., 2013; Kan et al., 2013) and bilateral vocoder listeners (Siciliano, Faulkner, Rosen, & Mair, 2010). Similarly, binaural fusion has been found to be disrupted by spectral compression for NH listeners presented with bilaterally vocoded signals (Aronoff et al., 2015). For bilateral CI listeners, interaural spectral mismatch reduces binaural fusion, causing a single stimulus to be perceived as multiple sounds (Kan et al., 2013). Binaural fusion is an important

prerequisite to the proper grouping and perceptual separation of concurrent sounds in the environment (Bregman, 1994). In the current study, the perceptual fusion of the unprocessed interferers presented to one ear and the vocoded interferers presented to the other ear would be required to allow the listener to perceive these sounds as a single auditory object and perceptually separate them from the monaural target speech. One way of interpreting the results is that spectral mismatch disrupted the interaural envelope correlation that is required to facilitate fusion. Correlated envelope information between the ears has been shown to facilitate auditory object formation and fusion (Carrell & Opie, 1992).

These results are also consistent with a recent vocoder study that examined the role of spectral mismatch and its effect on integration of speech information across ears. Ma, Morris, and Kitterick (2016) found that bilateral presentation of the vocoded speech resulted in better performance than with stimuli presented monaurally, but that this improvement was reduced by an interaural spectral mismatch. One caveat to the interpretation of these results is that Ma et al. (2016) presented target speech information to both ears, which makes it difficult to know whether the result reflects the effects of spectral shifts on binaural integration or better-ear listening. In the current study, no target speech information was presented to the vocoder ear, ensuring that the observed effects reflect the integration of information across the ears.

The current results regarding the effects of spectral mismatch contrast with the results of Bernstein et al. (2016) who found that most of the SSD-CI listeners in the study did obtain a substantial release from masking. The release occurred despite the likelihood that many of the listeners likely experienced an interaural spectral mismatch on the order

of 4 ERBs or more (Landsberger et al., 2015), which should have been large enough to extinguish any contralateral unmasking. The reason for this discrepancy between the studies is not clear, but perhaps it could be related to plasticity/adaptation effects for the CI listeners. The NH listeners in the current study were naïve to the mismatched frequency channels used in this experiment, whereas the CI listeners in the prior experiment had substantial experience listening to their own possibly mismatched frequency maps. Indeed, previous work has shown that CI listeners can adapt to frequency mismatches over time (Svirsky et al., 2004; Reiss et al., 2007). Alternatively, it is possible that the actual SSD-CI listeners had relatively poor spectral resolution, resulting in current spread and auditory-nerve activation across a broad swath of the cochlea for a given electrode. Current spread in CIs typically causes reduced speech understanding in noise (Srinivasan, Padilla, Shannon & Landsberger, 2013). However, in cases where the listener does not depend on the vocoded signals for information about the target speech, Experiment 2.3 showed that reduced frequency resolution could mitigate the negative effects of spectral mismatch on contralateral unmasking.

Impacts of spectral resolution.

Spectral resolution had relatively little effect on contralateral unmasking under conditions with no spectral mismatch, in that contralateral unmasking was observed even in conditions with few vocoder channels. This result is in qualitative agreement with Bernstein et al. (2015) who found that contralateral unmasking was maximal with six vocoder channels, was only modestly reduced with four channels, and did not disappear completely until processing was carried out in a single broadband vocoder channel. In

contrast, spectral resolution had a substantial impact on contralateral unmasking when there was also a spectral shift. Contrary to the typical improvement in speech perception associated with better resolution (e.g., Nie et al., 2006), in this case better spectral resolution actually led to *poorer* performance when there was also a spectral mismatch present.

Our interpretation of this result is that in conditions with broader channels, there was more resilience to spectral shift because at least some of the synthesis-filter bandwidth overlapped with the analysis-filter frequency range from which the signal was derived. This can be seen in Fig. 2.4, which shows the analysis and synthesis filter cutoff frequencies for the conditions with a 4-ERB spectral mismatch. For a vocoder with 3 channels, there is some overlap in the analysis and synthesis filter bandwidths for a given channel, which likely yielded some interaural correlation between the envelopes at a given cochlear place. In contrast, for a vocoder with 10 channels, there was no overlap between the analysis and synthesis filter bandwidths for a given channel. This would cause the envelope in a given frequency region to be decorrelated with the acoustic envelope in the other ear in the corresponding frequency region, thereby limiting contralateral unmasking. The bandwidths for the 5-channel vocoder are somewhat at odds with this interpretation, since the analysis and synthesis filters do not overlap, yet significant contralateral unmasking was observed. However, it is known that there is some envelope correlation between neighboring spectral bands (Buss, Whittle, Grose, & Hall, 2009); the analysis and synthesis filters might have still been close enough to one another in the 5-channel case to allow for some interaural correlation.

It should be noted that even though poorer vocoder resolution led to an advantage in this particular instance involving spectral mismatch, this does not necessarily mean that SSD-CI listeners might benefit in general from reduced spectral resolution. For example, if the experiment had been designed differently with the target presented to the vocoded ear, reduced spectral resolution is likely to have yielded a reduction in performance. Thus, in situations where SSD-CI listeners take advantage of a better SNR at the CI ear as a result of acoustic head shadow (e.g., Vermeire & van de Heyning, 2009; Arndt et al. 2011; Firszt et al. 2012), reduced spectral resolution would be more likely to harm performance.

Effects of temporal mismatch.

Experiments 2.2 and 2.4 examined the effect of interaural temporal mismatch on contralateral unmasking. Contralateral unmasking was only modestly affected by temporal mismatch in the expected range for SSD-CI listeners (0.5–12 ms), although larger mismatches of 24 ms or greater did substantially reduce the unmasking effect (Figure 2.2). A possible explanation for these results involves integration of early reflections to enhance perception of the speech signal. Basically reflections that occur within a small window (< 50ms) have been shown to enhance speech perception and reflections that occur later degrade a listener's ability to hear the speech signal. Essentially these early reflections (or delays) combine with the original signal and enhance the SNR, this could explain why small delays did not disrupt contralateral unmasking (Bradley, Reich, & Norcross, 1999; Soulodre, Popplewell, & Bradley, 1989). Related to the early reflections interpretation the robustness of contralateral unmasking to interaural delays less than 24 ms is consistent with the well-established precedence effect for speech, which has a high echo threshold — i.e.,

a long maximum duration over which stimulus echoes “fuse” with the direct sound and are perceived as part of a single auditory object (Litovsky et al., 1999). This process enables humans to hear in highly reverberant environments and is likely the result of cortical processing (Miller et al., 2009), with speech sounds arriving within a 30 – 40 ms window being perceived as one auditory object (Grant, Wassenhove, & Poeppel, 2004; Litovsky et al., 1999). An interesting caveat to the interpretation of these results being explained in terms of the precedence effect is that the drop off in performance after temporal mismatch was relatively symmetrical. If the results of this experiment truly did reflect a precedence mechanism, than listeners would have likely performed much better when the vocoder leading stimulus was applied. This would equate to the first vocoded sound taking “precedence” and the listeners would have received a clear spatial cue that the vocoded voices were on the right and the target was on the left. The opposite should have occurred in the acoustic ear leading conditions. The leading acoustic interferers should have sounded like they were primarily coming from the acoustic side, combined with the target in that ear. This would have strongly degraded performance in the acoustic ear leading vs the acoustic ear lagging conditions. This was not what occurred, therefore an alternative explanation based on interaural envelope correlation might be appropriate. The effects of temporal mismatch on contralateral unmasking might be thought of in terms of interaural coherence of the interferer envelopes. Speech contains inherent envelope fluctuations from 2–5 Hz for syllables and 15–30 Hz for phonemes (Elliott & Theunissen, 2009), corresponding to a modulation period of 200–500 ms for syllables and 30–60 ms for phonemes. For these slow modulations, some interaural temporal misalignment will have relatively little impact on interaural correlation of speech envelopes.

Effects of combined spectral and temporal mismatch.

SSD-CI listeners are likely to experience both a temporal and a spectral mismatch simultaneously. Experiment 2.4 investigated the interaction between these two distortions. We hypothesized that these distortions would be additive — that introducing a temporal mismatch in addition to a spectral mismatch (or vice versa) would cause an even larger reduction in contralateral unmasking. Temporal delays caused by frequency dependent differences in signal latency between the NH ear and CI ear could disrupt bilateral unmasking because common onset times are an important cue for grouping of sounds (Bregman, 1994). This disruption could be especially pronounced when accompanied with spectral compression, which is known to limit binaural fusion (Aronoff et al., 2015). However, the results did not support this hypothesis. In fact, the opposite occurred: mismatch in one dimension had a smaller additional effect if there was already mismatch in the other dimension (Fig. 2.5). One interpretation of this result is that if fusion is already disrupted, then further distortion does not have as much effect. From a clinical perspective, this result suggests that if one distortion is present, then reducing the other distortion through signal processing or re-programming the CI will only modestly improve contralateral unmasking. To yield the most possible contralateral unmasking, temporal and spectral interaural disparities must both be minimized.

Implications for SSD-CI listeners.

These results suggest that to maximize a listener's ability to use their two ears together to better understand speech in competing backgrounds, steps could be taken to minimize spectral and temporal distortions. Perhaps the most encouraging result from this paper is that the distortion that had the largest negative impact on contralateral unmasking — spectral mismatch — is also the distortion that can most readily be addressed clinically. Theoretically, a place-matched frequency mapping based on electrode location could be provided by an audiologist to better match the place of cochlear stimulation for a given CI electrode to the cochlear place of stimulation for an acoustic signal presented to the NH ear. This process would require estimates of the cochlear places of stimulation associated with individual electrodes in the array, which could be accomplished in one of several ways. Computerized tomography (CT) scans (Noble, Gifford, Hedley-Williams, Dawant, & Labadie, 2014) or radiographs (Landsberger et al., 2015) could be used to estimate the insertion angles of individual electrodes. Comparisons of ITD sensitivity for a given electrode and a range of acoustic stimulus frequencies might also provide information about which acoustic frequency would be best matched to a particular electrode (Goupell et al., 2013; Kan et al., 2013). Pitch matching between individual electrodes and acoustic stimuli (Carlyon et al., 2010) could also provide information about cochlear place of stimulation, although the pitch-matching estimates have been shown to be susceptible to adaptation effects (Reiss et al., 2014). Hu and Dietz (2015) compared pitch matching and ITD sensitivity in BICI users and found pitch matching preference was nearly identical as the programmed electrode frequency band, suggesting that pitch percepts adapt to the CI

processor allocation. There were also large differences between the chosen pitch-matched pairs and maximal ITD-sensitive pairs, as shown in previous studies (Long et al., 2003; Poon, Eddington, Noel, & Colburn, 2009; van Hoesel & Clark, 1997). Due to the adaptability of pitch percepts and the biological importance of ITD sensitivity for binaural hearing, Hu and Dietz (2015) conclude that identifying ITD-sensitive electrode pairs is the most promising method for remapping a CI, at least for BICI listeners.

Minimizing temporal mismatch is likely to be more difficult than minimizing spectral mismatch, given the expectation that CI stimulation will be delayed relative to the auditory-nerve response in the acoustic ear. In a group of bimodal CI listeners (CI in one ear and severely impaired acoustic hearing in the other ear), Francart and McDermott, (2013) used lateralization judgements to establish that the auditory-nerve response time is *faster* in the CI ear than in the acoustic ear (with no hearing aid worn) by about 1.5 ms, due to the delay associated with the cochlear traveling wave in the acoustic ear (Rasetshwane, Argenyi, Neely, Kopun, & Gorga, 2013b). However, this study used a method of direct-stimulation to control the electrical stimulation pattern on the array, and did not make use of the listener's external speech processors. In everyday listening conditions, the speech processor can add substantial delay to the overall processing time, resulting in a slower response auditory-nerve response time in the CI relative to the acoustic ear (Zirn et al. 2015). Thus, the only ways to reduce the interaural temporal mismatch are (1) to reduce the processing time for the CI external speech processor, or (2) to introduce a delay to the acoustic ear. Many individuals that might be classified as SSD have some hearing loss in the acoustic ear and wear a hearing aid to provide amplification (e.g., Vermeire & van de Heyning 2009; Firszt et al. 2012). Theoretically, the time delays in the hearing-aid and CI

ears could be adjusted to minimize interaural delay. However, this approach would not be reasonable for an SSD-CI listener with normal hearing in the acoustic ear, whereby adding any processing to the acoustic signal via a hearing-aid device is likely to be undesirable.

Study Limitations.

Vocoder simulations are imperfect estimates of the acoustic information that is delivered to the auditory nerve for a CI recipient (Freyman et al. 2008; Li & Loizou 2009). Nevertheless, Bernstein et al. (2016) employed the same contralateral unmasking paradigm as the current study, and found that NH listeners presented with vocoded signals yielded a similar qualitative pattern of results to actual SSD-CI listeners. Furthermore, the best performing SSD-CI listener obtained about the same amount of contralateral unmasking as the average vocoder listener. Thus, the substantial effects of spectral mismatch and to some extent, of temporal mismatch observed for NH listeners presented with vocoder simulations suggest the possibility that minimizing these particular distortions could improve performance for SSD-CI listeners.

Another important difference between CI and vocoder listeners is that the vocoder listeners did not have chronic exposure to the vocoded and interaurally mismatched stimuli, and therefore could not take advantage of any possible adaptation to distorted and mismatched inputs over time (Svirsky et al. 2004; Reiss et al. 2007). For CI users, it is possible that contralateral unmasking might emerge or improve following long-term exposure to mismatch. On the other hand, a subset of our vocoder listeners did take part in several of the experiments over time. We did not observe any evidence of training effects

over the course of the multiple experiments. In the monaural conditions, for which the stimulus parameters were identical across the individual experiments, there was no evidence of training effects for the four listeners who participated in experiments 2.2 (34% correct), 2.3 (23% correct) and 2.4 (27% correct). There was also no evidence that our vocoder listeners grew resistant to spectral mismatch as the study progressed. With the exception of the conditions with very poor vocoder spectral resolution in Experiment 2.3, no significant contralateral unmasking was observed for a spectral mismatch of 4 ERBs in any of the experiments. Still, this study was not specifically designed to examine effects of training and plasticity; we cannot rule out the possibility that with more extensive and controlled exposure to inter-aurally mismatched stimuli, adaptation could emerge.

An additional limitation of the current study is that it measured a very specific aspect of binaural hearing. These results might not generalize to all listening situations, and there could be listening environments and situations where remapping to reduce spectral mismatch might not be advantageous. While the contralateral unmasking paradigm demonstrates that CI and vocoder listeners are capable of experiencing squelch, the complete isolation of the interfering speech from the target in the CI or vocoder ear is an artificial situation that would not be encountered in real environments. Still, Bernstein et al. (2015) showed that the squelch effect was reduced when a more typical 6-dB of contralateral attenuation was employed. Previous studies have shown that SSD listeners can benefit from a CI for sound localization and for taking advantage of better-ear listening (Arndt et al., 2010; Buechner et al., 2010; Erbele et al., 2015; Firszt et al., 2012; Hansen et al., 2013; Vermeire & van de Heyning, 2009; Zeitler et al., 2015). Reducing spectral mismatch might improve localization since small interaural offsets can cause a

considerable disruption of the ITD and ILD cues needed for sound-source localization (Goupell et al., 2013; Kan, Stoelb, Litovsky, & Goupell, 2013b; Litovsky et al., 2012). On the other hand, altering the CI frequency allocation might impair speech perception in the implant ear, since the remapping would preclude the inclusion of low frequencies in the CI map. However, this loss of low-frequency information might not affect speech perception for a SSD-CI listener because head shadow is very limited at low frequencies (Bronkhorst & Plomp, 1988). As a result, in the free field, nearly identical low-frequency acoustic information should be available in the NH acoustic ear. In any case, further work is needed to investigate the possible impact of remapping to reduce interaural mismatch on a wider variety of speech-perception and sound-localization tasks before clinical recommendations can be made to take this approach for SSD-CI listeners.

Conclusions

The results of the experiments presented here demonstrate that spectral and temporal interaural mismatches reduce contralateral unmasking in a speech-identification task with interfering talkers for NH listeners presented with unprocessed signals in one ear and vocoded signals in the other. Spectral mismatches in the range that an average SSD-CI listener is likely to experience with standard frequency mapping (4–6 ERBs) were particularly detrimental to performance (Experiments 2.1, 2.3 and 2.4). The detrimental effect was mitigated to some extent by reducing the spectral resolution of the vocoder (experiment 2.2), although an approach that purposefully reduces frequency resolution is likely to impair speech-reception performance in other conditions not tested here. Temporal mismatches in the range expected for SSD-CI listeners (<12 ms) had a less

pronounced negative effect (experiments 2.3 and 2.4), although maximum contralateral unmasking was observed when signals were aligned across the ears in both time and frequency (experiment 2.4). Overall, the results of this study highlight the need for interaural alignment to maximize the use of interaural differences to parse a complex auditory scene involving multiple competing talkers when presented with unprocessed speech in one ear and only envelope information contralaterally. SSD-CI listeners might benefit from strategies to reduce interaural mismatch, such as frequency remapping or introducing a processing delay to the acoustic ear.

Footnotes Chapter 2

¹ Cochlear implants are not currently labeled by the United States Food & Drug Administration for use for the treatment of SSD.

² Bernstein et al., (2016) also showed a similar squelch benefit for bilateral CI listeners, but the SSD-CI configuration is the focus of the current study.

³ A Pearson type 7 distribution is similar to a normal distribution in that it can account for kurtosis and skew, thus the data could be more accurately represented using this type of fitting function. The free parameters used to fit the data were mean, variance (standard distribution), amplitude and skew.

Chapter 3: Effect of compression and expansion on binaural hearing in simulations of SSD-CI listeners

Introduction

Binaural hearing is integral for sound localization and hearing speech in noisy environments. Therefore, individuals without binaural hearing are at a severe disadvantage when it comes to listening in our complex, noisy world. A form of hearing loss with functional limitations that has been traditionally under-appreciated is SSD, which refers to a profound loss of hearing in one ear. Traditionally, SSD was not treated because it was not considered incapacitating. However, after it became apparent that SSD was a disability, hearing aid treatments became available. Treatments include a CROS hearing aid, which routes the signals from the deaf side via wireless or wired link to a hearing-aid transducer placed in the NH ear. Additionally, a BAHA that routes the signals from the deaf side of the head to the NH ear via a transducer surgically implanted to stimulate the listener's skull has been used as a treatment for SSD. These methods have been successful in alleviating some of the adverse effects of SSD mainly by giving access to signals presented on the deaf side (by delivering them to the NH side). However, these treatments do not restore binaural hearing, and these patients still have trouble with sound localization and hearing in noisy environments (Grantham et al., 2012; Linstrom et al., 2009).

In the past several years, CIs have been considered as a possible new treatment option for SSD. A CI is a surgically implanted device that treats deafness by bypassing the

dead or damaged hair cells in the cochlea by direct electrical stimulation of the spiral ganglion neurons in the auditory nerve. Historically, CIs have only been implanted in the profoundly deaf. Although CIs are not currently approved by the United States Food and Drug Administration for the treatment of SSD, criteria for implant candidacy at individual centers has been relaxed in the last few years and a substantial number of individuals with SSD in the U.S. have received CIs. Currently, the most compelling evidence for CIs aiding SSD listeners in spatial hearing comes from studies that have examined performance in localizing a sound source (Arndt et al., 2010; Firszt et al., 2012; Hansen et al., 2013) and from studies that have assessed the advantages for listening to speech in noise when there is a spatial separation between the two signals (Buechner et al., 2010; Firszt et al., 2012; Hansen et al., 2013).

Binaural hearing provides the listener with two main advantages: head-shadow and squelch. Head-shadow allows listeners to take advantage of listening to the ear with the better SNR, regardless of which side of the head receives the better SNR (Schleich et al., 2004). Squelch is a neural process that involves the use of differences in timing and level from sources originating in different locations to reduce the effective amount of masking. CIs mainly provide a benefit to SSD listeners in configurations where the signal is on the deaf side, and/or the masker is on the NH side of the head, consistent with the idea that the CI allows users to take advantage of head-shadow effects and a better-ear listening strategy (Arndt et al., 2010; Buechner et al., 2010; Firszt et al., 2012; Hansen et al., 2013). The magnitude of the head-shadow effect is approximately (2-5 dB) for SSD-CI listeners (Kamal, Robinson, & Diaz, 2012; Schleich, Nopp & D'Haese, 2004b). Until recently, there has been limited evidence that a CI for SSD can provide binaural squelch to the listener.

Squelch is defined operationally in this study as the improvement in speech understanding (relative to monaural performance) when the speech and noise are spatially separated and the ear with the poorer SNR is added. For the binaural squelch advantage there is no added target speech information that is not already available at the other ear; this kind of measurement isolates the component of the benefit related to binaural interactions. The binaural squelch benefit is different than the head-shadow benefit. The head-shadow advantage is defined operationally in this study as the improvement in speech understanding when the speech and noise are spatially separated and the ear with the better SNR is added.

A series of studies examining SSD-CI listeners (Bernstein et al., 2016; Bernstein et al., 2017) and NH listeners presented with vocoder simulations of SSD-CI listening (Bernstein et al., 2015; Wess et al., 2017) have demonstrated binaural squelch under certain conditions. Specifically, these studies found a significant squelch benefit in listening situations where the target and interfering speech were produced by talkers of the same gender. This was taken as evidence that having hearing restored in the deaf ear via a CI can provide spatial cues to perceptually separate competing talkers when they are difficult to perceptually separate based on monaural pitch and timbre cues alone.

Despite the fact that on average, SSD-CI and (vocoder simulated SSD-CI) listeners experience binaural squelch, there was considerable individual variability in the magnitude of the squelch benefit. Furthermore, the squelch benefit was considerably larger for NH listeners presented with vocoder simulations of SSD-CI listeners than for actual SSD-CI listeners (Bernstein et al., 2016). There are many factors that could cause this inter-subject variability and limited squelch benefit for SSD-CI listeners, such as neural survival (e.g.,

Maslin et al. 2013), limited cortical plasticity (e.g., Litovsky et al. 2012; Maslin et al. 2013), electrical current spread (e.g., van Hoesel, 2012) and duration of deafness before implantation (e.g., Blamey et al., 2012). Alternatively, certain distortions inherent in CI processing are likely to lead to large differences in processing between the NH and CI ears which could limit binaural integration required to generate squelch. Wess et al. (2017) (as reported in Chapter 2) showed that spectral mismatch, and to some extent temporal mismatch, between a NH ear and a vocoder-processed ear reduced the magnitude of the squelch benefit. This chapter examines another possible factor that may limit speech-perception benefits for SSD-CI listeners: a mismatch in loudness growth between the two ears.

CIs have a dramatically reduced DR relative to a NH acoustic ear, due in part to a poor electrode-neural interface (Kawano et al., 1995). CI users have a reduced total DR of about 40 dB (McDermott & Varsavsky, 2009), whereas the DR for a healthy NH ear is approximately 120 dB (Moore, 2003). Electrical stimulation allows for fewer just-noticeable differences (JNDs) between threshold and maximum acceptable loudness level than an acoustic ear. It has been estimated that the number of intensity steps for CI listeners ranges from 7 - 45 (compared to 83 in NH listeners) and this number is highly variable across CI listeners (Nelson et al., 1996). This means that CI technology cannot encode the full DR into an electrical representation. Compression is a necessary step in CI signal processing, however it does alter some important features present in the signal. Compression is usually implemented on the signal envelope via a static logarithmic function. A static compressive function has been shown to preserve semi-normal loudness growth in many CI users (Zeng & Shannon, 1992). Conversely, expansion relates to

expanding the fast amplitude fluctuations of a speech signal envelope to potentially increase the intelligibility of the signal (van Buuren, Festen, & Houtgast, 1999).

First, compression is known to smear acoustic landmarks important not only for vowel comprehension but for identification of word boundaries (Li & Loizou, 2009). Second, compression distorts binaural cues by raising thresholds for discriminating ILDs (Grantham et al., 2008). CI listeners must rely on ILDs for spatial hearing cues, because ITD fine structure cues are discarded in CI processing (Loizou, 2006). Thus, ILD threshold increments that occur with compression are likely to weaken a SSD-CI listener's ability to perceptually stream apart acoustic events based on location, and thereby reduce binaural squelch. As noted previously, the ability to achieve binaural squelch is based on ILDs, which provide information about where in space the target and interferers are located. Third, compression could alter the effective TMR, which could change the amount of speech information that comes through the CI ear.

The specific attribute of interaural difference in the loudness growth that this study examined was range of envelope compression and expansion factors in vocoder simulations of SSD-CI listening. To examine the effects of envelope compression on speech perception, this study took as a starting point the results of prior studies (Bernstein et al., 2015, 2016; Wess et al., 2017; Chapter 2), that were designed to measure binaural squelch in the absence of head-shadow effects. In those studies, the target was presented to only the unprocessed ear while the interferers were presented simultaneously to both ears (bilateral condition) or to only the acoustic ear (monaural condition). This contralateral unmasking paradigm allowed for an examination of the binaural squelch (i.e., reduction in masking effectiveness) provided by the vocoded ear because no target speech information

was provided to the vocoded ear. Moreover, the contralateral unmasking paradigm allowed for complete isolation of the interferer signals in the vocoder ear (i.e. no target energy in the vocoded ear), which is not realistic and would not happen in the real world. Furthermore, in real world competing talker situations, both ears contain mixtures of target and masker energy. However, in the current compression/expansion experiment, it was important to have a mixture of target and interferers in the vocoder ear to examine the effect of a change in the effective TMR in that ear. Therefore, compression/expansion is likely to affect spatial hearing in this more realistic paradigm by altering TMRs and ILDs. For these two reasons, the experiments described in this chapter examined the effects of compression/expansion using simulations of spatially separated target and interfering talkers.

HRTFs were used to simulate the effects of the head on the amplitude and phase characteristics of a signal coming from any given direction. An HRTF is an individualized frequency response describing how a sound signal is transformed by the head, external ear and to a certain extent, torso (Gardner, 1995). HRTFs permit researchers to represent any degree in the horizontal plane, rather than being restrained by physical speaker locations. They also allow for signal manipulations that could not be reliably represented in the free-field (i.e. SSD vocoder simulations). Most importantly, generalized HRTFs are generally reliable across individuals due to similarities in head size and shape among different listeners; this applies specifically to interaural cues and not pinna cues. Individualized HRTFs are better for localization tasks than generalized HRTFs, especially in the vertical plane. However, generalized HRTFs are consistent across individuals in producing spatial

perceptions in the horizontal plane and are valuable tools for studying hearing in spatial noise (Begault, Wenzel, & Anderson, 2001; Wenzel, Wightman, & Kistler, 1991).

As in the previous chapter, this study utilized vocoder simulations of SSD-CI listening. Vocoder simulations are used extensively in CI research and are an invaluable tool for studying aspects of CI processing without potential sources of uncontrolled inter-subject variability often found in experiments involving CI listeners. Although vocoder processing is an imperfect estimation of what CI users might hear (Freyman et al., 2008; Ihlefeld & Litovsky., 2012; Li & Loizou, 2009), the key advantage of using vocoder simulations in this set of experiments is it allows for more direct control over the relative compression in the two ears than what could be achieved with actual CI listeners.

The goal of this study was to examine how interaural differences in loudness growth in vocoder simulations of SSD-CI listening would affect two main benefits to speech perception that CIs are known to provide to listeners with SSD: binaural squelch and head-shadow. More specifically, this study examined the effect of envelope compression or expansion on binaural squelch (Experiment 3.1) and on head-shadow benefit (Experiment 3.2).

Experiment 3.1: The effect of envelope compression and expansion on squelch in simulations of cochlear implants for SSD listeners

Experimental question. How do envelope compression and expansion affect squelch in a HRTF generated virtual free-field environment?

Hypothesis. The relative effects of both compression and expansion on squelch will likely depend on the TMR. The hypothesized effects of compression and expansion are summarized in Table I below. At positive TMRs, squelch was not expected based on results from previous studies (Bernstein et al., 2015, 2016) that indicated the most squelch occurs at negative TMRs where monaural cues are insufficient for the perceptual separation of competing talkers. Therefore, we did not expect an effect of compression or expansion at positive TMRs. At negative TMRs, compression will amplify quiet sounds (i.e., the target) relative to louder sounds (i.e., the interferers), which will effectively decrease the amplitude differences between target and interferers. Because in this case the target is the quieter sound, the low-level amplitude compression should make the ILD for the interferers and target more similar. This would reduce the perceived spatial difference between them and potentially reduce squelch. On the other hand, expansion should exaggerate the relative difference in amplitude between the target and interfering speech in the vocoder ear, thereby increasing the perceived spatial separation.

Table I. Hypothesis table for Experiment 3.1

TMR	Compression @ vocoded ear	Expansion @ vocoded ear
+	Unlikely to see unmasking benefit No effect	Unlikely to see unmasking benefit No effect
-	↑ Target level relative to masker ↓ Performance	↓ Target level relative to masker ↑ Performance

Table I. Hypothesis table indicating the predicted outcome after compression and expansion. Based on previous experiments, when the TMR is positive, listeners are unlikely to demonstrate squelch. Therefore, it was predicted that compression and/or expansion would not have any effect on performance. However, when TMRs are negative, an unmasking benefit is likely, and therefore compression and expansion are likely to affect performance. We predicted compression would increase the level of the target in the vocoded ear, making the target and masker level more similar, essentially reducing the ILD between the target and maskers, disrupting performance. Expansion should exaggerate the level difference between the target and masker and improve performance.

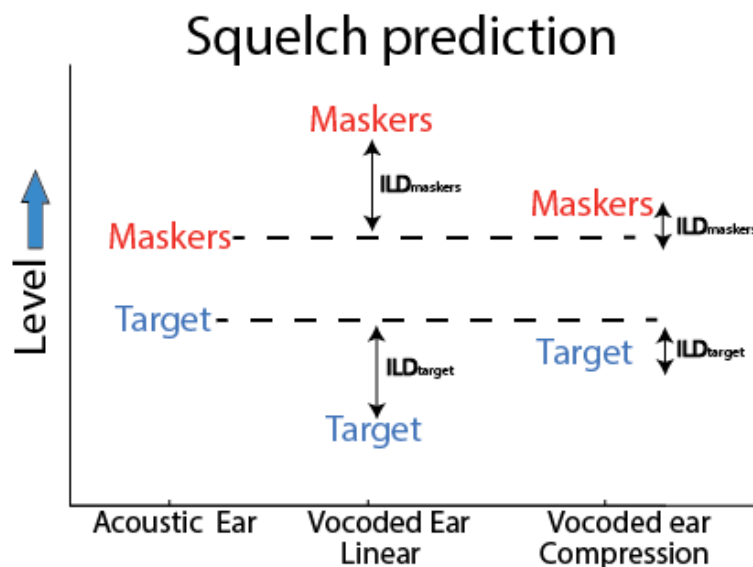


Figure 3.1. Prediction of what might happen to the squelch advantage after compression in Experiment 3.1.

Methods.

Approach. This experiment employed an HRTF-based simulation of spatially separated targets and maskers. This paradigm allowed for more realistic presentation of competing speech signals, with both ears receiving a mixture of target and masker energy as would occur in the free-field. This experiment was similar to the contralateral-unmasking paradigm of Bernstein et al. (2015, 2016) and Wess et al. (2017) in that it was used to measure the squelch benefit provided by a second (vocoded) ear in perceptually separated concurrent streams of speech. An additional similarity was that the target was located closest to the acoustic ear and the two same gender maskers were located closer to the vocoded ear. As in these previous studies, monaural performance was compared to bilateral performance. The amount of binaural squelch benefit was calculated as the difference between monaural (unprocessed ear only) and binaural performance – i.e., the magnitude of benefit the listener receives from the addition of the vocoded ear. In this experiment, the interfering speech was located closest to the vocoded ear, and therefore had a poorer TMR than the unprocessed ear. Thus, the addition of the vocoder ear had the opportunity to provide a squelch benefit, but not a head-shadow advantage. The target was presented virtually on the left side (-60 degrees) and two interfering talkers were presented on the right (+60 degrees) using HRTFs (see Figure 3.1). Compression and expansion were varied parametrically in the vocoder processing. This configuration was chosen because the largest possible head-shadow effects arise for sources originating from a 60 degree azimuth (Culling, Jelfs, Talbert, Grange, & Backhouse, 2012), on the order of 9 dB. Therefore, for a target and interfering sources at ± 60 degrees, the TMR of a target source is about 18 dB higher to the ear closer to the target source than the ear closer to the interferer. However, this is the theoretical maximum and is unlikely to occur in the real

world, where reverberation will decrease the SNR difference between the ears (Culling, et al., 2012). These spatial configurations were specifically chosen to maximize spatial differences (head-shadow), while also providing enough spatial separation to facilitate the perceptual separation of target and interfering voices based on perceived differences in spatial location.

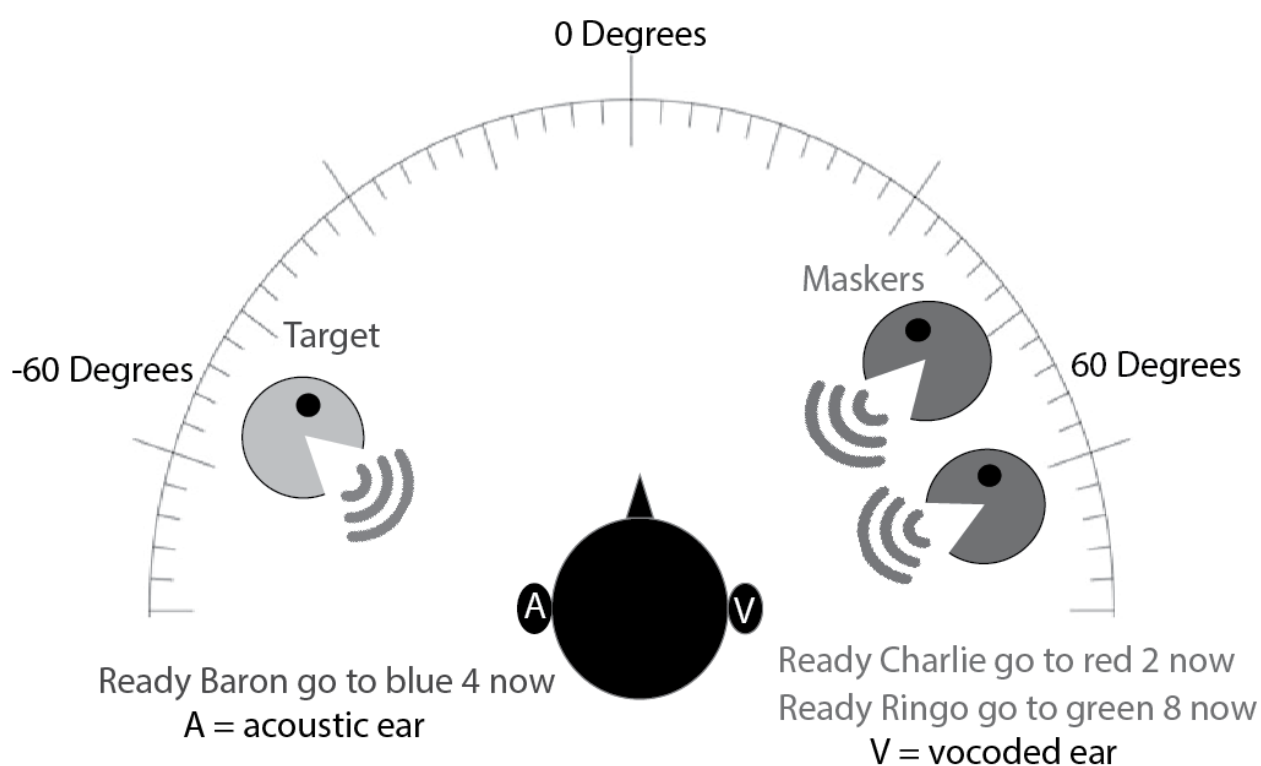


Figure 3.2. Schematic of the squelch experimental setup in experiment 3.1. The stimuli were presented over headphones, and each spatial configuration was created by convolving the speech with a generalized HRTF before additional processing. The target talker is located at -60 degrees, closer to the acoustic ear. The two same-gender maskers are located at +60 degrees, closer to the vocoded ear.

Participants. Experiment 3.1 was carried out at Walter Reed National Military Medical Center, Bethesda Maryland. Seven paid listeners (age range 18-39) participated in

this experiment. All listeners had NH (defined as symmetrical thresholds equal to or better than 20 dB hearing level at octave frequencies between 125 and 8000 Hz) and were free from cognitive and neurological disorders. All listeners were native English speakers. The listeners that participated in Experiment 3.1 also participated in Experiment 3.2.

Stimuli. The target and interfering speech were taken from the CRM speech corpus for multi-talker communication research (Bolia et al., 2000; Brungart, 2001). The CRM corpus consists of phrases of the form “Ready (call sign) go to (color) (number) now.” There were eight possible call signs (“Arrow,” “Baron,” “Charlie,” “Eagle,” “Hopper,” “Laker,” “Ringo” and “Tiger”), four possible colors (“blue,” “green,” “red” and “white”), and eight possible integer numbers (one through eight). A typical sentence would be “Ready Charlie go to white five now.” The target sentence call sign was always “Baron,” which provided the cue for the listener to identify which of the concurrent talkers was the target. The interferers used other call signs (e.g., “Arrow” or “Ringo”). Eight speakers (four females, four males) were used to record all possible combinations. To maximize the difficulty in perceptually separating concurrent talkers in the target (acoustic) ear, this experiment used all same-gender interferers and targets (Brungart et al., 2001). These are the conditions that produced the most masking release for SSD-CI listeners (Bernstein et al. 2016) and for single-sided vocoder listeners (Bernstein et al. 2015, 2016). The three talkers (target and two interferers) in a given trial were always of the same gender, although the gender varied randomly from trial to trial. The three simultaneous sentences were constrained such that they were always spoken by a different talker and had a different call sign, color and number. The target speech was presented at 60 dB SPL, with the interferer level adjusted

to yield the desired TMR. The following TMRs were tested: -16, -12, -8, -4, 0 and +4 dB. The TMRs were defined before the HRTFs and vocoding were applied. The TMR varied from -16 dB to +4 dB across the frequency range, with a speech importance-function weighted average +9 dB higher than the original TMR in one ear, and -9 dB in the other. The left ear was the unprocessed ear and the right ear was always the vocoded ear in this experiment. In the monaural condition, only the signals at the left (unprocessed) ear were presented.

Generation of HRTFs. We used HRTFs recorded at Oldenburg University (Wierstorf, Geier, Raake, & Spors, 2011). HRTFs were generated using an in-the-ear (ITE) microphone and a behind-the-ear (BTE) microphone (Siemens) on a Knowles Electronic Manikin for Acoustic Research (KEMAR). The HRTFs generated from the ITE microphone were used for stimuli presented to the unprocessed (left) ear and the BTE HRTFs were used for stimuli presented to the vocoded (right; CI simulation) ear. The excitation signal for the impulse response measurement used to create the HRTFs was presented from a loudspeaker at a distance of 80 cm from the center of the mannequin's head. This study used HRTFs that were recorded at -60 and +60 degrees.

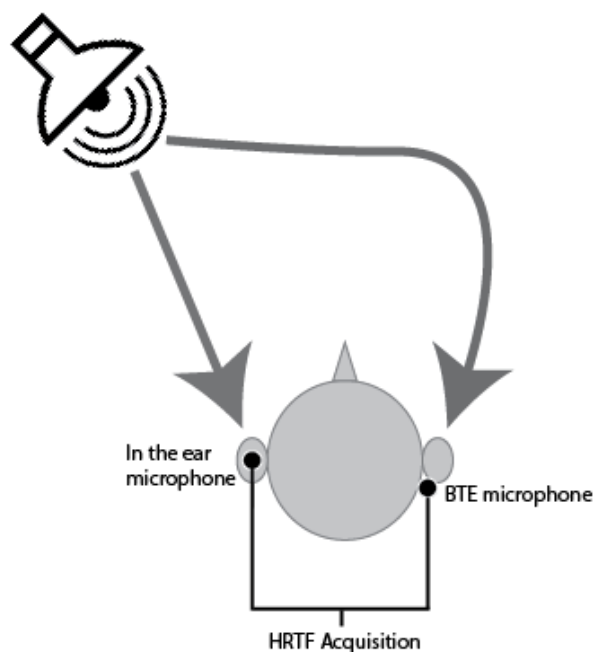


Figure 3.3. Schematic of HRTF acquisition for an ITE microphone and a BTE microphone.

Noise vocoding. An 8-channel noise vocoder was used to extract speech envelopes in a number of frequency channels and the envelopes were used to modulate bands of noise. First, stimuli were passed through a bank of fourth-order Butterworth (-24 dB roll off) “analysis” filters, the frequency range of the analysis filters was 100 to 10000 Hz. The envelope of the signal in each channel was extracted via a Half-wave rectification, then low-pass filtered at 400 Hz with a second-order Butterworth filter. Compression or expansion was applied (if applicable) at this stage in vocoder processing. Each envelope was then multiplied by a white noise carrier, with the resulting signal then passed through a series of bandpass “synthesis” filters. In this study, no frequency mismatch was applied. Frequency content was always delivered to the correct cochlear place, with the synthesis-

filter cutoff frequencies matching those of the analysis filters. Finally, the signals were summed across channels to create the noise-vocoded signal.

Loudness manipulations. Compression and expansion were implemented using a power-law function: $y = Ax^P + B$, A and B are constants where A is the threshold and B is the max comfortable level (expressed as micro amps in electric hearing). Since this formula was used to manipulate the envelope of an acoustic signal, no noise floor or max comfortable level was necessary for our simulation, therefore A and B are not included and the formula becomes, $y = x^P$. Instantaneous compression or expansion was applied to signal envelopes with compression exponents of 0.25, 0.50, 1 (linear, no compression), 1.5 and 2 (expansion) before the envelopes were applied to the noise carriers. The input/output function and a stem plot for the compression, linear and expansion conditions are shown in Figure 3.3. The nonlinear transformation (compression or expansion) was applied to each vocoder channel independently. The level of the resulting signal in each channel was adjusted to be equal to the RMS level of the input signal for that channel, and the delays associated with the filtering process were removed.

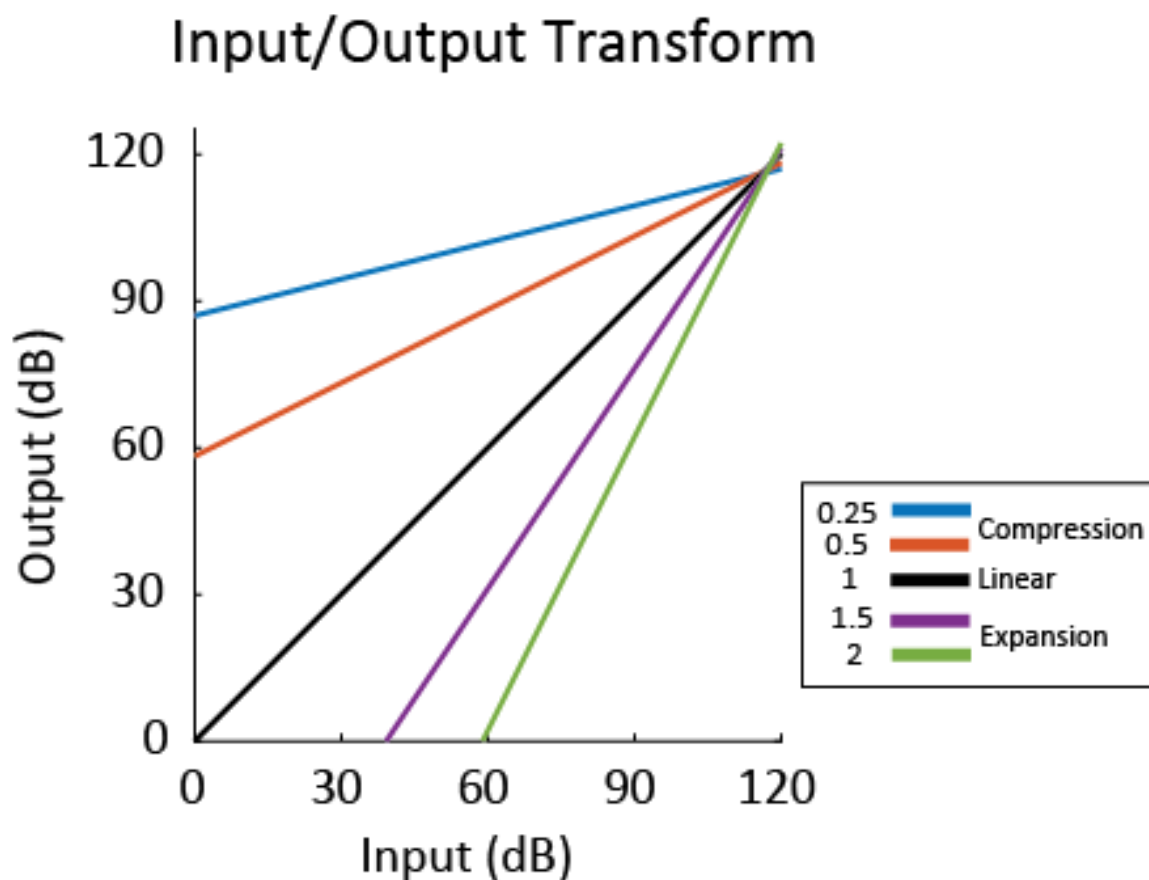


Figure 3.4. Input/output function for the compression, linear and expansion conditions.

Procedure. Listeners were instructed to identify the target talker based on the call sign “Baron” and repeat back the color and number reported by the target talker, while ignoring two interfering talkers who used different call signs. Listeners were told which side of the head the target would be presented. For the training portion of the experiment, 12 trials were presented for each TMR tested for the monaural condition and for the linear condition with no compression applied to the vocoder. Listeners were presented with blocks of 30 trials for a total of 120 trials in the training condition. The TMR varied

randomly from trial to trial. The monaural and linear bilateral conditions were fixed for all trials in a block.

For the experimental portion of the experiment, listeners were presented with 6 different vocoder-compression conditions: two compression conditions (compression factors = 0.25 and 0.5), two expansion conditions (compression factors = 1.5 and 2.0), a linear condition (compression factor = 1.0), and a monaural condition where no stimulus (interferers) was presented to the vocoder ear. Listeners completed 18 trials for each combination of TMR (-16, -12, -8, -4, 0 and +4 dB) and vocoder-compression condition. Listeners were presented with blocks of 36 trials for a total of 648 trials for each listener. Within each block, only one vocoder-compression was presented but the TMR varied randomly from trial to trial.

Participants were seated in a sound booth and directed their attention to a computer screen. The speech stimulus was generated by MATLAB, played via a RME Hammerfall (Haimhausen, Germany) sound card, and presented over Sennheiser HD 280 headphones. The RMS of the signal varied depending on the TMR and spatial location. The RMS of the signals ranged from 60 – 75 dB SPL. The signal level was not fixed because preservation of the ILDs created by the HRTFs was necessary for this experiment. The computer screen displayed an eight-column, four-row array of colored digits corresponding to the response set of the CRM. The listener used the mouse to select the colored digit corresponding to the number and color spoken by the target talker who used the call sign “Baron”. After each response, the subject received feedback, with the button associated with the correct answer flashing briefly. For the response to be considered correct, both the color and number needed to be correctly identified.

Results.

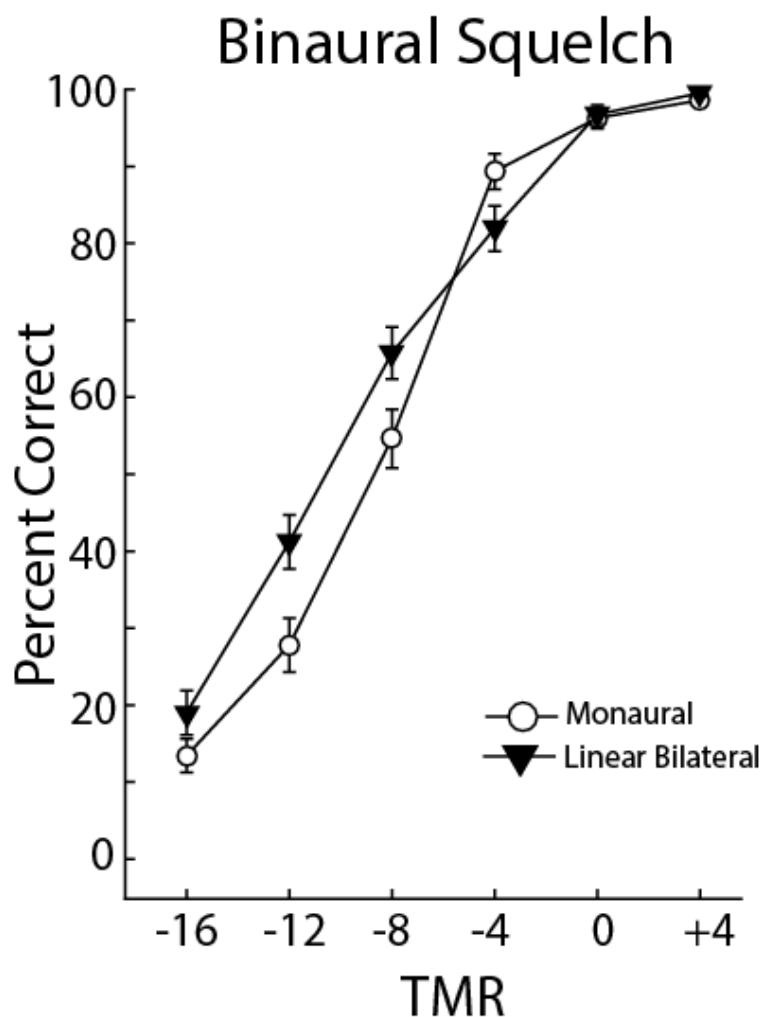


Figure 3.5. The linear bilateral and monaural data from experiment 3.1. The monaural conditions are depicted with white circles and the linear bilateral conditions are depicted with black triangles. These data indicate that the listeners did not receive a significant squelch benefit ($p > 0.05$).

Figure 3.5 plots the mean proportion of trials where the color and number were both correctly identified as a function of TMR for the monaural and linear vocoder conditions.

Percent correct was determined by summing the correct responses for each keyword (color and number) separately, then that number was divided by two.

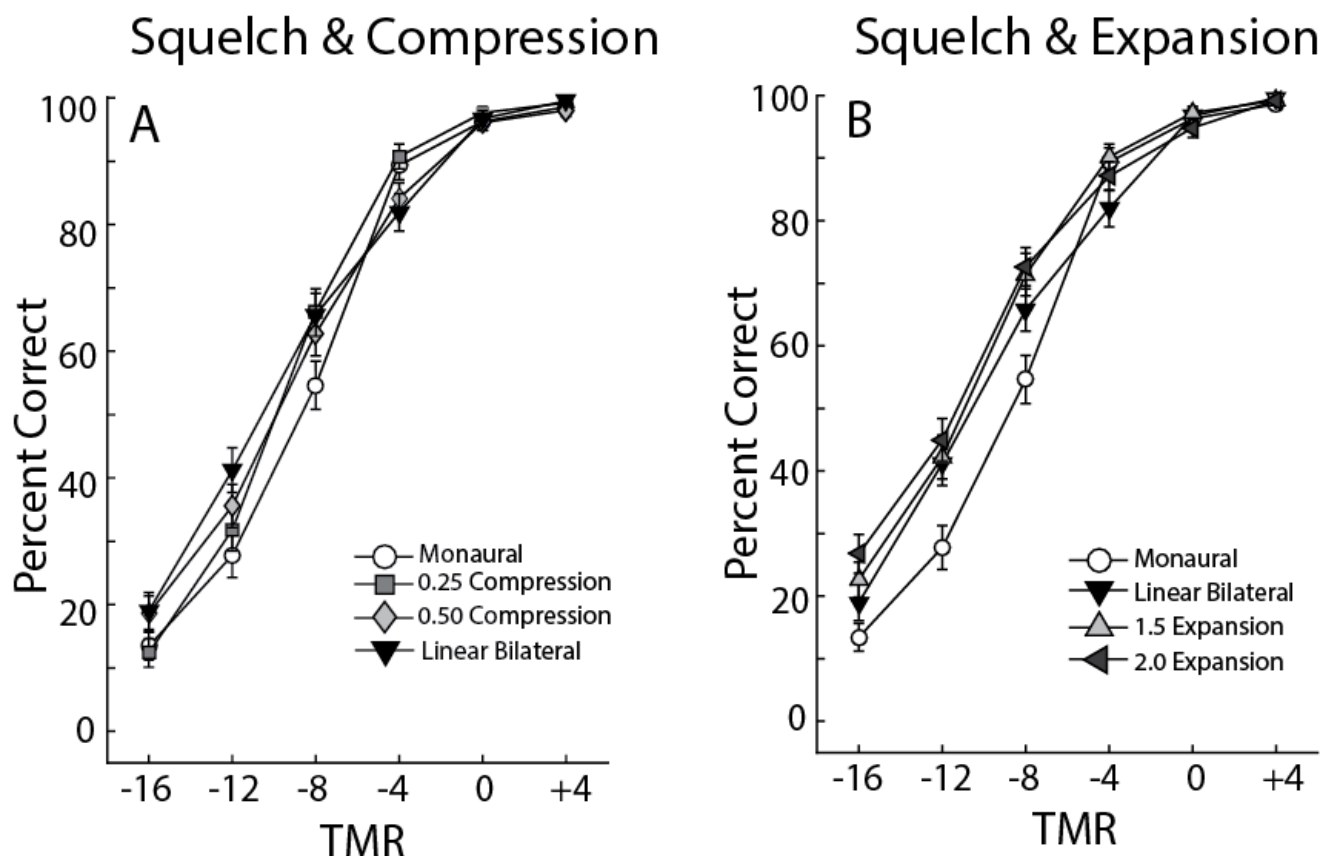


Figure 3.6. Compression and expansion data from experiment 3.1. The left figure 3.6A, shows the effect of compression on squelch compared to monaural and linear bilateral performance. The right figure 3.6B shows the effect of expansion on squelch. The negative effect of compression compared to linear bilateral is clear at lower TMRs. The slightly positive effect of expansion is also evident at lower TMRs.

Figure 3.6 plots the mean proportion of trials where the color and number were both correctly identified as a function of TMR for all six vocoder-compression conditions tested in the experiment. Figure 3.6.A. shows the effect of compression, plotting the results for

the two compression conditions ($\text{exp} = 0.25$ and 0.50) along with the monaural and linear-vocoder conditions replotted from Fig. 3.5. Figure 3.5.B plots the results for the two expansion conditions ($\text{exp} = 1.5$ and 2.0) together with the monaural and linear-vocoder conditions replotted from Fig. 3.5.

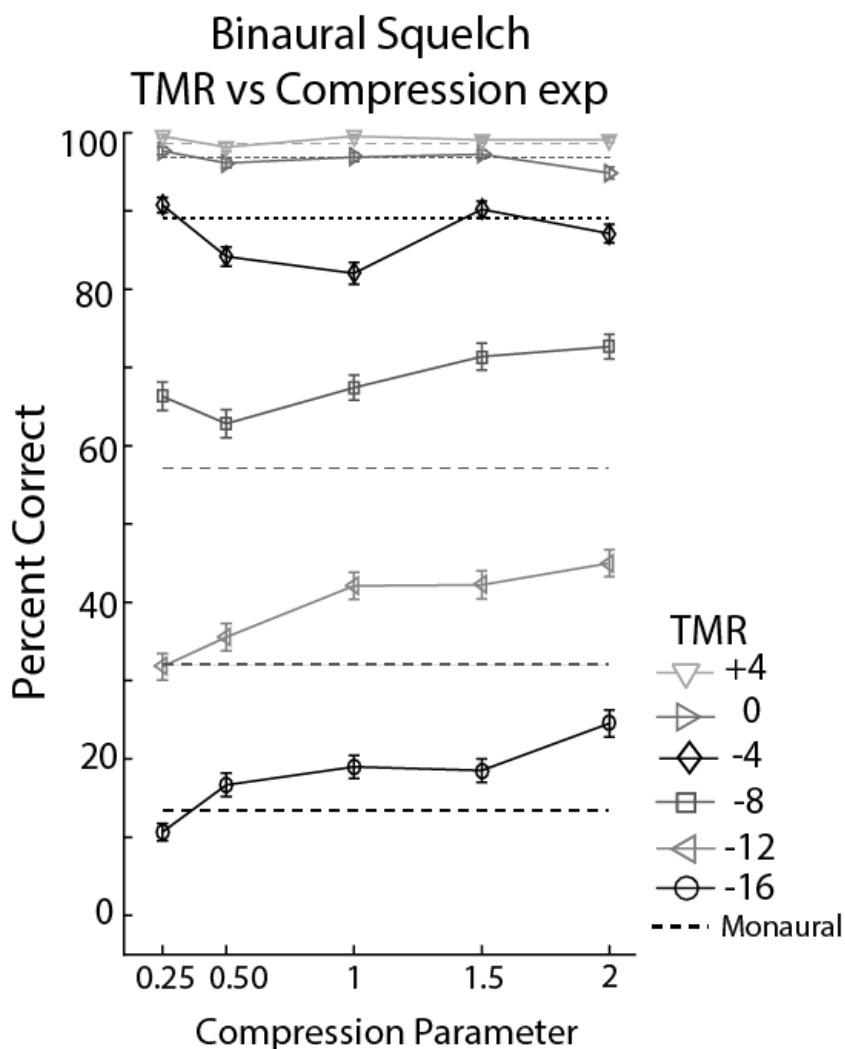


Figure 3.7. Data from experiment 3.1 plotted as a function of the compression parameter and TMR. The trend of improvements in squelch from compression to expansion are more clearly represented in this graph. Especially at low TMRs, listeners' performance improves as the vocoded signal moves from highly compressive to expansive. The dashed lined indicated monaural performance for a given TMR.

Figure 3.7 plots the same data as in Figure 3.5 and 3.6, but instead as a function of the compression parameter and TMR. The horizontal dashed lines indicate monaural performance for a given TMR. Plotted this way, these data more clearly show an effect of compression/expansion condition on performance, especially for lower TMR.

The data were analyzed using a repeated-measures binary-logistic regression analysis with two within-subject factors (compression parameter and TMR). This analysis was used because the data were binary in nature (correct or not) and the analysis takes into account the likelihood that percentage-correct scores are different based on the number of trials presented. The initial analysis included all the vocoder conditions (plus monaural) as well as all the TMRs tested. There was no significant main effect of condition, but there was a significant main effect of TMR [$\chi^2(5) = 10555.7, p < 0.001$] and a significant interaction between TMR and condition [$\chi^2(6) = 148.5, p < 0.001$]. A subsequent series of binary-logistic regression analyses was conducted to determine the source of the significant interaction by examining the effect of compression/expansion condition on performance at each TMR. Bonferroni corrections for 6 comparisons (TMR) were applied. There was a significant main effect of compression/expansion condition for all TMRs except -4 dB. The statistical results for each TMR are as follows: TMR -16 dB = [$\chi^2(5) = 99.1, p < 0.005$], TMR -12 dB = [$\chi^2(5) = 41.9, p < 0.005$], TMR -8 dB = [$\chi^2(5) = 32.5, p < 0.005$], TMR -4 dB = [$\chi^2(5) = 11.7, p = 0.25$], TMR 0 dB = [$\chi^2(5) = 23.9, p < 0.005$] and TMR +4 dB = [$\chi^2(5) = 161.4, p < 0.005$].

For the TMRs that showed a significant main effect, a series of post-hoc pairwise comparisons was conducted to determine which pairs of conditions were significant.

Bonferroni corrections for 15 comparisons were applied after statistical analysis. The results of the post-hoc tests are summarized in the table below Table II. Basically, compression disrupts performance compared to expansion at -12 and -16 dB and compared to linear (-16 dB TMR) performance.

Comparison	TMR	p value
0.25 vs Linear	-16 dB	$p=0.03$
0.25 vs 2.0	-16 dB	$p<0.001$
0.25 vs 2.0	-12 dB	$p=0.01$

Table II. Results from the post hoc tests from experiment 3.1. These results indicate that high compression (exp= 0.25) disrupts performance compared to a linear (exp=1.0) vocoded signal or an expanded one (exp =2.0).

Summary.

In summary, the results of experiment 3.1 demonstrated an effect of the compression coefficient. As the exponent moved from exp = 0.25 (highly compressed) to linear (1.0) and expanded (2.0) performance improved. The largest performance difference was observed between the highly expansive and highly compressive conditions at lower TMRs.

Experiment 3.2: The effect of compression and expansion on head-shadow benefit in simulations of cochlear implants for SSD listeners.

Introduction

The previous experiments described in Chapters 1 and 2 focused on binaural squelch. Head-shadow effects are also important for hearing speech in spatial noise. One of the most frequently reported speech-perception benefits for SSD-CI listeners is that the CI in the deaf ear allows them to take advantage of situations where the TMR is better at the CI ear (i.e., the head-shadow advantage). Here, the effect of compression and expansion on the head-shadow benefit were examined. The magnitude of the head shadow advantage provided by the CI will depend on the TMR at the CI. Therefore, we hypothesized this advantage would likely be impacted by compression or expansion due to changes imposed on the effective TMR at the CI ear.

Experimental question. What role do amplitude compression and expansion have on head-shadow benefit for vocoder-simulated SSD-CI listening in an HRTF generated virtual free-field environment?

Hypothesis. The effect of compression on head-shadow benefit will likely depend on the TMR at the vocoded ear after HRTF processing. Table III summarizes the hypothesized effects of compression and expansion on TMR and performance. Compression will tend to increase the level of the softer speech relative to the louder

speech. Therefore, at negative TMRs, compression should increase the effective TMR and therefore improve performance, while at positive TMRs, compression should decrease the effective TMR and decrease performance. Expansion is expected to have the opposite effect. This can be seen in Figure 3.8, compression will potentially reduce the TMR but expansion should expand it and performance should improve. The TMR at the ear is the relevant quantity, since that is what the vocoder receives as its input signal. The TMR at the loudspeakers and the effective TMR at the ears are different. Although the relationship between the TMR at the ear and the TMR of the original signal is frequency dependent. Culling et al. (2012) showed that for the spatial configuration tested here, the speech-weighted average TMR at the vocoded ear is about 9 dB higher than the signal TMR, after being convolved with an HRTF.

In addition to the possible role of the effective TMR, compression or expansion could also distort the envelope and subsequent speech cues which could also disrupt performance. This distortion of acoustic features and speech cues is more likely to negatively affect performance in this experiment (3.2) than in the previous squelch experiment, (3.1) because the target is presented on the vocoded side of the head and therefore the listener will rely more heavily on target speech information contained in the vocoded signal. This is in contrast to Experiment 3.1 where the vocoder ear mainly contained interfering speech information and was used to primarily provide spatial information for the listener to perceptually separate the target from the interfering signals.

MR	Compression @ vocoded ear	Expansion @ vocoded ear
+	↓ TMR ↓ Performance	↑ TMR ↑ Performance
-	↑ TMR ↑ Performance	↓ TMR ↓ Performance

Table III. Hypothesis table indicating the predicted outcome after compression and expansion. Since the talker is now located on the vocoded side of the head, predictions differ from experiment 3.1. Here in experiment 3.2, compression and expansion are likely to influence the effective TMR at the vocoded ear. At a positive TMR, compression should lower the TMR and disrupt performance. At negative TMRs, compression should increase the TMR and improve performance. At a positive TMR, expansion should increase the TMR and improve performance and at negative TMRs, expansion should decrease the TMR and disrupt performance relative to the linear bilateral condition.

Head-shadow prediction

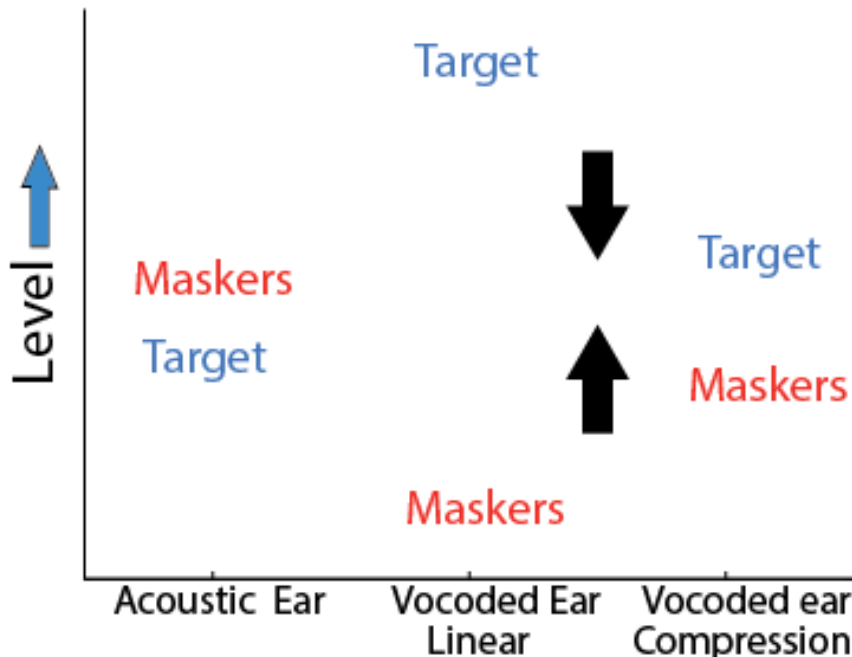


Figure 3.8. Prediction of what might happen to the head-shadow advantage after compression in Experiment 3.2.

Methods.

Approach. This experiment employed the same spatial paradigm as in Experiment 3.1 to maximize head-shadow effects, except that the locations of the target and interfering speech were reversed. Target and interferer azimuths of +60 and -60 degrees were used to maximize the potential for head-shadow benefit produced by adding the vocoder ear relative to the monaural condition (Culling et al., 2012). The target was presented on the vocoded, right side (+60 degrees) and the maskers were presented on the acoustic, left side (-60 degrees) using HRTFs to simulate the level and timing effects of the physical barrier created by the head. Comparing the monaural performance (unprocessed ear only, poorer TMR) to the bilateral performance (both ears) gives us a measure of head-shadow benefit afforded by the vocoder ear (better TMR). As in Experiment 3.1, the right ear was the vocoded ear and the left ear was unprocessed.

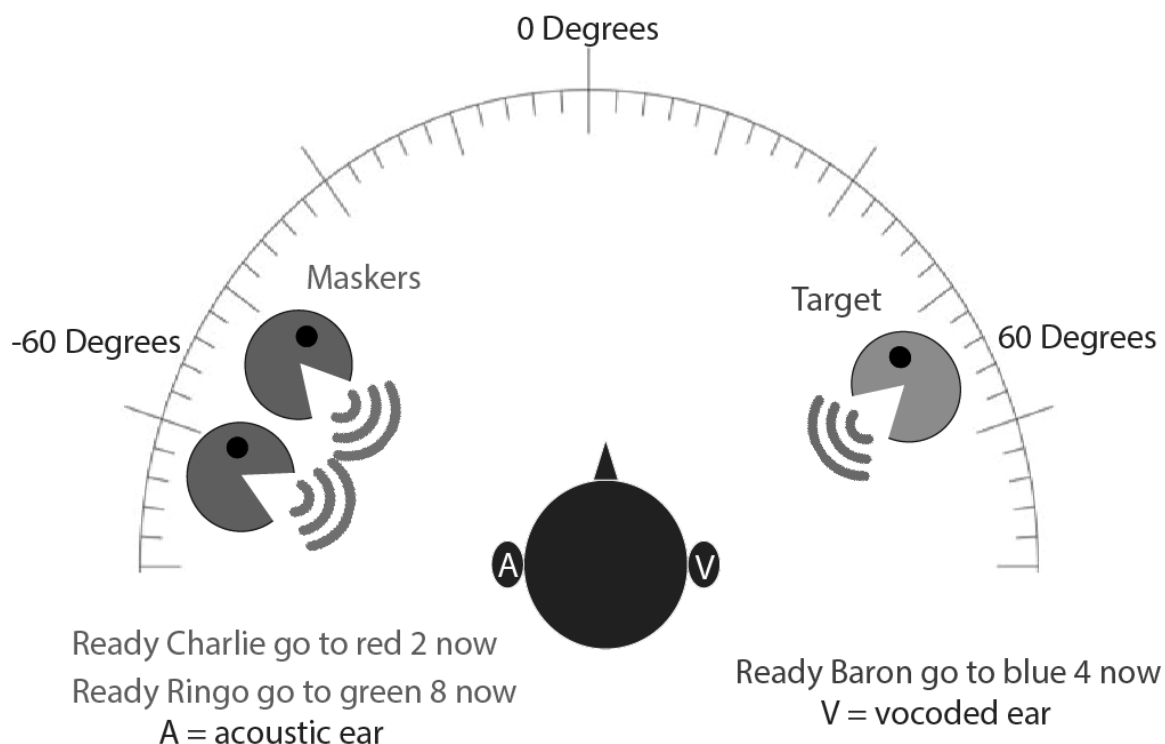


Figure 3.9. Schematic of the head-shadow experimental setup in experiment 3.2. As in experiment 3.1 the stimuli are presented over headphones and each spatial configuration is created by convolving the speech with a generalized HRTF before additional processing. Now the target talker was located at 60 degrees, closest to the vocoded ear. The two same gender maskers were located at -60 degrees, closest to the acoustic ear.

Participants. The same 7 NH listeners that completed Experiment 3.1 also participated in Experiment 3.2. The order of experiments was randomized between participants to reduce any chance of order effects.

Stimuli. The stimuli for experiment 3.2 were similar to those used in Experiment 3.1, except for the range of TMRs tested and the spatial locations of the target and maskers. Specifically, the range of TMRs (-4, 0, +4, +8, +12 and +16 dB) was higher to offset the large amount of attenuation of the target in the baseline monaural condition. This assured the TMR was almost always positive at the vocoded ear (i.e., about 9 dB higher than the

TMR before HRTF filtering), so that the task was possible, even in the monaural conditions. Therefore, the prediction (Figure 3.8) is that compression will reduce the TMR and thereby reduce performance. Expansion should increase the TMR and increase performance in this task.

Procedure. This experiment required the listener to identify the target talker based on the call sign “Baron” and repeat back the color and number reported by the target talker, while ignoring two interfering talkers who used different call signs. Listeners were told to which side of the head the target would be presented. For the training portion of the experiment, 12 trials were presented for each TMR tested for both the monaural condition and the bilateral condition with no compression applied to the vocoder (linear). Listeners were presented with blocks of 30 trials for a total of 120 trials in the training condition. The TMR varied randomly from trial to trial. The monaural and linear bilateral conditions were fixed for all trials in a block.

For the experimental portion of the experiment, listeners were presented with 6 different vocoder-compression conditions: two compression conditions (compression factors = 0.25 and 0.5), two expansion conditions (compression factors = 1.5 and 2.0), a linear condition (compression factor = 1.0), and a monaural condition where no stimuli (interferers) were presented to the vocoded ear (i.e., a simulation of SSD without a device intervention). Listeners completed 18 trials for each combination of TMR and vocoder-compression condition. Listeners were presented with blocks of 36 trials for a total of 648 trials for each listener. Within each block, only one vocoder-compression was presented, but the TMR varied randomly from trial to trial. As in Experiment 3.1, signals presented

to the left ear were unprocessed, while signals presented to the right ear were vocoded. But in contrast to experiment 3.1, the target speech was presented from the right (vocoded) side, so the vocoder ear had the better TMR.

Results.

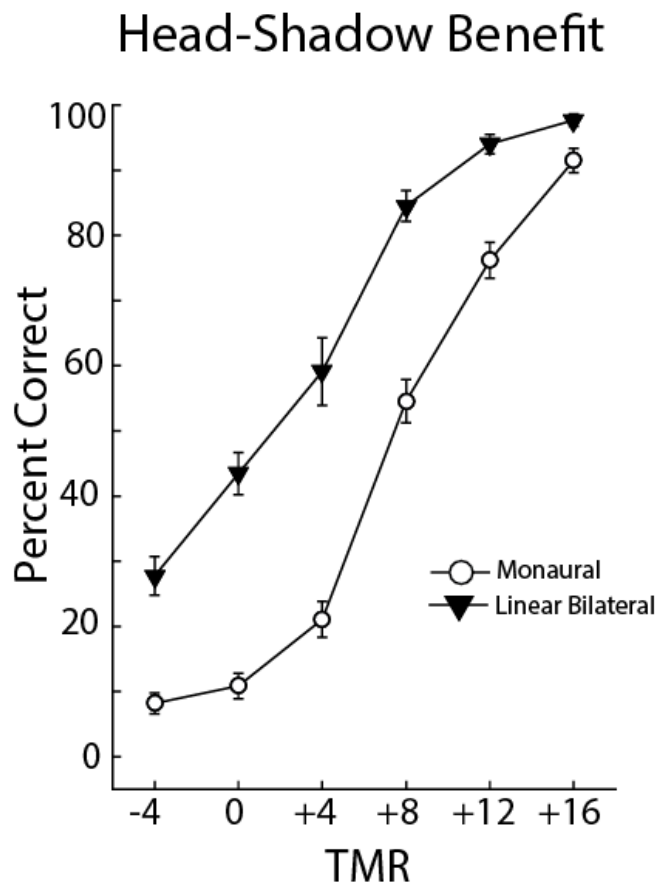


Figure 3.10. Plots the linear bilateral and monaural data from experiment 3.2. The monaural conditions are depicted with white circles and the linear bilateral conditions are depicted with black triangles. These data indicate that the listeners receive a head-shadow benefit at all TMRs except +16 ($p=0.06$), where performance between monaural and bilateral is not significantly different.

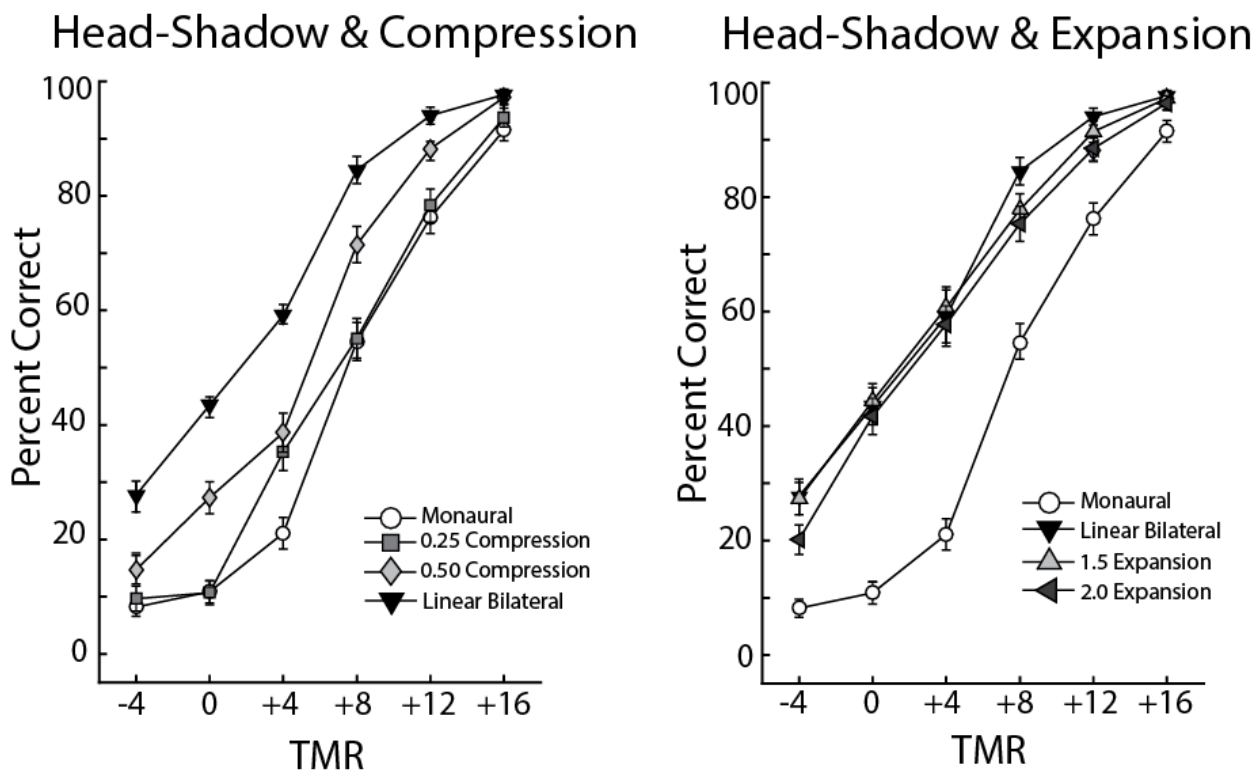


Figure 3.11. Compression and expansion data from experiment 3.2. The left figure 3.11A shows the effect of compression on head-shadow benefit compared to monaural and linear bilateral performance. The right figure 3.11B shows the effect of expansion on head-shadow benefit. The negative effect of compression compared to linear bilateral is clear at nearly all the TMRs tested (3.11A).

Figure 3.10 plots the mean proportion of trials where the color and number were both correctly identified as a function of TMR for the monaural and bilateral (linear condition). The data in Figure 3.10 show a clear head-shadow advantage for all TMRs tested except +16 dB. To illustrate the effect of compression and expansion on performance, Figure 3.11 plots the mean proportion of trials where the color and number were both correctly identified as a function of TMR for the monaural and all of the bilateral conditions. Figure 3.11.A. plots the data for the two bilateral conditions with compression (exp = 0.25 and 0.50). Figure 3.11.B. plots the results for the two bilateral conditions with

expansion (exp = 1.5 and 2.0). The monaural (white circles) and linear bilateral data (green triangles) from Figure 3.10 are replotted in both panels of Figure 3.11 for comparison.

Table IV: Head-shadow benefit experiment post-hoc results

Comparison	TMR	<i>p</i> value
0.25 vs Linear	-4 dB	$p < 0.001$
0.25 vs 1.50	-4 dB	$p = 0.01$
0.50 vs Linear	-4 dB	$p = 0.004$
M vs 1.5	-4 dB	$p = 0.02$
M vs Linear	-4 dB	$p = 0.006$
0.25 vs 0.50	0 dB	$p = 0.03$
0.25 vs Linear	0 dB	$p = 0.001$
0.25 vs 1.5	0 dB	$p < 0.001$
0.25 vs 2.0	0 dB	$p < 0.001$
M vs 2.0	0 dB	$p < 0.001$
M vs 1.5	0 dB	$p = 0.003$
M vs Linear	0 dB	$p = 0.002$
0.25 vs Linear	+4 dB	$p = 0.005$
0.25 vs 1.5	+4 dB	$p = 0.01$
0.25 vs 2.0	+4 dB	$p = 0.006$
M vs 2.0	+4 dB	$p < 0.001$
M vs 1.5	+4 dB	$p < 0.001$
M vs Linear	+4 dB	$p < 0.001$
0.25 vs 0.50	+8 dB	$p = 0.002$
0.25 vs Linear	+8 dB	$p < 0.001$
0.25 vs 1.5	+8 dB	$p < 0.001$
0.25 vs 2.0	+8 dB	$p = 0.005$
M vs 1.5	+8 dB	$p < 0.001$
0.50 vs Linear	+8 dB	$p < 0.001$
M vs Linear	+8 dB	$p < 0.001$
0.25 vs Linear	+12 dB	$p < 0.001$
0.25 vs 1.5	+12 dB	$p = 0.001$
M vs 2.0	+12 dB	$p = 0.049$

Table IV. Significant results from the post-hoc tests from experiment 3.2. The “M” refers to monaural performance. These results indicate that high compression (exp= 0.25) disrupted performance compared to all other conditions (exp= 0.50, linear, (exp=1.50) and (exp=2.0), at least at some TMRs. These data indicate a large negative effect of compression on head-shadow benefits.

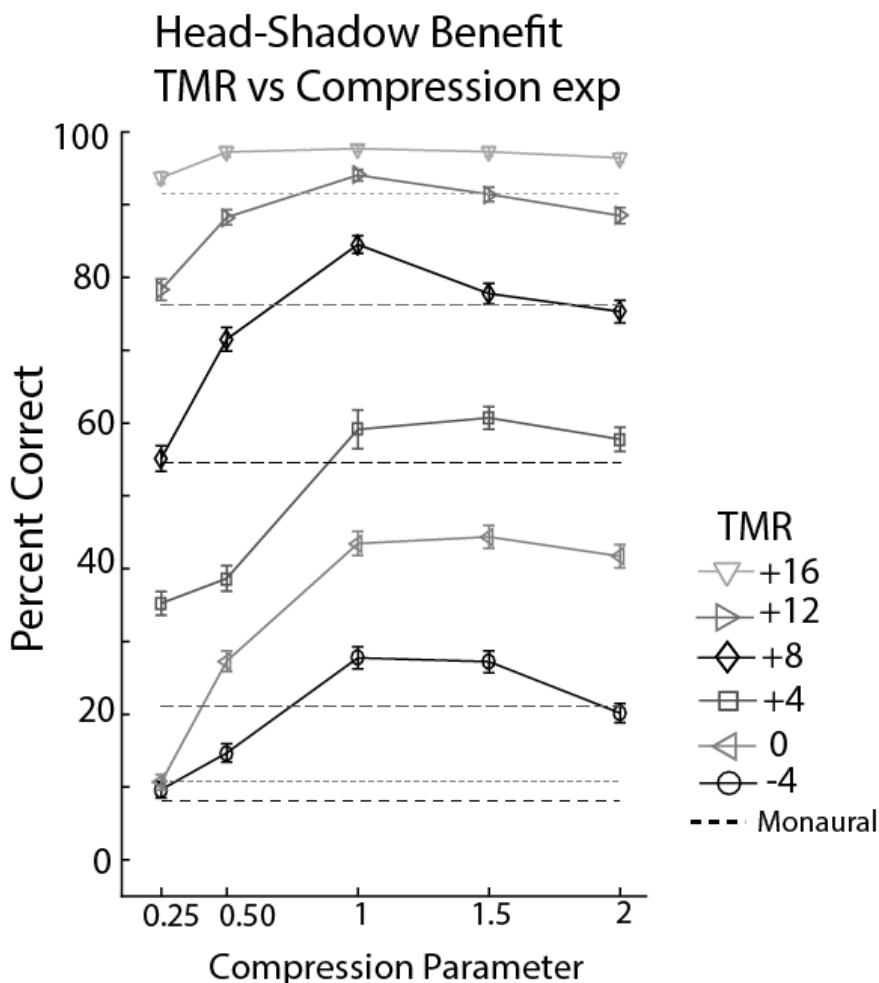


Figure 3.12. Data from experiment 3.2 plotted as a function of the compression parameter and TMR. The trend of improvements in head-shadow from compression to expansion is more clearly represented in this graph. At low TMRs, the data appear in a bell shape, with performance being worse than linear bilateral with both compression and expansion.

Figure 3.12 plots the same data as in Figure 3.10 and 3.11, as a function of the compression parameter. The horizontal dashed lines indicate monaural performance for a given TMR. Plotted this way, the data more clearly show a clear effect of the compression parameter on performance at all but the highest TMR tested (+16 dB). The data were analyzed using a repeated-measures binary-logistic regression analysis with two within-subject factors (compression parameter and TMR). The initial analysis included all the vocoder conditions (plus monaural) as well as all the TMRs tested. There was a significant

main effect of condition [$\chi^2 (5) = 1923.7, p < 0.001$], a significant main effect of TMR [$\chi^2 (5) = 28856.6, p < 0.001$] and a significant interaction between TMR and condition [$\chi^2 (6) = 28.0, p < 0.001$].

The initial analysis revealed a main effect of condition, a main effect of TMR and an interaction between TMR and vocoder condition, therefore a subsequent analysis was conducted at each TMR separately to determine the source of these significant results. Bonferroni corrections for 6 comparisons (TMR) were applied after statistical analysis. The statistical results for each TMR are as follows: TMR -4 dB = [$\chi^2 (5) = 251.2, p < 0.005$], TMR 0 dB = [$\chi^2 (5) = 242.4, p < 0.005$], TMR +4 dB = [$\chi^2 (5) = 64.7, p < 0.005$], TMR +8 dB = [$\chi^2 (5) = 667.4, p < 0.005$] and TMR +12 dB = [$\chi^2 (5) = 660.2, p < 0.005$]. Thus, a significant main effect of compression and expansion was found at each TMR except +16 dB.

For the TMRs that showed a significant main effect, a series of post-hoc pairwise comparisons were conducted to determine the significant interactions between the vocoder and monaural conditions. Bonferroni corrections for multiple (15) comparisons were applied for each analysis. The results of the post-hoc tests are summarized in Table IV. In summary, these results showed significant differences (a) between the 0.25 and linear conditions at most of the TMRs, (b) between 0.25 and expansion (1.5 or 2.0) at some TMRs, and (c) between the monaural condition and the linear or expanded (1.5 or 2.0) vocoder conditions at some TMRs.

In summary, the results of Experiment 3.2 show a significant negative effect of compression on head-shadow benefit at all TMRs except +16 dB. Compression completely eliminated the head-shadow benefit in many cases. Expansion had no significant effect on performance, although there was a non-significant trend for expansion to slightly reduce performance relative to the linear condition.

Discussion

The goal of this study was to examine the effects of compression and expansion on squelch and head-shadow benefit in vocoder simulations of SSD-CI listening in virtual auditory environments. The results of these experiments show that compression was detrimental to performance in both the squelch and in the head-shadow experiment. Expansion afforded a modest benefit in the squelch experiment when compared to performance in the highly compressed condition. There was a trend toward expansion having a slight negative effect on performance in the head-shadow experiment (albeit not significantly). The impact of envelope compression and expansion could be attributed to changes in the effective TMRs or ILDs or to the distortion of envelope speech cues. TMR and ILD are related quantities, both are determined by the interaction between head-shadow and spatial origin. However, they likely played different roles in determining outcomes in the squelch and head-shadow experiments. In this study, TMR is a monaural quantity referring to the level of the target and masker level in one ear. The TMR is especially important in the head-shadow experiment (3.2) since any listening advantage is

based on monaural listening to the ear with the better TMR. The essential cue for the squelch experiment was the ILD, which is the difference in loudness between the target and the maskers across the ears. In the squelch Experiment (3.1), the ILD provided the spatial cue necessary to perceptually segregate the target from the masker. Finally, envelope distortion after compression/expansion could have reduced the intelligibility of the target. This likely would have been particularly detrimental in the head-shadow experiment, because the listener depended on the vocoder ear for the extraction of target speech information. This is in contrast to the squelch experiment where the vocoder ear provided spatial cues.

The effect of compression and expansion on squelch.

The effects of compression and expansion in this experiment can be understood in terms of changes in the effective ILDs of the target and interfering speech. Referenced to the unprocessed ear, the target ILD was positive (i.e., louder at the acoustic ear) and the masker ILD was negative, due to the spatial locations of the target and maskers. Since compression amplifies quieter sounds relative to louder ones, the effect of compression in the vocoder on the target and masker levels depends on the TMR at the vocoder ear. Therefore, at negative TMRs, compression would amplify the target. The TMR at the vocoded ear was very negative: the TMRs tested were negative (at the level of the loudspeakers) and the HTRFs exacerbated this effect. Therefore, in this experiment compression would have amplified the target relative to the masker in the vocoded ear. This effectively caused the ILD between the target and masker to become more similar,

thereby reducing the perceived spatial difference and ultimately minimizing squelch. This is what occurred in this experiment: squelch was reduced as the compression exponent decreased from expansive ($\text{exp} = 2.0$) to compressive ($\text{exp} = 0.25$). In contrast, expansion should have exaggerated the difference between target and masker in the vocoder ear, thereby increasing the effective ILD difference and improving unmasking. This is indeed what occurred: expansion provided a small listening benefit (i.e., more squelch) compared to that provided by the compressed vocoder signal. The HRTF-generated virtual auditory environment utilized in this experiment allowed for differences in spatial cues to be represented to the listener over headphones. In CI and vocoder processing alike, the major inter-aural spatial cue is loudness differences or ILDs. A study by Grantham et al., (2008) examined the role of compression on ITD and ILD thresholds in BICI listeners. They measured ILD thresholds (with Gaussian noise bursts) with compression turned on and off, and found that compression drastically raised ILD thresholds for 10 out of 12 CI listeners. They found a mean ILD threshold of 3.8 dB with compression on to 1.9 dB with compression off. Similar results of the negative effects of compression on ILD threshold in CI listeners have been found by other researchers (Laback, Egger, & Majdak, 2014; Senn, Kompis, Vischer, & Haeusler, 2005).

The use of primarily ILD cues by CI and vocoder listeners has been widely reported in the literature (Buss et al., 2009; Garadat, Litovsky, Yu, & Zeng, 2009; Li & Loizou, 2009; Schleich et al., 2004b; van Hoesel, 2008). Mechanistically, this experiment causes ILD differences between the target and the interferers, and provides the listeners with a cue to differentiate the target from the masker based on the perceived difference in spatial location. Using the squelch paradigm described in Chapter 2, Bernstein et al., (2015)

investigated a situation (albeit artificial) where the masker ILD was held at 0dB and the target ILD was adjusted from negative infinity to 0 dB by mixing target energy with the masker energy in the vocoded ear. They found that as the target ILD decreased and became more similar to the masker ILD, the squelch benefit started to disappear. A target ILD of 6 dB or less completely eliminated squelch.

Another factor that might have influenced the results is the envelope distortion caused by compression and expansion. Distortion of the signal envelope could have led to decorrelation of the signals between the acoustic ear and the vocoded ear, which has been previously shown to limit unmasking (van de Par & Kohlrausch, 1998). However, both compression and expansion could have distorted the envelope relative to the unprocessed ear and had a similar effect on performance. This is not what was found in the results, suggesting the decrease in performance after compression had more to do with ILDs than with the distortion of speech information.

The effect of compression and expansion on head-shadow benefit.

In the head-shadow experiment (Experiment 3.2), performance generally improved when the linearly processed vocoder signal was provided to the listener (compared to monaural). This is because in this paradigm the target talker was located closest to the vocoded ear. Therefore, this result reflects a head-shadow advantage when the stimuli were provided to the vocoded ear, because they could now listen to the vocoded ear with the better TMR to hear the target speech.

The results of this experiment can be thought of in terms of changes in the TMR at the vocoded ear after compression and expansion. The TMR is the important parameter here, since the head-shadow benefit is based on monaural listening to the ear with the better TMR. Since the TMR was always positive at the vocoded ear, we hypothesized that compression should have amplified the masker relative to the target thereby reducing performance. Conversely, expansion should have increased the level of the target in the vocoded ear, which should have improved performance. However, this is not what occurred. Whereas compression did reduce performance, expansion did improve it, and there was even a non-significant trend toward a reduction in performance. This suggests that the distortion of speech cues might have offset any TMR advantage that expansion might have provided. It is likely that a disruption of intelligibility via envelope distortion caused by compression and expansion likely contributed to the observed decrease in performance (for compression) and the lack of improvement in performance (for expansion).

Envelope distortion and loss of intelligibility likely played a larger role in this experiment than in Experiment 3.1 since the listener had to primarily rely on the vocoder signal to hear the target. There is evidence in the literature that implicates that compression and expansion distort speech cues, and this is particularly relevant for the head-shadow case because listeners are relying on speech cues in the vocoder ear. According to the lexical access (i.e. speech recognition) model suggested by Stevens (2002), the first component in successful speech perception involves breaking down a speech signal into “acoustic landmarks” based on frequency features and amplitude peaks in the signal. If detection of these acoustic landmarks is compromised, then the listeners will have

difficulty perceiving the speech because word boundaries and syllable onsets will be misconstrued. Envelope compression has been shown to skew acoustic landmarks and subsequent word boundaries in speech, especially in noise. Combined with the poor spectral resolution of the vocoder, which reduces speech redundancy, compression causes the listener to lose the reliable cues required to correctly hear speech (Li & Loizou, 2009). Envelope distortion is also known to occur after envelope expansion (Clarkson & Bahgat, 1991; Fu & Shannon, 1998; Lorenzi, Berthommier, Apoux, & Bacri, 1999). The effect of expansion on intelligibility is not expected to be as great as with compression, at least for NH listeners as predicted by the Speech Transmission Index (Steeneken & Houtgast, 1980). The head-shadow experiment specifically called for the listener to primarily attend to the signal in their vocoded ear, since the target was located at 60 degrees, closest to that ear. It is likely that any alteration of the speech envelope in the head-shadow experiment could reduce performance. Expansion was not as detrimental as compression, possibly due to offsetting effects, whereby expansion might have increased TMR but also caused distortion in the signal. Alternatively, it could be that expansion is not as detrimental to the signal as envelope compression

Implications for CI listeners.

In this study, nonlinear loudness growth was found to affect head-shadow advantage and squelch for the NH listeners presented with vocoder simulations of SSD-CI listening. This suggests that adjusting the compression function is a potential candidate for

optimization for SSD-CI listeners. The ideal clinical solution for SSD-CI would be to have perfectly matched loudness growth between a CI and acoustic ear for SSD-CI populations. Given the reduced DR of a CI, this would be nearly impossible to achieve. However, audiologists do have some control over the details of the CI compressive function. This could allow the opportunity to test different strategies to offset the limitations imposed by envelope compression. For instance, envelope distortion via compression has been shown to severely limit the peak-to-trough ratio in the signal, which is a proxy for acoustic (obstruent) landmarks. The obstruent landmarks are consonants created by obstruction of vocal airflow. A study by Li and Loizou (2009) measured the peak-to-trough ratio for linear and compressed speech and found a decrease of 7.6 dB in the ratio after compression (down from 10dB). They concluded that CI listeners will not likely be able to perceive such a small ratio (2.4 dB) and acoustic landmark identification will be greatly reduced.

To address the distortion caused by CIs, researchers have suggested implementing different types of compression. For example, one type of compressive function uses an s-shaped input-output function, which would expand low level input, up to a certain point (knee point), then compression would turn on after the knee point is reached. Theoretically, the audiologist could adjust the knee point based on estimated noise levels. This could enable expansion of acoustic landmarks in speech while compressing less important features of speech rather than louder more salient portions. This is accomplished by amplifying the portion of the DR where speech features are more likely to occur. Kasturi and Loizou (2007) implemented a sigmoid-shaped compressive function and found that this more sophisticated compression improved speech perception for CI listeners, but only when it was optimized for each listener individually. The CI listeners showed improved

sentence recognition in noise when using a dynamic s-shaped function compared to a logarithmic compressive function, which they attributed to less distortion of critical speech features in the speech envelope. The success of a dynamic compressive function suggests that this might be an important potential target for optimization to reduce envelope distortion and subsequent speech perception distortions for SSD-CI listeners.

Study Limitations.

This study was conducted on NH participants listening to vocoder simulations of SSD-CI listeners. The power-law compression implemented in this study was very basic and likely didn't capture the complexity of compression in CI processing. For example, our vocoder processing did not include any pre-emphasis speech amplification or gain control mechanisms. Nevertheless, the objective of this study was to examine spatial speech perception with simple CI compression and expansion, and this study was a critical first step in examining the spatial hearing outcomes of amplitude distortion in CI listeners. Another possible limitation to this vocoder experiment is that it does not capture the effects of plasticity that can occur after altered sensory input. More specifically, CI listeners have time to adapt and get accustomed to listening to their CI which contains compression distortions. Therefore, CI listeners who have substantial listening experience with their implant might not experience the same decrease in speech perception that occurred in our vocoder experiments. Although vocoder experiments are imperfect estimates of what actual CI listeners' experience, they are valuable in that they allow for specific and independent manipulation of certain aspects of CI processing in a very precise way. The

vocoder results alone do not determine what might happen in actual SSD-CI users, but these results suggest that effects of compression should be examined in future studies in SSD-CI listeners.

A potential limitation of this study is that envelope compression was examined in isolation of any loudness specific conditions examining level differences between targets and maskers and possible outcomes on binaural squelch. In the real-world, a CI has very different loudness growth than an NH ear. The consequence of this is that the perceived loudness of the CI will change relative to the acoustic ear as a function of level. These differences in loudness growth would affect both the target and masker level (since HRTFs were used in this study). For actual SSD-CI listeners in the squelch experiment, where the target is located at -60 degrees and the maskers are located at 60 degrees, relative loudness differences between the target and maskers might profoundly affect performance in a non-standard way. This problem can be understood in terms of loudness growth curves from acoustic and CI ears adapted from McDermott & Varsavsky (2009) (Figure 3.13). For instance, for masker levels below the noise floor (below 25 dB SPL), SSD-CI listeners would not receive any binaural unmasking benefits, since the sound will not be transduced by the CI. The perceptual separation cue will not be provided by the CI and all talker energy will be relayed to the acoustic ear. For a masker level falling slightly above the noise floor (but less than 45 dB SPL), the SSD-CI listener might still not receive an unmasking benefit because the masker energy will not be loud enough to combine with masker energy in the acoustic ear and no listening benefit will be obtained (location on growth curved marked with number 1). However, a masker level of about more at about 50 dB (first cross-over point) would provide an unmasking benefit since the loudness between the masker will be

matched across the ears (Figure 3.13, point “2”). The situation changes for a masker sound that is loud (around 75 dB SPL, point “3”). In this situation, the masker will sound much louder in the CI ear relative to the acoustic ear. This situation might improve performance due to a very large ILD difference between the masker and target. Conversely, the louder maskers sound in a CI ear relative to the target dominated mixture in the acoustic ear could vastly increase the salience of the masker signals, which would limit a listener’s ability to hear the target. Finally, at the second cross over point (about 85 dB SPL, point “4”), the loudness between the target and maskers should match again and the listeners should receive an unmasking benefit. Taken together, the location of targets and maskers and the relative level differences of targets and maskers (due to CI loudness growth) could have profound effects on spatial hearing for SSD-CI listeners in real world situations.

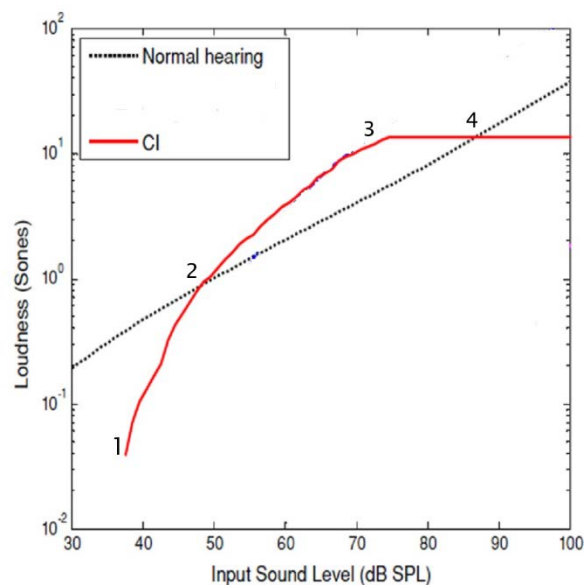


Figure 3.12. Loudness growth curve for a CI and acoustic ear adapted from McDermott & Varsavsky (2009).

Optimizing compression in CIs has become a focus of recent research efforts. Lopez-Poveda et al. (2016) developed a compression strategy that was meant to mimic the efferent

olivocochlear reflex (OCR) found in normal hearing. The OCR is important for NH listeners because this pathway allows for dynamic manipulation of the physical properties of the basilar membrane, which can effectively adjust the gain of signals reaching the brain. The OCR can be initiated by ipsilateral and contralateral input, and has been shown to aid in speech perception in noise (Mishra & Lutman, 2014). Unlike for NH listeners, compression is fixed in the CI processor and is likely inferior to the dynamic, adjustable compression seen in a NH listener's basilar membrane. Lopez-Poveda et al. (2016) examined the effect of compression on speech intelligibility in bilateral and SSD-CI listeners. They implemented a dynamic compressive function and compared performance to a standard logarithmic compressive function. They found a significant improvement in SRM and improved speech recognition for spatially separated speech and noise in both bilateral and SSD-CI listeners. These results are promising since the strategy is only an approximation of what a NH basilar membrane is doing at any given time. Ultimately, to ameliorate some of the CI amplitude distortions, steps needed to be taken to address the lack of front end compression (compression occurring at the microphone before additional signal processing) and the addition of artificial back-end compression (compression occurring later on in each individual channel), which can raise ILD thresholds and distort speech envelopes.

An additional limitation of this study was the small number of subjects that were tested ($n = 7$). In spite of this, many of the comparisons between the compression and expansion conditions were significant. It was difficult to compare linear/compressive and

or linear/expansive conditions because of small effect sizes. More subjects would help differentiate these nuances in the data.

Conclusions

HRTF-generated virtual auditory environments were used to test whether compression or expansion had an impact on spatial hearing in vocoder simulations of SSD-listening. In both the squelch and head-shadow experiments, the linear vocoder provided a listening advantage over monaural. This was especially true in the head-shadow experiment in which bilateral performance was better than monaural in all TMRs tested. Compression disrupted performance in both experiments but to a larger extent in the head-shadow experiment. Expansion caused a slight improvement in performance in the squelch experiment (relative to compression), but showed no significant effect on performance in the head-shadow experiment. Taken together, these results suggest that compression and expansion had two effects on performance: (i) changed the relative loudness of target and maskers and (ii) introduced envelope distortions. The results of the squelch experiment are likely attributed to changes in ILDs, with compression causing target and masker ILDs to become more similar to one another and thereby reducing perceived spatial separation.

In the head-shadow experiment, the results indicate that distortion of speech cues via envelope manipulation likely contributed most to the observed outcomes. Since both compression and expansion disrupted performance, this suggests that envelope distortion likely played a role in the results. Additionally, compression and expansion could have changed the TMR in the vocoded ear. Compression should have reduced the loudness of

the target in the vocoded ear, diminishing performance. Expansion should have amplified the target in the vocoded ear, increasing performance. The compressive function is an aspect of CI programming processing that clinicians have access to control to some extent. Therefore, compression may be an important target for optimization to improve speech perception outcomes for SSD-CI listeners.

Chapter 4: The role of spectral mismatch on perceived binaural fusion in vocoder simulations of SSD-CI listening

Introduction

Binaural hearing improves the ability to hear in noisy, complex environments. For those with SSD (one normal-hearing ear and one deaf ear), this loss of binaural hearing can be challenging. CIs can restore some of the benefits of having two ears and facilitate spatial hearing for SSD listeners in a number of ways, such as improving sound localization and speech perception in noise. Most of the previous work examining spatial hearing in SSD-CI listeners has found a clear head-shadow benefit after implantation, as evidenced by the fact that the CI provides improved speech perception when a target talker is on the deaf side or the noise is on the acoustic side of the head (Arndt et al. 2011; Bernstein et al., 2017; Buechner et al., 2010; Firszt et al., 2012; Hansen et al., 2013; Zeitler et al., 2015;). For NH listeners, the head-shadow effect arises from the physical acoustic barrier created by the head and allows listeners to attend to the ear with the better SNR for the signal of interest. For SSD-CI listeners, this head-shadow benefit arises because the CI allows the listener to take advantage of the fact that the deaf ear has a better SNR. This benefit does not require any binaural computations and arises solely from the physical barrier of the listener's head.

In addition to head-shadow benefits, individuals with two normal-hearing ears also receive an additional advantage that can improve speech perception in noise: binaural squelch. Binaural squelch involves neural computations based on differences in timing and

level across the two ears to reduce the effective amount of masking in situations involving spatially separated sound sources (Drullman & Bronkhorst, 2000). Although most studies of speech perception in noise have not identified a binaural squelch benefit for SSD-CI listeners, a series of recent studies has shown that SSD-CI listeners (Bernstein et al., 2016; 2017) and NH listeners presented with vocoder simulations of SSD-CI listening (Bernstein et al., 2015, 2016; Wess et al., 2017) can benefit from binaural squelch in situations with multiple competing talkers. In particular, binaural squelch is observed for SSD-CI and SSD-vocoder listeners when the target speech and interfering voices are all of the same gender, such that they are difficult to perceptually separate based on monaural cues such as voice pitch and timbre.

The binaural squelch benefit is believed to arise out of binaural fusion ability, such that if listeners can fuse signals across the ears, they will receive a squelch benefit. In the contralateral unmasking paradigm discussed in Chapter 2, listeners received a binaural squelch benefit (relative to monaural listening) when a copy of the interfering speech was added to the vocoded ear. The improvement in performance in this paradigm is thought to occur because the listener can perceptually fuse the maskers presented to the vocoded ear and unprocessed ear. This fusion then led to the perceived spatial separation between target (perceived at the unprocessed ear) and masker (perceived as a diffuse image or in the center of the head), and thereby improved speech intelligibility.

This binaural fusion hypothesis may also explain some of the results from Chapters 2 and 3. We found that certain vocoder distortions reduced the amount of binaural squelch. According to the fusion hypothesis, the introduction of misalignments between the unprocessed ear and vocoded ear impaired the ability to perceptually fuse the diotic speech

signals, thus eliminating the perceived spatial separation of the target and interferer stimuli, which reduced binaural squelch. In particular, spectral and temporal mismatch were found to reduce binaural squelch, with spectral mismatch causing the most detriment to performance (Chapter 2). Envelope compression was also detrimental to contralateral unmasking and head-shadow benefit compared to linearly processed vocoded signals (Chapter 3). However, the results of the compression experiments were interpreted in terms of TMR and ILD distortions and were not obviously attributable to binaural fusion mechanisms. Overall, the results from the CI distortion vocoder experiments implicated spectral mismatch as the largest potential cause of the performance decrease in our experiments, and hence, potentially most deleterious for actual SSD-CI listeners.

Mechanistically, we sought to explain the deleterious effect of spectral mismatch on binaural hearing. This led to the current set of experiments. We hypothesized that the loss of contralateral unmasking benefit after spectral mismatch could be caused by a loss of binaural fusion between the stimuli presented to the ears. Normally, listeners are able to integrate the sounds in their two ears together to hear a single voice, allowing it to be perceptually separated from other voices in a mixture based on spatial differences. However, frequency mismatch could have distorted the vocoded signal to such a degree that the acoustic ear was no longer able to integrate and fuse bilaterally presented signals with the vocoded ear.

The loss of binaural fusion hypothesis is a plausible explanation for the results of the previous experiments. However, the speech intelligibility measure is an indirect test of binaural fusion, and it has not been verified that these listeners have difficulty integrating signals across the ears after interaural misalignment. Additionally, it remains to be

elucidated if interaural alignment promotes fusion in this population. The goal of these experiments was to more directly assess and measure binaural fusion of speech signals in a multi-talker mixture using vocoder simulations of SSD-CI listening. This was accomplished by either asking the listener how many voices they hear (Experiment 4.1) or asking the listener to discriminate between a case with diotic speech (the same signal presented to both ears) and dichotic speech (different signals presented to both ears) presented to the two ears (Experiment 4.2).

Binaural fusion is the subjective experience of the perception of one sound rather than two that occurs when listeners are presented with signals to both ears (Steel, Papsin, & Gordon, 2015). Binaural fusion is essential for hearing in noisy environments and is encountered in nearly every real life listening situation. Integration of information across the ears into a cohesive, specific and continuous percept is an indispensable prerequisite to properly analyze an auditory environment and group sounds into distinct sources.

Despite the importance of fusion for listening in complex auditory scenes, binaural fusion is difficult to measure. The perceptual measurement of binaural fusion is usually accomplished by eliciting subjective reports of how “fused” the inputs to the two ears sound. This can be done using basic signals such as tones or noise, or by presenting speech to both ears (Aronoff et al., 2015; Reiss et al., 2014). For example, dichotic speech tasks involve presenting different verbal stimuli to each ear simultaneously and asking the listener whether the sounds are fused into one auditory image or object (integration) or if they are perceived as two separate sounds (separation). Binaural fusion has been investigated in CI listeners, with a focus on how spectral mismatch affects fusion. Fusion ability appears to be limited in BICI listeners, even with a small degree of spectral

mismatch between their processors. Listeners report unfused auditory images and often perceive multiple auditory images when there should only be one image (Kan et al., 2013). Goupell et al. (2013) also found that spectral mismatch impaired CI listeners' ability to achieve auditory image fusion, and images that were perceived correctly were often lateralized incorrectly after spectral mismatch. Further study of fusion, but in bimodal CI listeners (CI in one ear, hearing aid in the other), came from work by Reiss et al. (2014) who measured very wide fusion ranges in her listeners; these same listeners were more readily able to fuse pitch-matched signals across the ears than BICI users. The ability to pitch-match may have occurred because the bimodal listeners had some acoustic hearing in their hearing-aid ear. Abnormally large fusion ranges could likely lead to interference of speech perception for these listeners. Finally, a vocoder study by Aronoff et al. (2015) examined the effect of CI distortion on binaural fusion in SSD-vocoder listeners by applying either spectral or temporal compression to the vocoded signal. They found that both distortions disrupted fusion, but spectral compression was far more detrimental to binaural fusion.

Most of these previous studies focused on tonal or single electrode stimuli or directly asked about "fusion," which relies on the participants' subjective understanding of the meaning of this terminology.

To be successful, a subjective measurement that asks the listener to report how "fused" the ears sound requires the listener to understand what is meant by the question. Many individuals might not be able to characterize their auditory perceptions at this level of abstraction. Therefore, this study took a different approach to addressing the fusion question by directly asking the listener to report the number of concurrent voices they heard

in a mixture (Experiment 4.1) or by asking listeners to discriminate two mixtures based on how many voices were presented (Experiment 4.2). The idea was that a lack of fusion should lead to an increase in the number of voices reported or poor performance in discriminating a single diotic voice from two dichotic voices. In contrast, complete fusion should lead to an accurate estimate of the number of voices present in a mixture or good performance in discriminating diotic from dichotic voices.

In Experiment 4.1, listeners were asked to report the total number of talkers they heard in the scene (called a “numerosity” judgment). The key condition included was a diotic condition, where the same voice was presented bilaterally to the unprocessed and vocoded ears. This was the “fusion” condition, and it was paired with additional voices in the unprocessed or vocoded ear. The other key conditions were the control (foil) conditions, which were designed to be equivalent to the fusion conditions in all other respects except that the diotic voice was replaced by two different (i.e., dichotic) voices to represent the situation in which listeners were unable to perceptually fuse the diotic voice. If the listener was able to fuse the diotic stimulus, they should have reported it as one voice. Conversely, if the listeners were not able to fuse the diotic stimulus, they should report two voices for that stimulus (one voice in each ear). Listeners were only asked to report the number of voices they heard. They were not asked to recall any of the spoken speech.

Listeners were tested in two vocoder conditions: with a “mismatched” vocoder where speech information was delivered to the wrong cochlear place (as would be expected with a standard CI allocation and incomplete electrode insertion) and with a “matched” vocoder where the frequency content was spectrally aligned in the two ears. This was done to investigate whether spectral mismatch, which was shown to impair contralateral

unmasking in Chapter 2, also affected numerosity judgments of fusion. Moreover, if the mismatched vocoder negatively affected fusion, then the listeners might report the diotic stimulus as two voices. Additionally, conditions were included that just required the listeners to segregate all NH or all vocoded talkers presented either monaurally or bilaterally. These additional conditions were run in an attempt to explain the results from the diotic fusion conditions of interest (Experiment 4.1).

The paradigm used in experiment 4.1 introduced a potential confound relating to perceptual limits of voice counting in listeners. The perceptual limits of voice counting is broadly referred to as the limit in numerosity judgments. Knowing the numerosity limits of the listener's ability to count the number of voices in a scene is principal to understanding perceptual limits in multi-talker environments. A recent study by Kawashima and Sato (2015) investigated the numerosity judgement limit for multiple concurrent talkers. They found that listeners were generally accurate in the range of 3 to 5 voices, and accuracy increased when talkers were spatially separated. Knowing numerosity limits is important to understanding the results from the first fusion experiment in this study (Experiment 4.1).

The paradigm in Experiment 4.1 was ultimately found to not be sensitive to spectral mismatch. This could have occurred because the listeners found it difficult to accurately count the number of voices in the mixture. There was also no feedback provided to guide them to learn what was being asked of them. Experiment 4.2 was designed to ask a similar question about whether listeners were able to fuse two copies of a speech signal presented to each ear. But in this case, listeners were asked to discriminate between "diotic" and "foil" stimuli in a two-alternative forced-choice (2AFC) task. One ear was always

presented with unprocessed speech (acoustic ear) and the number of voices varied from one to six. The other ear was always presented with a vocoded stimulus (the vocoded ear) and only one voice was presented to the vocoded ear at a time. The vocoded speech was either the same voice and the same speech segment as one of the voices in the acoustic ear or was a completely different voice and speech segment. The listener was instructed to pick the interval that contained a “fused” or “stereo” voice (i.e. the interval that contained the same voice presented to the vocoded and acoustic ear). With only a single voice presented to the unprocessed ear, this was a trivial task. However, the task became more difficult with the systematic addition of unprocessed voices to the NH ear. The key question was whether the ability to discriminate between the “fused” and “unfused” mixtures was affected by spectral mismatch. In contrast to Experiment 4.1, this did not require listeners to count voices. Additionally, listeners received correct-answer feedback to train them on the discrimination judgments. We hypothesized that the matched vocoder would give rise to a fused percept and the listener would have an easier time selecting the correct “fusion” interval than with a mismatched vocoder.

Numerosity judgments of binaural fusion: Experiment 4.1.

Study Objectives. The goal of this study was to develop a test of the perceptual binaural fusion of speech stimuli—based on counting or discriminating the number of voices in a mixture—that was sensitive to changes in interaural spectral mismatch. Experiment 4.1A was designed to evaluate how many individual talkers the listener heard in a mixture when one or two of the talkers were presented concurrently to the unprocessed and vocoder ears. A second control experiment (4.1B) evaluated the number of total voices (either unprocessed or vocoded) that listeners could reliably count (i.e. numerosity judgments). Knowing the maximum number of voices that can be counted allowed us to determine whether any lack of difference between conditions could be ascribed to a limit in the number of perceptible voices in the mixture.

Experiment 4.1A presented combinations of one or more concurrent talkers, with each talker in the mixture presented just to the left ear (normal unprocessed speech), just to the right ear (vocoded speech) or diotically to both ears (normal speech in the left ear and vocoded speech in the right ear). Participants listened to a short segment of speech and then reported how many total voices they heard (0-6). In the “matched vocoder” conditions, the vocoder used the same synthesis and analysis filters thereby yielding a match in the cochlear place of stimulation across the ears. In the “standard vocoder” conditions, radiographic insertion depth data taken from Landsberger et al. (2015) were used to approximate the average spectral mismatch between the frequency allocation of the CI electrode array and basilar membrane for a typical CI listener. In cases where the diotic voice was fused, we expected listeners to report the correct number of voices in the mixture.

In cases where the diotic voice was not fused, we expected listeners to report one extra voice because the diotic signal would be perceived as two separate voices. We hypothesized that the standard vocoder would give rise to the unfused perception and the matched vocoder would lead to fusion of the diotic stimulus.

Experiment 4.1B was a control experiment designed to determine the perceptual limits of accurate numerosity judgments for the NH listeners participating in this study. Experiment 4.1B also presented combinations of talkers to one or both ears, but in this case either all of the voices were vocoded or none of the voices were vocoded (i.e., unprocessed). This provided a set of control data that established how many talkers the listeners could count, using the same basic experimental procedure and stimuli.

Experimental Questions. 4.1A) Will a more accurate “place-matched” vocoder mapping facilitate better binaural fusion, leading listeners to better identify the correct number of talkers in a scene over a standard vocoder mapping?

4.1B) How many total talkers (unprocessed or vocoded) can people segregate in an auditory scene?

Hypotheses. 4.1A) A more accurate “place matched” map compared to a standard map will facilitate the listener’s ability to correctly identify the number of talkers in the acoustic scene. 4.1B) Listeners will be better able to identify the correct number of talkers in an acoustic scene when the talkers are presented acoustically. Numerosity judgments

will likely be worse when the talkers are vocoded. Accuracy will diminish in both situations when the number of talkers in the scene increases.

Methods.

Participants. There were 10 paid listeners (age range 18-30) in this experiment. All listeners had NH, defined as symmetrical thresholds equal to or better than 20 dB hearing level at octave frequencies between 125 and 8000 Hz and were free from cognitive and neurological disorders. Listeners were tested at the Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio. The listener panel consisted of professional listeners, in that they are paid to conduct multiple psychoacoustic experiments.

Stimuli. This experiment utilized the CUNY topic sentence corpus (Boothroyd., et al., 1988). The corpus was originally developed using just two different talkers discussing 12 different topic areas such as food, work, family, weather etc. An example sentence would be “The thunder and lightning from the storm last night woke up all of us.” In order to create more than two talkers, the original corpus was modified using Praat software (Boersma & Weenink, 2007) to change the fundamental frequency, the intensity contours and other speech features. A total of 8 voices were used in this study: the 2 original talkers (1 Female and 1 Male) and 3 additional male and 3 additional female talkers that were created based on the original two recordings.

Sentences from the corpus were concatenated by topic area and talker. After combining all of the sentences from a given talker and topic, 90-second paragraphs were created. Two-second samples of the concatenated paragraph were chosen randomly for each talker in each trial.

Procedure. Experiment 4.1: The experiment involved having a listener report how many total talkers they heard in an auditory scene presented over headphones. Trials consisted of combinations of multiple concurrent talkers, each talking for 2 seconds. Each individual talker was presented at a level of 60 dB SPL (acoustic ear) or the matched level to the vocoded ear. The experimental conditions are summarized in Table V below. The experiment was divided into two sets of conditions: the “diotic” conditions of interest and the “foil” conditions. The diotic “fusion” conditions included diotic presentation of the same talker signal, therefore the number of voices heard should depend on the amount of binaural fusion between the ears. The foil conditions are conditions where each voice is only presented to one ear, so there should be no impact of the effect of binaural fusion on the results. The foil conditions served as a control, to examine whether any differences measured in the test conditions can be attributed to differences in the amount of binaural fusion and not to other perceptual differences imposed by the vocoder frequency mapping.

To create the percept of a talker originating from either the left, right or center of the listener’s head, each individual talker was presented to one ear or both ears simultaneously. The key condition was a diotic condition in which the same stimulus was presented simultaneously to both ears. The main question addressed in this experiment was whether or not listeners perceptually fused the diotic signal presented to both ears to hear a single voice. (NH listeners presented with diotic speech perceive a single talker in the center of the head.) The diotic “fusion” conditions were chosen to examine binaural fusion and numerosity judgments with increasing numbers of total talkers. Four sets (A, B, C and D) of conditions (fusion condition plus two foils) were presented (Table VI). The table

provides details about the arrangement of the speech stimuli in each condition. Each “X” denotes one voice. For example, Set A had the fewest number of total voices. The diotic condition consisted of a single voice presented to both the vocoded and unprocessed ears. The listener therefore should report one voice if the stimulus was perceptually fused and two if not. There were two controls included in Set A. The “unfused” foil condition represented a situation where the listener could not fuse two voices across the ears: two different voices which were presented to the vocoded and unprocessed ears. Because different speech segments and voices were presented to the two ears, the listeners should always report two voices. The “fused” foil condition consisted of a single voice presented vocoded to the right ear. This condition served as a control for what listeners would report if they heard the diotic voice in the test condition as a single fused voice. Although Set A was relatively easy with just a single voice in the test condition, Sets B-D introduced additional diotic or monaural voices into the mixture. The configuration of the Sets corresponded to the location of the “fused” voice and the additional acoustic voices added to the mix.

Table V. Experimental conditions for Experiment 4.1A.

Set	Configuration	Condition	Total # talkers	Left ear only (unprocessed)	Right ear only (vocoded)	Diotic
A	Center	Test A	1			X
A		Foil (Unfused)	2	X	X	
A		Foil (Fused)	1		X	
B	Center/Left	Test B	2	X		X
B		Foil (Unfused)	3	XX	X	
B		Foil (Fused)*	2	X	X	
C	2 Center/Left	Test C	3	X		XX
C		Foil (Unfused)	5	XXX	XX	
C		Foil (Fused)	3	X	XX	
D	Center/2 Left	Test D	3	XX		X
D		Foil (Unfused)	4	XXX	X	
D		Foil (Fused)*	3	XX	X	

Table V. Experimental conditions for experiment 4.1A. Each set contains the test (fusion) conditions of interest and two control conditions. The “X” denotes one talker. The unfused foil represents a control condition where the listener is not expected to receive fusion because the voices are presented dichotically. The fused control condition represents a condition of what the test condition might sound like if it was actually fused. (i.e., it would sound like a single vocoded voice).

*Set B fused is repeat of Set A unfused: Was not repeated twice

*Set D fused is same as Set B unfused: Was not repeated twice

The experiment began with a 15 minute training session. The training session was identical to the experimental session, except that only foil conditions were presented (no test conditions) and listeners were provided feedback at the end of each trial. After each trial, a GUI appeared which displayed a sentence asking the listeners, “How many total talkers do you hear?” The listener used the mouse to select the button corresponding to the number of talkers they heard (1-6). During training, the listeners were provided feedback

(via blinking of the numbered button corresponding to the total number of talkers in the trial) about the correct answer. During the main experiment, feedback was not provided, since the goal of the experiment was to measure listeners' subjective impression of the number of voices in the mixture. Experimental conditions were presented randomly, with 10 trials per condition, per experiment.

Participants were seated in a sound booth and directed their attention to a computer screen. The speech stimulus was generated by MATLAB and played via a RME Hammerfall (Haimhausen, Germany) sound card and presented over Sennheiser HD 280 headphones at a comfortable presentation level of 60 dB SPL.

Noise Vocoding: Noise vocoding was used to extract speech envelopes in seven frequency channels and then used the envelopes to excite specified regions of the cochlea (via synthesis filters). First, stimuli were passed through a bank of “analysis” filters, the frequency range of the analysis filters was 100 to 10000 Hz. The envelope of the signal in each channel was extracted via a Half-wave rectification then low-passed filtered at 400 Hz with a second-order Butterworth filter. Each envelope was then multiplied by a white noise carrier, with the resulting signal then passed through a series bandpass “synthesis” filters. The level of the resulting signal in each channel was adjusted to be equal to the RMS level of the input signal for that channel, and the delays associated with the filtering process were removed. Finally, the signals were summed across channels to create the noise-vocoded signal.

Interaural spectral mismatch was introduced through the use of synthesis filters that did not match the analysis filters used to extract the envelope, thereby stimulating a different cochlear place than would be stimulated by an unprocessed acoustic signal. This particular spectral mismatch is different from the spectral mismatch implemented in experiments in Chapter 2 (2.1-2.4). Instead of using a linear spectral shift, this experiment employed a more realistic spectral mismatch that took into account published mismatch measurements for CI listeners. Radiographic insertion depth data from Landsberger et al. (2015) were used to estimate the average mismatch that would occur for an average CI listener. These data were combined with clinical frequency allocations to create the corresponding standard and place-matched vocoder mappings summarized in Table IV. For a typical CI listener, the electrode is fixed and is stimulating a specific place of the cochlea; this cannot be changed. In a vocoder simulation, this is emulated by having a fixed set of synthesis bands, so that the vocoder is always stimulating the same set of locations on the cochlea. For CI users, the electrode array is not fully inserted along the length of the cochlea. Therefore, the basilar membrane can only be stimulated down to the ~400 Hz place in the cochlea. This was represented by the low end of the lowest synthesis band being set to 438 Hz. An audiologist has control over the analysis bands, which dictate which acoustic frequencies get delivered to each band. In our vocoder simulation for the “place-matched” case, the analysis bands are set to equal the synthesis bands, thereby providing a frequency match between analysis and synthesis channels. In the “standard” case, the typical 100-8500 Hz frequency range is input to the available channels, generating a place mismatch. In the standard map, the upper frequency cutoff for the analysis bands was set to 3548 Hz (Table VI). This was done to ensure that extra channels were not added for synthesis filters

above 8500 Hz both for audibility reasons (hard to excite frequencies that high) and to not include extra channels of auditory information. This ensured the two vocoder conditions had the same number of active “electrodes” (i.e., synthesis bands).

Table VI. Frequency allocation for the place-matched and standard vocoder map.

Channel #	Synthesis Bands: (Hz)	Analysis Bands: Place-Matched (Hz)	Analysis Bands: Standard Map (Hz)
1	438 — 576	438 — 576	100 — 237
2	576 — 757	576 — 757	237 — 431
3	757 — 1238	757 — 1238	431 — 710
4	1238 — 2072	1238 — 2072	710 — 1115
5	2072 — 3548	2072 — 3548	1115 — 1707
6	3548 — 5623	3548 — 5623	1707 — 2574
7	5623 — 8500	5623 — 8500	2574 — 3849

Table VI. Frequency allocation for the place-matched and standard vocoder map. The spectral mismatch was created using radiographic insertion depth data from Landsberger et al. (2015) to estimate the average mismatch for a typical CI listener. These data were combined with clinical frequency allocations to create the corresponding standard analysis bands. The synthesis bands remained fixed for both vocoder conditions, since the synthesis bands represent the physical location of the electrode array on the basilar membrane in this vocoder simulation. For the place-matched map, the analysis bands are equal to the synthesis bands. For the standard map, the analysis bands are set as the standard CI frequency allocation.

In control experiment 4.1B, all vocoded stimuli were presented with the place-matched vocoder. The goal of this part of the experiment was to determine how many total voices listeners were able to accurately count in the mixture with vocoded or unprocessed stimuli. Conditions are summarized in Table VII below. The total number of talkers varied from 2-

6, and were either all vocoded or all unprocessed. In the bilateral conditions the total number of talkers was presented roughly evenly between the ears. In the monaural conditions, all the voices were presented to one ear. Experimental conditions were presented randomly, with 10 trials per tracked condition per experiment.

Configuration	Condition	Total # talkers	Voices Left ear	Voices Right ear
Bilateral	Vocoded	2	X	X
Bilateral	Unprocessed	2	X	X
Bilateral	Vocoded	3	XX	X
Bilateral	Unprocessed	3	XX	X
Bilateral	Vocoded	4	XX	XX
Bilateral	Unprocessed	4	XX	XX
Bilateral	Vocoded	5	XXX	XX
Bilateral	Unprocessed	5	XXX	XX
Bilateral	Vocoded	6	XXX	XXX
Bilateral	Unprocessed	6	XXX	XXX
Monaural	Vocoded	2		XX
Monaural	Unprocessed	2	XX	
Monaural	Vocoded	3		XXX
Monaural	Unprocessed	3	XXX	
Monaural	Vocoded	4		XXXX
Monaural	Unprocessed	4	XXXX	
Monaural	Vocoded	5		XXXXX
Monaural	Unprocessed	5	XXXXX	
Monaural	Vocoded	6		XXXXXX
Monaural	Unprocessed	6	XXXXXX	

Table VII. Experimental conditions for numerosity experiment 4.1.B. Stimuli were presented as either unprocessed or vocoded. The “X” denotes one talker. Two through six voices were presented either monaurally or roughly spilt between the left and right ears.

Results.

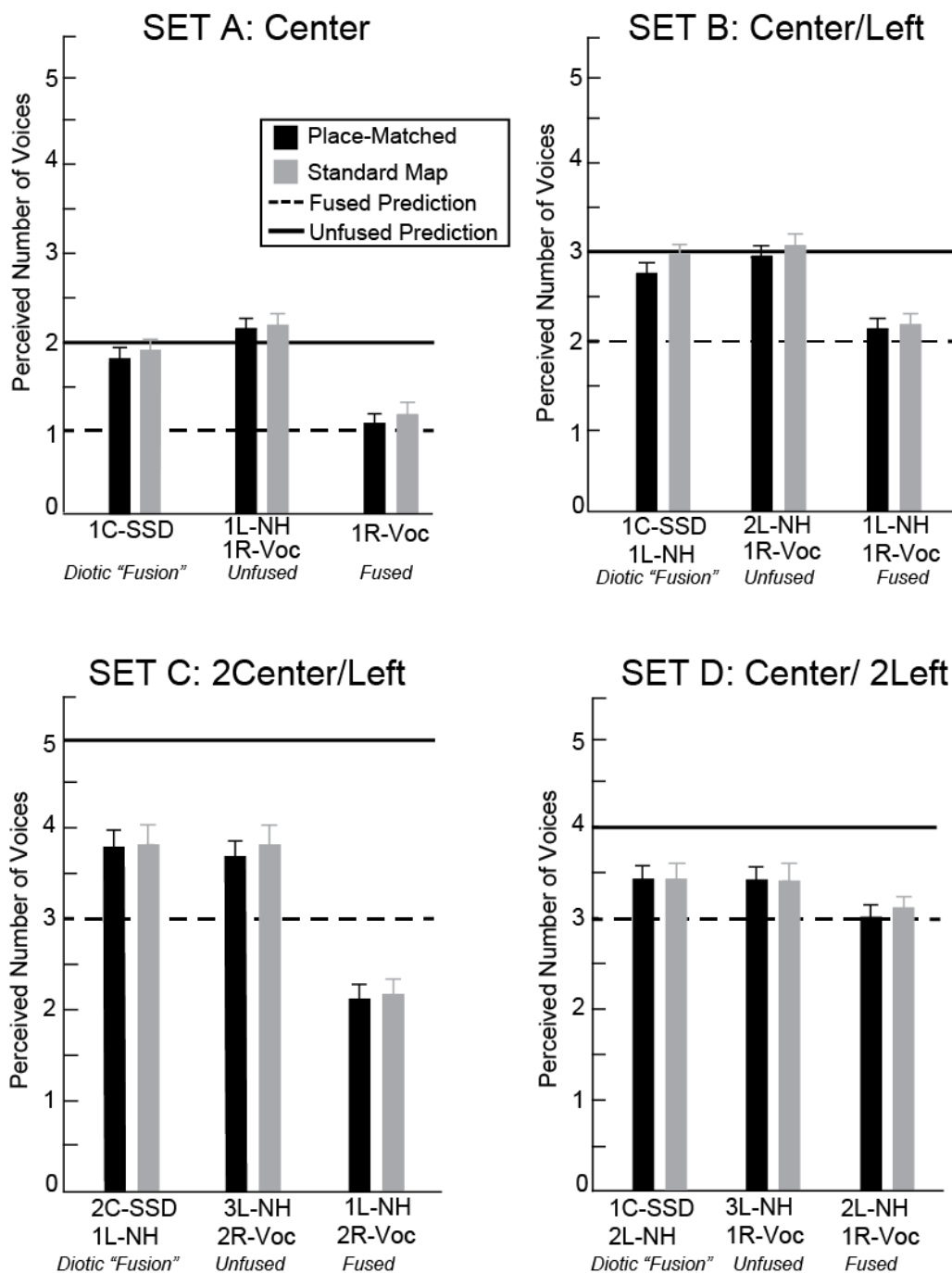


Figure 4.1. Results from Experiment 4.1A: Sets A, B, C and D plotting perceived number of talkers vs experimental condition. The black bar represents the place-matched vocoder, the grey bar represents the standard vocoder map. The dashed line indicates the expected number of voices if the listener is receiving fusion. The solid black line represents the expectation if the listener does not receive fusion. The results indicate that listeners were more likely to report the test condition stimulus as unfused. Error bars represent \pm one standard error of the mean.

The results of Experiment 4.1A are shown in Figure 4.1. Each panel plots the mean perceived number of talkers as a function of listening condition for one set of diotic “fusion” and foil conditions. The two bars in each pair are the different vocoder conditions, the standard (mismatched) map in grey and the place-matched map in black. The first pair of bars in each panel represents the diotic fusion condition which included one or two diotically presented talkers. The two other pairs of bars in each panel represent the two associated foil conditions, one representing an unfused percept and the other representing a fused percept. The horizontal lines in each plot represent the number of talkers the listener would have reported if they had perceived the diotic voice(s) as completely unfused (upper solid line) or completely fused (lower dashed line).

In all four sets, as the number of voices included in the mixture increased, the number of perceived voices increased. The listeners reported more talkers in the unfused controls than in the fused controls. Implementation of the standard vocoder had a marginal effect on the number of perceived talkers. Listeners generally reported a greater number of talkers than were presented, indicating a lack of fusion. This is evident by the fact that the unfused control and test conditions show a similar response. This was true for sets B, C and D. For each of the four sets of conditions, a repeated-measures two-way analysis of variance (ANOVA) was conducted to examine the effects of vocoder condition and experimental condition on the reported number of talkers in the scene. Vocoder condition contained two levels (place-matched vs standard map) and experimental condition contained three levels (test, fused foil and unfused foil).

For Set A, there was a significant main effect of experimental condition [$F= (2, 16) = 64.4, p<0.001$]. Post-hoc tests found differences between the fused foil and the test condition ($p< 0.001$) and between the fused and unfused foil ($p< 0.001$) condition.

For Set B, there was a significant main effect of experimental condition [$F= (2, 16) = 64.4, p<0.001$]. Post hoc tests found differences between the test condition and fused foil condition ($p<0.01$), and between the test condition and the unfused foil conditions ($p<0.01$). There was also a small (2.93 vs. 3.10 voices) but significant difference between the test condition and the unfused foil condition ($p<0.01$). This difference suggests that the listeners might have experienced very slight partial fusion of the diotic voice in this condition.

For Set C, there was a significant main effect of experimental condition [$F= (2, 16) = 55.5, p<0.001$]. Post hoc tests found differences between the unfused foil and the test conditions ($p< 0.001$) and between the unfused and fused foil ($p< 0.001$) conditions.

For Set D, there was a significant main effect of experimental condition [$F= (2, 16) = 26.6, p<0.05$]. Post hoc tests found differences between the unfused foil and the test condition ($p<0.05$) and between the unfused and fused foil ($p< 0.001$) conditions.

The results of these experiments found no effects of vocoder on the perceived number of voices, but strong effects of experimental condition were found. Generally, there were no differences in perceived number of voices between the test condition and unfused foil condition. But listeners reported more voices for the test condition than for the fused foil condition, and more voices for the unfused foil than for the fused foil condition.

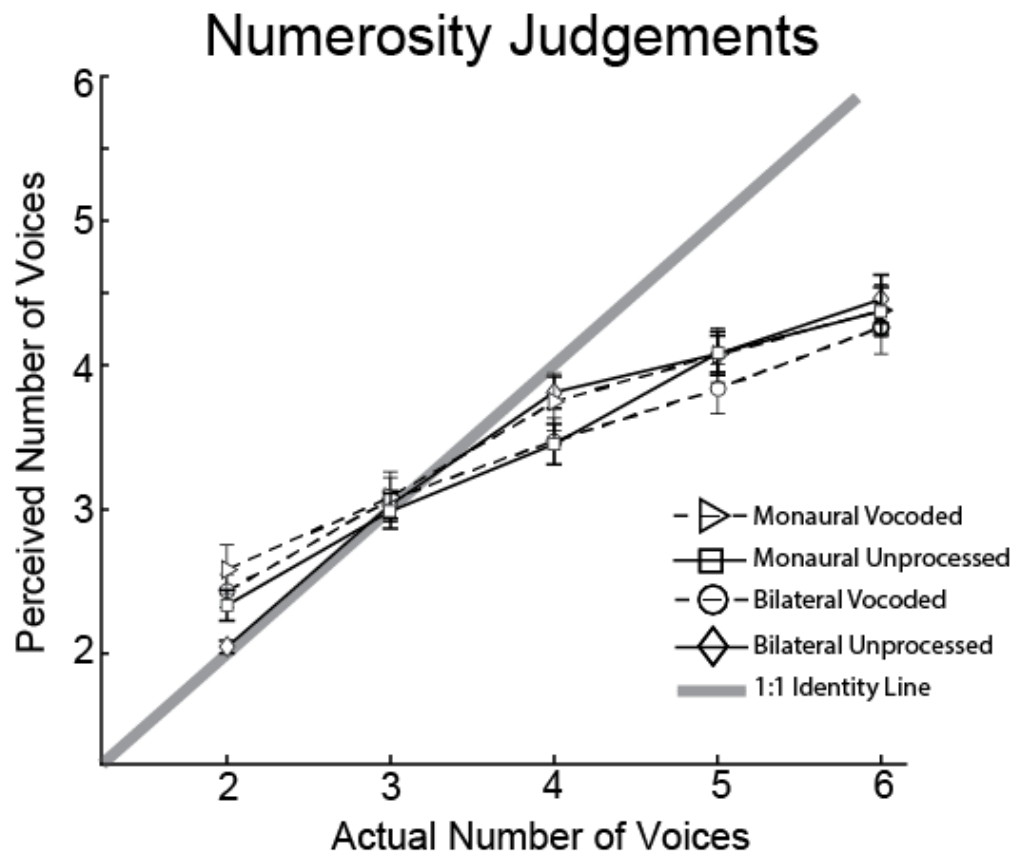


Figure 4.2. Results from the numerosity experiment from Experiment 4.1B. The solid lines represent the unprocessed speech stimuli and the dashed lines indicate the vocoder conditions. The solid grey line is the identity line, which represents if the listener's perceived number of voices matched the actual number of voices. Error bars represent \pm one standard error of the mean.

The results of Experiment 4.1B are shown in Figure 4.2. Figure 4.2 plots the perceived number of talkers as a function of the number of voices in the mixture. For this portion of the experiment, listeners had to count the number of voices all presented to one ear or distributed relatively evenly to both ears. All of the voices in a mixture were unprocessed or they were all vocoded. Only the place-matched vocoder was used for this portion of the experiment.

For the numerosity experiments, a repeated measures three-way ANOVA was conducted to compare the main effects of listening condition (monaural vs bilateral),

processing condition (acoustic vs vocoded) and number of voices (two through six voices), and the interactions between all three variables. There was a significant main effect of number of voices condition [$F(4, 28) = 82.2, p < 0.001$], no other significant main effects or interactions were found. Four main trends are apparent in the data. Just as in Experiment 4.1A, the perceived number of talkers increased with the actual number of talkers. However, listeners tended to report more voices than were actually presented for two talkers. With three total voices, listeners were accurate in all conditions and under reported for four talkers and above... Surprisingly, there was no significant difference in performance between the vocoded and unprocessed conditions. There was also no significant difference between monaural and bilateral presentation. Overall, numerosity judgments were relatively accurate in the two-four talker range, but listeners underestimated the number of talkers when there were more than four talkers in the mixture. Listeners perceived fewer voices than are actually presented when the number of voices increases beyond four. Four voices may be the numerosity limit for unprocessed and vocoded voices, at least in this experiment with limited spatial conditions (monaural or bilateral).

Interim discussion: Experiment 4.1.

The results of experiment 4.1A demonstrated that listeners almost always reported the “diotic” stimuli as two separate voices. In every case except “Set A,” the number of

perceived voices was equal to the predictions based on the “unfused” foil, and substantially greater than predictions for the “fused” foil. In this paradigm, listeners were not fusing the diotic “fusion” condition voices, and the listeners reported the fusion stimulus as two separate voices, instead of one fused voice.

These results are counter to the contralateral unmasking results from Bernstein et al. (2015, 2016) and Wess et al. (2017). Listeners must have been combining the diotic sounds from the ears in some way to achieve the contralateral unmasking advantage, but perhaps they were not fusing the voices to the point where they heard one single sound. This “incomplete fusion” could have been enough to provide a spatial listening advantage but the percept was not fused enough that the listeners freely reported one single fused voice.

Experiment 4.2 took a slightly different approach to asking the question of whether listeners were able to perceive the same stimulus presented to the unprocessed and vocoded ear as a perceptually fused entity. In this experiment, listeners discriminated between two sequential speech mixtures: a diotic mixture (same voice presented to the unprocessed and vocoded ear) and a dichotic mixture (two different voices presented to each ear). This paradigm did not rely on the subjective report of the listener. The idea was that if the listener received even partial fusion from the fusion interval, they should have been able to report the correct interval. Another possible explanation of the results from 4.1 was that there was no feedback provided to the listener. The discrimination paradigm employed in experiment 4.2 allowed for feedback (correct/incorrect) to be provided to the listener after each response.

Discrimination, spectral mismatch and binaural fusion: Experiment 4.2

Study Objectives. The goal of this study was to examine whether a discrimination-based perceptual test of binaural fusion of speech stimuli is sensitive to changes in interaural spectral mismatch. As in Experiment 4.1, vocoder simulations of SSD-CI listening were used to investigate the effect of spectral mismatch on the perception of fusion in NH listeners. A virtual cocktail party was created by presenting combinations of one or more concurrent talkers to the left ear (normal unprocessed speech), to the right ear (vocoded speech) or to both ears (normal speech and vocoded speech).

The experiment was a 2AFC task where the listener was required to identify which interval contained a diotic speech signal. The signal interval contained the same speech waveform, presented unprocessed to the left and vocoded to the right ear (fusion possible). The reference interval presented an unrelated voice to the left and right ears (no fusion possible). With only a single diotic voice in the signal interval, this task was trivial, and most listeners could easily determine which interval contained the diotic voice. The task was made more difficult by systemically adding additional unprocessed voices to the mixture in the left ear.

Experimental question. Are NH listeners presented with vocoder simulations of SSD-CI listening better able to identify the correct interval in which a diotic speech signal is presented with a place-matched vocoder frequency map than with a standard map?

Hypothesis. The hypothesis was that listeners will be more likely to choose the correct fusion interval with a place-matched map. This will facilitate binaural fusion because the matching frequency bands across the ears should result in more interaural correlation between the signals. The place-matched map should facilitate more fusion over the standard map, no matter how many additional voices are added. However, performance could decline in both vocoder conditions with an increasing number of additional unprocessed acoustic voices added to the mixture in the left ear.

Methods.

Participants. There were nine paid listeners (age range 18-30) participating in this experiment. All listeners had NH, defined as symmetrical thresholds equal to or better than 20 dB hearing level (HL) at octave frequencies between 125 and 8000 Hz and were free from cognitive and neurological disorders. Listeners were tested at the Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio. Seven out of the nine listeners who participated in Experiment 4.1 also participated in Experiment 4.2.

Stimuli. The stimuli used in this experiment were the same that were used in Experiment 4.1 (the CUNY topic sentence corpus with 8 different talkers).

Procedure. The experiment used a 2AFC paradigm to assess binaural fusion. The signal interval always contained a 2-second segment of speech that was presented diotically

(unprocessed in one ear, vocoded in the other). The reference interval always contained two different segments of speech produced by two different talkers, saying two different things, one unprocessed, one not. These intervals were similar to the “test” and “fused foil” conditions from Experiment 4.1 (See Figure 4.3 for two example trials). Example trial one is the easiest case, where the listener should have no issue selecting the correct interval. Example trial two is harder and the listener might have more trouble selecting the correct interval.

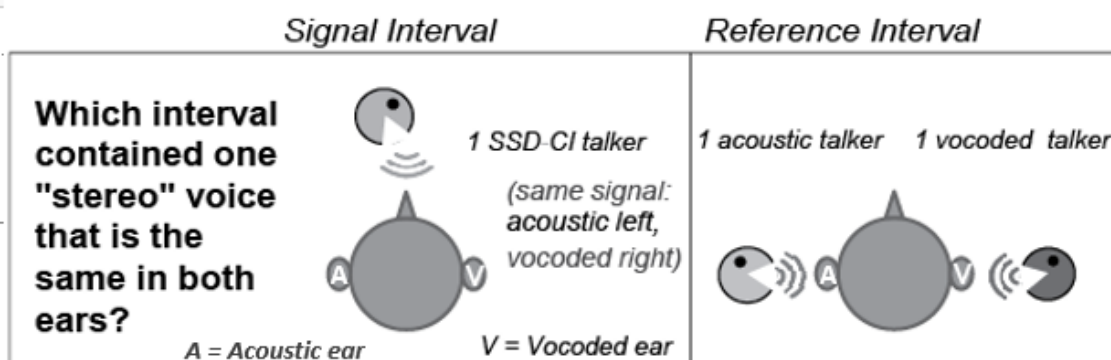
After each trial, a GUI window appeared and the following text appeared “Which interval contained one ‘stereo’ voice that was the same in both ears?” The listener’s task was to identify the interval where one vocoded speech signal in the right ear matched one of the unprocessed speech signals in the left ear. The listener used the computer mouse to select the button corresponding to the first or second interval. Blocks consisted of combinations of multiple concurrent talkers. The listeners were provided feedback (via blinking of the numbered button corresponding to the interval that contained the diotic stimulus).

The training portion of the experiment contained three multi-talker conditions (one, two or three voices in the acoustic ear) and two vocoder conditions (standard vs place-matched). The six combinations in the training were presented randomly in each block. The vocoder condition was fixed for each block. Listeners were presented with 30 trials per block and 20 trials for each combination of number of talkers and vocoder condition, for a total of 120 trials for the training portion of the experiment.

The experimental portion of the experiment consisted of six talker conditions (one, two, three, four, five and seven voices in the acoustic ear) and two vocoder conditions

(standard vs place-matched). The 12 conditions in the experimental blocks were presented randomly. The vocoder condition was held fixed for each block. Listeners were presented with 100 trials per block and 20 trials per tracked condition, for a total of 240 trials for the experimental portion of the experiment.

Example trial 1: Easiest case, compare 1 vs. 2 talkers



Example trial 2: Harder case, compare 3 vs. 4 talkers

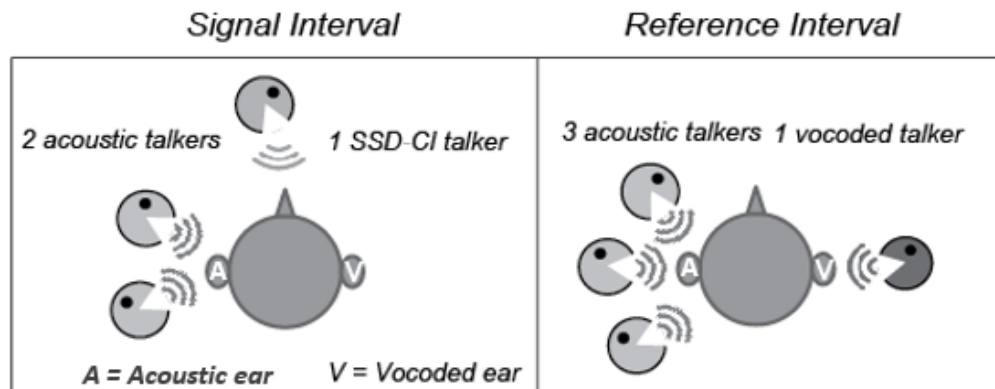


Figure 4.3. Schematic of possible perception for two example trials from experiment 4.2. Example 1 depicts the easiest scenario, where the signal interval should be easily distinguishable from the reference interval. In the signal interval, a diotic voice is presented to the listener, with the same speech segment spoken by the same talker presented unprocessed to the left ear and vocoded to the right ear. In the reference interval, two different speech segments produced by two different talkers' voices are presented to the vocoded and unprocessed ears. Example 2 is similar to example 1, except the task is made more difficult by presenting two additional unprocessed speech segments to the left ear in both intervals.

Procedure. Participants were seated in a sound booth and directed their attention to a computer screen. The speech stimuli were generated by MATLAB and played via an RME Hammerfall (Haimhausen, Germany) sound card and presented over Sennheiser HD 280 headphones at a comfortable presentation level of 60 dB SPL. The RMS was set fixed to 60 dB regardless of the number of talkers in the condition.

Results.

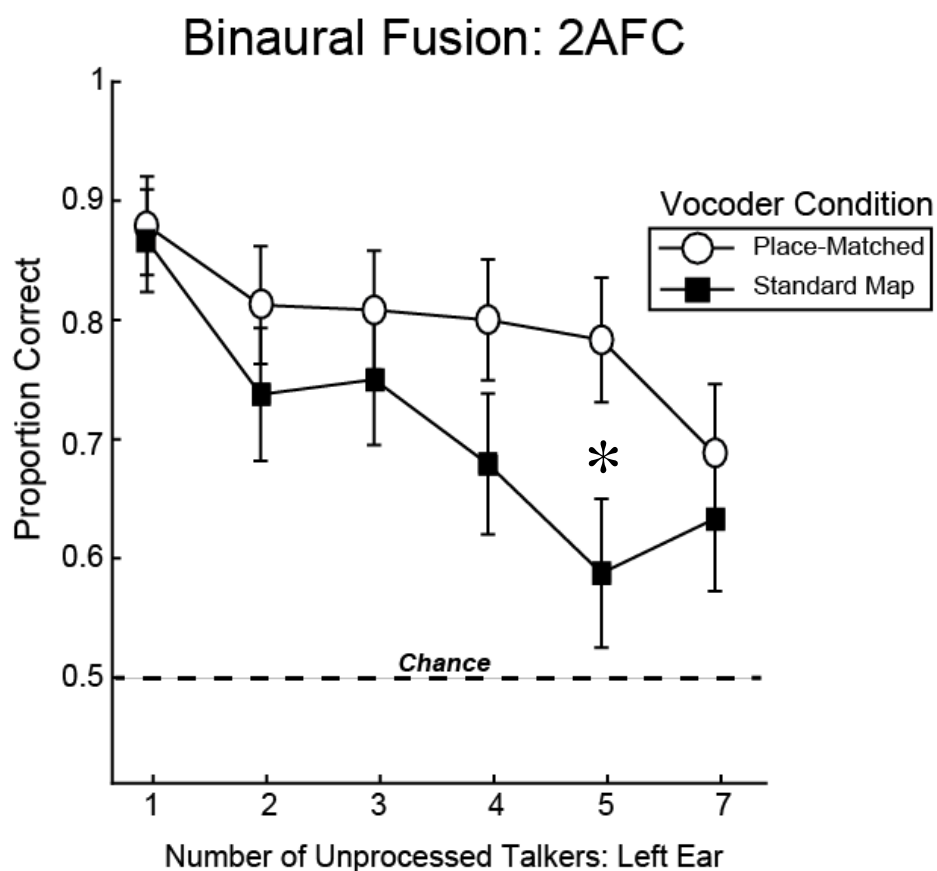


Figure 4.4. Results from fusion experiment 4.2. The dashed line indicated chance performance. The place-matched vocoder is represented by the white circles. The standard condition is represented by the black squares. Taken as a whole the place-matched vocoder yielded a higher proportion of correct answers relative to the standard vocoder. Post-hoc analysis revealed listeners were significantly more likely to guess the signal interval with 5 talkers in the left ear with the place-matched vocoder over the standard vocoder. These results indicate better

performance when there was more of a spectral match between the diotic stimuli presented to each ear. Error bars represent \pm one standard error of the mean.

The results of Experiment 4.2 are shown in Figure 4.4. The mean percentage correct in identifying the interval containing a diotic voice is plotted as a function of the number of unprocessed talkers presented to the left ear. With only a single talker, performance was very high regardless of the vocoder type. With an increasing number of unprocessed talkers in the left ear, performance decreased, but more rapidly for the standard (mismatched) vocoder frequency map than for the matched vocoder. A binary-logistic regression analysis revealed significant main effects of vocoder condition [$\chi^2(1) = 10.40$ $p < 0.001$] and number of talkers [$\chi^2(5) = 89.89$ $p < 0.001$] and a significant two-way interaction between vocoder condition and number of talkers [$\chi^2(5) = 205.87$ $p < 0.001$]. Post-hoc tests were performed to determine the difference between the vocoder conditions at each number of talkers condition. After Bonferroni corrections for 6 comparisons, the vocoder conditions were significantly different from each other only in the case of 5 talkers ($p = 0.002$).

Discussion

The goal of this series of experiments was to further explore the possible mechanisms behind the contralateral unmasking results from experiments in Chapters 2 and 3 in terms of fusion and object formation. In Experiment 2.1 (Chapter 2) the contralateral unmasking benefit for NH listeners presented with vocoder simulations of

SSD-CI listening disappeared after a modest spectral mismatch of 3.6 mm or more. These results led to the current set of experiments whose goal was to determine if the loss of squelch benefit after frequency mismatch in experiment 2.1 could be attributed to disruption of object formation (forming discrete auditory percepts) and binaural fusion. Binaural fusion was assessed by measuring how many voices listeners heard in a mixture. This was measured in two ways, either by having the listeners count voices or discriminate between diotic and dichotic voices in the vocoder simulations. In Experiment 4.1, listeners reported a number of voices that indicated they were not fusing the diotic stimulus (the same speech segment presented unprocessed to one ear and vocoded to the other), regardless of the vocoder condition. This result indicated the listeners had difficulty fusing an identical acoustic stimulus with the vocoded one. The goal of Experiment 4.2 was to examine the effect of a realistic spectral mismatch on discrimination of a diotic fusion stimulus from a dichotic reference stimulus. More specifically, the aim was to determine if listeners were more likely to achieve successful binaural fusion between their acoustic ear and vocoded ear when the vocoded ear's frequency allocation more closely matched their acoustic ear. In contrast to the lack of an effect of vocoder mismatch in Experiment 4.1, Experiment 4.2 demonstrated that listeners were more likely to correctly select the interval containing the diotic stimulus with a place-matched vocoder mapping than with a mapping that was based on a standard CI frequency map. The contrasting results from Experiments 4.1 and 4.2 lead to several interesting interpretations about fusion and spectral mismatch. When simply asking listeners to count the number of voices in an acoustic scene, there was no difference between vocoder conditions (Experiment 4.1). However, when listeners were asked to discriminate between diotic fused and non-fused intervals, performance was

sensitive to spectral mismatch, i.e., the listeners performed worse with the mismatched vocoder (Experiment 4.2). One possible interpretation of these contrasting results is that listeners were experiencing partial or incomplete fusion. On one hand, the diotic stimulus might not have been sufficiently fused for a listener to identify it as one voice instead of two when asked for a free answer (Experiment 4.1). On the other hand, the diotic signal might have been sufficiently fused for listeners to detect interaural coherence in the signals (Experiment 4.2). These results suggest that traditional subjective measures of fusion might be less sensitive to changes in interaural coherence than an objective discrimination task. Another important difference between the current study and previous studies of fusion in real or simulated CI listeners is that these experiments involved stimuli that included extra voices in the mixture. Most studies of binaural fusion involve only a single sound presented to each ear such as a tone or noise, or a segment of speech presented to each ear. The complex mixture of extra voices included with the fusion stimulus in the current studies likely required the use of stream segregation and object formation in addition to binaural fusion. By evaluating fusion in the context of a complex mixture of voices, this study revealed the negative effects of the spectrally mismatched vocoder on binaural hearing processes. This result was also not apparent in the simple condition (i.e., one voice) but emerged when a complex mixture of voices was presented to the listener.

Impacts of spectral mismatch.

The results of these experiments are in agreement with other recent binaural fusion data in CI and vocoder listeners. Aronoff et al. (2015) examined the effect of spectral compression on binaural fusion for NH listeners listening to vocoder simulations where one ear was spectrally mismatched relative to the other. They also concluded that spectral mismatch resulted in significantly less fusion. The test of fusion in Aronoff et al. (2015) was a basic subjective test of fusion, where they simply asked if the listeners heard the same sound in both ears or a different sound. The conclusions of the Aronoff study were similar to the conclusions in this study although the methodology used herein was more quantitative. Determining the effect of frequency distortion on discrimination is a more objective way to measure the functional limitations of a spectral mismatch for a SSD-CI listener. This is because discrimination is a pivotal step in segregating different voices in an acoustic scene. Goupell et al. (2013) examined the effect of spectral mismatch on binaural fusion in vocoder stimulations of BICI listeners. Goupell et al. (2013) measured fusion by varying the spectral mismatch and measuring the perceived image location on a GUI that displayed a face, which the listener could click on. The researchers predicted fused stimuli would cause the listeners to choose a location near the center of the face, and partially fused or unfused stimuli would cause the listener to choose a location that was diffuse or off center. They found that as spectral mismatch increased between the ears, so did the likelihood that the listeners would report more than one auditory image and that the perception was biased towards the ear where the stimulus had a higher carrier frequency. In a follow up study conducted in BICI listeners, Kan et al. (2013) performed the same experiment but

controlled spectral mismatch by selecting single electrode pairs to present the bilateral stimuli. They found similar results: as mismatch between electrodes increased so did the propensity for the listeners to report multiple auditory images. The reduced binaural fusion that occurred after processing with the standard mismatched vocoder is consistent with other studies examining the effects of frequency mismatch on binaural processing.

An interesting result from this current study was that the difference between vocoder conditions was only revealed as the number of concurrent talkers in the mixture was increased. Therefore, when many multiple talkers are present, a spectrally matched vocoder was critical for achieving fusion. In this current study, successful selection of the fusion interval depended on the listener's ability to combine information from their vocoded ear and the acoustic ear to create a single auditory object (i.e. the fused voice).

Ma et al. (2016) examined the role of frequency mismatch on binaural integration in vocoder simulations of SSD listening. They found that perception of speech presented bilaterally was better than speech presented monaurally, but that this effect was largely reduced by a spectral mismatch. However, a caveat of their work was that they presented target speech to both ears, which prevented the results from being interpreted purely in terms of binaural squelch. Taken together, it is clear that spectral mismatch is detrimental to binaural fusion and binaural processing in general. Fortunately for CI listeners, spectral mismatch can be clinically addressed with current technology and slight adjustments of mapping procedures.

Disruption of temporal processing.

In the current experiment, loss of fusion sensitivity after spectral mismatch could have disrupted temporal grouping cues needed for binaural fusion, which could have contributed to the listeners' poor performance in the standard vocoder mapping case. Accurate auditory grouping is an integral step for fusion of binaural stimuli. Auditory grouping refers to the processes of breaking down a complete auditory scene into its constituent components, or auditory objects and then connecting these objects together into streams. Spatial cues are powerful grouping cues when other more salient cues are unavailable, such as pitch cues which are mostly absent in CI listeners and vocoder listeners alike. Grouping based on temporal coherence is much more likely to occur when there is spectral overlap between binaural stimuli, since it provides continuity to the listener (Shamma, Elhilali, & Micheyl, 2011). Temporal coherence is an integral first step for auditory grouping. This is because linking dynamic, rapidly changing speech into the appropriate auditory object will enable the listener to stream said object (such as a target talker in a backdrop of competing talkers). Related to coherence is temporal integration which refers to the neural process of integrating sounds in a certain temporal window. Temporal integration has been shown to be negatively impacted by spectral mismatch in CI listeners (Poon et al., 2009). This could be due to miscalculations in anatomical areas that process both spectral and temporal information, namely the brainstem. Moreover, the tonotopic nature of the auditory system has been shown to extend to the auditory cortex and association areas. It is probable that binaural processing could be affected by spectral mismatches originating in the periphery and further propagated to higher areas in the

auditory system. Therefore diminished temporal grouping might have played a role in these experiments.

An additional way to think about the decline in performance with the standard mapping condition is in terms of envelope correlation between the ears. Spectral mismatch might have disrupted the interaural envelope correlation between the vocoded signal and the acoustic signal. Correlated envelope information has been shown to facilitate auditory object formation and binaural fusion (Carrell & Opie, 1992). Frequency mismatch could shift the envelope to higher frequencies, thus lowering interaural correlation and inhibiting fusion.

Implications for SSD-CI listeners.

In this study, our listeners were more likely to pick the correct fusion interval with a place-matched map over a standard map. Not only was spectral mismatch shown to be detrimental to binaural fusion in this study, spectral mismatch caused the largest decrease in performance on contralateral unmasking in the experiments in Chapter 2 (2.1, 2.3 and 2.4). Currently, CIs are programmed for the profoundly deaf with little or no attention given to binaural hearing ability after CI implantation. However, for those with SSD-CI, a more place-matched map based on the electrode location on the basilar membrane could be implemented by an audiologist. This type of change is readily realizable from a technological standpoint, since it would only require a shift in the speech-processor frequency-to-electrode allocation table (similar to the change in vocoder analysis filters in the current study, Table VII). However, an accurate interaural frequency match is not trivial

to accomplish, because it requires knowledge regarding the characteristic frequencies of the auditory nerve fibers being stimulated by each electrode in the array. This could be accomplished in a number of ways. CT scans (Noble et al., 2014) or radiographs (Landsberger et al., 2015) could be used to estimate the insertion angles of individual electrodes. In fact, the radiographic data from Landsberger et al. (2015) was used to generate an estimated average spectral mismatch map and corresponding place-matched map in this study. Individualized CT scans after implantation could give clinicians a good approximation of the various electrode locations for their patient and this information can help guide an individualized place-matched remapping protocol. However, these CT scans would not inform the audiologist about the best frequency of the neurons located below the individual electrodes. Psychoacoustic methods might be used to try to determine electrode location. However, many psychoacoustic measures are very laborious and long, and thus not clinically feasible.

The 2AFC paradigm utilized in Experiment 4.2 could potentially be used to determine efficacy of new “place-matched” maps for CI listeners. The speech task used in Experiment 4.2 is unique in that it does not rely on speech intelligibility. This is important because speech intelligibility with a CI is plastic and changes over time with adaptation, which makes intelligibility measures difficult to use as an acute tool for comparing maps. Using the paradigm introduced in Experiment 4.2, an audiologist could potentially fit a listener with a few different maps and determine which map is best for binaural fusion. This would eliminate the added confound of reduced intelligibility that might occur after acute map changes. Additionally, Experiment 4.2 is a good candidate for a clinical tool to build new maps, because relative to the above psychophysical techniques, it is fast and

easy for subjects to understand, and sensitive to mismatch. However, it still remains to be elucidated whether binaural fusion is solely brainstem-mediated or if the cortex plays a role. If fusion does instead involve cortical-mediated processing, then listeners could possibly adapt to their mismatched maps and achieve partial fusion without remapping (Svirsky et al., 2004). However, adaptation might occur faster if interaural frequency alignment occurs right at implant activation (Svirsky, Talavage, Sinha, Neuburger, & Azadpour, 2015). Regardless, there are many potential avenues for clinicians and researchers alike to potentially reduce the hearing impacts of spectral mismatch for SSD-CI listeners and BICI listeners alike.

Study Limitations.

The current study utilized vocoder simulation of SSD-CI listening. Although valuable for studying the various effects of CI processing on auditory perception, vocoder simulations are an imperfect approximation of what actual CI listeners hear (Freyman et al., 2008; Li & Loizou, 2009). Duration of deafness, spiral ganglion nerve survival, amount of time after implantation, listeners' age and electrode placement and programming can all impact CI listeners' outcome after implantation. Vocoder simulations allow for the reduction of these confounds and enable researchers to study specific aspects of CI processing on auditory perception, without all of the potential confounds.

The use of multiple different talkers was necessary for these experiments. The CUNY topic sentence corpus only contained two original talkers, the additional six were created using Praat synthesis techniques. However, this strategy of creating multiple might

not have been ideal for these experiments. One of the features of Praat is that it can change the fundamental frequencies of the talkers. In order to implement the spectral mismatch in these studies, the new talkers were created by shifting the fundamental frequency of the original two talkers down in frequency. This method could have created new talkers that still sounded like the original talkers. However, this was likely not a significant confound in this study, since speech segments and talkers were all randomized. The odds that the same talker was presented saying the same thing in a trial was extremely low. A related issue with this approach is that Praat software alters the formant frequencies (shifted up or down), which could have reduced the difference between one of the original voices and a 'new' voice; basically the concern was that the Praat software shifts the voice and the vocoder shifts it back again. Because of the randomness of the speech samples, however, this is also likely of minor significance. Despite this limitation, in future fusion experiments, a corpus with multiple (actually) different talkers should be used.

A potential issue with Experiment 4.2 relates to the possible cues that listeners might have used to complete the task. This experiment was designed to measure binaural fusion. However, the listeners could have possibly been using additional cues to help them complete the experiments. In particular since the fusion interval contained a talker saying the same words in both ears, the listeners could have potentially been monitoring each ear for common words to complete the task. This strategy would likely only work when the total number of talkers was low, since individual words in the mixture would become more difficult to understand when the total number of words presented is high (as would occur with multiple talkers). Additional research could investigate whether listeners could have used this potential alternative cue. For instance, this fusion experiment can be manipulated

such that a listener must choose the interval where the same person is saying the same sentence in both ears (fusion interval) rather than choose the interval where *different* people are saying the same sentence in both ears. This alternative technique would isolate the fusion cue from any potential word cue strategy and remove this potential confound from the experiment.

An important shortcoming of the vocoder approach is that actual CI listeners have time to adapt to their mismatched frequency maps, whereas vocoder listeners do not, at least in these acute experiments. A study by Siciliano et al. (2010) found that even after 10 plus hours of training with a unilateral frequency shifted vocoder, listeners received no binaural benefits, i.e. they could not learn the shifted map. This indicates a constraint on the limits of binaural plasticity, at least in NH vocoder listeners. There is some limited evidence that post-lingually deafened monaural CI listeners and bimodal users can adapt to this frequency mismatch between their implanted ear and their acoustic ear (Svirsky et al. 2004; Reiss et al. 2007). However, this plasticity is likely to be incomplete, and these listeners would probably benefit from a more aligned spectral mapping, especially as it pertains to binaural hearing. With regards to those with SSD, plasticity mechanisms could theoretically overcome some of the limitations imposed by spectral mismatch on binaural fusion. This would be more likely if the fusion percept is occurring at the cortical level. However, most of the evidence points to fusion occurring at a lower level of the auditory pathway where envelope coherence sensitivity is highly sensitive to frequency mismatch (Buss et al., 2009).

Conclusions

The present set of experiments examined how spectral mismatch in vocoder simulations of SSD-CI listening affected the ability of listeners to fuse binaural stimuli and form auditory objects. Experiments 4.1, A and B were designed to measure fusion in the context of the formation of perceptual objects in a multi-talker environment. The results of these experiments indicated that neither vocoder provided enough fusion cues for the listener to report the diotic stimuli as one voice. This could have occurred because the listeners were achieving partial fusion and the percept was not strong enough for the listeners to report the fusion stimulus as one voice. Experiment 4.2 measured a listener's ability to choose the interval that contained a fused voice (signal interval) from a reference interval that contained two different voices (1 vocoded, 1 NH). The 2AFC task in experiment 4.2 was much more sensitive to spectral manipulation. If the listeners in this study were achieving partial fusion, it could have been enough to discriminate between the two intervals in Experiment 4.2, but not enough for the listeners to report one voice in Experiment 4.1. Similarly, the previous results of spectral mismatch on contralateral unmasking probably reflect some degree of fusion – spectral mismatch leads to enough fusion to identify when the sounds from the two ears originate from the same source (Experiment 4.2), and gives rise to binaural squelch under some conditions (Bernstein et al., 2015; 2016; 2017; Wess et al., 2017). This percept did not provide enough fusion to sound like one voice (this study) or to produce binaural squelch in non-informational masking conditions (Bernstein et al. 2015; 2016).

Taken together, these results suggest that a typical mismatch associated with the average insertion angle of the CI electrode array may have a substantial effect on the ability

to perceptually fuse a diotic speech signal in the acoustic and CI ear for SSD-CI listeners, and limit the ability to correctly parse the auditory scene. Still, the place-matched vocoder did not provide listeners with enough interaural coherence to achieve full fusion of the diotic stimuli, indicating that even under ideal circumstances, a crude vocoder signal (or CI signal) might yield, at best, a partially fused percept. Overall, the fusion paradigm in Experiment 4.2 was sensitive to interaural mismatch and was relatively easy for participants to understand. This makes the measurement technique developed for Experiment 4.2 a potentially useful clinical tool to determine optimal frequency mapping or to evaluate outcomes of binaural integration for SSD-CI users. What is clear from the data is that for a SSD-CI listener to have the best chance of binaural hearing with their implant, steps need to be taken to reduce spectral mismatch between their acoustic ear and their CI.

Chapter 5: Summary of dissertation and general discussion

The goal of this dissertation was to examine the effect of common CI distortions on binaural hearing in vocoder simulations of SSD-CI listening. Individuals with SSD are at a severe disadvantage when it comes to listening in noisy environments due to lack of binaural hearing. This dissertation was primarily concerned with how CI distortions affect binaural squelch and perceptual fusion. Vocoder simulations in NH listeners enabled the selective manipulation of certain aspects of CI processing as a first step in determining how these distortions might disrupt binaural hearing for SSD-CI listeners. This work was motivated in part based on results from Bernstein et al. (2015, 2016). In these previous studies, SSD-CI listeners were found to receive a contralateral unmasking benefit from their implant. In this dissertation, contralateral unmasking is defined as the improvement in speech perception associated with adding the interfering voices to the ear contralateral to the target speech. SSD-CI listeners demonstrated highly variable performance, poorer than for vocoder listeners performing the same task, and some SSD-CI listeners did not receive a binaural benefit at all (Bernstein et al., 2016). This may be explained by a key difference between SSD-CI listeners and SSD-vocoder listeners—that is, vocoder simulations in NH listeners do not fully capture the distortion and functional limitations of a CI. Nevertheless, experimental manipulations that create the effect of these distortions are much more easily controlled in vocoder simulations of SSD-CI listeners; therefore, the entirety of this dissertation was conducted using vocoder simulations. Using these manipulations, this dissertation addressed questions to examine the relative importance of specific sources of the variability in binaural hearing outcomes for SSD-CI listeners. The

results of these dissertation studies identified the dimensions that should be studied and manipulated in actual CI listeners to see if binaural outcomes can be improved, and ultimately enable clinicians to make better programming choices for SSD-CI listeners.

A series of experiments tested the effects of frequency mismatch, temporal disparities and amplitude compression on the ability to binaurally integrate unprocessed speech in one ear and vocoded speech in the other. The over-arching goal was to better understand binaural perception of speech in the presence of interfering talkers. To elucidate some of the effects of CI distortions on contralateral unmasking, the experiments in Chapter 2 manipulated three variables related to CI processing: interaural temporal and spectral mismatch and spectral resolution. Spectral mismatch was chosen as the first variable to study since a SSD-CI listener will typically possess a frequency mismatch between their acoustic ear and implant. This is because (i) the implant array is not fully inserted into the cochlea and (ii) an implant is normally programmed between 100-8500 Hz to cover the most important frequencies for speech perception. Spectral mismatch was applied by linearly shifting the vocoded signal up and down in frequency in the range of 1.8 - 7.4 mm. The results from Experiment 2.1 found that contralateral unmasking was completely eliminated with a negligible mismatch of 4-6 ERBs (3.6 - 5.4 mm). This is at the low end of the expected spectral mismatch for an average CI listener.

The next distortion examined was temporal mismatch. Although CI manufacturers do not have a uniform delay in their processors, on average the speech processor in a CI is ~10 ms slower than the traveling wave latency in an acoustic ear. This is because a CI contains some delay that is associated with the speech processor. A range of temporal delays was applied to the vocoded speech presented to one ear. We found that contralateral

unmasking was not negatively affected by timing differences between the vocoder and acoustic ear up until about 24 ms, which is well beyond the timing discrepancy that would occur in actual SSD-CI listeners (Experiment 2.2). Thus, the findings suggest that the interaural temporal mismatch is most likely not an important source contributing to the limited binaural unmasking observed in SSD-CI listeners.

Next the effect of spectral mismatch was examined along with changes in spectral resolution. This was done because CIs have reduced spectral resolution and, due to current spread, only have about 8 functional channels at a time. The interaction between spectral mismatch and spectral resolution of the vocoder was examined by implementing a vocoder with either 3, 5, 8 or 10 channels, while systemically shifting the frequency allocation of the vocoder. Spectral resolution only affected performance when it accompanied a frequency mismatch, such that performance was more robust to spectral mismatch when the resolution of the vocoder was reduced (Experiment 2.3). This was a somewhat surprising result, the interpretation is that a lower number of frequency channels made the listener more immune to spectral shifts. In other words, broader channels allowed for more interaural correlation between the ears. Even after frequency manipulation, some interaural correlation was perceived. When frequency resolution was high, even a small change in frequency would cause the bands in both ears to become decorrelated.

Finally, the interaction between temporal and spectral resolution was also examined, since both distortions are likely to coexist in a typical CI listener. The results from the temporal-spectral mismatch interaction experiment found that performance was best when the signals were aligned in frequency and in latency. In cases where a mismatch was present in one dimension, the additional mismatch did not further disrupt performance

(Experiment 2.4). The distortions in Experiment 2.4 were not additive as expected. This result was encouraging because CI listeners will likely have both a temporal and spectral mismatch between their CI and acoustic ear. Taken together, the results of Chapter 2 indicated that spectral mismatch was by far the largest disruptor to binaural squelch.

Chapter 3 examined the effect of amplitude compression and expansion on head-shadow benefit and squelch. CI listeners have fewer discriminable intensity steps than are available to NH listeners (Nelson et al., 1996). Additionally, the large DR enjoyed by NH listeners is dramatically reduced for CI listeners, so compression must be applied to the signals in order to deliver a wide range of amplitudes into a much smaller range. Chapter 3 used HRTFs (horizontal spatial cues provided to the listener) to examine two spatial configurations to study the effects of compression on binaural squelch (3.1) and on head-shadow benefit (3.2). The effect of expansion was also examined to determine if the opposite distortion (i.e. exaggerating the amplitude of the signal) could enhance performance in this spatial listening task. Compression was shown to have a negative effect on head-shadow benefit and binaural squelch. The results of Chapter 3 indicate that compression likely reduced ILD cues in the squelch experiment and reduced the effective TMRs in the head-shadow experiment, which reduced perceived spatial separation of the target and maskers. A direct comparison between the results of Experiment 2.1 (spectral mismatch and contralateral unmasking) and the experiments in Chapter 3 is difficult. The paradigms used in Chapter 2 and 3 are different, with the experiments in Chapter 3 providing spatial cues to the listeners. Compression disrupted performance in both spatial conditions in Chapter 3, having a larger detrimental effect in the head-shadow case. In the head-shadow experiment (Experiment 3.2), the listener was required to listen to the

vocoded speech to adequately perform the task. Therefore, any further manipulation to the vocoded signal (compression/expansion) potentially corrupted the speech and reduced intelligibility. In contrast, the contralateral unmasking experiment required the listener to primarily attend to the acoustic ear and ignore the vocoded ear. Based on that distinction, it's acceptable to conclude that the mechanisms pertaining to the effects of compression and spectral mismatch on binaural hearing are different and spectral mismatch might be more detrimental to binaural hearing.

Finally, the experiments in Chapter 4 sought to determine if the decrease in contralateral unmasking (measured in chapter 2) after spectral mismatch was related to a loss of binaural fusion ability. The results of this dissertation implicated spectral mismatch as causing a large hindrance to contralateral unmasking. An interpretation of these results from the contralateral unmasking experiments was that frequency mismatch disrupted binaural fusion between the signals in the ears. In other words, listeners might not have been able to use spatial cues to perceptually pull the target talker from the maskers, if the maskers were not perceived as distinct, separate fused voices. Chapter 4 more directly investigated whether listeners could integrate signals between their two ears to hear a single voice in the context of multiple interfering talkers. Binaural fusion ability was examined in two different experiments. The first experiment examined numerosity judgments and the second examined binaural fusion in a discrimination task. These two experiments produced divergent results. When listeners were asked to freely report the number of voices they heard, their responses suggested that they always reported the diotic stimulus as unfused, with no effect of vocoder mismatch. On the other hand, when listeners were asked to discriminate between a diotic fusion interval and a non-fusion interval, they performed

significantly better with the place-matched vocoder than the standard vocoder. Additionally, the listeners were better able to determine when there was a stereo voice present in the mixture with the place-matched vocoder. The interpretation of these studies is that the listeners might have been achieving partial fusion. This partial fusion was enough for the listeners to identify the correct fusion interval (Experiment 4.2) but not enough to report the diotic signal as one voice (Experiment 4.1). The idea behind this is that interaural frequency alignment facilitates identification of the correct fusion interval (Experiment 4.2) and enables listeners to receive a binaural benefit to better understand speech in noise i.e., receive partial fusion to facilitate contralateral unmasking (Experiment 2.1).

General Discussion.

Despite the negative impact CI distortions have on spatial hearing, many SSD-CI listeners receive binaural benefits such as squelch from their implant. This seems to contradict the results of this dissertation, which would predict that typical CI programming would greatly reduce or eliminate binaural benefit for SSD-CI listeners. In our vocoder simulations, level and frequency distortions of similar magnitude to what many SSD-CI listeners likely experience substantially reduced or eliminated binaural benefits. However, despite the likely presence of these distortions, many actual SSD-CI listeners achieve partial restoration of binaural function. This suggests that over time, individuals' auditory systems might be compensating for these mismatched inputs (Reiss et al., 2007; Svirsky et al., 2004). Still, for those SSD-CI listeners who achieve some binaural hearing after implantation, they might nevertheless benefit from remapping strategies to diminish the effects of spectral mismatch and compression. These listeners might still not be

maximizing the potential power of their implant to provide binaural hearing. It is an open question as to what kind of hearing benefits SSD-CI listeners might achieve with CI mapping that more closely reflects the needs of SSD individuals. In contrast, for SSD-CI users who do not exhibit any binaural hearing benefits after implantation, it is possible that a modified frequency allocation and compression algorithm would restore some aspects of binaural hearing for these individuals.

Frequency mismatch and compression are viable and realistic targets for optimization because they can be minimized with current CI technology and techniques. Simply changing the frequency allocation of a SSD-CI listener's electrode array has the potential to reduce spectral mismatch if it can be determined what the optimal map is. Additionally, dynamic and adaptive compression algorithms could be implemented that might be less disruptive to spatial hearing than static envelope compression. Reduction of the frequency mismatch between an acoustic ear and an implanted ear has the most potential for improved spatial hearing for those with SSD. Although the data are sparse, the binaural system in the brainstem is believed to be based on coincidence detection by spectrally matched inputs coming from each ear (Joris et al., 1998). Therefore, simply providing a better interaural frequency match might restore many binaural hearing benefits for SSD-CI listeners via improved alignment of subcortical circuitry. Additionally, most plasticity is seen during development and normally there exists no reason to rewire subcortical binaural circuits in adulthood (King, Parsons, & Moore, 2000). Once the head and ears reach adult size, these brainstem-mediated binaural circuits are essentially stable. Therefore, plasticity mechanisms cannot necessarily be relied on to remedy misalignment

in subcortical circuits. Thus, providing the binaural system with a more accurate alignment between the implant and acoustic ear is of principal importance for binaural hearing.

Frequency mismatch between the CI and acoustic ear can be diminished by innovative mapping techniques. This type of change is readily realizable from a technological standpoint, since it would only require a shift in the speech processor frequency-to-electrode allocation table (similar to the change in vocoder analysis filters in Experiment 4.2). However, a completely accurate interaural frequency match is not trivial to accomplish, because it requires knowledge regarding the characteristic frequencies of the auditory nerve fibers being stimulated by each electrode in the array. Determining the location of the electrode array could be accomplished in a number of ways. CT scans (Noble et al., 2014) or radiographs (Landsberger et al., 2015) could be used to estimate the insertion angles of individual electrodes. Individualized CT scans after implantation could give clinicians a good approximation of the various electrode locations for their patient and this information can help guide an individualized place-matched remapping protocol. However, CT scans would only provide the location of the electrode array and would not inform the audiologist about the characteristic frequency of the neurons located below the individual electrodes, neural survival and potential electric field interactions.

Alternatively, psychoacoustic methods might be used to try to determine electrode location. Pitch matching between acoustic and electrical stimuli could be used to determine electrode location. However, pitch-matching procedures can be susceptible to methodical bias (Carlyon et al., 2010) and have been shown to be susceptible to adaptation effects (Reiss et al., 2014). Pitch perception changes reflect cortical plasticity instead of brainstem relative alignments and would likely not be implicated in optimized binaural function. ITD-

sensitivity comparisons between a given electrode and a limited range of acoustic stimuli could be used to approximate the location of a listener's electrodes on the basilar membrane (Goupell et al., 2013; Kan et al., 2013). Identifying ITD-sensitive pairs of electrodes might be the most direct way to determine the best frequency allocation for pairs of electrodes. This approach directly engages the binaural system and has been shown to be a promising psychoacoustic method for determining electrode location for bilateral CI listeners (Hu & Deitz, 2015; Kan et al., 2013). ITD sensitivity could also be measured in SSD-CI listeners by presenting a narrow-band acoustic stimulus to the NH listener paired with a sensitive single electrode in the CI ear. ITD sensitivity might not be as susceptible to adaptation effects, as is pitch matching. This is because ITD processing is a brainstem-mediated computation, and therefore is less vulnerable to plasticity mechanisms than the perception of pitch, which could be subject to cortical plasticity (Weinberger, 1995). However, these ITD measurement experiments take a very long time to complete and some SSD-CI listeners are unable to complete the task at all. The fusion paradigm utilized in Experiment 4.2 could be a good candidate for determining optimal binaural sensitivity in SSD-CI listeners. A clinician could test different maps to determine which map facilitates fusion for the SSD-CI listener. The experiment is relatively straightforward, quick to administer and has the potential to determine whether a particular map can lead to binaural fusion. Based on these putative brainstem mechanisms for ITD and the results presented in Chapter 2 that demonstrate the largest detriments from spectral mismatch, more accurate alignment

between the basilar membrane and implant array would likely lead to the largest improvements in binaural hearing for SSD-CI listeners.

Distortion caused by compression was found to be substantially detrimental to binaural hearing in the experiments in Chapter 3. Due to the integrity of the cochlea after deafness and the limitations in how electric current can encode level, compression is necessary in CI processing. However, just as with spectral mismatch, several possible remedies are available. The most innovative of solutions comes from a study by Kasturi and Loizou (2007) who implemented a dynamic compressive function to determine the effects of a rapidly changing compression function on speech understanding in CI listeners. Static envelope compression falls short when background noise is increased; therefore, the authors aimed to determine what effect a sigmoid-shaped compression function might have on perception of speech in noise. This innovative technique involves suppressing any signals that fall below the noise floor and retaining any signals that fall above the noise floor (likely speech). The sigmoid function likely works well because the knee point (compression threshold) was set to change depending on the listening environment (dynamic compression based on the current noise floor). After examining speech perception, they found that the sigmoid compressive function produced significantly lower speech reception thresholds over the standard logarithmic compression algorithm. A follow up study by the same research group (Hu, Loizou, Li, & Kasturi, 2007) compared their sigmoid compressive function to CI listeners using their own daily strategy and obtained the same result—that is, the dynamic sigmoid compressive function outperformed the standard compressive function in every noise condition tested.

Additional research examining adaptive compression strategies in vocoder simulations corroborated the findings of previous research. Lai, Tsao, and Chen (2015) implemented an envelope compression strategy which enhanced the modulation depth of the vocoded signals and compared speech perception performance to that of a standard static compression algorithm. They found that the adaptive strategy substantially improved speech intelligibility in noise. They conclude that this type of adaptive strategy could show real promise in actual CI listeners by enhancing signal envelopes while reducing the impacts of background noise. Taken together, the research on dynamic, adaptive compression algorithms shows real promise in improving speech perception in noise for CI listeners. Adaptation of sigmoid-shaped compression might facilitate hearing speech in noise for SSD-CI listeners. This type of compression is thought to improve spectral contrast (as would be needed in competing talker situations) without disrupting loudness. Sigmoid-shaped compression also should attenuate more spatial noise due to the adaptive noise floor, which would facilitate hearing in noisy environments. Although promising, it remains to be elucidated whether or not these adaptive compression strategies could improve spatial hearing outcomes for SSD-CI listeners.

The future is bright for optimization of compression CI processing for all CI users. Ultimately, for CI compression to be enhanced, it needs to be adaptive to the listening environment (noisy or quiet) and situation. Additionally, the compression parameter should be adjusted for each individual listener based on his or her unique needs and limitations. New technologies are being developed to restore normal loudness growth for CI listeners and potentially the number of discriminable intensity steps as well.

Taken together, the results of this dissertation indicate that common CI distortions can impose some listening challenges for SSD-CI listeners. The principal findings of this dissertation identify frequency mismatch and compression as important possible targets for optimization to facilitate binaural hearing for SSD-CI listeners. Follow-up studies should specifically target these two CI distortions in actual SSD-CI listeners to determine what effect they have on binaural hearing. Fortunately, these distortions can likely be minimized by innovative mapping and signal programming techniques in order to ensure that SSD-CI listeners receive binaural hearing benefits from their implant. Given the importance of verbal communication in our society, better spatial hearing in noise for SSD-CI listeners will undoubtedly improve their quality of life. More broadly, better hearing outcomes for current SSD-CI listeners will motivate more individuals who are suffering with SSD to seek out CIs as a treatment option.

References

Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society*

of America, 112(5), 2086. <http://doi.org/10.1121/1.1510141>

- Arndt, S., Aschendorff, A., Laszig, R., Beck, R., Schild, C., Kroeger, S., ... Wesarg, T. (2010). Comparison of Pseudobinaural Hearing to Real Binaural Hearing Rehabilitation After Cochlear Implantation in Patients With Unilateral Deafness and Tinnitus.
- Aronoff, J. M., Freed, D. J., Fisher, L. M., Pal, I., & Soli, S. D. (2011). The Effect of Different Cochlear Implant Microphones on Acoustic Hearing Individuals' Binaural Benefits for Speech Perception in Noise. *Ear & Hearing*, 468–484. <http://doi.org/10.1097/AUD.0b013e31820dd3f0>
- Aronoff, J. M., Shayman, C., Prasad, A., Suneel, D., & Stelmach, J. (2015). Unilateral spectral and temporal compression reduces binaural fusion for normal hearing listeners with cochlear implant simulations. *Hearing Research*, 320, 24–29. <http://doi.org/10.1016/j.heares.2014.12.005>
- Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society. Audio Engineering Society*, 49, 904–916.
- Bernstein, J.G.W., Goupell, M.J., Iyer, N., Schuchman, G.I., Rivera, A.L., and Brungart, D. . (2013). Binaural speech stream segregation for single-sided deaf and bilateral cochlear implantees. Poster presentation, Conference on Implantable Auditory Prostheses.
- Bernstein, J. G. W., Goupell, M. J., Schuchman, G. I., Rivera, A. L., & Brungart, D. S. (2016). Having Two Ears Facilitates the Perceptual Separation of Concurrent Talkers

- for Bilateral and Single-Sided Deaf Cochlear Implantees. *Ear and Hearing*, 289–302. <http://doi.org/10.1097/AUD.0000000000000284>
- Bernstein, J. G. W., Iyer, N., & Brungart, D. S. (2015). Release from informational masking in a monaural competing-speech task with vocoded copies of the maskers presented contralaterally. *The Journal of the Acoustical Society of America*, 137(2), 702–13. <http://doi.org/10.1121/1.4906167>
- Bess, F. H., & Tharpe, A. M. (1984). Unilateral hearing impairment in children. *Pediatrics*, 74(2), 206–16. Retrieved from <http://pediatrics.aappublications.org/content/74/2/206.abstract>
- Best, V., Thompson, E. R., Mason, C. R., & Kidd, G. (2013). An energetic limit on spatial release from masking. *JARO - Journal of the Association for Research in Otolaryngology*, 14, 603–610. <http://doi.org/10.1007/s10162-013-0392-1>
- Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., ... Lazard, D. S. (2012). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology and Neurotology*, 18(1), 36–47. <http://doi.org/10.1159/000343189>
- Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.5.)[Computer program]. Retrieved from <Http://www.praat.org/>, 5(9/10), 341–345. Retrieved from <papers3://publication/uuid/AF582E4D-2F7A-409E-B4F1-7A10385D9135>
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*. <http://doi.org/10.1121/1.428288>

- Bradley, J. S., Reich, R. D., & Norcross, S. G. (1999). On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. *The Journal of the Acoustical Society of America*, *106*(4 Pt 1), 1820–8. <http://doi.org/10.1121/1.427932>
- Bregman, A. S. (1994). The Auditory Scene. In *Auditory Scene Analysis: The perceptual organization of sound* (pp. 1–45).
- Bronkhorst, A. W. (2000). The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. *Acustica*, *86*, 117–128. <http://doi.org/10.1306/74D710F5-2B21-11D7-8648000102C1865D>
- Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *83*, 1508–1516. <http://doi.org/10.1121/1.395906>
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. <http://doi.org/10.1121/1.1345696>
- Brungart, D. S., Simpson, B. D., Ericson, M. a., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527. <http://doi.org/10.1121/1.1408946>
- Buechner, A., Brendel, M., Lesinski-Schiedat, A., Wenzel, G., Frohne-Buechner, C., Jaeger, B., & Lenarz, T. (2010). Cochlear implantation in unilateral deaf subjects associated with ipsilateral tinnitus. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otolology and Neurotology*, *31*(9), 1381–5.

<http://doi.org/10.1097/MAO.0b013e3181e3d353>

- Buss, E., Whittle, L. N., Grose, J. H., & Hall, J. W. (2009). Masking release for words in amplitude-modulated noise as a function of modulation rate and task. *The Journal of the Acoustical Society of America*, *126*(1), 269–80. <http://doi.org/10.1121/1.3129506>
- Cai, Y., Zheng, Y., Liang, M., Zhao, F., Yu, G., Liu, Y., ... Chen, G. (2015). Auditory spatial discrimination and the mismatch negativity response in hearing-impaired individuals. *PLoS ONE*, *10*(8). <http://doi.org/10.1371/journal.pone.0136299>
- Carlyon, R. P., Macherey, O., Frijns, J. H. M., Axon, P. R., Kalkman, R. K., Boyle, P., ... Dauman, R. (2010). Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear. *Journal of the Association for Research in Otolaryngology: JARO*, *11*(4), 625–40. <http://doi.org/10.1007/s10162-010-0222-7>
- Carrell, T. D., & Opie, J. M. (1992). The effect of amplitude comodulation on auditory object formation in sentence perception. *Perception & Psychophysics*, *52*(4), 437–45. <http://doi.org/10.3758/BF03206703>
- Chermak, G., & Lee, J. (2005). Comparison of children's performance on four tests of temporal resolution. *Journal of the American Academy of Audiology*, *16*(8), 554–563. <http://doi.org/10.3766/jaaa.16.8.4>
- Clarkson, P. M., & Bahgat, S. F. (1991). Envelope expansion methods for speech enhancement. *The Journal of the Acoustical Society of America*, *89*(3), 1378–82. <http://doi.org/10.1121/1.400538>
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*, 1562–1573. <http://doi.org/10.1121/1.2166600>

- Crew, J. D., Galvin, J. J., & Fu, Q.-J. J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *J. Acoust. Soc. Am.*, *132*(October), EL429. <http://doi.org/10.1121/1.4758770>
- Culling, J. F., Jelfs, S., Talbert, A., Grange, J. a, & Backhouse, S. S. (2012). The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise. *Ear and Hearing*, *33*(6), 673–82. <http://doi.org/10.1097/AUD.0b013e3182587356>
- Culling, J. F., Jelfs, S., Talbert, A., Grange, J. a, & Backhouse, S. S. (2012). The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise. *Ear Hear.*, *33*(6), 673–682. <http://doi.org/10.1097/AUD.0b013e3182587356>
- Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *The Journal of the Acoustical Society of America*, *103*, 1080–1084. <http://doi.org/10.1121/1.421221>
- de Cheveigné, A., McAdams, S., & Marin, C. M. H. (1997). Concurrent vowel identification. II. Effects of phase, harmonicity, and task. *The Journal of the Acoustical Society of America*. <http://doi.org/10.1121/1.419476>
- DeVries, L., Scheperle, R., & Bierer, J. A. (2016). Assessing the Electrode-Neuron Interface with the Electrically Evoked Compound Action Potential, Electrode Position, and Behavioral Thresholds. *JARO - Journal of the Association for Research in Otolaryngology*, *17*(3), 237–252. <http://doi.org/10.1007/s10162-016-0557-9>
- Dong, S., Mulders, W. H. a M., Rodger, J., & Robertson, D. (2009). Changes in neuronal activity and gene expression in guinea-pig auditory brainstem after unilateral partial hearing loss. *Neuroscience*, *159*(3), 1164–74.

<http://doi.org/10.1016/j.neuroscience.2009.01.043>

Dooley, G. J., Blarney, P. J., Seligman, P. M., Alcantara, J. I., Clark, G. M., Shallop, J. K., ... Menapace, C. M. (1993). Combined Electrical and Acoustical Stimulation Using a Bimodal Prosthesis.

Dorman, M. F., Zeitler, D., Cook, S. J., Loiselle, L., Yost, W. A., Wanna, G. B., & Gifford, R. H. (2015). Interaural level difference cues determine sound source localization by single-sided deaf patients fit with a cochlear implant. *Audiology and Neurotology*, 20(3), 183–188. <http://doi.org/10.1159/000375394>

Drullman, R., & Bronkhorst, a W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *The Journal of the Acoustical Society of America*, 107(4), 2224–2235. <http://doi.org/10.1121/1.428503>

Dunn, C. C., Tyler, R. S., Witt, S., Ji, H., & Gantz, B. J. (2012). Sequential bilateral cochlear implantation: Speech perception and localization pre-and post-second cochlear implantation. *American Journal of Audiology*, 21, 181–189. [http://doi.org/10.1044/1059-0889\(2012/12-0004\)](http://doi.org/10.1044/1059-0889(2012/12-0004))

Durlach, N. I. (1963). Equalization and Cancellation Theory of Binaural Masking-Level Differences. *The Journal of the Acoustical Society of America*. <http://doi.org/10.1121/1.1918675>

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, 114(1), 368. <http://doi.org/10.1121/1.1577562>

- Eapen, R. J., Buss, E., Adunka, M. C., Pillsbury, H. C., & Buchman, C. A. (2009). Hearing-in-noise benefits after bilateral simultaneous cochlear implantation continue to improve 4 years after implantation. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 30, 153–159. <http://doi.org/10.1097/MAO.0b013e3181925025>
- Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology*, 5(3), e1000302. <http://doi.org/10.1371/journal.pcbi.1000302>
- English, K., & Church, G. (1999). Unilateral hearing loss in children: An update for the 1990s. *Language, Speech, and Hearing Services in Schools*, 30(1), 26–31. Retrieved from <http://lshss.asha.org/cgi/content/abstract/30/1/26>
- Erbele, I. D., Bernstein, J. G. W., Schuchman, G. I., Brungart, D. S., & Rivera, A. (2015). An initial experience of cochlear implantation for patients with single-sided deafness after prior osseointegrated hearing device. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 36(1), e24-9. <http://doi.org/10.1097/MAO.0000000000000652>
- Firszt, J. B., Holden, L. K., Reeder, R. M., Cowdrey, L., & King, S. (2012). Cochlear implantation in adults with asymmetric hearing loss. *Ear and Hearing*, 33(4), 521–33. <http://doi.org/10.1097/AUD.0b013e31824b9dfc>
- Francart, T., & McDermott, H. J. (2013). Psychophysics, fitting, and signal processing for combined hearing aid and cochlear implant stimulation. *Ear and Hearing*, 34(6), 685–

700. <http://doi.org/10.1097/AUD.0b013e31829d14cb>

- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA*, *109*(5), 2112–2122. <http://doi.org/10.1121/1.1354984>
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2008). Spatial release from masking with noise-vocoded speech. *The Journal of the Acoustical Society of America*, *124*(3), 1627–37. <http://doi.org/10.1121/1.2951964>
- Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2005). Spatial and spectral factors in release from informational masking in speech recognition. *Acta Acustica United with Acustica*, *91*, 537–545. <http://doi.org/10.1121/1.1354984>
- Freyman, R. L., Helfer, K. S., McCall, D. D., & CLIFTON, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America*, *106*(6), 3578–3588. <http://doi.org/10.1121/1.428211>
- Fried, D. L. (1990). Greenwood frequency measurements. *Journal of the Optical Society of America A*. <http://doi.org/10.1364/JOSAA.7.000946>
- Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150. <http://doi.org/10.1121/1.1381538>
- Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology: JARO*, *6*(1), 19–27. <http://doi.org/10.1007/s10162-004-5024-3>
- Fu, Q. J., & Shannon, R. V. (1998). Effects of amplitude nonlinearity on phoneme

- recognition by cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 104(5), 2570–2577. <http://doi.org/10.1121/1.423912>
- Gallun, F. J., Mason, C. R., & Kidd, G. (2005). Binaural release from informational masking in a speech identification task. *The Journal of the Acoustical Society of America*, 118(3), 1614. <http://doi.org/10.1121/1.1984876>
- Garadat, S. N., Litovsky, R. Y., Yu, G., & Zeng, F.-G. (2009). Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech. *The Journal of the Acoustical Society of America*, 126(5), 2522–2535. <http://doi.org/10.1121/1.3238242>
- Gardner, W. G. (1995). HRTF measurements of a KEMAR. *The Journal of the Acoustical Society of America*. <http://doi.org/10.1121/1.412407>
- Gelfand, S. (2004). *Hearing- An Introduction to Psychological and Physiological Acoustics* (4th ed.). New York, Marcel Dekker.
- Glasberg, B. R., & Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79, 1020–1033. <http://doi.org/10.1121/1.393374>
- Goksoy, C., Demirtas, S., Yagcioglu, S., & Urgan, P. (2005). Interaural delay-dependent changes in the binaural interaction component of the guinea pig brainstem responses. *Brain Research*, 1054(2), 183–191. <http://doi.org/10.1016/j.brainres.2005.06.083>
- Gordon, K. A., Valero, J., van Hoesel, R., & Papsin, B. C. (2008). Abnormal timing delays in auditory brainstem responses evoked by bilateral cochlear implant use in children. *Otology & Neurotology: Official Publication of the American Otological Society*,

American Neurotology Society [and] European Academy of Otology and Neurotology, 29, 193–198. <http://doi.org/10.1097/mao.0b013e318162514c>

Goupell, M. J., & Litovsky, R. Y. (2015). Sensitivity to interaural envelope correlation changes in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, 137(1), 335–349. <http://doi.org/10.1121/1.4904491>

Goupell, M. J., Stoelb, C., Kan, A., & Litovsky, R. Y. (2013). Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening. *The Journal of the Acoustical Society of America*, 133(4), 2272–87. <http://doi.org/10.1121/1.4792936>

Grant, K. W., Wassenhove, V. Van, & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication*, 44(1–4), 43–53. <http://doi.org/10.1016/j.specom.2004.06.004>

Grantham, D. W., Ashmead, D. H., Haynes, D. S., Hornsby, B. W. Y., Labadie, R. F., & Ricketts, T. A. (2012). Horizontal Plane Localization in Single-Sided Deaf Adults Fitted With a Bone-Anchored Hearing Aid (Baha). *Ear and Hearing*. <http://doi.org/10.1097/AUD.0b013e3182503e5e>

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Haynes, D. S., & Labadie, R. F. (2008). Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using CIS+ processing. *Ear and Hearing*, 29, 33–44. <http://doi.org/10.1097/AUD.0b013e31815d636f>

Green, T., Faulkner, A., & Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants.

- The Journal of the Acoustical Society of America*, 112(5), 2155.
<http://doi.org/10.1121/1.1506688>
- Greenwood, D. D. (1961). Auditory Masking and the Critical Band. *The Journal of the Acoustical Society of America*. <http://doi.org/10.1121/1.1908699>
- Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3), 983–1012.
<http://doi.org/10.1152/physrev.00026.2009>
- Hall, J. W., Buss, E., & Grose, J. H. (2005). Informational masking release in children and adults. *The Journal of the Acoustical Society of America*, 118, 1605–1613.
<http://doi.org/10.1121/1.1992675>
- Hansen, M. R., Gantz, B. J., & Dunn, C. (2013). Outcomes After Cochlear Implantation for Patients With Single-Sided Deafness , Including Those With ` re ' s Disease Recalcitrant Me.
- Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115(2), 833. <http://doi.org/10.1121/1.1639908>
- Hoesel, R. Van. (2012). *Auditory Protheses*. (F.-G. Zeng, A. N. Popper, & R. R. Fay, Eds.) (Vol. 39). New York, NY: Springer New York. <http://doi.org/10.1007/978-1-4419-9434-9>
- Hopkins, K., & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *The Journal of the Acoustical Society of America*, 125(1), 442–6. <http://doi.org/10.1121/1.3037233>
- Hu, H., & Dietz, M. (2015). Comparison of Interaural Electrode Pairing Methods for

- Bilateral Cochlear Implants. *Trends in Hearing*, 19, 233121651561714.
<http://doi.org/10.1177/2331216515617143>
- Hu, Y., Loizou, P. C., Li, N., & Kasturi, K. (2007). Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *The Journal of the Acoustical Society of America*, 122(4), EL128-L134. <http://doi.org/10.1121/1.2772401>
- Ihlefeld, A., & Litovsky, R. Y. (2012). Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening. *PLoS One*, 7(9), e45296. <http://doi.org/10.1371/journal.pone.0045296>
- Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, 123(6), 4369–79. <http://doi.org/10.1121/1.2904826>
- Jones, G. L., Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2013). Relationship between channel interaction and spectral-ripple discrimination in cochlear implant users a). *The Journal of the Acoustical Society of America*, 133(1), 425–433. <http://doi.org/10.1121/1.4768881>
- Joris, P. X., Smith, P. H., & Yin, T. C. T. (1998). Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron*. [http://doi.org/10.1016/S0896-6273\(00\)80643-1](http://doi.org/10.1016/S0896-6273(00)80643-1)
- Kamal, S. M., Robinson, A. D., & Diaz, R. C. (2012). Cochlear implantation in single-sided deafness for enhancement of sound localization and speech perception. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 20(5), 393–397. <http://doi.org/10.1097/MOO.0b013e328357a613>
- Kan, A., Stoelb, C., Litovsky, R. Y., & Goupell, M. J. (2013a). Effect of mismatched place-

- of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, 134(4), 2923–36. <http://doi.org/10.1121/1.4820889>
- Kan, A., Stoelb, C., Litovsky, R. Y., & Goupell, M. J. (2013b). Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, 134(4), 2923–36. <http://doi.org/10.1121/1.4820889>
- Kasturi, K., & Loizou, P. C. (2007). Use of S-shaped input-output functions for noise suppression in cochlear implants. *Ear and Hearing*, 28(3), 402–411. <http://doi.org/10.1097/AUD.0b013e31804793c4>
- Kawano, A., Seldon, H. L., Pyman, B., & Clark, G. M. (1995). Intracochlear factors contributing to psychophysical percepts following cochlear implantation: A case study. In *Annals of Otology, Rhinology and Laryngology* (Vol. 104, pp. 54–57). <http://doi.org/10.1080/00016489850183386>
- Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., & Neely, J. G. (1998). In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. *Annals of Otology, Rhinology and Laryngology*, 107, 1–16.
- Kidd, G., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, 111(3), 1367. <http://doi.org/10.1121/1.1448342>
- Kidd, G., Mason, C. R., Best, V., & Marrone, N. (2010). Stimulus factors influencing spatial release from speech-on-speech masking. *The Journal of the Acoustical Society of America*, 128, 1965–1978. <http://doi.org/10.1121/1.3478781>

- Kidd, G., Mason, C. R., & Deliwala, P. S. (1994). Reducing informational masking by sound segregation, *95*(June 1994), 3475–3480.
- Kidd, G., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *104*, 422–431.
<http://doi.org/10.1121/1.423246>
- King, a J., Parsons, C. H., & Moore, D. R. (2000). Plasticity in the neural coding of auditory space in the mammalian brain. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11821–11828.
<http://doi.org/10.1073/pnas.97.22.11821>
- Laback, B., Egger, K., & Majdak, P. (2014). Perception and coding of interaural time differences with bilateral cochlear implants. *Hearing Research*, 1–13.
<http://doi.org/10.1016/j.heares.2014.10.004>
- Lai, Y. H., Tsao, Y., & Chen, F. (2015). Effects of adaptation rate and noise suppression on the intelligibility of compressed-envelope based speech. *PLoS ONE*, *10*(7), 1–19.
<http://doi.org/10.1371/journal.pone.0133519>
- Landsberger, D. M., Svrakic, M., Roland, J. T., & Svirsky, M. (2015). The Relationship Between Insertion Angles, Default Frequency Allocations, and Spiral Ganglion Place Pitch in Cochlear Implants. *Ear Hear2*, *36*, 207–213.
<http://doi.org/10.1097/AUD.0000000000000163>
- Landsberger DM, Svrakic M, Roland JT Jr, S. M. (2015). The Relationship Between Insertion Angles, Default Frequency Allocations, and Spiral Ganglion Place Pitch in Cochlear Implants. *Ear and Hearing*.

- Leek, M., Brown, M., & Dorman, M. (1991). Informational masking and auditory attention. *Perception & Psychophysics*, *50*(3), 205–214. Retrieved from <http://link.springer.com/article/10.3758/BF03206743>
- Li, N., & Loizou, P. C. (2009). Factors affecting masking release in cochlear-implant vocoded speech. *The Journal of the Acoustical Society of America*, *126*(1), 338–46. <http://doi.org/10.1121/1.3133702>
- Lieu, J. E. C., Tye-Murray, N., Karzon, R. K., & Piccirillo, J. F. (2010). Unilateral hearing loss is associated with worse speech-language scores in children. *Pediatrics*, *125*(6), e1348–e1355. <http://doi.org/10.1542/peds.2009-2448>
- Linstrom, C. J., Silverman, C. a., & Yu, G.-P. (2009). Efficacy of the bone-anchored hearing aid for single-sided deafness. *The Laryngoscope*, *119*(4), 713–20. <http://doi.org/10.1002/lary.20164>
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, *106*, 1633–1654.
- Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., ... Misurelli, S. (2012). Studies on bilateral cochlear implants at the University of Wisconsin's Binaural Hearing and Speech Laboratory. *Journal of the American Academy of Audiology*, *23*(6), 476–94. <http://doi.org/10.3766/jaaa.23.6.9>
- Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., & Yu, G. (2004). Bilateral Cochlear Implants in Adults and Children. *Archives of Otolaryngology–Head & Neck Surgery*, *130*(5), 648. <http://doi.org/10.1001/archotol.130.5.648>
- Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances*

in Oto-Rhino-Laryngology. <http://doi.org/10.1159/000094648>

- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., & Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, *125*(1), 372–83. <http://doi.org/10.1121/1.3036175>
- Long, C. J., Eddington, D. K., Colburn, H. S., & Rabinowitz, W. M. (2003). Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user. *The Journal of the Acoustical Society of America*, *114*(3), 1565. <http://doi.org/10.1121/1.1603765>
- Lopez-Poveda, E. A., Eustaquio-Martín, A., Stohl, J. S., Wolford, R. D., Schatzer, R., & Wilson, B. S. (2016). A Binaural Cochlear Implant Sound Coding Strategy Inspired by the Contralateral Medial Olivocochlear Reflex. *Ear and Hearing*, *37*(3), e138-48. <http://doi.org/10.1097/AUD.0000000000000273>
- Lorenzi, C., Berthommier, F., Apoux, F., & Bacri, N. (1999). Effects of envelope expansion on speech recognition. *Hearing Research*, *136*(1–2), 131–138. [http://doi.org/10.1016/S0378-5955\(99\)00117-3](http://doi.org/10.1016/S0378-5955(99)00117-3)
- Ma, N., Morris, S., & Kitterick, P. (2015). Benefits to speech perception in noise from the binaural integration of electric and acoustic signals in unilateral deafness. *Ear and Hearing*.
- Manuscript, A., & Listeners, C. (2013). NIH Public Access, *33*(5), 645–659. <http://doi.org/10.1097/AUD.0b013e318252caae>. Timbre
- Maslin, M. R. D., Munro, K. J., & El-Deredy, W. (2013). Evidence for multiple mechanisms of cortical plasticity: a study of humans with late-onset profound

- unilateral deafness. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 124(7), 1414–21.
<http://doi.org/10.1016/j.clinph.2012.12.052>
- McDermott, H. J., McKay, C. M., Richardson, L. M., & Henshall, K. R. (2003). Application of loudness models to sound processing for cochlear implants. *The Journal of the Acoustical Society of America*, 114(4), 2190.
<http://doi.org/10.1121/1.1612488>
- McDermott, H., & Varsavsky, A. (2009). Better fitting of cochlear implants: modeling loudness for acoustic and electric stimuli. *Journal of Neural Engineering*, 6, 65007.
<http://doi.org/10.1088/1741-2560/6/6/065007>
- McKinney, C. (2002). Hear the other side – a report on Single Sided Deafness. *Entific Medical Systems*.
- Middlebrooks, J. C. (1999). Individual differences in external-ear transfer functions reduced by scaling in frequency. *The Journal of the Acoustical Society of America*, 106(3 Pt 1), 1480–1492. <http://doi.org/10.1121/1.427176>
- Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, 42, 135–159.
<http://doi.org/10.1146/annurev.ps.42.020191.001031>
- Middlebrooks, J. C., Macpherson, E. A., & Onsan, Z. A. (2000). Psychophysical customization of directional transfer functions for virtual sound localization. *The Journal of the Acoustical Society of America*. <http://doi.org/10.1121/1.1322026>
- Mills, A. W. (1960). Lateralization of High-Frequency Tones. *The Journal of the Acoustical Society of America*, 32(1), 132. <http://doi.org/10.1121/1.1907864>

- Mishra, S. K., & Lutman, M. E. (2014). Top-down influences of the medial olivocochlear efferent system in speech perception in noise. *PloS One*, 9(1), e85756. <http://doi.org/10.1371/journal.pone.0085756>
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*. Boston Academic Press (Vol. 3). <http://doi.org/10.1016/j.tins.2007.05.005>
- Moore, J. K. (2000). Organization of the human superior olivary complex. *Microscopy Research and Technique*, 51(4), 403–412. [http://doi.org/10.1002/1097-0029\(20001115\)51:4<403::AID-JEMT8>3.0.CO;2-Q](http://doi.org/10.1002/1097-0029(20001115)51:4<403::AID-JEMT8>3.0.CO;2-Q)
- Nelson, D. a, Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., & Javel, E. (1996). Intensity discrimination as a function of stimulus level with electric stimulation. *The Journal of the Acoustical Society of America*, 100(4 Pt 1), 2393–2414. <http://doi.org/10.1121/1.417949>
- Nie, K., Barco, A., & Zeng, F.-G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear and Hearing*, 27(2), 208–217. <http://doi.org/10.1097/01.aud.0000202312.31837.25>
- Noble, J. H., Gifford, R. H., Hedley-Williams, A. J., Dawant, B. M., & Labadie, R. F. (2014). Clinical evaluation of an image-guided cochlear implant programming strategy. *Audiology and Neurotology*, 19(6), 400–411. <http://doi.org/10.1159/000365273>
- O'Donoghue, G. M., Nikolopoulos, T. P., & Archbold, S. M. (2000). Determinants of speech perception in children after cochlear implantation. *Lancet*, 356, 466–468. [http://doi.org/10.1016/S0140-6736\(00\)02555-1](http://doi.org/10.1016/S0140-6736(00)02555-1)
- Pelizzone, M., Kasper, A., & Montandon, P. (1990). Binaural interaction in a cochlear

- implant patient. *Hearing Research*, 48(3), 287–290. [http://doi.org/10.1016/0378-5955\(90\)90069-2](http://doi.org/10.1016/0378-5955(90)90069-2)
- Poon, B. B., Eddington, D. K., Noel, V., & Colburn, H. S. (2009). Sensitivity to interaural time difference with bilateral cochlear implants: Development over time and effect of interaural electrode spacing. *The Journal of the Acoustical Society of America*, 126(2), 806–815. <http://doi.org/10.1121/1.3158821>
- R.C., S., J.W., S., J.D., W., P.C., L., & M.D., K. (2014). Vocoder simulations of highly focused cochlear stimulation with limited dynamic range and discriminable steps. *Ear and Hearing*, 35(2), 262–270. <http://doi.org/10.1097/AUD.0b013e3182a768e8>
- Rasetshwane, D. M., Argenyi, M., Neely, S. T., Kopun, J. G., & Gorga, M. P. (2013a). Latency of tone-burst-evoked auditory brain stem responses and otoacoustic emissions: level, frequency, and rise-time effects. *The Journal of the Acoustical Society of America*, 133(5), 2803–17. <http://doi.org/10.1121/1.4798666>
- Rasetshwane, D. M., Argenyi, M., Neely, S. T., Kopun, J. G., & Gorga, M. P. (2013b). Latency of tone-burst-evoked auditory brain stem responses and otoacoustic emissions: level, frequency, and rise-time effects. *The Journal of the Acoustical Society of America*, 133, 2803–17. <http://doi.org/10.1121/1.4798666>
- Rayleigh, Lord. (1907). On our perception of sound direction. *Philosophical Magazine Series 6*, 13, 214–232. <http://doi.org/10.1080/14786440709463595>
- Reiss, L. A. J., Ito, R. A., Eggleston, J. L., & Wozny, D. R. (2014). Abnormal binaural spectral integration in cochlear implant users. *JARO - Journal of the Association for Research in Otolaryngology*, 15(2), 235–248. <http://doi.org/10.1007/s10162-013-0434-8>

- Reiss, L., Turner, C. W., Erenberg, S. R., & Gantz, B. J. (2007). Changes in pitch with a cochlear implant over time. *Journal of the Association for Research in Otolaryngology : JARO*, 8(2), 241–57. <http://doi.org/10.1007/s10162-007-0077-8>
- Reynolds, G. S., & Stevens, S. S. (1960). Binaural Summation of Loudness. *The Journal of the Acoustical Society of America*, 32(10), 1337–1344. <http://doi.org/10.1121/1.1907903>
- Riedel, H., & Kollmeier, B. (2002). Comparison of binaural auditory brainstem responses and the binaural difference potential evoked by chirps and clicks. *Hearing Research*, 169(1–2), 85–96. [http://doi.org/10.1016/S0378-5955\(02\)00342-8](http://doi.org/10.1016/S0378-5955(02)00342-8)
- Roberts, M. T., Seeman, S. C., & Golding, N. L. (2013). A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry. *Neuron*, 78(5), 923–935. <http://doi.org/10.1016/j.neuron.2013.04.022>
- Rubinstein, J. T., & Miller, C. A. (1999). How do cochlear prostheses work? *Current Opinion in Neurobiology*. [http://doi.org/10.1016/S0959-4388\(99\)80060-9](http://doi.org/10.1016/S0959-4388(99)80060-9)
- Schleich, P., Nopp, P., D’Haese, P., & D’Haese, P. (2004a). Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear Hear.*, 25, 197–204. <http://doi.org/10.1097/01.AUD.0000130792.43315.97>
- Schleich, P., Nopp, P., D’Haese, P., & D’Haese, P. (2004b). Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear Hear.*, 25(3), 197–204. <http://doi.org/10.1097/01.AUD.0000130792.43315.97>
- Schroder, A. C., Viemeister, N. F., & Nelson, D. A. (1994). Intensity discrimination in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 96(5), 2683. <http://doi.org/10.1121/1.411276>

- Senn, P., Kompis, M., Vischer, M., & Haeusler, R. (2005). Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors. *Audiology and Neurotology*, *10*(6), 342–352. <http://doi.org/10.1159/000087351>
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*. <http://doi.org/10.1016/j.tins.2010.11.002>
- Shannon, R. V, Fu, Q.-J., & Galvin, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica. Supplementum*, (February), 50–54. <http://doi.org/10.1080/03655230410017562>
- Shinn, J. B., Baran, J. a, Moncrieff, D. W., & Musiek, F. E. (2005). Differential attention effects on dichotic listening. *Journal of the American Academy of Audiology*, *16*(4), 205–18. <http://doi.org/10.3766/jaaa.16.4.2>
- Siciliano, C. M., Faulkner, A., Rosen, S., & Mair, K. (2010). Resistance to learning binaurally mismatched frequency-to-place maps: implications for bilateral stimulation with cochlear implants. *The Journal of the Acoustical Society of America*, *127*(3), 1645–1660. <http://doi.org/10.1121/1.3293002>
- Sinopoli, T. (2003). Single Sided Deafness: Issues and Alternatives. www.Audiologyonline.com.
- Soulodre, G. A., Popplewell, N., & Bradley, J. S. (1989). Combined effects of early reflections and background noise on speech intelligibility. *Journal of Sound and Vibration*, *135*(1), 123–133. [http://doi.org/10.1016/0022-460X\(89\)90759-1](http://doi.org/10.1016/0022-460X(89)90759-1)

- Stakhovskaya, O., Sridhar, D., Bonham, B. H., & Leake, P. a. (2007). Frequency map for the human cochlear spiral ganglion: implications for cochlear implants. *Journal of the Association for Research in Otolaryngology: JARO*, 8(2), 220–33. <http://doi.org/10.1007/s10162-007-0076-9>
- Stecker, G. C., & Hafter, E. R. (2002). Temporal weighting in sound localization. *The Journal of the Acoustical Society of America*, 112, 1046–1057. <http://doi.org/10.1121/1.1497366>
- Steel, M. M., Papsin, B. C., & Gordon, K. A. (2015). Binaural fusion and listening effort in children who use bilateral cochlear implants: A psychoacoustic and pupillometric study. *PLoS ONE*, 10(2), 1–29. <http://doi.org/10.1371/journal.pone.0117611>
- Steeneken, H. J. M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality, 67, 318–326.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111(4), 1872–1891. <http://doi.org/10.1121/1.1458026>
- Stewart, C. M., Clark, J. H., & Niparko, J. K. (2011). Bone-anchored devices in single-sided deafness. In *Implantable Bone Conduction Hearing Aids* (Vol. 71, pp. 92–102). <http://doi.org/10.1159/000323589>
- Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S.-W., & Suárez, H. (2004). Long-term auditory adaptation to a modified peripheral frequency map. *Acta Oto-Laryngologica*, 124, 381–386. <http://doi.org/10.1080/00016480310000593>
- Svirsky, M. A., Talavage, T. M., Sinha, S., Neuburger, H., & Azadpour, M. (2015). Gradual adaptation to auditory frequency mismatch. *Hearing Research*, 322, 163–

170. <http://doi.org/10.1016/j.heares.2014.10.008>

- Tyler, R. S., Noble, W., Dunn, C., & Witt, S. (2006). Some benefits and limitations of binaural cochlear implants and our ability to measure them. *International Journal of Audiology, 45 Suppl 1*, S113-9. <http://doi.org/10.1080/14992020600783095>
- van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. *The Journal of the Acoustical Society of America, 105*(5), 2903–2913. <http://doi.org/10.1121/1.426943>
- Van de Heyning, P., Vermeire, K., Diebl, M., Nopp, P., Anderson, I., & De Ridder, D. (2008). Incapacitating Unilateral Tinnitus in Single-Sided Deafness Treated by Cochlear Implantation. *Annals of Otology, Rhinology & Laryngology, 117*(9), 645–652. <http://doi.org/10.1177/000348940811700903>
- van de Par, S., & Kohlrausch, a. (1998). Comparison of monaural (CMR) and binaural (BMLD) masking release. *The Journal of the Acoustical Society of America, 103*(3), 1573–1579. <http://doi.org/10.1121/1.421292>
- van Hoesel, R. J., & Clark, G. M. (1997). Psychophysical studies with two binaural cochlear implant subjects. *The Journal of the Acoustical Society of America, 102*(1), 495–507. <http://doi.org/10.1121/1.419611>
- van Hoesel, R. J. M. (2008). Observer weighting of level and timing cues in bilateral cochlear implant users. *The Journal of the Acoustical Society of America, 124*, 3861–3872. <http://doi.org/10.1121/1.2998974>
- van Hoesel, R. J. M., & Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society*

- of America*, 113(3), 1617–1630. <http://doi.org/10.1121/1.1539520>
- Vermeire, K., & Van de Heyning, P. (2009). Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiology & Neuro-Otology*, 14(3), 163–71. <http://doi.org/10.1159/000171478>
- Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica United with Acustica*, 91(3), 502–512.
- Weinberger, N. M. (1995). Dynamic regulation of receptive fields and maps in the adult sensory cortex. *Annual Review of Neuroscience*, 18, 129–158. <http://doi.org/10.1146/annurev.ne.18.030195.001021>
- Welsh, L. W., Rosen, L. F., Welsh, J. J., & Dragonette, J. E. (2004). Functional impairments due to unilateral deafness. *Annals of Otology, Rhinology and Laryngology*, 113(12), 987–993.
- Wenzel, E. M., Wightman, F. L., & Kistler, D. J. (1991). Localization with non-individualized virtual acoustic display cues. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Reaching through Technology - CHI '91*, 351–359. <http://doi.org/10.1145/108844.108941>
- Wierstorff, H., Geier, M., Raake, A., & Spors, S. (2011). A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. *Audio Engineering Society Convention*, 130, 3–6. Retrieved from <https://dev.qu.tu-berlin.de/projects/measurements/>
- Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91, 1648–1661.

- Zeitler, D. M., Dorman, M. F., Natale, S. J., Loisel, L., Yost, W. A., & Gifford, R. H. (2015). Sound Source Localization and Speech Understanding in Complex Listening Environments by Single-sided Deaf Listeners After Cochlear Implantation. *Otology & Neurotology*, *36*(9), 1467–1471. <http://doi.org/10.1097/MAO.0000000000000841>
- Zeng, F. G., & Shannon, R. V. (1992). Loudness balance between electric and acoustic stimulation. *Hearing Research*, *60*(2), 231–235. [http://doi.org/10.1016/0378-5955\(92\)90024-H](http://doi.org/10.1016/0378-5955(92)90024-H)
- Zhou, J., & Durrant, J. D. (2003). Effects of interaural frequency difference on binaural fusion evidenced by electrophysiological versus psychoacoustical measures. *The Journal of the Acoustical Society of America*, *114*(3), 1508–1515. <http://doi.org/10.1121/1.1600718>
- Zirn, S., Arndt, S., Aschendorff, A., & Wesarg, T. (2015). Interaural stimulation timing in single sided deaf cochlear implant users. *Hearing Research*, *328*, 148–156. <http://doi.org/10.1016/j.heares.2015.08.010>
- Zurek, P. M. (1993). A note on onset effects in binaural hearing. *The Journal of the Acoustical Society of America*, *93*, 1200–1201. <http://doi.org/10.1121/1.405516>