THE EFFECTS OF HEARING IMPAIRMENT ON THE ABILITY TO GLIMPSE SPEECH IN A SPECTRO-TEMPORALLY COMPLEX NOISE

Erol James Ozmeral

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology (Cognitive Program).

Chapel Hill
2013

Approved by:

Emily Buss

Peter C. Gordon

Joseph W. Hall, III

Mark Hollins

Joseph Hopfinger
ABSTRACT

Erol James Ozmeral: The effects of hearing impairment on the ability to
glimpse speech in a spectro-temporally complex noise
(Under the direction of Joseph W. Hall, III and Emily Buss)

The aim of this project was to investigate the effects of hearing impairment on speech
perception in spectro-temporally complex noise. The specific objective of the project was to
psychophysically and computationally assess speech reception in the presence of a masker
that fluctuates both in time and frequency. The experiments were designed to compare
hearing-impaired and normal-hearing listeners on a task which has been shown to highlight
the effect of spread of masking. Through dichotic stimulation, a previous study had shown a
sizeable benefit when compared to monaural stimulation. Experiment 1 tested normal-
hearing and hearing-impaired listeners on consonant recognition in the presence of an
asynchronously modulated noise. We tested the primary hypotheses that spread of masking
reduces available glimpsing opportunities for hearing-impaired listeners, and that removing
spread of masking enhances performance relative to normal-hearing listeners. Results
showed greater masking release in normal-hearing listeners compared to hearing-impaired
listeners, but all listeners achieved some benefit of reducing the effects of spread of masking.
Experiment 2 tested consonant recognition in similar masking conditions as Experiment 1,
testing normal-hearing listeners with simulated reduced audibility and reduced frequency
resolution. We tested the primary hypothesis that reduced audibility is not the only limiting
factor for hearing-impaired listeners to glimpse speech, but rather, that reduced frequency
resolution also plays an important role in the ability to glimpse speech in spectro-temporally complex noise. Results showed that while reduced audibility was a key factor, reduced frequency resolution also contributes to deficits seen in Experiment 1. Experiment 3 tested a computational glimpsing model. We tested the hypotheses that spectral resolution plays a key role in glimpsing for both normal-hearing and hearing-impaired listeners; by analyzing dichotically presented stimuli, the model was expected to predict the benefit seen in the behavioral data. Results indicated that the behavioral data could be accurately predicted by the model, although in some cases, the model out-performed listeners with simulated hearing loss. These studies contribute to a better understanding of factors responsible for hearing-impaired listeners’ reduced ability to follow speech in complex backgrounds, with implications for auditory prosthesis design.
To Alisha and Reyna
ACKNOWLEDGEMENTS

There are a number of influential people I would like to thank for supporting my training, without whom, I never could have been as successful. Dr. Joseph Hall, my graduate research supervisor, for giving me the chance to prove myself and always pushing me to be better. I am thankful for his guidance and leadership needed to complete this dissertation. Dr. Emily Buss, my graduate research mentor, for making me a better psychophysicist. My writing will always need improvement and my experiment designs will always need tweaking, but I am grateful to her for strengthening the tools I need to succeed in those endeavors. Dr. Barbara Shinn-Cunningham, my first mentor and biggest supporter, for believing in me no matter what and for persuading me to leave the nest to pursue my dreams. I am thankful for the introductions she has made and countless recommendations she has provided. Above all, I am grateful to have Barb as a friend, who is always there to say hello. Dr. Virginia Best, for allowing me to ride her coattails at BU, while also instilling in me a sense of work ethic and due diligence in the laboratory. Other faculty members of the BU Hearing Research Center: Dr. Steve Colburn, Dr. Gerald Kidd, Chris Mason, and Dr. Kamal Sen, for establishing an environment that hooked me to the field of auditory neuroscience and never made a young scientist feel unwelcome to the discussion. Dr. Peter C. Gordon, my academic graduate advisor at UNC, for providing me the opportunity to learn the important facets of Cognitive Psychology, which will always have a major influence on my approach to the study of hearing. Past collaborators: Drs. Micheal Dent, Frederick Gallun, and Norbert Kopčo, for
their generous contributions to designing experiments and writing manuscripts. Each is a notable scientist that I hope to have the chance to collaborate with again in the future. My academic big brothers and sister in the Auditory Neuroscience Lab at BU: Dr. Adrian KC Lee, Dr. Antje Ihlefeld, Dr. Eric Thompson, Scott Bressler, and Tim Streeter for the times spent scheming and enjoying the nightlife in Boston, New Orleans, Baltimore, and elsewhere. The UNC Hearing Research Lab: Dr. John Grose, Dr. Sara Mamo, Dr. Angela Bonino, Dr. Heather Porter, Tara Steplowski, and Madhu Dev, for visiting me at the water cooler and listening to countless presentations of the checkerboard study. My fellow students in the Cognitive Program at UNC, especially Jason Kahn, for the late-night board games and many words of wisdom. And last but certainly not least, I want to thank my family for their endless love and support. Alex, my closest friend for over thirty years; Kaan, Katherine, and Ali, for sibling rivalry and love that cannot compare; Mom and Dad, for putting education above all else and giving me the freedom to find my own path. Alisha, my love, my wife, and the mother of my happy daughter, Reyna. Thanks, sweets!
PREFACE

Erol J. Ozmeral attended Boston University (Boston, MA; 2000-2004) and earned a Bachelor of Science in Biomedical Engineering, minoring in Philosophy. He then went on to earn a Master of Arts in Cognitive and Neural Systems, also at Boston University (2004-2007), under the supervision of Dr. Barbara Shinn-Cunningham and with the support of a research assistantship in the Auditory Neuroscience Lab. During his tenure at BU, including as a research specialist (2007-2008), he co-authored seven peer-reviewed manuscripts and submitted twelve abstracts to international and domestic research meetings.

In August 2008, Erol began his doctoral studies at the University of North Carolina at Chapel Hill in the Cognitive Program in the Department of Psychology, minoring in Neurobiology. He completed his masters and doctoral theses under the supervision of Drs. Joseph Hall and Emily Buss, with support by a research assistantship in the Psychoacoustics Lab and a National Research Service Award (NRSA) from the National Institutes of Health (NIH; F31 DC01269). He published a first-authored manuscript and presented his research at five research meetings during his tenure at UNC.

At the conclusion of his doctoral studies, Erol accepted a post-doctoral fellowship at the University of South Florida (Tampa, FL) under the supervision of Drs. David Eddins and Ann C. Eddins, where he currently studies the effects of aging on spatial hearing. That work is supported by an NRSA from the NIH (F32 DC013724).
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>X</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>XI</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>XIII</td>
</tr>
<tr>
<td>CHAPTER 1: GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Glimpsing of Speech in Fluctuating Noise</td>
<td>1</td>
</tr>
<tr>
<td>1.2 What is Frequency Resolution and Why is it Important?</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Motivation for Experiments</td>
<td>8</td>
</tr>
<tr>
<td>1.4 Outline of Dissertation</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER 2: EXPERIMENT 1 – EFFECTS OF HEARING IMPAIRMENT</td>
<td>13</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Method</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>23</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>33</td>
</tr>
<tr>
<td>CHAPTER 3: EXPERIMENT 2 – EFFECTS OF SIMULATED HEARING LOSS</td>
<td>38</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>38</td>
</tr>
<tr>
<td>3.2 Method</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>47</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>58</td>
</tr>
<tr>
<td>CHAPTER 4: EXPERIMENT 3 – A BINAURAL GLIMPISING MODEL</td>
<td>63</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>63</td>
</tr>
<tr>
<td>4.2 Method</td>
<td>66</td>
</tr>
<tr>
<td>4.3 Results</td>
<td>78</td>
</tr>
<tr>
<td>4.4 Discussion</td>
<td>83</td>
</tr>
<tr>
<td>CHAPTER 5: GENERAL DISCUSSION</td>
<td>87</td>
</tr>
<tr>
<td>5.1 Summary of Experiments</td>
<td>87</td>
</tr>
<tr>
<td>5.2 Clinical Applications</td>
<td>90</td>
</tr>
<tr>
<td>5.3 Limitations and Possible Improvements to the Study</td>
<td>91</td>
</tr>
<tr>
<td>5.4 Conclusions</td>
<td>93</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>94</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1: Experiment 1 listener demographics and audiometric profiles ..................15
Table 2.2: Mean speech reception thresholds from Experiment ........................................21
Table 2.3: Mean speech reception thresholds for the dichotic and monaural controls in Experiment 1 ..........................................................21
Table 3.1: Mean speech reception thresholds from Experiment 2 .................................46
Table 4.1: Mean speech reception thresholds without a noise floor in Experiment 3 ........77
Table 4.2: Mean raw speech reception thresholds with a noise floor in Experiment 3 ........78
LIST OF FIGURES

Figure 1.1: Spectrograms of the Sync and Async maskers.................................................2

Figure 2.1: Audiograms for normal-hearing and hearing-impaired listeners ....................14

Figure 2.2: Schematic of masker conditions for all experiments......................................19

Figure 2.3: Schematic of the frequency spectrum for the notched-noise method.............20

Figure 2.4: Mean masking release in Experiment 1.............................................................22

Figure 2.5: Dichotic advantage in Experiment 1 ...............................................................27

Figure 2.6: Consonant accuracy for each band number (by row: 2, 4, 8, and 16).............28

Figure 2.7: The difference between thresholds in the Async-D and Control-D conditions...30

Figure 2.8: Similar to Figure 2.4 except masking release is normalized to the Sync MR....34

Figure 2.9: Similar to Figure 2.5 except dichotic advantage is normalized to MR in the Sync condition........................................................................................................37

Figure 3.1: Spectrograms of an Async masker with 8 frequency bands after spectral smearing ..........................................................................................................................43

Figure 3.2: Estimated thresholds (in dB SPL) with no masker present for each smearing group ..................................................................................................................................46

Figure 3.3: Masking release in the average Sync conditions for each smear factor group....49

Figure 3.4: Mean masking release for Async-D and Async-M conditions.......................49

Figure 3.5: Masking release re-plotted for normal-hearing group in Experiment 1 (85 dB SPL) and group SF1 in Experiment 2 (55 dB SPL) .................................................................51

Figure 3.6: Dichotic advantage for all three smear factor groups at 4 and 8 bands........53

Figure 3.7: The masking release difference between Async-D and Control-D for all three smear factor groups at 4 and 8 bands.................................................................55

Figure 3.8: Speech reception thresholds for the Unmod condition in each smear factor group .................................................................................................................................57

Figure 3.9: Masking release in the Sync condition for each smear factor .......................58
Figure 4.1: Outputs of the spectro-temporal excitation pattern processor for the speech token /aga/ and three maskers
Figure 4.2: Processing architecture of the STEP processor
Figure 4.3: Probabilistic parameters of a hidden Markov model of a series of coin
Figure 4.4: Processing architecture of the decision model
Figure 4.5: Each panel represents the spectro-temporal excitation pattern of the signal
Figure 4.6: Speech reception thresholds in the Unmod and Sync conditions as a function of the noise floor parameter in the model
Figure 4.7: Masking release in the Sync condition as a function of the local SNR used for glimpse detection in the model
Figure 4.8: Model results: masking release in the synchronously-modulated noise condition for each level of smearing without a noise floor
Figure 4.9: Model results: masking release in the synchronously-modulated noise condition for each level of smearing with a noise floor
Figure 4.10: Data from Experiment 1 and the model with no smearing and no noise floor
Figure 4.11: Dichotic advantage for SF groups in Experiment 2 and each SF model with a noise floor
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Articulation Index</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>Async</td>
<td>asynchronously modulated noise</td>
</tr>
<tr>
<td>D</td>
<td>dichotic</td>
</tr>
<tr>
<td>ERB</td>
<td>equivalent rectangular bandwidth</td>
</tr>
<tr>
<td>FFT</td>
<td>fast-Fourier transform</td>
</tr>
<tr>
<td>HMM</td>
<td>hidden Markov model</td>
</tr>
<tr>
<td>L</td>
<td>left</td>
</tr>
<tr>
<td>M</td>
<td>monaural</td>
</tr>
<tr>
<td>MR</td>
<td>masking release</td>
</tr>
<tr>
<td>R</td>
<td>right</td>
</tr>
<tr>
<td>rms</td>
<td>root mean squared</td>
</tr>
<tr>
<td>SF</td>
<td>smear factor</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>STEP</td>
<td>spectro-temporal excitation pattern</td>
</tr>
<tr>
<td>Sync</td>
<td>synchronously modulated noise</td>
</tr>
<tr>
<td>Unmod</td>
<td>unmodulated noise</td>
</tr>
</tbody>
</table>
1.1 Glimpsing of Speech in Fluctuating Noise

In everyday listening environments, following a conversation amidst numerous competing sounds can seem very difficult. Interfering background sounds energetically mask target speech sounds by adding energy to shared frequency bands at the periphery. On a noisy city street or at a crowded party, there is a blend of interfering sounds that fluctuate in time and frequency depending on their sources. Due to inherent fluctuations in the spectro-temporal structure of these maskers, the brain is sometimes able to take advantage of the redundancy in speech across time and frequency to piece together a coherent message. One view of how this takes place is that listeners are able to listen in the dips, or glimpse, sparse speech cues at brief moments of favorable signal-to-noise ratios (SNRs; Miller and Licklider, 1950; Dirks and Bower, 1970; Howard-Jones and Rosen, 1993; Peters et al., 1998; Li and Loizou, 2007). The series of experiments reported here explores the ability to recognize speech presented in spectro-temporally complex maskers and the factors that influence this ability, including those associated with hearing impairment.

Because steady-state noise is rather uncommon in natural settings, amplitude-modulated maskers are often argued to better reflect real-world masking scenarios (Nelken et al., 1999). Target identification in amplitude-modulated noise has been studied extensively in both normal-hearing and hearing-impaired listeners (Wilson and Carhart, 1969; Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Takahashi and Bacon, 1992; Howard-Jones and Rosen, 1993; Eisenberg et al., 1995; Bacon et al., 1998). Relative to steady noise conditions, speech
reception thresholds in fluctuating maskers are typically better. During the “off” phase of amplitude modulation, target speech has the highest SNR, and taking advantage of the high SNR at the masker minima typically leads to the improved thresholds, also known as masking release (MR; Miller and Licklider, 1950; Wilson and Carhart, 1969). Normal-hearing listeners have been shown to achieve MR as large as 23 dB in an amplitude-modulated noise (Howard-Jones and Rosen, 1993; Ozmeral et al., 2012). The size of MR gained by replacing steady noise with an amplitude-modulated masker, however, can vary depending on the masker bandwidth (Hall and Grose, 1989; Bacon et al., 1997), modulation rate (Carlyon et al., 1989; Peters and Hall, 1994), and intensity (Zwicker and Schorn, 1982; Moore and Shailer, 1991).

Figure 1.1: Spectrograms of the Sync (left panel) and Async maskers with 2, 4, 8 and 16 filtering bands (from left to right). Dark regions represent masker energy, while white areas indicate regions of masker minima (i.e., dips). Frequency bands are logarithmically equal in bandwidth, and rate of temporal modulation is 10 Hz corresponding to 50-ms dips. Band numbers are indicated along the right-hand side of each Async masker spectrogram.
Howard-Jones and Rosen (1993) tested the hypothesis that MR associated with fluctuating maskers relies on periods of favorable SNR coinciding across frequency. Their innovative design measured speech reception thresholds with background maskers that fluctuated in both time and frequency. The left panel in Figure 1.1 shows the spectrogram of an amplitude-modulated noise in which all masker minima (white regions) at all frequencies occur at the same time. In contrast, the four right panels in Figure 1.1 show the spectrograms of a masker that fluctuate in time and frequency, and consequently, not all masker minima coincide across frequency. In our current study, we refer to this less conventional masker as asynchronously modulated noise (Async), and we refer to the amplitude-modulated noise as synchronously modulated noise (Sync) in order to highlight the relationship between neighboring frequency regions (i.e., modulated out-of-phase or in-phase, respectively).

Howard-Jones and Rosen (1993) found an interesting result for speech identification in the presence of a diotic Async masker.\(^1\) Consonant identification was tested in three primary diotic conditions: steady noise, Sync, and Async (with 2, 4, 8, or 16 frequency bands). In Async conditions, there was some MR when noise was filtered into only 2 or 4 frequency bands, but no MR was observed in the 8- or 16-band conditions. This was surprising because the cumulative spectro-temporal glimpsing area – regions of masker minima – appear physically equivalent across all modulated maskers (Figure 1.1; compare total white area in each panel).

A recent study from our lab (Ozmeral et al., 2012) hypothesized that the MR differences between Sync and Async conditions in Howard-Jones and Rosen’s (1993) study were due to peripheral processing causing energy in masked frequency regions to spread into neighboring, unmasked frequency regions. Because listeners have been shown to integrate

\(^1\) Diotic presentations mean simply that both ears receive the equivalent stimulus.
speech information distributed across a large number of asynchronously modulated speech bands under some conditions (Buss et al., 2004), we speculated that Howard-Jones and Rosen (1993) may have failed to show MR at greater numbers of bands because spread of masking degraded the quality of the available speech in the modulation minima. In contrast to the Sync condition, glimpses available to the listeners in the Async conditions are flanked by frequency regions with masker maxima (Figure 1.1; Async dark regions above and below white regions). To relieve listeners of the effects of spread of masking, we adopted an approach that has previously been used in bilateral hearing aid and cochlear implant studies (Loizou et al., 2003; Tyler et al., 2010; Kulkarni et al., 2012; Zhou and Pfingst, 2012), in which the bands were distributed across ears, such that the odd-numbered bands were presented to one ear and the even-bands to the other – termed dichotic stimulation, because each ear received different information. By separating the bands across the ears, the effects of spread of masking were avoided, and listeners had a better opportunity to identify the speech. Ozmeral et al. (2012) showed between 5 and 8 dB greater MR in the dichotic Async condition than the monaural Async condition with the same number of frequency bands.

The purpose of the current study was to characterize hearing-impaired listeners’ abilities to understand speech in a complex background such as the Async noise and to understand the factors that may limit those abilities. Along with reduced audibility, hearing-impaired listeners also suffer from poorer frequency resolution than normal-hearing listeners. The long-term goal of this project was to gain a better understanding of the role of frequency resolution in the ability to benefit from spectro-temporal masker modulation. In addition, it was of interest to evaluate the effects of reduced audibility in conjunction with reduced frequency resolution in hearing-impaired listeners, because previous studies have noted that
reduced audibility alone cannot explain all of the difficulties hearing-impaired listeners have in fluctuating noise (e.g., Eisenberg et al., 1995).

1.2 What is Frequency Resolution and Why is it Important?

The sounds that reach a listener’s ears can be the combination of a number of natural and mechanical signals. What allows the listener to detect one particular signal in the presence of another can come down to whether or not the two sounds share overlapping or neighboring frequency components. In principle, spectrally neighboring sounds can mask each other because the ear does not have perfect frequency resolution.

By using masking as a tool in the laboratory, researchers have been able to model frequency resolution to predict signal detection in the presence of noise (for a review, see Moore, 1995). For tone detection in the presence of Gaussian noise, it is assumed that perception is based on the SNR at the output of a filter with a center frequency near the tone frequency. When noise power density is held constant for different bandwidths of the noise, threshold for the tone can vary depending on the bandwidth of the noise. Thus, for noise bands that are narrower than the filter width, detection thresholds will increase as the noise bandwidth increases because of an increase in the total noise power passed by the filter. However, as long as the noise bandwidth is at least as broad as the filter, the threshold of the tone will not change for wider noise bands because the filter does not pass the noise outside of its passband. The filter width is hence named the “critical bandwidth” because it is

---

2 Increasing detection thresholds by adding masking energy to shared frequencies is known as energetic masking, but in some cases, maskers that do not overlap or neighbor the target frequency can also disrupt detection in what is known as informational masking (e.g., Kidd et al., 1994). Unless otherwise stated, the following paper will refer to energetic masking as simply, masking.

3 Total noise power increases proportionally with increasing bandwidth.
equivalent to the bandwidth at which detection thresholds will no longer change (Fletcher, 1940).

Fletcher (1940) described basilar membrane function as consisting of many of these overlapping, band-pass auditory filters. For tones with maskers occupying neighboring frequencies, listeners with good frequency resolution are able to make use of narrow filters that pass primarily target information, whereas listeners with poor frequency resolution are more susceptible to effects of masking spread – the phenomenon in which masker frequency components are not overlapping the target, but the auditory filters nonetheless pass their energy along with that of the target.

Auditory filter bandwidths describe a listener’s frequency resolution, and a number of studies have measured normal auditory filter bandwidths at different center frequencies (Patterson, 1976; Dubno and Dirks, 1989; Glasberg and Moore, 1990; Moore et al., 1990; Shailer et al., 1990; Zhou, 1995). From the measured filter bandwidths, Glasberg and Moore (1990) derived a formula that calculates the equivalent rectangular bandwidth (ERB) across the range of frequencies relevant to human hearing:

\[
ERB_N = 24.7(4.37F + 1)
\]  

(1.1)

where the subscript \(N\) refers to the distinction that filter bandwidth corresponds to normal-hearing listeners tested at medium sound levels, and where \(F\) is the center frequency of the filter (in kHz). Variability in the predicted \(ERB_N\) is minimal, but it is less variable in the middle of the range of human sensitivity than at very low frequencies (Moore et al., 1990) and very high frequencies (Patterson et al., 1982; Shailer et al., 1990; Zhou, 1995). Predicted auditory filter bandwidths are important to the current project because efforts will be made to simulate different degrees of frequency resolution relative to normal-hearing listeners. For
example, if an auditory filter is modeled by a filter bandwidth that is greater than an ERB$_N$, the energy passed by the filter is assumed to reflect the output by an auditory filter of a listener with poorer-than-normal frequency resolution.

Until now, we have discussed how auditory filter widths determine the degree of frequency resolution of the listener, but more importantly, it is of interest to understand how frequency resolution can affect speech perception. If we assume that poor frequency resolution can degrade the perceived spectral quality of speech, then we can test the effects of poor frequency resolution by purposefully degrading the spectral quality of speech. In some studies, this approach has shown that the spectral quality of speech is not necessarily crucial in quiet (e.g., Shannon et al., 1995), but others have shown that spectral quality can affect speech perception differently depending on the difficulty of the speech materials (Fu et al., 1998; Smith et al., 2002), and the age – and arguably the amount of language experience – of the listener (Eisenberg et al., 2000). Finally, in cases where noise is present, spectral quality appears to play an important role in speech identification (Rosen and Fourcin, 1986).

Baer and Moore (1993) demonstrated that good frequency resolution is important for speech perception in the presence of interfering background noise. Their experiments measured speech recognition in quiet and in noise for normal-hearing subjects with simulated reduced frequency resolution. The method employed for their simulations was a previously validated form of spectral smearing (Moore et al., 1992) – a process which blurs the spectral envelope but leaves lower harmonic resolution intact. In quiet, intelligibility of the spectrally smeared speech stimuli was hardly affected, even when simulated filters were effectively six times broader than normal (i.e. an ERB$_N$). However, results for speech presented in noise indicated greater-than-normal masking for high degrees of smearing, an indication that the
effects of spectral smearing, as with poor frequency resolution, are more deleterious in noise than in quiet.

1.3 Motivation for Experiments

Hearing-impaired listeners tend to perform worse overall than normal-hearing listeners in spectro-temporally fluctuating noise (Hall et al., 2012), but it is not certain whether poorer performance is solely a result of reduced audibility or also includes another factor, like broader-than-normal auditory filters at the periphery (see Moore, 2007, for a review). A number of studies that have measured the psychophysical tuning curves in both normal-hearing and hearing-impaired listeners have shown broader tuning (Zwicker and Schorn, 1978; Florentine et al., 1980; Festen and Plomp, 1983), and measured filter shapes have also confirmed a general result of broader-than-normal tuning (Tyler et al., 1984; Dubno and Dirks, 1989; Leek and Summers, 1993). A consequence of broader auditory filters is that hearing impairment is often associated with greater effects of spread of masking (Leshowitz, 1977; Florentine et al., 1980; Gagné, 1988), although not all studies have observed differences (Martin and Pickett, 1970; for a review, see Tyler, 1986).

A common criticism of these reported differences between normal-hearing and hearing-impaired listeners is that stimuli are often not presented at comparable sensation levels\(^4\) to each other (Gagné, 1988), so caution must be made when comparing listeners in studies that do not control for the sensation level of the stimuli across groups. It is often impractical, however, to test hearing-impaired listeners at the same sensation levels as normal-hearing listeners because presentations could reach uncomfortable levels for the hearing-impaired listeners. In order to compare across listeners, some studies have included conditions that are

---

\(^4\) Sensation level refers to the presentation level in dB above threshold in quiet.
meant to equate audibility through manipulation of overall stimulus level (Zurek and Delhorne, 1987), inclusion of a threshold-elevating background noise (Moore et al., 1995), or other signal-processing strategies (e.g., amplitude expansion; Lum and Braida, 2000).

In an early study on the relationship between frequency resolution and consonant recognition in quiet, Dubno and Dirks (1989) used the Articulation Index (AI; Galecki et al., 2009) to predict recognition in normal-hearing and hearing-impaired listeners. Estimates of auditory filter shapes were collected for all subjects, and speech recognition was measured at presentation levels tailored to each individual’s hearing loss, such that the AI predicted equivalent consonant recognition across listeners. Hearing-impaired listeners performed roughly as the AI model predicted, and frequency resolution did not correlate well with performance differences. The study suggests that hearing-impaired listeners are mostly susceptible to poor speech performance in quiet due to reduced audibility, but the evidence does not rule out that poor frequency resolution in hearing-impaired listeners may have added effects beyond those of reduced audibility for speech in noise (e.g., Baer and Moore, 1993).

The current project tested whether frequency resolution could account for some of the difficulty that hearing-impaired listeners have understanding speech in noise, especially spectro-temporally modulated noise. Generally, hearing-impaired listeners benefit very little from masker modulations when compared to normal-hearing listeners (Carhart and Tillman, 1970; Duquesnoy and Plomp, 1983; Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Gustafsson and Arlinger, 1994; Eisenberg et al., 1995; Peters et al., 1998). This discrepancy between hearing-impaired and normal-hearing listeners has been attributed to the combined

---

5 The AI is a tool used for measuring the proportion of a speech signal that is audible (on a range from 0 to 1). By normalizing to the AI, other factors of hearing impairment beyond audibility can arguably be assessed.
effects of reduced audibility – due to audiometric threshold differences between groups (Dubno and Dirks, 1989) – and reduced temporal (Festen and Plomp, 1990; Dubno et al., 2003; George et al., 2006) and frequency (Baer and Moore, 1993; ter Keurs et al., 1993; Baer and Moore, 1994) resolution – which can introduce greater effects of spread of masking.

Evidence from normal-hearing listeners suggests that effects of spread of masking can have deleterious effects on speech reception in Async noise (Howard-Jones and Rosen, 1993; Ozmeral et al., 2012). The current study tested monaural and dichotic maskers, following the same procedures used in our previous experiment evaluating only normal-hearing listeners (Ozmeral et al., 2012), to investigate the effect of spread of masking on asynchronous glimpsing for hearing-impaired listeners. In addition to testing hearing-impaired listeners, normal-hearing listeners were tested using the Baer and Moore (1993) method of spectral smearing in order to dissociate key impairments associated with hearing loss: reduced frequency resolution and reduced audibility. Finally, a computational model based on a glimpsing algorithm (Cooke, 2006) was created to explore the relative roles of reduced audibility and reduced frequency resolution in the asynchronous glimpsing paradigm.

The central hypothesis was that the benefit associated with dichotic presentation is comparable to or larger in hearing-impaired than normal-hearing listeners because spread of masking should have a greater effect on hearing-impaired than normal-hearing listeners. Spread of masking is a peripheral phenomenon, so a model that accounts for peripheral processing of the auditory input was used to quantify the effects of spread of masking, allowing us to see whether this factor fully accounted for the obtained data patterns. From a clinical perspective, we propose that a better understanding of the role of spread of masking
for speech perception in spectro-temporally complex backgrounds could lay the foundation for more effective pre-processing procedures in auditory prostheses.

1.4 Outline of Dissertation

Along with the background information put forth in Chapter 1, the following chapters report on three independent but inter-related experiments conducted as part of the author’s degree requirements. In Chapter 2, Experiment 1 is reported. The purpose of this experiment was to test hearing-impaired listeners on speech recognition in the presence of various masking noises. We examined speech recognition in noise that was unmodulated, synchronously modulated, or asynchronously modulated. The asynchronous masker conditions were of key interest because they reflect the ability to integrate glimpsed speech across time and frequency. Consistent with the view that spread of masking has a negative effect on speech recognition, we expected hearing-impaired listeners to perform poorly in monaural asynchronous conditions. However, separating frequency bands that are associated with out-of-phase masker modulation, via dichotic presentation, should have limited the effects of spread of masking and improved performance. Importantly, the improvement between monaural and dichotic conditions was expected to be comparable to or greater than that found with normal-hearing listeners, to the extent that spread of masking was more detrimental for hearing-impaired listeners.

In Experiment 2 (Chapter 3), we attempted to dissociate the roles of reduced audibility and reduced frequency resolution in hearing-impaired listeners’ performances in the asynchronous glimpsing task. Along with reduced audibility, hearing-impaired listeners are also limited by greater spread of masking due to broader auditory filters. We simulated reduced audibility and reduced frequency resolution in normal-hearing adults and tested the
benefit of dichotic presentation in the asynchronous glimpsing task. The relative impact of reduced frequency resolution was assessed by comparing the data between reduced audibility alone and a combination of reduced audibility and spectral smearing, simulating a loss of frequency resolution. Based on findings in the literature and from our first experiment, reduced audibility likely limited the MR achieved in either dichotic or monaural conditions. In addition, benefit from dichotic presentation was expected to be more pronounced for spectrally smeared stimuli than when only reduced audibility was taken into account.

In the final experiment (Experiment 3), reported in Chapter 4, we tested the impact of audibility and frequency resolution in a computational glimpsing model. There has been some success in modeling speech recognition in fluctuating noise based on the combination of sparse glimpses of the signal and prior template learning (Cooke, 2006). Because windows prime for glimpsing are subjected to spread of masking, it follows that a binaural model based on Cooke’s glimpsing model would also account for the benefit of dichotic listening. Consistent with the effect spectral smearing has in noise, it was expected that the glimpsing model would be consistent with the data obtained from normal-hearing listeners and from normal-hearing listeners with reduced presentation levels and spectrally smeared stimuli in the behavioral experiments. However, no attempt was made to account for higher-level processing of the brain, so the possibility of additional cognitive factors remained a possible explanation of the results.

In Chapter 5, there is a general discussion of the relationship between the results of these studies and those in the literature. Additionally, we summarize the project and briefly discuss future directions for this line of research.
CHAPTER 2: EXPERIMENT 1 – EFFECTS OF HEARING IMPAIRMENT

2.1 Introduction

Studies have shown that taking advantage of high SNRs at masker envelope minima, also known as glimpsing, is important to understanding speech in fluctuating noise. Because the spectral regions associated with high SNRs can vary dynamically in time, glimpsing requires the integration of speech cues across both frequency and time (Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Buss et al., 2004; Cooke, 2006; Hall et al., 2008). Only a few studies have demonstrated hearing-impaired listeners’ abilities to integrate sparse speech cues across frequency and time (Hall et al., 2008; Hall et al., 2012). Hall et al. (2008) concluded that hearing-impaired listeners did not have an inherent deficiency integrating speech cues across frequency, while Hall et al. (2012) demonstrated less MR for hearing-impaired listeners with spectro-temporally-modulated maskers than for normal-hearing listeners. In the presence of noise, hearing-impaired listeners may have to overcome additional factors which normal-hearing listeners are not susceptible to.

In the following experiment, we test the hypothesis that hearing-impaired listeners can benefit from a dichotic presentation of the Async masker used in our previous study (Ozmeral et al., 2012). Our research indicates that spread of masking limits a normal-hearing listener’s ability to fully benefit from glimpses in spectro-temporally complex maskers, and that removing the effect of spread of masking on neighboring bands via dichotic presentation results in improved performance (Ozmeral et al., 2012). Sensorineural hearing loss is
characterized by reduced sensitivity to sound as well as broader auditory filters, and therefore greater susceptibility to spread of masking. As a result, techniques to limit spread of masking may be particularly effective in improving glimpsing of speech for hearing-impaired listeners. Such an outcome would suggest that attempts to improve hearing-aid technology should include efforts to reduce the effects of peripheral spread of masking.

Figure 2.1: Audiograms for normal-hearing (NH; n = 7; dotted lines) and hearing-impaired (HI; n = 9; solid lines) listeners for both left (squares) and right (circles) ears in dB hearing level (HL). HI listeners were screened to have roughly flat and symmetric mild-to-moderate hearing loss. Error bars represent the standard error of the mean.

2.2 Method

2.2.1 Listeners

Sixteen native English-speaking adults were recruited from the local and surrounding communities. Listeners had either normal-hearing (n = 7) with a criterion of 20 dB hearing level or better at octave frequencies from 0.25 to 8 kHz in each ear (ANSI, 2010), or listeners had mild-to-moderate sensorineural hearing loss (n = 9) with mostly flat and symmetric loss.
no worse than 60 dB hearing level. Ages ranged from 21 to 68 years old and were roughly matched across groups (normal-hearing group: $\mu = 42.9$, s.d. = 14.4; hearing-impaired group: $\mu = 46.6$, s.d. = 18.5). Listeners over the age of 60 yrs (1 normal-hearing and 2 hearing-impaired listeners) were asked to complete a cognitive assessment before completing the experiment (Montreal Cognitive Assessment; Nasreddine et al., 2005). Inclusion criteria were set at a score of 26 or better, and all older subjects who met audiometric criteria for inclusion in the study also passed this assessment. Demographic information is reported in Table 2.1, and Figure 2.1 presents the average audiograms for each group.

Table 2.1: Normal-hearing (NH) and hearing-impaired (HI) subject information, including age, pure-tone average (PTA) for 0.5, 1, and 2 kHz for the left (L) and right (R) ears, notched-noise differences (NND) for each ear, and speech reception thresholds of vowel-consonant-vowels (VCV) in quiet. In-quiet speech reception thresholds are not available for two hearing-impaired listeners due to a previous version of the training only requiring 100% accuracy of stimuli at a comfortably loud level. Means are displayed under each column for each group with the standard error of the mean in parentheses.

<table>
<thead>
<tr>
<th>Group- ID</th>
<th>Age (yrs)</th>
<th>PTA-L (dB)</th>
<th>PTA-R (dB)</th>
<th>NND-L (dB)</th>
<th>NND-R (dB)</th>
<th>VCV in Quiet (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH01</td>
<td>18.4</td>
<td>13.3</td>
<td>15.0</td>
<td>5.0</td>
<td>5.0</td>
<td>28.0</td>
</tr>
<tr>
<td>NH03</td>
<td>44.8</td>
<td>6.7</td>
<td>6.7</td>
<td>6.5</td>
<td>6.5</td>
<td>5.4</td>
</tr>
<tr>
<td>NH04</td>
<td>38.1</td>
<td>3.3</td>
<td>-3.3</td>
<td>4.7</td>
<td>4.7</td>
<td>4.8</td>
</tr>
<tr>
<td>NH06</td>
<td>48.3</td>
<td>8.3</td>
<td>3.3</td>
<td>5.5</td>
<td>5.5</td>
<td>7.2</td>
</tr>
<tr>
<td>NH07</td>
<td>45.0</td>
<td>8.3</td>
<td>6.7</td>
<td>4.4</td>
<td>4.4</td>
<td>10.8</td>
</tr>
<tr>
<td>NH11</td>
<td>38.8</td>
<td>1.7</td>
<td>5.0</td>
<td>6.0</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>NH12</td>
<td>66.8</td>
<td>8.3</td>
<td>11.7</td>
<td>4.0</td>
<td>4.0</td>
<td>4.8</td>
</tr>
<tr>
<td>MEAN</td>
<td>42.9 (5.9)</td>
<td>7.1 (1.6)</td>
<td>6.4 (2.4)</td>
<td>5.2 (0.4)</td>
<td>6.1 (0.9)</td>
<td>30.6 (1.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group- ID</th>
<th>Age (yrs)</th>
<th>PTA-L (dB)</th>
<th>PTA-R (dB)</th>
<th>NND-L (dB)</th>
<th>NND-R (dB)</th>
<th>VCV in Quiet (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI02</td>
<td>34.8</td>
<td>43.3</td>
<td>50.0</td>
<td>5.3</td>
<td>-3.5</td>
<td>n/a</td>
</tr>
<tr>
<td>HI08</td>
<td>21.2</td>
<td>55.0</td>
<td>55.0</td>
<td>-0.5</td>
<td>-1.7</td>
<td>n/a</td>
</tr>
<tr>
<td>HI09</td>
<td>56.1</td>
<td>43.3</td>
<td>40.0</td>
<td>3.7</td>
<td>6.8</td>
<td>59.6</td>
</tr>
<tr>
<td>HI10</td>
<td>53.7</td>
<td>41.7</td>
<td>38.3</td>
<td>4.9</td>
<td>4.0</td>
<td>57.5</td>
</tr>
<tr>
<td>HI13</td>
<td>41.5</td>
<td>45.0</td>
<td>41.7</td>
<td>1.0</td>
<td>5.8</td>
<td>67.0</td>
</tr>
<tr>
<td>HI14</td>
<td>23.6</td>
<td>35.0</td>
<td>30.0</td>
<td>2.0</td>
<td>3.8</td>
<td>51.3</td>
</tr>
<tr>
<td>HI16</td>
<td>68.9</td>
<td>33.3</td>
<td>28.3</td>
<td>2.7</td>
<td>6.9</td>
<td>48.4</td>
</tr>
<tr>
<td>HI17</td>
<td>67.1</td>
<td>23.3</td>
<td>30.0</td>
<td>4.8</td>
<td>7.4</td>
<td>46.2</td>
</tr>
<tr>
<td>HI18</td>
<td>52.8</td>
<td>40.0</td>
<td>45.0</td>
<td>4.2</td>
<td>7.4</td>
<td>63.8</td>
</tr>
<tr>
<td>MEAN</td>
<td>46.6 (6.2)</td>
<td>40.0 (3.1)</td>
<td>39.8 (3.3)</td>
<td>3.1 (0.7)</td>
<td>3.4 (1.4)</td>
<td>56.2 (3.2)</td>
</tr>
</tbody>
</table>
2.2.2 Stimuli

Stimuli were identical to the ones found in Ozmeral et al. (2012). The speech material included five recordings each for 12 vowel-consonant-vowels ([b d f g k m n p s t v z] as in /lagal/) spoken by an adult female speaker from this lab and recorded at 44.1 kHz sampling rate. Stimulus duration ranged from 528 to 664 ms, with a mean duration of 608 ms. Each token was normalized to equal root-mean-square (rms) level and filtered into 2, 4, 8, or 16 frequency bands using sixth-order Butterworth band-pass filters. For a given number of bands, filter bandwidths were equivalent in logarithmic units, with bands spanning 0.1 to 10 kHz.

Maskers were based on broadband pink noise samples that, by definition, contained equal energy per octave band. Each masker sample was generated digitally with duration equal to the longest possible speech token plus 300 ms (964 ms total duration). Presentations of speech stimuli began 150 ms after the onset of the noise masker. Masker modulation could be performed either synchronously or asynchronously (i.e., in-phase or out-of-phase across frequency, respectively). Sync maskers were modulated in the time-domain with a 10-Hz square wave. To create Async maskers, the pink noise was filtered into 2, 4, 8, or 16 bands using sixth-order Butterworth band-pass filters. Then, a 10-Hz square wave was applied to each noise band via multiplication, with a starting phase alternating between starting on and starting off in neighboring bands. In order to limit spectral energy to the specified frequency region, 10-ms raised cosines were used to smooth these modulation transitions. The level of the full stimulus was fixed at 85 dB SPL. To achieve this, speech and masker were independently filtered and processed, as described above, then summed at the desired SNR.
(in dB). Before stimulus presentation, speech and noise signals were up-sampled to 48,828 Hz to conform to hardware specifications.

Either monaural (left [L] or right [R] ear only) or dichotic (D) stimuli were presented in a single block of trials. Dichotic stimulation included the odd-numbered bands of the combined speech and noise to the left ear and even-numbered bands to the right ear. In some cases, masker bands were presented to a single ear without the associated speech bands (see dichotic controls described below).

2.2.3 Procedure and conditions

Procedures were similar to those used by Ozmeral et al. (2012). Speech reception thresholds were measured using an adaptive up-down tracking algorithm estimating 50% correct identification (Levitt, 1971). The adaptive computer-controlled test procedure used a custom graphical user interface administered through Matlab (Mathworks, Inc, Natick, MA) on a personal computer. Stimuli were presented through a pair of insert earphones (Etymotic ER-2, Elk Grove Village, IL), and listeners were seated in a single-wall, sound-treated booth. The initial SNR level was set to 10 dB for each condition, and SNR increased or decreased by 4 dB, depending on whether a response from the listener was incorrect or correct, respectively. Step size did not change over the course of the block of trials. Speech reception thresholds (in dB SNR) were determined by computing the mean SNR at the last 24 of 26 track reversals. Trials were blocked by condition, and the order of conditions was quasi-randomly selected for each listener to avoid order effects. In general, each listener performed between three and four tracks for each condition. The fourth estimate was obtained if the first three thresholds were not all within 3 dB of each other. Overall testing time was roughly 5 h, typically spread out over five sessions on multiple days.
Listeners first performed the test with no masking (i.e., in quiet). These measures primarily served as a familiarization tool, but data were also included as another measure of hearing ability (results reported in Table 2.1). For each noise-masking trial, a speech token was randomly selected with replacement, and a masker was generated and summed with the speech signal. Listeners indicated their response by selecting 1 of 12 buttons on the computer screen using a mouse.

Figure 2.2 illustrates the key features of the 28 total conditions described below. The baseline condition was the unmodulated noise condition and was presented monaurally in each ear (Unmod-L and Unmod-R). The Sync condition was presented monaurally to each ear as well (Sync-L and Sync-R). For each Async monaural and dichotic condition (Async-L, Async-R and Async-D, respectively), stimuli were processed into 2, 4, 8, or 16 bands for a total of twelve Async test conditions. Additionally, there were two types of control condition for the Async-D conditions. The first set of control conditions presented the Async-D masker (with 2, 4, 8, or 16 bands) but included only half of the speech bands: in Async-D-EVEN, the even speech bands were presented to the right ear, and in Async-D-ODD, the odd speech bands were presented to the left ear. These control conditions were intended to reveal whether performance in the Async-D conditions could be accounted for solely by either the even or odd speech bands alone. By including the masker in both ears but speech in only one ear, we were also able to test the possibility that maskers could have an across-ear affect (i.e., contralateral effects). Two additional control conditions in each ear were also run that removed the possibility of contralateral effects. In these conditions, only half of the Async-D masker and speech bands were presented (either Async-L/R-ODD in which only the left or

---

6 For the first two hearing-impaired subjects, only 100% accuracy in quiet was required for comfortably loud stimuli, and therefore, in-quiet speech reception thresholds are not available for these subjects.
right ear received odd-numbered frequency bands or Async-L/R-EVEN in which only the left or right ear received even-numbered frequency bands). These conditions were only run using 8 band-pass filters (i.e., 4 bands per ear).

Figure 2.2: Schematic of masker conditions for all experiments. Primary conditions are represented on the top row, and controls are shown below. Only left-ear (L) schematics are visually depicted for monaural conditions, but right (R) ear conditions were also tested. As the legend indicates, each condition is represented as a 2-by-2 box in which the left and right columns represent stimulation of the left and right ears, respectively, and the top and bottom rows represent the speech and noise stimuli, respectively. In each box, frequency from 0.1 to 10 kHz is represented vertically, and a time span of 200 ms is represented horizontally. Speech is represented via spectrogram, and noise is represented by black spectro-temporal regions indicating the “on” periods of masker modulation. Amplitude modulation is performed at a rate of 10 Hz, and frequency bands are filtered in equal widths on a logarithmic scale. The order of the primary conditions in the top row is an indication of the expected ranking in thresholds, with the worst performance starting on the left, with the Unmod-L and Unmod-R conditions, and the best performance on the right, with the Sync-L and Sync-R conditions. The numbers of bands tested per condition are given below each condition schematic. Asterisks indicate bands used only in experiment 1.

To simplify the report on the monaural and dichotic controls, data were analyzed for the better of the associated conditions (ODD versus EVEN bands). For instance, on a subject-by-subject basis, the better threshold in either the Async-D-ODD or Async-D-EVEN was the only dichotic control threshold used to assess the performance on control conditions. The
better of the dichotic controls is reported in Table 2.3 as Control-D, and the better of the monaural controls is reported as either Control-L or Control-R. We used the lower (better) of the two control thresholds to evaluate performance in the primary Async conditions because it would provide the most conservative measure of integration when all bands were available.

![Figure 2.3](image)

**Figure 2.3:** Schematic of the frequency spectrum for the notched-noise method for estimating a listener’s frequency resolution. Listeners performed a 3-AFC task in which they needed to detect a tone at 1500 Hz (vertical line) in the presence of a band-pass noise centered at the signal frequency (top; no-notch condition) or in the presence of two band-pass noises, one on either side of a 450 Hz-wide protected region (bottom; notch condition).

Frequency resolution of each listener was measured by estimating a pure tone threshold in the presence of a either a spectrally contiguous masker or two bands of noise separated by a protected region around the probe frequency (Glasberg and Moore, 1990). Figure 2.3 illustrates the two test conditions. The signal was a 1500-Hz pure tone. In the no-notch condition (top), the masker was a 2100-Hz wide noise centered at the signal frequency. In the notch condition (bottom), a 450-Hz wide protected region centered on 1500 Hz was introduced, and a 1050-Hz band of noise flanked either side of the protected region. The difference between thresholds in the no-notch and notch conditions (notched-noise difference) is inversely proportional to the estimated width of the auditory filter at the signal frequency.
frequency. A large notched-noise difference would be consistent with relatively good frequency resolution, whereas a small notched-noise difference would suggest relatively poor frequency resolution. Individual notched-noise difference values for each ear are reported in Table 2.1.

Table 2.2: Mean speech reception thresholds (in dB SNR) from Experiment 1 are reported for each primary test condition. The standard error of the mean for normal-hearing (n = 7) and hearing-impaired (n = 9) groups is shown in parentheses next to the associated mean. Monaural (L and R) thresholds for Unmod, Sync, and Async conditions were averaged and given the labels, Unmod-M, Sync-M, and Async-M, respectively. The dichotic asynchronously modulated noise condition is labeled Async-D. See Figure 2.2 for visual representation of key conditions.

<table>
<thead>
<tr>
<th>Normal Hearing</th>
<th>Number of Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Unmod-M</td>
<td>-0.3 (0.6)</td>
</tr>
<tr>
<td>Sync-M</td>
<td>-25.5 (1.5)</td>
</tr>
<tr>
<td>Async-M</td>
<td>-17.8 (1.3)</td>
</tr>
<tr>
<td>Async-D</td>
<td>-24.9 (1.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hearing Impaired</th>
<th>Number of Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Unmod-M</td>
<td>0.3 (0.5)</td>
</tr>
<tr>
<td>Sync-M</td>
<td>-7.7 (1.2)</td>
</tr>
<tr>
<td>Async-M</td>
<td>-5.6 (0.9)</td>
</tr>
<tr>
<td>Async-D</td>
<td>-9.1 (1.3)</td>
</tr>
</tbody>
</table>

Table 2.3: Mean speech reception thresholds (in dB SNR) for the dichotic and monaural controls in Experiment 1. Control conditions included the same masker as in the primary test condition but only half of the speech bands; for example, the 8-band Async-D-ODD controls included the Async-D masker with odd bands in the left ear and even bands in the right ear, while only the odd speech bands were included in the left ear and no speech was present in the right ear. For the dichotic control conditions, Control-D was the average the lower threshold in the Async-D-ODD and Async-D-EVEN conditions for each subject. For the monaural controls, the better of the Async-L/R-ODD and Async-L/R-EVEN for each subject is given by Control-L and Control-R, respectively. See Figure 2.2 for visual representation of conditions.

<table>
<thead>
<tr>
<th>Normal Hearing</th>
<th>Number of bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Async-D-ODD</td>
<td>3.1 (2.0)</td>
</tr>
<tr>
<td>Async-D-EVEN</td>
<td>-19.0 (1.4)</td>
</tr>
<tr>
<td>Async-L-ODD</td>
<td>-12.6 (1.9)</td>
</tr>
<tr>
<td>Async-R-ODD</td>
<td></td>
</tr>
<tr>
<td>Async-L-EVEN</td>
<td></td>
</tr>
<tr>
<td>Async-R-EVEN</td>
<td></td>
</tr>
</tbody>
</table>
### Table

<table>
<thead>
<tr>
<th></th>
<th>Number of bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td><strong>Hearing Impaired</strong></td>
<td></td>
</tr>
<tr>
<td>Async-D-ODD</td>
<td>5.0 (2.1)</td>
</tr>
<tr>
<td>Async-D-EVEN</td>
<td>-2.7 (1.8)</td>
</tr>
<tr>
<td>Async-L-ODD</td>
<td></td>
</tr>
<tr>
<td>Async-R-ODD</td>
<td></td>
</tr>
<tr>
<td>Async-L-EVEN</td>
<td></td>
</tr>
<tr>
<td>Async-R-EVEN</td>
<td></td>
</tr>
<tr>
<td>Control-D</td>
<td>-2.8 (1.8)</td>
</tr>
<tr>
<td>Control-L</td>
<td></td>
</tr>
<tr>
<td>Control-R</td>
<td></td>
</tr>
</tbody>
</table>

### Figure 2.4

Figure 2.4: Mean masking release in Experiment 1 is plotted for modulated noise conditions relative to the unmodulated condition for normal-hearing (left panel) and hearing-impaired (right panel) groups. The differences in mean thresholds relative to the Unmod condition at 2, 4, 8, or 16 bands are plotted for the monaural asynchronous condition (Async-M; circles), the dichotic asynchronous condition (Async-D; triangles), the better of dichotic control conditions (Async-B; bowties), and the mean of the synchronous conditions (Sync-M; straight line). Error bars indicate standard error of the mean (n = 7 for normal-hearing group; n = 9 for hearing-impaired group).
2.3 Results

2.3.1 Normal-hearing group speech reception thresholds and masking release (MR)

Mean speech reception thresholds for normal-hearing listeners are presented in Table 2.2 (top) for all primary test conditions. The speech reception thresholds for control conditions are presented in Table 2.3 (top). Thresholds in the Unmod-L and Unmod-R were not significantly different ($t[6] = -0.4, p > .05$), so the average is reported (Unmod-M). Speech reception thresholds in the Sync-R and Sync-L conditions were also not significantly different from each other ($t[6] = 0.6, p > .05$), so the average of these two conditions is shown in Table 2.2 (Sync-M). The speech reception thresholds in the Async-R and Async-L conditions at each band were not significantly different from each other either (2 band: $t[6] = 1.7, p > .05$; 4 band: $t[6] = -0.4, p > .05$; 8 band: $t[6] = 0.9, p > .05$; 16 band: $t[6] = 1.5, p > .05$), so the average of threshold for each ear is shown in Table 2.2 for each band number.

To measure the ability to glimpse speech in a fluctuating masker, data were analyzed in terms of MR, quantified as the difference in speech reception thresholds between a condition with modulated noise and the Unmod-M case. The speech reception thresholds for all modulated masker conditions were significantly better than Unmod-M (paired $t$-tests; $p < 0.05$) – an indication of positive MR for modulated masker conditions. Figure 2.4 (left panel) shows the mean MR (in dB) of the normal-hearing group for the average of the monaural Async conditions (Async-M), the dichotic condition (Async-D), the average of the Sync conditions (Sync-M), and better of the Async-D control conditions (Control-D), expressed relative to the speech reception threshold for the Unmod-M reference value. Error bars show one standard error of the mean, and symbols indicate the masker condition, as defined in the legend. The MR for normal-hearing listeners was greatest for Sync-M (average of 25.3 dB),
intermediate for Async-D (ranging from 18.6 to 24.6 dB), and least for Async-M (ranging from 3.9 to 17.5 dB). A two-way repeated-measures analysis of variance (ANOVA) was performed to compare performance in Async-D and Async-M, with two levels of condition and four levels of band number. This analysis yielded a main effect of condition ($F[1,6] = 108.8$, $p < .001$), a main effect of the number of bands ($F[3,18] = 64.5$, $p < .001$), and an interaction ($F[3,18] = 15.1$, $p < .001$). The interaction is explained by the greater separation between conditions as the band number increases. In all, the results replicate the findings in (Ozmeral et al., 2012) that MR is higher in the Async-D than the Async-M conditions at all bands.

2.3.2 Hearing-impaired group speech reception thresholds and masking release (MR)

Mean speech reception thresholds for the hearing-impaired listeners are presented in Table 2.2 (bottom) for all conditions. Thresholds in the Unmod-L and Unmod-R were not significantly different ($t[8] = 0.7$, $p > .05$) and were averaged for subsequent analysis. The speech reception thresholds for the Sync-R and Sync-L conditions were also not significantly different ($t[8] = 2.1$, $p > .05$), so the average speech reception thresholds of these two conditions is reported in Table 2.2. Finally, speech reception thresholds in the Async-R and Async-L conditions for each band number were not significantly different either (2 band: $t[8] = -0.2$, $p > .05$; 4 band: $t[8] = 2.1$, $p > .05$; 8 band: $t[8] = 1.5$, $p > .05$; 16 band: $t[8] = 1.2$, $p > .05$), so the average speech reception thresholds of these two conditions is reported in Table 2.2 for each band number.

Hearing-impaired listeners’ speech reception thresholds for all modulated masker conditions, reported in Table 2.2, are significantly better than the threshold in the Unmod-M reference case (paired $t$-test; $p < 0.05$). That is, MR was positive for all modulated maskers,
as was also seen with the normal-hearing listeners. Figure 2.4 (right panel) shows MR (in dB) for Async-M, Async-D, Sync-M, and Control-D for the hearing-impaired group, measured relative to the speech reception threshold for the Unmod-M reference. Error bars show one standard error of the mean, and symbols indicate the masker condition, as defined in the legend. The MR for hearing-impaired listeners was greatest for Sync-M (average of 8.0 dB) and for Async-D (ranging from 6.0 to 9.4 dB), while MR was consistently smaller for Async-M (ranging from 2.4 to 5.9 dB). A two-way repeated-measures ANOVA was performed to compare performance in the Async-D and Async-M conditions, with two levels of condition and four levels of band number. This analysis yielded a main effect of condition (F[1,8] = 10.2, p < .05), a main effect of the number of bands (F[3, 24] = 15.7, p < .001), but no interaction (F[3, 24] = 0.4, p > .05), indicating that thresholds were similarly affected in the two conditions as band number changed, which was contrary to what was seen for normal-hearing listeners. In general, the speech reception was better, as indicated by the larger MR, in the dichotic Async condition than the monaural Async condition (i.e., dichotic advantage), as was the case for the normal-hearing listeners.

2.3.3 Between-group analysis of speech reception thresholds and masking release (MR)

Speech reception thresholds in the Unmod-M case were submitted to a one-way ANOVA. This analysis showed no significant difference between the normal-hearing and hearing-impaired listeners (F[1,14] = 0.53, p = .48), which indicated that at an overall presentation level of 85 dB SPL, hearing-impairment did not affect speech reception in steady noise.

On the other hand, it was evident from Figure 2.4 that normal-hearing listeners had greater MR in most modulated-noise conditions compared to the hearing-impaired group. A
one-way ANOVA for the Sync-M case indicated that normal-hearing listeners had significantly greater MR ($F[1,14] = 75.8, p < .001$). With respect to the Async noise conditions, data were submitted to a three-way ANOVA with two levels of presentation type (dichotic and monaural), four levels of number of bands (2, 4, 8, and 16), and two levels of listener group (normal-hearing and hearing-impaired). This analysis showed significant main effects of presentation type ($F[1,14] = 87.4, p < .001$), number of bands ($F[3,42] = 77.9, p < .001$), and listener group ($F[1,14] = 58.1, p < .001$). There were also significant interactions between condition and band number ($F[3,42] = 12.4, p < .001$), between condition and group ($F[1,14] = 23.6, p < .001$), and between band number and group ($F[3,42] = 27.7, p < .001$). Lastly, the three-way interaction was significant ($F[3,42] = 8.8, p < .001$), indicating that effects of band number on the MR in the two Async masker conditions were different between normal-hearing and hearing-impaired listeners. From Figure 2.4, we can see that while normal-hearing listeners tend have less MR in both noise conditions as the number of bands increases, hearing-impaired listeners show relatively flat MR for all numbers of bands. The absence of an effect of band number in the hearing-impaired data may be influenced by the smaller range of MR relative to the normal-hearing listeners.

2.3.4 Dichotic advantage between groups

The differences in MR between Async-D and Async-M conditions – referred to as dichotic advantage – is presented in Figure 2.5 for normal-hearing (striped bars) and hearing-impaired (white bars) listeners. For hearing-impaired listeners, there was a consistent dichotic advantage across all band numbers, whereas normal-hearing listeners generally had greater dichotic advantage as band number increased. The dichotic advantage was between 7.1 and 15.3 dB for the normal-hearing group, and between 2.9 and 4.2 dB for the hearing-
impaired group. A two-way ANOVA with two levels of group and four levels of number of bands showed a clear between-subjects group effect (F[1,14] = 87.4, p < .001), indicating that the normal-hearing group had significantly greater dichotic advantage than the hearing-impaired group. Using a linear-contrast model, a main effect was found for number of bands (F[1,14] = 25.5, p < .001), and there was an interaction with group (F[1, 14] = 8.8, p < .001) – supporting the observation that the dichotic advantage increased with band number for the normal-hearing group, but did not increase as much (if at all) for the hearing-impaired group.

Figure 2.5: Dichotic advantage (i.e., the difference between Async-D and Async-M conditions) for both normal-hearing (NH) and hearing-impaired (HI) groups. Whereas the dichotic advantage increased significantly for the normal-hearing listeners as band number increased, hearing-impaired listeners received similar benefit at all band numbers. Stars indicate significant differences of means at each band number (p < .001).
Dichotic advantage and consonant errors

The observed benefits for dichotic listening are consistent with the ability to integrate speech information from the two ears. It is possible, however, that specific consonants were better positioned spectrally to be glimpsed when only odd- or even-numbered bands were

Figure 2.6: Consonant accuracy for each band number (by row: 2, 4, 8, and 16) for normal-hearing (NH; left column) and hearing-impaired (HI; right column) listeners. Filled triangles represent the Async-D condition and open circles represent the Async-M conditions. Each of the 12 consonants is along the x-axis.

2.3.5 Dichotic advantage and consonant errors

The observed benefits for dichotic listening are consistent with the ability to integrate speech information from the two ears. It is possible, however, that specific consonants were better positioned spectrally to be glimpsed when only odd- or even-numbered bands were
present, and therefore, listeners could have used one ear or the other to identify each consonant. If listeners were benefitting from the dichotic presentation due to certain consonants being easier in a subset of bands, then our interpretation of the dichotic advantage as reflecting integration of glimpses would be undermined. In other words, if the interpretation in terms of spread of masking is correct, that removing spread of masking would lead to better glimpsing opportunities conducive to better integration, then error patterns would be similar across conditions.

Data were analyzed on a consonant-by-consonant basis. Performance was determined for all trials after the second reversal in each condition block. Figure 2.6 shows the accuracies for each consonant for Async-M (circles) and Async-D (triangles) broken down by normal-hearing (left column) and hearing-impaired (right column) listeners and each band number (by row). Data shown in Figure 2.6 were submitted to a four-way ANOVA with two levels of presentation type, four levels of number of bands, twelve levels of consonants, and two levels of listener group. Of interest, was whether the monaural and dichotic presentations interacted with the consonants, because an interaction would indicate that cues to identify consonants were different between dichotic and monaural presentations of the Async masker. The results of the analysis showed a significant interaction (F[11,154] = 2.2, p < .05), and the three-way interaction between presentation type, consonant, and listener group was also significant (F[11,154] = 1.9, p < .05), which indicated that the interaction between the presentation and consonant was different for the normal-hearing and hearing-impaired listeners. Qualitatively, from Figure 2.6, we notice that in most instances, consonants are recognized with similar accuracy between the Async-D and Async-M conditions. However, for normal-hearing listeners, it is possible that in the 8-band condition, some consonants (e.g., /p/, /s/, and /t/)
were more or less difficult depending on the presentation type. There are less visible instances where that is the case for hearing-impaired listeners, which may explain the three-way interaction. The results of this analysis suggests that for 2, 4, and 16 bands, listeners appear to use similar cues to recognize speech in the Async-D and Async-M conditions, but it is possible, that with 8-bands, normal-hearing listeners were utilizing different cues in the two conditions.

![Figure 2.7](image)

Figure 2.7: The difference between thresholds in the Async-D and Control-D conditions for both normal-hearing (NH) and hearing-impaired (HI) groups. All means are significantly different than zero (p < .001), indicating some integration between ears across time and frequency for both groups. For bands 4 and 8, there is a significant difference between groups, with NH listeners higher than HI listeners, suggesting that NH listeners may benefit from having all the speech bands more than HI listeners.

2.3.6 Between-group analysis of controls

Control measures taken in the study were useful in assessing the possibility that a listener was simply attending to a subset of bands – either the even or the odd bands – in the Async conditions, thereby not actually integrating across time and frequency. For both the 2- and 4-
band conditions in Howard-Jones and Rosen’s (1993) original study, there was some question as to whether listeners were simply attending to half of the frequency bands with modulation in phase, but control data indicated that in the 2-band condition, listeners were integrating glimpses across time neighboring frequencies. However, the same control masker conditions in which one band was modulated and the other was steady resulted in comparable speech reception thresholds for the 4-band condition, suggesting listeners could theoretically rely on a portion of the stimulus and were not necessarily integrating glimpses across time and frequency in the Async condition.

In the current experiment, performance in the Async-D conditions was uniformly better than the either Async-D-ODD or Async-D-EVEN control conditions for both groups. Figure 2.7 presents the difference in MR between Async-D and Control-D (the better of the two dichotic controls) for both groups. Values ranged from 5.9 to 10.2 dB for normal-hearing subjects and from 3.5 to 6.3 dB for hearing-impaired subjects, depending on the number of bands. Two-tailed t-tests indicated that Async-D thresholds were consistently better than Control-D thresholds at each band number (2-tailed t-tests, p < .001), consistent with an interpretation that listeners were making use of information from spectral regions associated with both the even and odd bands.  

Recall that in the Async-D-EVEN and Async-D-ODD conditions, the noise-only ear received bands of noise that were modulated out-of-phase relative to the masker modulation in the ear with a speech signal. We compared the Control-D and Control-L/R measures at 8 bands to assess the effect of having the modulated masker in the opposite ear of the ear with speech. For normal-hearing listeners, the monaural controls had 4 dB greater MR than the

---

7 Although integration appears to be independent of the number of bands (F[3,42] = 0.4, p = 0.7), there is a main effect of group (F[1,14] = 292.3, p < .001) and an interaction (F[3, 42] = 4.3, p < .01). The interaction is likely due to a greater difference between groups for the middle band numbers (4 and 8) than either 2 or 16 bands.
dichotic control. Similarly, hearing-impaired listeners had roughly 4.5 dB greater MR in the monaural controls. From these results, it appears that having a modulated masker in non-overlapping frequency regions in one ear can mask speech in the other ear. This is remarkable because in most cases, listeners do not achieve as high of MR in the Async-D conditions as they do in the Sync condition, indicating that although the effects of masking spread have been removed in the Async-D condition, there appears to be some other factor limiting performance in the dichotic condition. We will address these other possible factors in the discussion.

2.3.7 Notched-noise test of frequency resolution

To address the role of frequency resolution in the speech perception results, an estimate of frequency resolution was measured for each listener using the notched-noise method. This frequency resolution measure is taken as the difference in threshold between no-notch and notch conditions (in dB). From Table 2.1, it can be seen that the normal-hearing group had an average of roughly 3-dB greater notched-noise difference than the hearing-impaired group, which is consistent with the notion that for a fixed notch-width, hearing-impaired listeners have a greater-than-normal level of masking energy at the output of an auditory filter centered on the probe-tone frequency. Data were submitted to a two-way repeated-measures ANOVA with a between-subjects factor of group and within-subjects factor of stimulus ear. The effect of group was significant (F[1,14] = 89.5, p < .05), while the two ears did not differ significantly, and there was no interaction. These results support the underlying assumption that frequency resolution in the hearing-impaired listeners was indeed poorer than normal.
2.4 Discussion

2.4.1 Energetic masking release for normal-hearing and hearing-impaired groups

Previous studies by Howard-Jones and Rosen (1993) and our lab (Ozmeral et al., 2012) have shown MR for speech in the presence of Async maskers. In the Howard-Jones and Rosen study, speech and maskers were presented diotically (i.e., identical stimuli presented to each ear). In their study, MR was present for two bands and four bands, but not for greater numbers of bands. Ozmeral et al. (2012) aimed to reduce the deleterious effects of spread of masking by presenting neighboring spectral regions in the noise dichotically (i.e., odd-numbered bands in the left ear and even-numbered bands in the right ear). The result was an average of 7.3 dB better speech reception thresholds across all band conditions in the Async-D condition relative to a monaural Async condition, and data from control conditions suggested that the benefit was a result of integrating speech information from both ears across time and frequency. The current study replicated much of the Ozmeral et al. (2012) experiment for normal-hearing listeners and added a hearing-impaired group to determine whether listeners with sensorineural hearing loss could also benefit from dichotic listening in an asynchronous glimpsing task. Because hearing-impaired listeners are known to have poorer-than-normal frequency resolution, it was hypothesized that MR would be severely impaired in a monaural asynchronous masker, so dichotic listening benefit had the potential to be comparable or even greater than in normal-hearing listeners.

As in our previous study (Ozmeral et al., 2012), the current normal-hearing group showed a large benefit in consonant identification when the signal and Async maskers bands were presented dichotically. On average, this dichotic advantage relative to the monaural conditions was roughly 11.2 dB, which is 3.9 dB larger than the benefit seen in the previous
study. The difference between data from normal-hearing listeners in the current and previous studies could be due to methodological differences. Specifically, the current procedure called for all stimuli to be presented at an overall level of 85 dB SPL, whereas the previous experiment fixed the target level at 55 dB SPL and varied the masker level to measure threshold. Because this task has been shown to be susceptible to changes in presentation level (see Experiment 2 in Ozmeral et al., 2012), and sensation level can influence speech understanding in modulated noise (e.g., George et al., 2006), it is not surprising that we see minor differences in results between the two normal-hearing groups.

The hearing-impaired listeners did not achieve comparable dichotic advantage to normal-hearing listeners in terms of magnitude (only 3.5 dB dichotic advantage on average).

Although this could be seen as being counter to our prediction that hearing-impaired listeners would achieve comparable benefit from dichotic presentation, it is probably inappropriate to make direct comparisons between the dB values of the two groups in the current
configuration. Not only was dichotic advantage significantly smaller in the hearing-impaired group, MR for all modulated maskers was significantly smaller. This result is generally consistent with previous findings showing that hearing-impaired listeners tend to have smaller MR in fluctuating maskers compared to normal-hearing listeners when stimuli are presented at equal levels (Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Gustafsson and Arlinger, 1994; Peters et al., 1998; Snell et al., 2002), especially for single syllable stimuli (Jin and Nelson, 2006). Although some of this result may be explained by lower sensation in the masker dips or possibly poorer temporal resolution (cf. George et al., 2006; Jin and Nelson, 2006), it may still be that MR is also limited by poor frequency resolution (Jin and Nelson, 2010). Given the overall difference of MR between the groups across all conditions, it may be instructive to examine group differences in the Async conditions when groups are normalized with respect to MR in the Sync condition.

### 2.4.2 Normalization to Sync MR

The average dichotic advantage for hearing-impaired listeners was significantly positive, and although it was considerably less than shown by the normal-hearing group, other studies have also shown less MR in hearing-impaired subjects relative to normal-hearing listeners when presentation levels were held constant (e.g., Festen and Plomp, 1990). Therefore, dichotic advantage was also unlikely to reach levels seen in the normal-hearing group because it is associated with the size of the MR in the Async-M and Async-D conditions. In order to interpret dichotic advantage differences between normal-hearing and hearing-impaired groups, it was worthwhile to examine group and condition effects for data analyzed after normalization. Due to the greatest average MR occurring in the Sync condition for normal-hearing listeners, normalization was anchored to MR in the Sync condition.
Figure 2.8 displays the transformed Async-M (circles) and Async-D (triangles) MR data, plotted as a percent of the MR in the Sync condition for the normal-hearing (left panel) and hearing-impaired (right panel) groups. Figure 2.9 displays the corresponding dichotic advantage for transformed data for both groups. Statistical analysis of the normalized data yields group differences in normalized MR ($F[1,14] = 5.1, p < .05$) due to overall greater values for hearing-impaired listeners. However, the dichotic advantage was not statistically different ($F[1,14] = .004, p = .95$), suggesting that each group received a similar benefit with dichotic listening after normalization. The results of this analysis are consistent with an interpretation that while hearing-impaired listeners are limited by their ability to glimpse overall, dichotic stimulation provided a benefit proportionally similar to that achieved by normal-hearing listeners. The consequence of normalization was that, in one way, the two groups could be viewed as having similar benefits of dichotic listening.

2.4.3 Summary

Masking release in all conditions was significantly smaller for hearing-impaired listeners when compared to normal-hearing listeners. The results of Experiment 1, however, indicated that hearing-impaired listeners still benefited from dichotic presentation of stimuli when maskers were asynchronously modulated. Although hearing-impaired listeners appear to have less MR at current stimulus intensities, the amount of benefit from dichotic listening was comparable to that of normal-hearing listeners when analyzed relative to MR in the Sync condition.

The benefits of dichotic listening can be attributed to the reduced effects of spread of masking at the periphery. In hearing-impaired listeners, spread of masking was expected to be especially limiting due to the reduced frequency resolution that is common for listeners.
with sensorineural hearing loss. The evidence suggests that dichotic listening was beneficial for both listening groups. Because hearing-impaired listeners suffer from both reduced audibility and reduced frequency resolution, though, it is difficult to dissociate the roles each have in the current design. In Experiment 2, hearing impairment was simulated in normal-hearing listeners with varying degrees of frequency resolution in order to make this dissociation.

Figure 2.9: Similar to Figure 2.5 except dichotic advantage is normalized to masking release (MR) in the Sync condition. Statistical analyses indicated that each group received comparable dichotic advantage when normalized to MR in the Sync condition.
CHAPTER 3: EXPERIMENT 2 – EFFECTS OF SIMULATED HEARING LOSS

3.1 Introduction

Early studies of normal speech perception in noise observed that performance was better for amplitude-modulated maskers when compared to steady maskers (i.e., MR; Miller and Licklider, 1950). The ability to take advantage of robust speech cues in fluctuating backgrounds has been attributed to integration of glimpses – the available speech information associated with spectro-temporal epochs of favorable local SNR (Howard-Jones and Rosen, 1993; Cooke, 2006; Ozmeral et al., 2012). Studies show that MR is typically smaller in hearing-impaired listeners relative to normal-hearing listeners (as in Experiment 1; Carhart and Tillman, 1970; Duquesnoy and Plomp, 1983; Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Gustafsson and Arlinger, 1994; Peters et al., 1998; Lorenzi et al., 2006; Bernstein and Grant, 2009; Strelcyk and Dau, 2009a). Because fluctuating maskers are common in natural settings, it is unsurprising that hearing-impaired listeners often report problems following conversations where normal-hearing listeners show little difficulty (Peters et al., 1998; Hopkins et al., 2008).

Studies that test speech perception with equal presentation levels for normal-hearing and hearing-impaired listeners are unable to control for audibility of the stimulus as a single factor. It is often impractical to test hearing-impaired listeners at equal sensation levels as their normal-hearing counterparts because presentations can reach uncomfortable levels for the hearing-impaired listeners. In order to control for reduced audibility in the hearing-
impaired listeners, some studies have included conditions that are meant to simulate hearing loss through manipulation of overall stimulus level (Zurek and Delhorne, 1987), inclusion of a threshold-elevating background masking noise (Moore et al., 1995), or other signal-processing strategies (e.g., amplitude expansion; Lum and Braida, 2000).

The results of studies that have made efforts to account for differences in audibility between normal-hearing and hearing-impaired listeners have not always been in agreement. For consonant identification in high-pass noise, normal-hearing listeners and normal-hearing listeners with simulated hearing loss were still found to have significantly more MR than hearing-impaired listeners (Eisenberg et al., 1995). In contrast, Bacon et al. (1998) found through hearing-loss simulations via noise masking, that reduced audibility could account for the relatively small MR in some of their hearing-impaired listeners. Moreover, George et al. (2006) found that hearing-impaired listeners mostly performed similar to normal-hearing listeners with artificially elevated thresholds, and both groups performed worse than normal-hearing listeners. However, small MR was shown to persist in hearing-impaired listeners when noise was high-pass filtered, but this was not the case for the normal-hearing listeners with simulated hearing loss. The results of these studies suggest that audibility can only partly account for the reduced MR benefit in hearing-impaired listeners.

Although reduced temporal resolution has been proposed to limit MR in hearing-impaired listeners (Dubno et al., 2003; George et al., 2006), some investigators have suggested that reduced frequency resolution may play a stronger role (Rosen and Fourcin, 1986; Baer and Moore, 1993; ter Keurs et al., 1993; Baer and Moore, 1994; Peters et al., 1998). Baer and Moore described three significant differences between a normally-functioning ear and an ear with reduced frequency resolution due to hearing impairment: 1)
poorer resolution of harmonics in complex tones and speech; 2) poorer detection and
discrimination of spectral features in speech (i.e., formants), due to a smeared spectral
envelope at the output of the auditory filter (Leek et al., 1987); and 3) a lower ratio of vowel
representation to noise when a wideband noise is centered at the vowel formant energy.

Because the frequency regions associated with high SNRs vary dynamically in time for
spectro-temporally complex maskers, glimpsing requires the integration of speech cues
across both frequency and time. Our research indicates that spread of masking limits a
normal-hearing listener’s ability to fully benefit from glimpses in spectro-temporally
complex maskers, and that removing the impact of spread of masking via dichotic
presentation results in improved performance (see Experiment 1 and Ozmeral et al., 2012).
Hearing-impairment is characterized by reduced sensitivity to sound and broader auditory
filters, and therefore greater susceptibility to spread of masking. As a result, techniques to
limit spread of masking in the asynchronous glimpsing task were effective in improving MR
in hearing-impaired patients, but overall MR was reduced in Experiment 1 relative to normal-
hearing listeners.

Experiment 1 tested normal-hearing and hearing-impaired listeners at the same
presentation level, but it was not clear whether the reduced MR in the hearing-impaired
group was due to reduced audibility, reduced frequency resolution, or a combination of these
factors. The current experiment tested the hypothesis that, for hearing-impaired listeners,
small MR in the asynchronous glimpsing task was due to some combination of the reduced
audibility and reduced frequency resolution. To assess the effects of reduced audibility alone,
normal-hearing listeners were given the asynchronous glimpsing task at a reduced
presentation level relative to the normal-hearing listeners in Experiment 1. The relative
impact of reduced frequency resolution was assessed by comparing thresholds between normal-hearing listeners who received attenuated signals and normal-hearing listeners who received a combination of attenuated and spectrally smeared stimuli. We employed a spectral smearing algorithm (Baer and Moore, 1993) that has previously been shown to successfully mimic the effects of reduced frequency resolution in hearing-impaired listeners and thus reduce speech intelligibility in noise. We hypothesized that lower presentation levels limited the amount of MR overall (Bacon et al., 1998; George et al., 2006). And because spread of masking has a larger effect when frequency resolution is poorer than normal, benefit from dichotic presentation was expected to be more pronounced for spectrally smeared stimuli than when only reduced audibility was simulated (e.g., Leger et al., 2012).

3.2 Method

3.2.1 Listeners

Twenty-four native English-speaking adults (μ = 30.5 yrs., s.d. = 2.6) with no history of hearing loss or ear problems were recruited from the local and surrounding communities. Listeners were randomly assigned to one of three groups. All listeners were screened for normal-hearing, with a criterion of pure-tone thresholds of 20 dB hearing level or better at octave frequencies from 0.25 to 8 kHz in both ears (ANSI, 2010).

3.2.2 Stimuli

Speech tokens were the same as in Experiment 1. Maskers were created in the same fashion as in Experiment 1, except the Async conditions were only tested with 4 and 8 filtered bands. In order to simulate hearing loss, stimuli underwent additional processing before being presented to the listeners. First, a general reduction in audibility was simulated by attenuating the overall level to 55 dB SPL (i.e., 30 dB attenuation relative to Experiment
No effort was made to spectrally shape the simulated hearing loss because hearing-impaired listeners from Experiment 1 had generally flat hearing losses. Secondly, reduced frequency resolution was simulated through a spectral smearing algorithm described by Baer and Moore (1993).

To model the effects of frequency resolution, the signals were submitted to a custom digital processor in Matlab, which simulated the decomposition of spectral bands at the basilar membrane. For signals intended to simulate the effects of normal auditory filters, spectral bands were based on the frequency-dependent equivalent rectangular bandwidth (ERB) given above in Equation 1.1. To produce the spectrally smeared signal, the bandwidth of the normal filter was widened by a multiplier, called the smearing factor (SF). Normal-hearing listeners were assigned to groups based on which SF was used to process the stimuli. The first group received stimuli passed through the smearing algorithm with no change to the filter bandwidths (abbreviated SF1). This group, therefore, was subjected to an overall reduction in stimulus level, but not simulated impaired frequency resolution. In addition to a reduced presentation level, stimuli presented to the next group were processed to simulate a moderately reduced frequency resolution (SF1.7), which has previously been shown to closely resemble effects seen in much of the hearing-impaired population (Tyler et al., 1984; Glasberg and Moore, 1986; Dubno and Dirks, 1989; Peters and Moore, 1992; Stone et al., 1992; Gnansia et al., 2009; Leger et al., 2012). Finally, the third group listened to stimuli which tested the limits of the asynchronous glimpsing task when frequency resolution was severely impaired (SF3). Figure 3.1 shows spectrograms of the Async masker with 8 frequency bands. The effect of spectral smearing on the noise present in glimpsing windows
is evident in Figure 3.1 by comparing different masker energy (gray regions) among the three panels. That is, less glimpsing area (white regions) exists as SF increases.

The sequence of operations for the spectral smearing algorithm was as follows (for additional detail, see Baer and Moore, 1993). First, the stimuli were time-windowed using a Hamming window with a frame size corresponding to 8 ms of the signal. Next, the fast-Fourier transform (FFT) was calculated for each frame, and the power of each time frame was convolved with a smearing function. The smearing function was a 128-by-128 point matrix representing a bank of auditory filters\(^8\) normalized by an ERB multiplied the desired SF. The resulting smeared spectrum was recombined with the original phase spectrum, and an inverse FFT was performed. Finally, the sequence of windows was recombined using the overlap-add method (Allen, 1977).

![Figure 3.1: Spectrograms of an Async masker with 8 frequency bands after spectral smearing by smear factors (SFs) of 1, 1.7, and 3 (left to right).](image)

\(^8\) Each row in the matrix corresponded to the center frequency of the auditory filter. Center frequencies ranged from roughly 0.1 to 10 kHz.
3.2.3 Procedure and conditions

Procedures were the same as in Experiment 1. Listeners were seated in a single-wall, sound-treated booth. An adaptive computer-controlled test procedure used a custom graphical user interface administered through Matlab on a personal computer. The estimated speech reception thresholds corresponded to 50% correct identification (Levitt, 1971). Stimuli were presented through a pair of insert headphones (Etymotic ER-2, Elk Grove Village, IL) at a presentation level of 55 dB SPL. Speech and maskers were independently filtered, mixed at the appropriate SNR (in dB), and sent through the spectral smearing processor. The starting SNR was set to 10 dB for each block. The SNR was increased or decreased by 4 dB, depending on whether a response from the listener was incorrect or correct, respectively. The listener’s estimated threshold (in dB SNR) was determined by computing the mean SNR at the last 24 of 26 track reversals. Thresholds were blocked by condition, and the order of conditions was quasi-randomly selected for each listener to avoid order effects. Each listener performed between three and four tests for each condition. The fourth estimate was obtained if the first three thresholds were not all within 3 dB of each other. Overall testing time per subject was roughly 3h, typically spread out over three 1h sessions on multiple days.

During a trial, the speech token was randomly selected with replacement. Listeners responded by selecting a button with a mouse on the computer screen corresponding to the consonant heard, out of a possible 12 consonants. In all, there were 14 test conditions (Figure 2.2; 4 and 8 bands only for Async conditions and controls), and thresholds were also measured for targets in quiet in order to assess the effect of smearing without noise present. All test condition thresholds were referenced to the average of the unmodulated noise
conditions (Unmod-L and Unmod-R) to calculate MR. Two monaural Sync conditions were tested (Sync-L and Sync-R). For each asynchronous monaural and dichotic condition (Async-L, Async-R and Async-D, respectively), stimuli were processed into 4 or 8 bands for a total of six Async test conditions. The key distinction between monaural (L or R) and dichotic (D) configurations was that the former had stimulus bands presented to only the left or right ear, respectively, whereas the latter had just the even bands presented to the right ear and just the odd bands presented to the left ear.

We also tested the dichotic controls, as in Experiment 1. In the Async-D-EVEN, the even speech bands were presented to the right ear, and the even and odd noise bands were presented to the right and left ears, respectively; in Async-D-ODD, the odd speech bands were presented to the left ear, and the even and odd noise bands were presented to the right and left ears, respectively. These control conditions were intended to reveal whether performance in the Async-D conditions could be accounted for solely by either the even or odd speech bands. These controls were run for both of the Async band number conditions (4 or 8 bands). Note that in these control conditions, one of the ears received no speech signal, but did receive masking bands that were “on” when the speech bands in the other ear were unmasked (i.e., the speech-side maskers were in the “off”-phase). Unlike Experiment 1, monaural controls were not included because it was already confirmed that some masking occurred from the opposite ear from the speech side (see Section 2.3.5).
Figure 3.2: Estimated thresholds (in dB SPL) with no masker present for each smearing group. Smearing groups SF_{1.7} and SF_3 had an average of 5 dB greater thresholds relative to SF_1 which had no smearing.

Table 3.1: Mean speech reception thresholds (in dB SNR) from Experiment 2 are reported for each stimulus condition. The standard error of the mean for each group (n=8) is shown in parentheses next to the associated mean. Recall that control conditions included only half of the speech bands of the associated Async condition; for example, the 8-band controls included only 4 bands of speech.

<table>
<thead>
<tr>
<th>SF_1</th>
<th>Number of Bands</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data</td>
<td>Unmod-M</td>
<td>-2.25 (0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sync-M</td>
<td>-12.40 (1.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Async-M</td>
<td>-9.62 (1.2)</td>
<td>-5.32 (1.2)</td>
</tr>
<tr>
<td></td>
<td>Async-D</td>
<td>-11.58 (1.4)</td>
<td>-9.98 (1.7)</td>
</tr>
<tr>
<td>Controls</td>
<td>Async-D-ODD</td>
<td>-1.51 (2.4)</td>
<td>1.11 (1.9)</td>
</tr>
<tr>
<td></td>
<td>Async-D-EVEN</td>
<td>-2.82 (1.4)</td>
<td>-3.00 (1.8)</td>
</tr>
<tr>
<td></td>
<td>Control-D</td>
<td>-3.98 (1.5)</td>
<td>-3.26 (1.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SF_{1.7}</th>
<th>Number of Bands</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data</td>
<td>Unmod-M</td>
<td>-1.81 (0.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sync-M</td>
<td>-9.82 (0.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Async-M</td>
<td>-7.05 (0.8)</td>
<td>-5.04 (0.7)</td>
</tr>
<tr>
<td></td>
<td>Async-D</td>
<td>-8.24 (0.7)</td>
<td>-7.12 (1.1)</td>
</tr>
<tr>
<td>Controls</td>
<td>Async-D-ODD</td>
<td>1.14 (1.2)</td>
<td>-0.07 (1.4)</td>
</tr>
<tr>
<td></td>
<td>Async-D-EVEN</td>
<td>-1.41 (1.13)</td>
<td>-4.46 (1.4)</td>
</tr>
<tr>
<td></td>
<td>Control-D</td>
<td>-1.78 (1.0)</td>
<td>-4.84 (1.2)</td>
</tr>
</tbody>
</table>
3.3 Results

3.3.1 Speech reception thresholds by number of bands and group

As can be seen in Figure 3.2, which shows the speech-in-quiet thresholds for each group, spectral smearing impaired speech reception when no maskers were present. Both SF1.7 and SF3 led to poorer thresholds in quiet than SF1, which was not spectrally smeared (p < .05 and p < .01, respectively).9

Mean speech reception thresholds are presented in Table 3.1 for all conditions or the average of the monaural conditions. The mean speech reception threshold in the reference (Unmod-M) conditions for SF1, SF1.7, and SF3 were -2.25, -1.81 and 0.13 dB SNR, respectively. Post-hoc multiple comparisons indicate that SF1 and SF1.7 were both significantly different from SF3 (Tukey HSD; p < .05), but not each other. As in Experiment 1, performance in the reference condition was generally poorest. Thresholds in the Sync-R and Sync-L conditions were not significantly different (SF1: t[7] = 1.45, p > .05; SF1.7: t[7] = .033, p > .05; SF3: t[7]=1.58, p > .05) for any group, so the average of these two conditions for each group is shown in Table 3.1 (Sync-M). Thresholds in the Async-R and Async-L

<table>
<thead>
<tr>
<th></th>
<th>SF3</th>
<th>Number of Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Primary data</td>
<td>Unmod-M</td>
<td>0.13 (0.5)</td>
</tr>
<tr>
<td></td>
<td>Sync-M</td>
<td>-6.94 (1.4)</td>
</tr>
<tr>
<td></td>
<td>Async-M</td>
<td>-3.63 (1.1)</td>
</tr>
<tr>
<td></td>
<td>Async-D</td>
<td>-6.94 (1.0)</td>
</tr>
<tr>
<td>Controls</td>
<td>Async-D-ODD</td>
<td>9.31 (1.9)</td>
</tr>
<tr>
<td></td>
<td>Async-D-EVEN</td>
<td>-0.47 (0.8)</td>
</tr>
<tr>
<td></td>
<td>Control-D</td>
<td>-0.47 (0.8)</td>
</tr>
</tbody>
</table>

---

9 This analysis omitted 2 subjects each from groups 1 and 2, and 1 subject from group 3 due to insufficient data. Early versions of the protocol did not include obtaining threshold in quiet. Listeners in that early protocol were only required to get 100% of the test stimuli correct at a presentation level of 55 dB SPL.
conditions at 4 and 8 bands were also not significantly different (all paired t-tests, p > .05), so
the average of these two conditions for each group is shown in Table 3.1 (Async-M).

As with Experiment 1, MR was calculated as the difference between thresholds in the
modulated-noise conditions and those in the reference, unmodulated-noise condition. The
MR for the Sync condition, reported in Figure 3.3, declined as spectral smearing increased. A
linear contrast model confirmed this result (F[1,21] = 6.34, p = .02), indicating that frequency
resolution can affect the ability of the listeners to benefit from masker fluctuation when
presentation level was held constant.

Figure 3.4 shows the mean MR (in dB) of each group for Async-M and Async-D. Error
bars show one standard error of the mean, and symbols indicate the masker condition and
band number, as defined in the figure legend. The data from the monaural and dichotic
Async conditions were submitted to an ANOVA with two levels of band number and three
levels of smear factor. The MR was greater in the Async-D conditions compared to Async-M
(F[1,21] = 77.0, p < .001). There was also a main effect of the number of bands (F[1,21] = 67.9, p < .001), seen by the generally larger MR at 8 bands than 4 bands. Also, there was an
interaction between the condition and number of bands (F[1,21] = 11.9, p < .005), which
appears to have been driven by a greater decline in MR as band number increased in the
Async-D than the Async-M conditions. Between-subject analysis showed that there was not a
main effect of SF group membership (F[2,21] = 2.0, p = .16), but there was a significant
interaction with condition (F[2,21] = 6.5, p < .01) and numbers of bands (F[2,21] = 8.45, p <
.005). The three-way interaction was not significant.

---

10 The better of the dichotic controls (Control-D) were omitted from the figure to better visualize the key
conditions. As can be seen in Table 3.1, these controls were considerably worse than either the Async-M or
Async-D thresholds.
Figure 3.3: Masking release (MR) in the average Sync conditions for each smear factor (SF) group. As smearing increased, MR declined, indicating that poorer frequency resolution can affect the benefit achieved in fluctuating noise when audibility is held constant.

Figure 3.4: Mean masking release (MR) for Async-D (triangles) and Async-M (circles) for each smear factor group at 4 (open markers) and 8 (filled markers) bands.
To better understand the interactions between smear factor and either condition or band number, univariate analyses were run on the individual conditions with a between-subjects factor of smear factor. Greater spectral smearing appeared to reduce MR in the Async-M, 4-band condition, whereas MR remained relatively constant in the 8-band Async-M condition. Post-hoc analysis for just the 4-band Async-M condition indicated that SF₃ was associated with a significantly smaller MR than SF₁ and SF₁.₇ (Tukey HSD; p < .01). Groups SF₁ and SF₁.₇ were not significantly different, however, suggesting that moderate smearing did not impact MR in the Async-M condition with 4 bands. The ANOVA for just the 8-band Async-M condition indicated that the SF did not affect MR (p > .05). As for the effect of spectral smearing on MR in the Async-D condition, a marginal effect was seen for the 4-band condition but not the 8-band condition. An ANOVA revealed that a difference between groups at 4 bands approached significance (F[2,21] = 3.4, p = 0.055), and post-hoc analyses showed that the primary difference was driven by a slightly larger MR in SF₁ relative to SF₁.₇ (Tukey HSD; p = 0.054). A linear contrast of the groups at 4 bands also approached significance (p = 0.068), indicating a trend for smaller MR in the 4-band Async-D condition as spectral smearing increased. The ANOVA did not show a significant effect of smear factor for the 8-band Async-D condition.
3.3.2 Comparison with normal-hearing listeners in Experiment 1

The normal-hearing group in Experiment 1 performed the same task as group SF\textsubscript{1} in the current experiment with the exception that the former group listened to stimuli at an overall level of 85 dB SPL, while the latter listened to stimuli at an overall level of 55 dB SPL. Figure 3.5 presents MR data from these two groups in a similar manner as Figure 3.3. As can be seen, the difference in presentation level led to smaller MR for the SF\textsubscript{1} group in each condition. This result is consistent with previous experiments with normal-hearing listeners, which showed MR for speech presented in an amplitude-modulated noise is dependent on the sensation level of the target (Bacon \textit{et al.}, 1997; George \textit{et al.}, 2006).

Figure 3.5: Masking release (MR) re-plotted for normal-hearing group in Experiment 1 (85 dB SPL) and group SF\textsubscript{1} in Experiment 2 (55 dB SPL), which receive equivalent stimuli with the exception of a level difference.
3.3.3 Dichotic Advantage

Dichotic advantage is plotted in Figure 3.6 for each group at both 4 and 8 bands. A two-way ANOVA with two levels of numbers of bands and three levels of group indicated a significant main effect of bands (F[1,21] = 11.9, p < .005), but no interaction between bands and group. There was a main effect of group (F[2,21] = 6.0, p < .01), which was evident by the greater dichotic advantage in the SF_3 group relative to the groups with less spectral smearing. Whereas all groups had significant dichotic advantage for both 4 and 8 bands (p < .001), results were mixed with respect to the magnitude of this advantage. Univariate analysis indicated no significant difference between groups at 4 bands (F[2, 21] = 2.0, p = .16), but there was a significant difference between groups at 8 bands (F[2,21] = 5.27, p < .05). At 8 bands, post-hoc analysis revealed that the dichotic advantage for SF_1.7 was significantly smaller than for SF_3 (p < .05), which indicated that dichotic listening in the 8-band Async noise is more beneficial when the smear factor was large. However, an interesting and significant U-shaped trend was observed (quadratic contrast model; p < .01), which showed that although spectral smearing affected dichotic advantage positively from SF_1.7 to SF_3, the opposite was true from SF_1 to SF_1.7. This effect was particularly odd considering the expected effects of spectral smearing. That is, spectral smearing was expected to have a monotonic effect on the performance in the task. Because the current study cannot directly compare the effects of spectral smearing to listeners with a wide range of frequency resolution, we were unable to determine whether the U-shape pattern reflected the comparable range from mildly to severely reduced frequency resolution.
Control measures taken in the study were useful in assessing the possibility that a listener was simply attending to a subset of bands – either the even or the odd bands – for the Async conditions. On average, SF$_1$, SF$_{1.7}$, and SF$_3$ had 7.2, 4.4, and 5.1 dB greater MR in the Async-D than the better of the dichotic control conditions (Control-D), respectively. Data for each group at each band are plotted in Figure 3.7. A two-way ANOVA with two levels of number of bands and three levels of smear factor indicated a significant main effect of bands (F[1,21] = 18.6, p < .001), but no interaction between bands and SF. There was also a marginally significant main effect of SF (F[2,21] = 3.4, p = .052). Multivariate analyses indicated that there was a sizeable difference between the three levels of SF at 8 bands (F[2,21] = 40.9, p < .005), but not at 4 bands. A post-hoc multiple comparisons test revealed that SF$_1$ was significantly different from SF$_{1.7}$ and SF$_3$ (Tukey HSD; p < .005 and p < .05,
respectively), but $SF_{1.7}$ and $SF_3$ were not significantly different from each other. The current analysis indicated that listeners performed better in the dichotic Async condition than with only a subset of bands, but it unclear whether differences between each measure are due to differences in integration or differences in the available speech cues in either the Async-D-EVEN or Async-D-ODD alone. That is, if useful speech cues are associated with both subsets of bands, then better performance for all bands present should reflect good integration, but if useful speech cues are associated with only one subset of bands, better performance for all bands present may be primarily due to those dominant speech cues alone (i.e., poor integration). Nevertheless, some degree of integration was evident for each Async-D condition because the difference with the better control was consistently positive.

3.3.5 Difference in baseline SNR

Generally, normal-hearing listeners achieve larger MR when the baseline SNR is more negative (Oxenham and Simonson, 2009), which may be influenced by the performance-intensity function of speech perception in noise (Bernstein and Grant, 2009). The performance-intensity function indicates how much change in speech reception is gained by a change in level – at medium levels, small changes in level will lead to large performance differences, whereas at low and high levels, small changes in level do not affect performance very much. Bernstein and Grant (2009) argued that comparisons of MR measured from different baseline SNRs can be misleading because the degree of MR is dependent on the underlying performance-intensity function.
Whereas numerous studies have shown that hearing-impaired listeners are less able to benefit from the introduction of masker fluctuation compared to normal-hearing listeners (e.g., Festen and Plomp, 1990; Peters et al., 1998; George et al., 2006), Bernstein and Grant (2009) note that these particular studies were undermined by a confound between group differences in the baseline SNR. In Experiment 1 (Chapter 2), baseline SNRs were not found to be significantly different between groups, so this issue was less of a concern; however, SF1 and SF1.7 in Experiment 2 were shown to have significantly different baseline SNRs compared to SF3. Further analysis was warranted in order to determine whether differences in the magnitude of MR were associated with the differences in the ability to listen in the dips or some other factor. While we cannot normalize the baseline SNR at threshold for 50%
correct in the three groups without substantially changing the task (e.g., degrading stimuli in the low-SF conditions), we can compare performance across groups at a common baseline SNR by adjusting the percent correct associated with threshold for each group. Specifically, because psychometric functions tend to have different slopes for modulated-masker conditions compared to steady-noise conditions (Festen and Plomp, 1990; Bernstein and Grant, 2009; Oxenham and Simonson, 2009), different baseline SNRs can make a difference in the amount of MR the listener can achieve.

In order to investigate possible effects related to group baseline differences, psychometric functions were fitted to individual listeners’ data using the psignifit toolbox version 2.5.6 for Matlab (see http://bootstrap-software.org/psignifit/), which implements the maximum-likelihood method described by Wichmann and Hill (2001). By fitting the psychometric functions we were able to estimate percent correct at a desired SNR. This allowed us to fix SNR in the baseline condition for all SF groups, extract the percent correct for that SNR, and use the same percent correct value to extract the corresponding threshold (in dB SNR) for each of the modulated-masker conditions.

To test whether differences in baseline SNR affected MR among the SF groups, the fitted functions for each group were used to determine MR when measured relative to the baseline SNR in the SF1 group.11 The purpose of anchoring to a single SNR was to allow comparisons across groups with a common baseline SNR. In Figure 3.8, reference threshold data estimated with the adaptive tracking method are displayed (squares; ‘data’), along with the threshold estimated from the fitted psychometric curve at 50% accuracy (stars; ‘fitted data’) and the threshold estimated from the fitted psychometric curve for a normalized SNR

---

11 Any anchor could be used here. The baseline SNR in group SF1 was chosen because this group represented listeners with only a reduced audibility but normal frequency resolution.
(diamonds; ‘fitted data – normalized’). This figure illustrates two points: 1) the fitting procedure produced estimating 50% correct produced thresholds similar to the adaptive procedure (Levitt, 1971; compare squares to stars); and 2) SF3 had a significantly higher reference threshold than SF1 and SF1.7 (Tukey HSD; p < .05) for the behavioral data and fitted psychometric, but not for the fitted psychometric with a normalized SNR. This shows that normalization achieved the intended effect: it normalized SNR at (normalized) threshold in the baseline condition.

The MR was computed based on thresholds estimated in one of two ways: using the adaptive method, the 50% correct point in the fitted function, and the normalized threshold in the fitted function. The goal was to show whether differences in baseline SNR using the adaptive methods fully accounted for the differences in MR observed between SF groups.

Figure 3.8: Speech reception thresholds (SRTs) for the Unmod condition in each smear factor group. Symbols represent mean thresholds estimated using the adaptive track procedure (squares; ‘data’), based on 50% correct in the fitted psychometric function (stars; ‘fitted data’), and based on the normalized threshold in the fitted psychometric function, corresponding to the baseline SNR in the SF1 group (diamonds; ‘fitted data – normalized’).
Figure 3.9 displays the MR for the Sync condition from the fitted data (stars; ‘fitted data’) and the fitted data with a fixed baseline SNR (diamonds; ‘fitted data – normalized’). The MR for the fitted data with and without a fixed baseline SNR were submitted to a two-way ANOVA with two levels of method of measuring MR and a between-subject factor of smear factor. The analysis showed that the three methods did not significantly differ (F[1,21]=2.1,p=.186), the groups did not differ (F[2,21]=1.8,p=.189), and there was no interaction (F[2,21]=.91,p=.455). This indicated that while the slopes of the underlying psychometric functions of the Sync and Unmod condition may have been different, a difference in baseline SNR for SF3, for example, did not affect the interpretation of MR for conditions with the higher smear factor.

3.4 Discussion

Figure 3.9: Masking release (MR) in the Sync condition for each smear factor. Symbols represent mean thresholds estimated based on 50% correct in the fitted psychometric function (stars; ‘fitted data’) and based on the normalized threshold in the fitted psychometric function, corresponding to the baseline SNR in the SF1 group (diamonds; ‘fitted data – normalized’).
In Experiment 1, we tested normal-hearing and hearing-impaired listeners on the asynchronous glimpsing task. Although normal-hearing listeners clearly displayed higher MR than hearing-impaired listeners, there was no way of discerning whether the hearing-impaired listeners were limited by reduced audibility or reduced frequency resolution or some combination of the two. In Experiment 2, we tested normal-hearing listeners with simulated hearing loss, including a reduced presentation level (relative to Experiment 1) and spectral smearing. By varying the level of spectral smearing and keeping sensation level constant, the current experiment displayed unique effects of reduced frequency resolution on speech perception in fluctuating maskers.

The results showed that spectral smearing can affect speech reception in noise as well as speech in quiet. The effect of spectral smearing on speech identification has been a matter of debate in the literature. Although a recent study by Leger et al. (2012) indicated that spectral smearing had deleterious effects on speech recognition in quiet, which is consistent with our data, several other studies have failed to show such an effect (ter Keurs et al., 1992; Baer and Moore, 1993; ter Keurs et al., 1993; Baer and Moore, 1994; Gnansia et al., 2009). The discrepancy with the earlier studies may be due to the fact that the earlier observations were measured near ceiling (Leger et al., 2012), which means spectral smearing may not have had as much of an effect as it would if thresholds were measured in a more sensitive region of the psychometric function. It is not entirely clear why spectral smearing would affect performance on consonant recognition in a vowel-consonant-vowels context in quiet. Consonant identification may be more susceptible to the negative effects of spectral smearing when no other cues or redundancies are available. Unlike our stimuli, sentence materials can be recognized predominantly by the vowel cues (Fogerty et al., 2012), which tend to be
higher in amplitude and longer in duration than consonants; contextual cues associated with sentences also provide robust cues for speech recognition (Miller et al., 1951).

In addition to effects on threshold in quiet, spectral smearing at the extreme (SF₃) also had a significant effect on threshold in steady noise (see Figure 3.9, filled squares). Similar to results found in Leger et al. (2012), there was a significant effect on threshold in the Unmod masker for SF₃ relative to both a SF₁ and SF₁.₇. Differences in baseline SNR between groups was also seen by Gnansia et al. (2009), albeit with slightly different smear factors tested. The implications of the differences between SNRs at threshold in the baseline condition were explored above using fixed baseline SNRs in fitted psychometric functions, and it was determined that differences in baselines did not alter the relationship between groups regarding MR differences.

In previous studies that simulated reduced audibility with noise elevation in normal-hearing listeners, MR for amplitude-modulated maskers was found to resemble that of hearing-impaired listeners (Eisenberg et al., 1995; George et al., 2006). The results of Experiment 2 suggest that audibility plays a major role in the size of MR for modulated maskers. An effect of level is evident when comparing the normal-hearing group in Experiment 1 to SF₁ in Experiment 2, in which the only stimulus difference was a presentation level of 85 dB or 55 dB SPL, respectively. Attenuation of the presentation level appears to account for most of the decline in MR for the Sync condition in hearing-impaired groups from Experiment 1, as indicated by the difference in MR between the normal-hearing group in Experiment 1 and the SF₁ group in Experiment 2 (compare an average 25.9 dB MR to 10.2 dB, respectively). The MR level dependence has also been shown for pure tones in a
broadband masker (Zwicker and Schorn, 1982; Moore and Shailer, 1991), although not all studies demonstrate level effects with speech stimuli in noise (cf. Summers and Molis, 2004).

The effect of spectral smearing approached significance for MR in the Sync condition (see Figure 3.3), and a linear contrast model confirmed the trend for reduced MR as spectral smearing increased. While this is contrary to the findings by Leger et al. (2012) using similar stimulus processing, other studies have provided evidence that poorer frequency resolution can impact speech understanding in amplitude-modulated noise (Gnansia et al., 2009; Jin and Nelson, 2010). Leger et al. (2012) have proposed that different baseline SNRs may account for the lack of MR differences among SF groups, but we have already established that this likely was not the case here. Although, it is hard to determine the MR in dB from the percent scores reported by Leger et al., it is possible that their study demonstrated overall smaller MR than ours, so group differences may have been difficult to show statistically. This is especially worth considering because their task was more difficult than the task in the current study. Because Leger et al. used four times as many vowel-consonant vowels as we did, MR was likely to be more limited (e.g., Experiment 2 in Ozmeral et al., 2012).

To understand the effects of spectral smearing on MR in the Async maskers, analyses were run on MR values as well as the dichotic advantage measure. Our hypotheses predicted greater dichotic advantage for listeners who were more impeded by spread of masking in the monaural asynchronous conditions (e.g., SF1.7 and SF3). In the case of dichotic advantage, the groups were significantly different at 8 bands, and a linear contrast model was consistent with a greater benefit as spectral smearing increased. This effect was primarily driven by the SF3 group, which had greater MR for the Async-D condition than the other groups, while MR in the Async-M condition was relatively constant across groups. At 4 bands, however,
there was less of an effect of spectral smearing. Greater spectral smearing led to smaller MR in both Async-M and Async-D condition, which led to a relatively similar dichotic advantage among the groups. This is consistent with the interpretation that smearing is more detrimental with larger numbers of more closely spaced bands.

The data in Experiment 2 suggest that reduced sensation levels led to smaller MR overall (relative to MR for normal-hearing listeners in Experiment 1). In addition, spectral smearing – processing which mimics impaired frequency resolution in hearing-impaired listeners – further reduced MR in the Sync condition, and other smearing effects were seen for dichotic advantage when frequency bands were narrow. Together, these results support roles for both audibility and frequency resolution in the ability to benefit from masker fluctuation – two factors that are impaired in listeners with sensorineural hearing loss. Whereas the effects of spectral smearing appeared to be consistent with the effects of reduced frequency resolution for performance in the Async-M condition, an unusual pattern emerged in the Async-D conditions, in which small MR was seen with moderate smearing but relatively large MR with severe smearing. The U-shaped pattern was an unexpected result and requires further studies to understand fully. Specifically, future work would need to test hearing-impaired listeners with a range of frequency resolution in order to determine whether there is a non-monotonic effect to MR in the Async-D condition as the severity of frequency resolution increases. The following chapter presents a computational glimpsing model that attempts to capture the data patterns from Experiments 1 and 2. Specifically, we were interested in whether a binaural glimpsing model would show a dichotic advantage for the Async masker, and whether the effects of spread of masking could account for this advantage.
CHAPTER 4: EXPERIMENT 3 – A BINAURAL GLIMPSING MODEL

4.1 Introduction

The previous experiments show that listeners are adept at glimpsing speech information in the presence of modulated maskers. In part, this very likely to be due to life-long experience of listening to speech in spectro-temporally complex backgrounds at a range of different SNRs (Hall et al., 2012). As we saw with the listeners in Experiments 1 and 2, some maskers have a greater effect on speech recognition than others. For example, Sync noise conditions generally led to better speech reception thresholds than Async noise conditions. The primary hypothesis of the previous experiments was that spread of masking could account for less MR in monaural Async than Sync noise. Experiments 1 and 2 found greater MR when Async noise conditions were presented dichotically – a method which limited the effect of spread of masking on neighboring frequency bands by separating the out-of-phase bands to opposite ears – than when Async noise conditions were presented monaurally. The purpose of Experiment 3 was to use a computational model that could capture the dichotic advantage, and to test the effects of reduced audibility and reduced frequency resolution – two impairments seen in listeners with sensorineural hearing loss.

It was recently shown that speech recognition in noise can be accurately modeled by providing a computer speech recognizer sparse speech information based on spectro-temporal epochs of favorable local SNR (Cooke, 2006). Cooke looked at three types of maskers with various levels of spectro-temporal complexity: steady-state noise, speech
babble, and a single talker background. Behavioral measurements in a related study (Simpson and Cooke, 2005) showed that at equivalent sound levels, performance began to degrade as the masker became more steady-state and lost glimpsing regions. The model analyzed separate speech and noise spectro-temporal excitation patterns (STEPs; Moore, 2003) – the output of a bank of auditory filters centered at a large range of frequencies. The model compared the separate speech- and noise STEPs to determine regions of favorable local SNR. The regions of the speech STEP that exceeded some glimpsing threshold (in dB) relative to the noise STEP were passed along to a computer speech recognizer, which had been previously trained on clean speech STEPs. The results of this study were clear – as the noise input approached steady state, fewer high-SNR glimpses were available for the speech recognizer, and consequently, the model accurately predicted a decline in performance.

Figure 4.1 provides a visual representation of the STEP of a target speech sound /afal/ and three different maskers of varying fluctuations at a long-term average (global) SNR of -6 dB. The bottom panels of Figure 4.1 shows potential glimpses – spectro-temporal windows exceeding a local SNR of 3 dB – for the maskers (middle row from left to right): Sync, 8-band Async, and Unmod. As noted by Cooke (2006), a depiction like Figure 4.1 demonstrates why global SNR is often a bad predictor of speech perception across different types of maskers. In this case, the total glimpsing area – the cumulative area of individual glimpsing windows – and the proximity of glimpses across time and frequency vary widely across the three maskers. Even though each has an equivalent average intensity, speech recognition performance is directly proportional to total glimpsing area (compare warm colors in three bottom panels of Figure 4.1).
Figure 4.1: Outputs of the spectro-temporal excitation pattern (STEP) processor for the speech token /aga/ (top-center panel) and three maskers (from left to right): Sync, 8-band Async, and Unmod noise (middle row). Each masker is scaled so that the long-term average SNR is -6 dB. The bottom row illustrates the output of the glimpsing model – regions of the speech STEP which exceed a local SNR of 3 dB re the noise STEP.

In the following design, the model was expected to provide a quantitative assessment of the hypothesis that reduced frequency resolution contributes to poor MR in spectro-temporally complex maskers in hearing-impaired listeners. Consistent with the effect of broader auditory filters on the STEP, it was expected that the glimpsing model would capture differences in performance found between normal-hearing and hearing-impaired listeners in the previous behavioral experiments (Experiments 1 and 2). We aimed to accurately predict poor performance in the monaural asynchronous conditions due to spread of masking’s influence on glimpsing in a single ear. If this interpretation of the behavioral data is correct,
then a similar result should be obtained when the model is modified to incorporate reduced frequency resolution.

Figure 4.2: Processing architecture of the STEP processor. See text for details.

4.2 Method

4.2.1 STEP processor

The STEP processor has been described in previous studies (cf. Moore, 2003; Cooke, 2006). As the flow chart in Figure 4.2 illustrates, the processor begins with a 40-channel gammatone filterbank (Patterson and Moore, 1986) implemented using a toolkit described by Slaney (1998). To model the effects of normal frequency resolution, gammatone filter widths are based on an equivalent rectangular bandwidth, given in Equation 1.1. For modeling the effects of reduced frequency resolution, gammatone filters can be widened by the smear factor (SF; also see Chapter 3), which corresponds to an ERB multiplier.
A Hilbert transform was applied to the output of each channel of the gammatone filterbank in order to extract the envelope of the waveforms. The envelopes were smoothed with a leaky integrator function with an 8-ms time constant (τ; Moore et al., 1988). Next, the channels were down-sampled so that each point was associated with a 10-ms window (W) of the envelope. Lastly, the signal was represented in dB through log-compression. The resulting output was 40 vectors representing the energy present in 40 frequency bands spanning 0.1 to 10 kHz, similar to a spectrogram but also accounting for processing by the auditory filter. Images in Figure 4.1 are sample outputs of the STEP processor for speech (top-center panel) and maskers (middle row).

![Hidden Markov Model Diagram](image)

Figure 4.3: Probabilistic parameters of a hidden Markov model (HMM) of a series of coin flips (e.g., heads, tails, tails, heads, etc...). In this example, there are 2 states (X1, X2 representing 2 concealed, weighted coins) each with an outcome probability (green lines indicate high probability, red lines indicate lower probabilities, and the two probabilities add up to 1), and a transition probability for either staying at its current state or changing to the other state. In this example, X1 is more likely to be tails, X2 is more likely to be heads, and changing states is less likely than staying at the same state as the previous state. In order to model a particular sequence of coin selections, the model evaluates all possible sequences of states and calculates the joint probabilities of both the state sequence and the observations for each event.

4.2.2 Hidden Markov models (HMMs)

The current model employed a dynamic Bayesian network commonly used for automatic speech recognition algorithms, called hidden Markov models (HMMs; Baum et al., 1970).
Because we used a freely available toolkit (HTK Toolkit v3.4.1; licensed by Microsoft Corp., Redmond, WA) to create the needed HMMs, it is beyond the scope of this chapter to report on the specific process needed for HMM creation; moreover, multiple resources are freely available for this purpose (e.g., Odell et al., 1994; Young et al., 2006). It is, however, necessary to report a basic structure of HMMs and how they can be utilized for automatic speech recognition.

Modeling with HMMs can be especially useful for applications with temporal pattern analyses, such as speech, handwriting or gesture recognition. The HMMs are based on unobserved (“hidden”) states that model a temporal sequence of events. For example, consider a scenario in which we are given a sequence of coin flips from two concealed coins. Next, let us assume that one coin is weighted towards tails while the other is weighted towards heads, and the probability for choosing one coin depends on the choice of the previous coin. Based upon the sequence of outcomes and these two assumptions, we can create an HMM that predicts the order of the coins used to produce the observed sequence.

Figure 4.3 shows a 2-state diagram for the coin-flip example. Each weighted coin is a hidden state (X) of the HMM. The state, X1, is weighted toward tails, and X2 is weighted toward heads (green lines indicate high probability, red lines indicate low probability, and the probabilities of the two lines add up to 1). The likelihood of changing coins is the transition probability, which in this example, is weighted toward staying in its current state. This indicates that the outcome of flipping depends on the outcome probability of the current state, which in turn, depends on the transition probability of the previous state. By solving the joint probabilities (i.e., transition and outcome probabilities) of different state sequences, the

---

12 The same probabilities (red and green lines) are used in the example for the outcomes and transitions, but this is for the sake of simplicity and not a requirement of HMMs.
HTK Toolkit can find some maximum likelihood estimation that best fits the sequence of the observed heads and tails. In the case of more than one model having equal likelihoods, the toolkit settles on the HMM with the fewest transitions.

For an HMM-based speech recognizer, such as the one used in the present model, each vowel-consonant-vowel (e.g., /agal, /azal, etc...) is first transformed into a sequence of time vectors representing the energy in each frequency channel of the waveform (i.e., the STEP). The stimuli are then used as training data for the HTK Toolkit. After a sufficient amount of training data is fed to the system, a unique sequence of states is identified and associated with each particular speech utterance. To test new sequences of observations, the speech recognizer uses the unique HMMs supplied from the training and weighs the likelihood that each HMM could have produced the observed sequence. The HMM associated with the highest likelihood is considered to be the solution for that trial (for additional detail, see Rabiner and Juang, 1986; Rabiner, 1989; Cooke, 2006).

4.2.3 Binaural adaptation

Figure 4.4 shows the architecture of the binaural glimpsing model adapted from the monaural glimpsing model by Cooke (2006). Input signals were the speech and masker channels for the left and right ears, as well as a noise floor. Stimuli were initially transformed to a spectro-temporal representation by the STEP processing stage (see Figure 4.1, top and middle rows), which can accommodate varying levels of frequency resolution. Speech inputs were sent to the HTK Toolkit to train the HMMs required for speech recognition. Before combining with maskers to form a composite representation, speech inputs were summed with the noise floor – a threshold-elevating Gaussian background intended to mimic the effects of reduced audibility. The composite representation was determined by the dominant
energy in the left and right ears, and was only consequential when the left and right ears received different signals, as in the dichotic conditions described in Experiment 1, 2, and below. Figure 4.5 shows the composite transform process for both the speech (top row) and noise (middle row) inputs, and the subsequent glimpses determined by the model (bottom row). This figure illustrates how combination across ears provides more information to the speech analyzer than either ear alone.

Figure 4.4: Processing architecture of the decision model, including the preprocessing of the signals, the glimpse detection model, and the speech recognizer. Inputs are the left (L) and right (R) ears for the speech (S) and noise maskers (N), as well as a noise floor (NF). Clean speech is used to train the HMMs (see text for details). Speech preprocessing includes the STEP processor and composite (C) transformation which combines information from the two ears based on the highest SNR in the corresponding frequency channels and time windows in each ear. Glimpse detection allows speech inputs to pass if spectro-temporal windows are greater than a user-defined local SNR.

The last stage before speech recognition was the glimpse detection model. This stage analyzed the local SNRs for the left ear, right ear, and the composite representation. The output of the glimpsing detection model was three matrices (left, right, and composite signals) representing only the sparse speech cues that exceeded the local SNR parameter.
Examples of the output are shown in the bottom rows of Figure 4.1 and Figure 4.5. The speech recognizer then used the HMMs based on training with unmasked speech to make a likelihood estimation.

Figure 4.5: Each panel represents the spectro-temporal excitation pattern (STEP) of the signal, with time on the x-axis and frequency on the y-axis. A binaural, 8-band example of a speech signal /agau/ (top row) presented with an Async-D masker (middle row) and the resulting glimpse detection by the model (bottom row). The left ear (first column), right ear (second column), and composite (third column) signals are all inputs to the glimpsing model. The composite transform can be clearly seen to incorporate the speech information from both the left and right ears. The corresponding composite noise incorporates only the noise present at the same spectro-temporal windows used for the composite speech.
4.2.4 Parameterization of the model

As can be seen from the description of HMMs and the binaural glimpsing model, there were a number of free parameters. The parameters for the STEP processor have already been reported above. As for creating the HMMs, it was necessary to find an optimal number of states to be associated with each speech utterance. In its current form, there were 40 states. While this number is relatively high for similar models using HMMs (e.g., Cooke [2006] required only 6 states), this number was needed to achieve better than 95% correct performance on speech tokens in quiet. The reason for so many states is likely due to the homogeneity across training tokens (i.e., the same female speaker and only 12 vowel-consonant-vowels with only a single vowel). We expect that as the training set becomes more diversified, fewer states would be required.

The next parameter was the noise floor needed to simulate the effects of reduced audibility. To model behavior in normal-hearing listeners, no noise floor was included in the model. In contrast, to model behavior in hearing-impaired listeners and normal-hearing listeners tested under conditions of reduced audibility, a noise floor was included to elevate thresholds in the model. Figure 4.6 shows the speech reception thresholds for the model with a noise floor presented at a range of intensities (in 5 dB steps) in both the Sync (diamonds) and Unmod (squares) conditions. The noise floor is expressed in dB relative to the speech. The left-hand side of the graph has the results for a noise-floor level that was infinitely (Inf) below the speech level (i.e., no noise floor), and we ultimately chose a noise floor of -35 dB SNR (in red) to model the effects of reduced audibility in listeners because the speech reception thresholds for both the Unmod and Sync case fit best with the normal-hearing behavioral data in Experiment 2, group SF1 (-2.25 dB and -12.40 dB, respectively). The final
parameter considered was the local SNR needed for glimpse detection. In our simulations, predicted thresholds in each masking condition were affected differently by adjusting this parameter. For example, if we excluded the noise floor, MR in the Sync condition was directly proportional to the local SNR parameter; however, if the noise floor was included, there was not a distinct effect of local SNR. Figure 4.7 shows the values of the Sync MR for a range of local SNRs from 0 to 20 dB (2 dB steps) without a noise floor (filled diamonds) and with a noise floor (open diamonds). A local SNR of 8 dB (indicated by red color in Figure 4.6: Speech reception thresholds (SRTs) in the Unmod (square markers) and Sync (diamond markers) conditions as a function of the noise floor parameter in the model. The noise floor parameter was intended to simulate elevated thresholds resulting from reduced audibility. The predicted SRT in a model omitting the noise floor is represented by –Inf, indicating that the noise floor was infinitely below the level of the speech. The smear factor was set to 1, so these data represent predicted data for listeners with normal frequency resolution. Local SNR was set to 8 dB. Red color indicates the chosen noise floor for subsequent model analyses that included a noise floor. Error bars represent the standard error of the mean (n = 100), but in most cases, were too small to exceed the size of the symbols.
Figure 4.7) was ultimately chosen because it produced roughly 23 dB of MR in the Sync condition with no noise floor, which was similar to the results in Experiment 1 for normal-hearing listeners.

![Figure 4.7: Masking release in the Sync condition as a function of the local SNR used for glimpse detection in the model. Smear factor was set to 1, and noise floor was either present (at -35 dB SNR; filled diamonds) or not (open diamonds). Red color indicates the chosen local SNR for subsequent model analyses. Error bars indicate the standard error of the mean (n = 25).](image)

4.2.5 Stimuli

Twenty-one tokens of the 12 vowel-consonant-vowels (including the 5 tokens of each vowel-consonant-vowel used in Experiments 1 and 2) were used for both training and testing the speech recognizer. Tokens were recordings by a female speaker and had an average duration of 641 ms. Each token was filtered into 4 and 8 bands and normalized to an rms
level of 1. Filter bandwidths were equivalent in logarithmic units, with bands spanning 0.1 to 10 kHz. Monaural, odd-numbered-channel, and even-numbered-channel speech tokens were passed through the STEP processor and saved as a Matlab matrix file. In all, there were 1512 files (12 consonants x 21 tokens x 2 bands x 3 channel configurations) used for training the HTK Toolkit, as well as testing in the simulated asynchronous glimpsing task.

A single masker was generated for each condition and SNR, spanning from -40 to 20 dB (in 2 dB steps). Each masker was generated with 6.3 s duration, normalized to an rms level of 1, multiplied by a scaling factor corresponding to a desired SNR (relative to a speech signal normalized to an rms level of 1), passed through the STEP processor, and saved as a Matlab matrix file. These files were later combined with a speech file and presented in the simulated behavioral task.

4.2.6 Procedures and conditions

Figure 2.2 illustrates the key features of the simulated masker conditions. Maskers were unmodulated (Unmod), synchronously modulated (Sync), or asynchronously modulated (Async) with 4 or 8 bands. Control conditions included all noise bands and only either the even-numbered or odd-numbered speech bands. Presentations could be monaural (M), or dichotic (D) with speech and maskers split between two ears (odd-numbered bands in the left ear and even-numbered bands in the right ear). In total there were 10 masking conditions (Unmod-M, Sync-M, and 4 and 8 bands each for Async-M, Async-D, Async-D-ODD, and Async-D-EVEN). Stimuli could be passed through the STEP processor with an SF of 1, 1.7, or 3 in order to simulate the effects of reduced frequency resolution. Lastly, for half of the simulations, a noise floor was included to simulate reduced audibility. The data reported are for a noise floor of -35 dB attenuation re the level of the target speech.
Simulations followed the same behavioral procedures described in Section 3.2. A threshold estimation track was initiated by defining the three necessary parameters: the masker type (e.g., Unmod-M, Sync-M, Async-M, etc...), the SF, and the level of the NF. The Gaussian noise floor (3s duration) was generated at the onset of each adaptive track and passed through the STEP processor. A random vowel-consonant-vowel STEP was chosen with replacement, summed with the noise floor STEP, and combined with a masker STEP of equal duration.\textsuperscript{13} The starting SNR was set to 10 dB. The stimuli were passed through the glimpsing model and speech recognizer for a solution. In the case of a correct response, the masker level was increased by 4 dB, and in the case of an incorrect response, the masker level was decreased by 4 dB. Threshold for a track was determined by the average SNR at the last 24 of 26 track reversals.

\textbf{4.2.7 Data analysis}

Data reported in the results are based on a total of 100 iterations of the experiment for each masking condition (10 total) and each SF (3 total), with and without a noise floor. Only the better threshold between the two control conditions (abbreviated as Control-D) at any band number was considered in subsequent analysis. The MR was measured as the absolute threshold difference between the Unmod-M condition and any other condition the associated iteration at a single SF and noise floor\textsuperscript{14}. Dichotic advantage was defined as the threshold difference between the Async-M and Async-D conditions for each iteration. Mean speech reception thresholds (in dB SNR) are reported in Tables 4.1 and 4.2 for all conditions with

\textsuperscript{13} A random starting point was chosen in the masker file, and the end of the file was truncated to match the length of the signal. This ensured a random starting phase of masker modulation and thus provided a unique presentation on every trial.

\textsuperscript{14} Because a separate model was created for a given frequency resolution, SF was treated as a between-subjects variable. In contrast, the same model was used to simulate thresholds for conditions with and without a noise floor, so noise floor was treated as a within-subjects variable.
and without a noise floor, respectively. In the simulations without a noise floor, we attempted to model the effects of spectral smearing alone. For $SF = 1$ and no noise floor, the model was expected to capture the normal-hearing listener data in Experiment 1. The simulations with a noise floor were intended to model the combined effects of both reduced audibility and reduced frequency resolution, similar to Experiment 2.

Table 4.1: Mean speech reception thresholds (in dB SNR) from simulation without a noise floor are reported for each stimulus condition. The standard error of the mean for each SF model ($n = 100$) is shown in parentheses next to the associated mean.

<table>
<thead>
<tr>
<th>SF&lt;sub&gt;1&lt;/sub&gt;</th>
<th>Without Noise Floor</th>
<th>Number of Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Primary data</td>
<td>Unmod-M</td>
<td>-0.31 (0.3)</td>
</tr>
<tr>
<td></td>
<td>Sync-M</td>
<td>-22.85 (0.4)</td>
</tr>
<tr>
<td></td>
<td>Async-M</td>
<td>-12.48 (0.3)</td>
</tr>
<tr>
<td></td>
<td>Async-D</td>
<td>-23.98 (0.4)</td>
</tr>
<tr>
<td>Controls</td>
<td>Async-D-ODD</td>
<td>3.30 (0.7)</td>
</tr>
<tr>
<td></td>
<td>Async-D-EVEN</td>
<td>-2.13 (0.7)</td>
</tr>
<tr>
<td></td>
<td>Control-D</td>
<td>-4.03 (0.7)</td>
</tr>
</tbody>
</table>

| SF<sub>1.7</sub> |                     |       |       |
| Primary data    | Unmod-M              | 1.10 (0.4) |       |
|                 | Sync-M               | -19.39 (0.5) |       |
|                 | Async-M              | -7.26 (0.4) | -2.82 (0.4) |
|                 | Async-D              | -22.30 (0.5) | -23.29 (0.5) |
| Controls        | Async-D-ODD          | 7.35 (0.6)   | -1.57 (0.6) |
|                 | Async-D-EVEN         | -3.71 (0.6)  | -3.11 (0.7) |
|                 | Control-D            | -4.15 (0.6)  | -6.41 (0.5) |

| SF<sub>3</sub> |                     |       |       |
| Primary data   | Unmod-M              | 2.89 (0.3) |       |
|                | Sync-M               | -21.36 (0.5) |       |
|                | Async-M              | -5.91 (0.4) | 3.03 (0.4) |
|                | Async-D              | -20.78 (0.5) | -24.55 (0.4) |
| Controls       | Async-D-ODD          | 7.96 (0.7)   | -5.31 (0.7) |
|                | Async-D-EVEN         | 0.21 (0.6)   | -4.92 (0.8) |
|                | Control-D            | -0.93 (0.6)  | -9.55 (0.6) |
Table 4.2: Mean raw speech reception thresholds (in dB SNR) from simulation with a noise floor are reported for each stimulus condition. The standard error of the mean for each group (n=100) is shown in parentheses next to the associated mean.

<table>
<thead>
<tr>
<th></th>
<th>With Noise Floor</th>
<th>Number of Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF_1</td>
<td>4</td>
</tr>
<tr>
<td>Primary data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmod-M</td>
<td>4.04 (0.5)</td>
<td></td>
</tr>
<tr>
<td>Sync-M</td>
<td>-11.58 (0.8)</td>
<td></td>
</tr>
<tr>
<td>Async-M</td>
<td>-0.52 (0.7)</td>
<td>1.08 (0.6)</td>
</tr>
<tr>
<td>Async-D</td>
<td>-18.75 (0.7)</td>
<td>-16.97 (0.6)</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Async-D-ODD</td>
<td>9.17 (0.6)</td>
<td>8.43 (0.6)</td>
</tr>
<tr>
<td>Async-D-EVEN</td>
<td>3.35 (0.7)</td>
<td>1.47 (0.7)</td>
</tr>
<tr>
<td>Control-D</td>
<td>1.89 (0.6)</td>
<td>0.44 (0.7)</td>
</tr>
<tr>
<td></td>
<td>SF_1.7</td>
<td></td>
</tr>
<tr>
<td>Primary data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmod-M</td>
<td>-1.06 (0.5)</td>
<td></td>
</tr>
<tr>
<td>Sync-M</td>
<td>-14.49 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Async-M</td>
<td>-3.64 (0.5)</td>
<td>-1.40 (0.6)</td>
</tr>
<tr>
<td>Async-D</td>
<td>-10.57 (0.9)</td>
<td>-14.65 (0.7)</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Async-D-ODD</td>
<td>14.89 (0.3)</td>
<td>6.76 (0.6)</td>
</tr>
<tr>
<td>Async-D-EVEN</td>
<td>6.75 (0.7)</td>
<td>6.55 (0.7)</td>
</tr>
<tr>
<td>Control-D</td>
<td>6.45 (0.7)</td>
<td>2.78 (0.6)</td>
</tr>
<tr>
<td></td>
<td>SF_3</td>
<td></td>
</tr>
<tr>
<td>Primary data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmod-M</td>
<td>10.03 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Sync-M</td>
<td>2.85 (0.8)</td>
<td></td>
</tr>
<tr>
<td>Async-M</td>
<td>9.41 (0.5)</td>
<td>11.27 (0.4)</td>
</tr>
<tr>
<td>Async-D</td>
<td>-0.88 (0.8)</td>
<td>-2.88 (0.8)</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Async-D-ODD</td>
<td>13.23 (0.4)</td>
<td>12.58 (0.3)</td>
</tr>
<tr>
<td>Async-D-EVEN</td>
<td>13.99 (0.2)</td>
<td>9.57 (0.6)</td>
</tr>
<tr>
<td>Control-D</td>
<td>11.91 (0.4)</td>
<td>8.26 (0.5)</td>
</tr>
</tbody>
</table>

4.3 Results

4.3.1 Comparisons to Experiments 1 and 2

Figure 4.8 presents MR from the Sync condition for each SF model without the noise floor as well as for the normal-hearing group in Experiment 1. Figure 4.9 presents MR from the Sync condition for each SF model with a noise floor as well as for the three SF groups in Experiment 2. With a no noise floor and no smearing, the model had roughly 22.5 dB of MR
in the Sync condition, which was not statistically different from the 25.3 dB of MR for the normal-hearing group in Experiment 1 (One-way ANOVA; F[1,105] = 1.88, p = .17). The addition of a noise floor led to a 6.9-dB MR reduction (compare SF1 model in Figures 4.8 and 4.9), which was a statistically significant change (One-way ANOVA; F[1,198] = 41.3, p < .001).

Figure 4.8: Model results: masking release (MR) in the synchronously-modulated (Sync) noise condition for each level of smearing (SF1, SF1.7, and SF3) without a noise floor (left columns). The mean speech reception threshold for the normal-hearing (NH) listeners tested in Experiment 1 is plotted in the right column for comparison.

The MR data in the Sync condition were also submitted to a two-factor univariate analysis with two levels of noise floor (with and without) and three levels of SF in order to compare the effects of SF with the effects of noise floor. Main effects were found for the noise floor (F[1,594] = 287.5, p < .001), which was evident in the greater MR without a noise floor than with a noise floor, and for smear factor (F[2,594] = 10.332, p < .001), which was
evident by the higher MR when no spectral smearing was applied to the stimulus. Post-hoc analysis (Tukey HSD) confirmed that MR in the Sync condition was higher for SF\(_1\) relative to SF\(_{1.7}\) (p < .05) and SF\(_3\) (p < .001), but was not statistically different between SF\(_{1.7}\) and SF\(_3\).

Lastly, an interaction between SF and noise floor was significant (F[2,594] = 10.3, p < .001). Whereas smearing appeared to have no effect when there was no noise floor (see Figure 4.8), having a noise floor led to smaller MR in the model as smearing increased (see Figure 4.9). Post-hoc analysis for data from the model with a noise floor revealed that masking release for SF\(_1\) and SF\(_{1.7}\) was significantly greater than those for SF\(_3\) (Tukey HSD; p < .001 and p < .001, respectively), while threshold for SF\(_1\) and SF\(_{1.7}\) were not significantly different from each other.

![Figure 4.9](image)

**Figure 4.9:** Model results: masking release (MR) in the synchronously-modulated (Sync) noise condition for each level of smearing (SF\(_1\), SF\(_{1.7}\), and SF\(_3\)) with a noise floor (left columns) of -35 dB re the level of the speech signal. Results from Experiment 2 are plotted in the right columns for comparison.

Figure 4.10 presents MR for Async-M and Async-D in both the model (SF\(_1\) with no noise floor) and the normal-hearing data from Experiment 1. Two distinct observations can be
made regarding the model results with respect to the normal-hearing listeners in Experiment 1: 1) the model over-predicted the MR observed in the Async-D condition; and 2) the model under-predicted the MR observed in the Async-M condition. These effects were similar for both 4 and 8 bands. The corresponding dichotic advantage, hence, was larger for the model than the behavioral data.

Figure 4.10: Data from Experiment 1 (filled markers) and the model (open markers) with no smearing (SF₁) and no noise floor. Masking release (MR) is plotted for Async-D (triangles) and Async-M (circles) conditions.

Dichotic advantage was measured as the difference between MR in the Async-D and Async-M masking conditions. Whereas the model with no noise floor predicted dichotic advantage to be somewhat larger than observed for normal-hearing data in Experiment 1, the model with a noise floor substantially over-predicted the dichotic advantage observed in Experiment 2. Figure 4.11 presents dichotic advantage for each SF model with a noise floor and the data from Experiment 2 re-plotted from Chapter 3. A two-way ANOVA was performed to assess the dichotic advantage predicted by the model with a noise floor, with
three levels of SF and two levels of numbers of bands. The analysis indicated a main effect of band numbers (F[1,297] = 18.7, p < .001), which was the result of greater dichotic advantage as the number of bands increased. Smearing was also a significant factor (F[2, 297] = 38.1, p < .001), which is evident from the decline in dichotic advantage as smearing increased from SF₁ to either SF₁.7 or SF₃. The differences between each SF model depending on the band number is also evidenced by a significant interaction (F[2, 297] = 6.1, p < .005). Post-hoc analysis revealed that the SF₁ model predicted the greatest dichotic advantage at both band numbers, whereas at 4 bands the SF₁.7 model predicted a significantly smaller dichotic advantage than the SF₃ model (Tukey HSD; p < .001) and roughly equivalent for 8 bands (Tukey HSD; p = .76).

Figure 4.11: Dichotic advantage for SF groups in Experiment 2 and each SF model (SF₁, SF₁.7 and SF₃) with a noise floor. The numbers of bands are displayed separately with 4 bands on the left and 8 bands on the right. Asterisks indicate significant differences (see text for p-values).
4.4 Discussion

The main purpose of the present study was to model the effects of reduced audibility and reduced frequency resolution when glimpsing speech in a spectro-temporally fluctuating masker. The glimpsing model was based on a previously verified monaural model, which successfully fitted behavioral data by providing only sparse speech cues to a speech recognizer (Cooke, 2006). The current design extends the earlier model to account for dichotic stimuli and STEP processing with broader auditory filters. In order to compare results with our behavioral experiments, the model simulated the asynchronous glimpsing task (Experiments 1 and 2) using spectrally smeared stimuli with a noise floor. Normal-hearing listeners were assumed to have normal frequency resolution (i.e., no spectral smearing) and no noise floor, whereas hearing-impaired listeners were modeled with different SF values and a noise floor.

4.4.1 Comparisons to Experiments 1 and 2

Behavioral data from Experiments 1 and 2 indicated that MR in amplitude-modulated noise decreases with reduced sensation levels and poorer frequency resolution. The model confirmed the effect of reduced sensation level (compare SF1 model in Figures 4.8 and 4.9). However, the model predicted spectral smearing to affect MR in the Sync condition only when in combination with a noise floor. Because Experiment 2 included only groups with attenuated stimuli, it is unknown whether spectral smearing would have affected MR in the Sync condition in which stimuli were presented at a higher level. In a related study, Gnansia et al. (2009) also tested speech recognition in modulated maskers and found that spectral smearing reduced MR. Precise comparison between our studies and theirs, however, is complicated by methodological differences, including overall presentation level and units of
measurement for MR. Specifically, the Gnansia et al. study presented stimuli at 75 dB SPL, and the task set-size was considerably larger, two differences we have shown to affect MR in modulated maskers (Ozmeral et al., 2012). Perhaps more importantly, MR was measured in percent correct for a fixed global SNR determined by 50% in the baseline, unmodulated noise. Although Gnansia et al. (2009) showed reduced MR as spectral smearing increased, it is unknown whether the effect would persist if the baseline SNRs were the same across SFs (see Section 3.3.5; Bernstein and Grant, 2009).

For Async conditions, the model successfully captured greater MR in the dichotic presentations compared to monaural presentations. Under conditions intended to be analogous to the normal-hearing listeners in Experiment 1, the dichotic advantage predicted by the model was somewhat larger than in the behavioral data due to an over-prediction of MR in the Async-D condition and an under-prediction of MR in the Async-M condition (see Figure 4.9). One possible explanation for the over-prediction in the Async-D condition is that the model is able to use every available glimpse whereas humans may still miss available glimpses. For example, we saw in Experiment 1 that the difference between monaural and dichotic control conditions was roughly 4 dB for both normal-hearing and hearing-impaired listeners. This indicated that masking can occur even for masking bands that are presented to the contralateral side and in different frequency regions than the target bands. Whereas the model can base its solution on perfect integration of the best target information available in either ear, human listeners will undoubtedly be limited by less-perfect integration. This is perhaps most evident at the narrow, 8-band condition, in which the model predicts an improvement in performance relative to the 4-band condition, where the behavioral
experiments show the opposite to be true (see Experiments 1 and 2; Howard-Jones and Rosen, 1993; Ozmeral et al., 2012).

Under conditions intended to be analogous to listeners with reduced audibility and reduced frequency resolution, the dichotic advantage was much larger in the model for all SF groups than observed in Experiment 2, and the effect of smearing in the model did not capture the effects seen in Experiment 2 entirely (Figure 4.11). Whereas the behavioral data and model both show relatively larger dichotic advantage in the SF3 versus SF1.7 groups, the model achieved the greatest dichotic advantage in the SF1 condition, while listeners in Experiment 2 achieved the least dichotic advantage in the SF1 group. As discussed in Chapter 3, the effect of spectral smearing on the dichotic advantage at reduced presentation levels was explained by the greater benefit of removing the deleterious effects of spread of masking in the monaural conditions. As with the model, this seems to be most beneficial when modulated frequency bands are relatively narrow. However, it is not as clear why the model has the greatest dichotic advantage with no smearing.

4.4.2 Summary

The adaptation of the Cooke (2006) glimpsing model to accommodate dichotic stimuli and poorer frequency resolution in the STEP processor has led to a greater insight into normal-hearing and hearing-impaired listeners’ abilities to integrate sparse speech cues in Async noise. As seen in the behavioral data in the previous chapters, the current model predicted better thresholds when normal auditory filter bandwidths were used and no noise floor was present. The addition of a noise floor, which simulated limited audibility, led to reduced MR in all modulated masker conditions. Although the model seemed to capture the negative effect of smearing for MR in the Sync noise, spectral smearing appeared to have
disadvantages for dichotic listening relative to no spectral smearing. Overall, the model similarly predicted a dichotic advantage as seen in the behavioral data from Experiments 1 and 2, and it partially predicted the deleterious effects of hearing impairment, consistent with reduced audibility and reduced frequency resolution simulated in normal-hearing listeners in Experiment 2.
CHAPTER 5: GENERAL DISCUSSION

5.1 Summary of Experiments

The goal of the current study was to investigate the effects of peripheral auditory processing on the ability to recognize speech in the presence of spectro-temporally fluctuating noise. To that end, we measured MR for modulated maskers in listeners with real and simulated hearing impairments in order to investigate the contributing factors. Better speech reception thresholds in a modulated-masker condition relative to a steady noise (i.e., MR) indicate a better ability to glimpse speech cues from spectro-temporal regions of masker minima. With Sync noise, listeners are able to glimpse speech by taking advantage of favorable SNRs in the brief masker dips (Miller and Licklider, 1950). Unlike Sync noise, however, Async noise fluctuates both temporally and spectrally, and it has been shown to mask speech more effectively than Sync noise (Howard-Jones and Rosen, 1993; Ozmeral et al., 2012). We hypothesized that epochs of the Async masker maxima energetically spread into spectral regions for which the masker envelope is at a minimum, thereby reducing opportunities to glimpse speech. To limit the effect of spread of masking, Async stimuli can be presented dichotically – a strategy that has previously led to between 5 and 8 dB additional MR relative to a monaural presentation (Ozmeral et al., 2012). Because limiting the effects of spread of masking can benefit normal-hearing listeners in the Async noise, we believe that a listeners’ frequency resolution is a significant factor in understanding speech recognition in this type of noise. Little was previously known about glimpsing speech in
spectro-temporally fluctuating noise for hearing-impaired listeners – a population which suffers from reduced audibility and poor frequency resolution – and one goal was to identify the relative contributions of the factors associated with sensorineural hearing loss. The behavioral experiments presented here tested hearing-impaired adults and normal-hearing adults with and without simulated hearing loss. The central hypothesis was that the benefit associated with dichotic presentation in normal-hearing listeners would be comparable or larger for hearing-impaired listeners because hearing-impaired listeners are more susceptible to effects of spread of masking. The results of the study can be summarized as follows.

In Experiment 1, normal-hearing and hearing-impaired listeners performed a speech recognition task in the presence of wide-band noise. Relative to a continuous noise, MR was measured for various monaural and dichotic modulated maskers. Dichotic conditions presented out-of-phase frequency bands to opposite ears so that neighboring frequency bands did not energetically mask each other. It was hypothesized that due to limiting the effects of peripheral spread of masking, all listeners would have larger MR in dichotic Async conditions than monaural Async conditions. The results showed that while normal-hearing and hearing-impaired listeners both benefit from dichotic listening, the extent of the dichotic advantage was only comparable between groups when MR was normalized. Nevertheless, the experiment showed that hearing-impaired listeners are aided by dichotic listening. The dichotic advantage was on the order of 4 dB for hearing-impaired listeners and at least 7 dB for normal-hearing listeners – both considerable increases in MR in terms of the possible improvements in percent correct for a speech identification task. Although no effort was made to present stimuli at similar sensation levels in Experiment 1, previous studies and the results of Experiment 2 suggest that audibility differences between normal-hearing and
hearing-impaired listeners could play a substantial role in MR differences between listener groups. However, the results of Experiment 1 do not rule out the possibility that hearing-impaired listeners were also limited by reduced frequency resolution.

Whereas Experiment 1 showed that hearing-impaired listeners are able to benefit from dichotic listening in the Async noise, Experiment 2 investigated the relative contributions of reduced audibility and reduced frequency resolution in the ability to glimpse speech in the asynchronous glimpsing task. Normal-hearing listeners were tested on stimuli that were processed to mimic the effects of varying levels of hearing loss. The effects of reduced audibility were simulated by attenuating the signals 30 dB re the presentation level in Experiment 1, and the effects of reduced frequency resolution were simulated by spectrally smearing the stimuli. The results were consistent with an interpretation that audibility could account for much of the reduced MR of the hearing-impaired listeners. Additionally, the results were consistent with an interpretation that reduced frequency resolution could result in further reduction in MR. Lastly, dichotic listening was found to be most beneficial for the 8-band cases, especially for stimuli with the greatest spectral smearing. Together, these results were consistent with a view that glimpsing speech in spectro-temporally fluctuating noise is impaired with reduced frequency resolution, and the use of dichotic listening strategies can provide some relief from the effects of reduced frequency resolution.

In Experiment 3, we modeled the behavioral data using a binaural glimpsing model adapted from a monaural version by Cooke (2006). It was hypothesized that the ability to identify speech in an Async masker could be accomplished from the integration of sparse spectro-temporal windows of favorable SNRs (i.e., glimpses). Moreover, it was expected that a binaural implementation could capture the advantages of dichotic listening seen in the
behavioral data from Experiments 1 and 2. Specifically, dichotic presentations were expected to show larger MR than monaural presentations due to the removal of spread of masking’s effect on available glimpses. Additionally, as the behavioral data suggested, MR was expected to be reduced overall for conditions with greater numbers of bands (i.e., narrow glimpsing windows), but the difference between monaural and dichotic Async conditions was expected to be greater at these narrow-band conditions. The results of the model confirmed an advantage to dichotic listening, especially in the 8-band condition, but MR in the dichotic conditions rose for the 8-band condition, which was contrary to the human data. Considering human listeners are not ideal observers, even with life-long experience listening to speech in noise, it was concluded that additional factors likely affect glimpsing in Async noise for narrow bandwidths. Possible factors include effects of temporal resolution, limits in central processing, and contralateral masking, which were not modeled in the current experiment.

5.2 Clinical Applications

Hearing impairment affects an estimated 29 million adults domestically (aged 20 to 69 years; Agrawal et al., 2008), and the present study has interesting implications for hearing-aid design. In quiet settings, most aided hearing-impaired listeners with mild-to-moderate sensorineural hearing loss have minor difficulty following a conversation. However, the same listeners often complain that it is difficult to follow speech in noisy environments. Traditional hearing aids with advanced noise-reduction processing (e.g., Hu and Loizou, 2007) have thus far been largely ineffective in improving speech understanding in noise. One of the obvious factors contributing to this phenomenon is that amplification has the negative effect of adding gain to all incoming sounds, including the unwanted noise. Therefore, new
strategies must be developed to limit the influence unwanted noise has on speech perception. The results of the present study suggest that additional speech understanding can be accomplished by presenting non-overlapping stimulus bands to opposite ears.

In addition to hearing-impaired listeners having naturally broader-than-normal auditory filters, hearing aids can cause additional filter broadening (Moore, 1996; Strelcyk and Dau, 2009b), and cochlear implant users experience degraded frequency resolution due to electrical spread (Chatterjee et al., 1998). Hearing-impaired listeners in Experiment 1 were not found to benefit from dichotic presentation as much as the normal-hearing listeners, but presentation levels did not permit equivalent sensation levels between the groups, so this was consistent with previous studies that have shown small MR for reduced sensation levels (e.g., Festen and Plomp, 1990). With the support of amplification, dichotic presentation could alleviate the detrimental effects of peripheral spread of masking. There have been some successful attempts at using dichotic listening paradigms to improve speech identification in hearing-aid or cochlear implants users (Loizou et al., 2003; Kulkarni et al., 2012; Zhou and Pfingst, 2012), and the current study provides additional support for this approach. However, there may be unintended consequences of removing crucial binaural spatial cues, such as interaural phase differences (Itoh et al., 2012), so further study in spatially diverse settings are still needed. In combination with amplification, dichotic processing for bilateral hearing prosthetics may benefit the listeners when listening to speech in adverse, spectro-temporally complex environments.

5.3 Limitations and Possible Improvements to the Study

In Experiment 1, we did not test hearing-impaired listeners at comparable sensation levels as the normal-hearing listeners; however, the data from the SF1 group in Experiment 2
suggest that MR is comparable if sensation levels are similar for the normal-hearing and hearing-impaired groups. It may have been worthwhile to test normal-hearing listeners with stimuli that were spectrally shaped to match the audibility profiles of individual hearing-impaired listeners. This could have provided better evidence that MR differences between listening groups can be accounted for by audibility differences.

One of the primary criticisms of a glimpsing model, like the one used in Experiment 3, is that it does not accurately model the nature of how sounds reach the ear because the noise and speech are known in advance. In addition to using speech cues in epochs associated with favorable signal SNRs, human listeners are quite adept at using auditory cues to parse apart sound streams associated with different sources: these include cues associated with spatial separation (Freyman et al., 1999; Arbogast et al, 2002), pitch and timbre differences (Plack and Oxenham, 2005), and other perceptual grouping factors (Bregman, 1990). In contrast to human hearing, the current machine-based speech recognizer is limited by a number of factors, including an inability to compensate for large variability in speech signals, including pitch differences, disfluency, variable speech rate, and perhaps other factors that humans encounter through life. Due to this limitation, it was necessary to provide the model with separate speech and noise signals that would otherwise be combined in a natural setting. Future models of glimpsing might implement some form of perceptual grouping (e.g., computational auditory scene analysis; cf. Wang and Brown, 2006) instead of relying on prior knowledge of the signal and masker, which could perhaps better represent real-world glimpsing of speech in noise.
5.4 Conclusions

Previous studies on glimpsing speech in fluctuating noise have determined that MR is influenced by the sensation level of the target (Festen and Plomp, 1990). However, some studies have shown that even when level effects are accounted for, normal-hearing listeners still tend to outperform hearing-impaired listeners (Eisenberg et al., 1995; Bacon et al., 1998; George et al., 2006). Although poor temporal resolution has been suggested as one factor that limits hearing-impaired listeners’ abilities to understand speech in fluctuating noise (Dubno et al., 2003; George et al., 2006), others have found reduced frequency resolution to correlate well with smaller MR after partialling out the effects of pure-tone sensitivity (Eisenberg et al., 1995). The current design of the asynchronous glimpsing task makes no effort to alter the natural effects of temporal resolution, but it does attempt to reduce the direct effects of poor frequency resolution. The evidence from the current study suggests that when audibility is held constant, reducing frequency resolution can impair a listener’s ability to glimpse speech in modulated maskers, both Sync and Async. Although the glimpsing model in Experiment 3 tended to predict an exaggerated dichotic advantage, listeners in Experiment 2 clearly showed that greater spectral smearing was associated with a larger dichotic advantage, especially when the Async masker was filtered into 8 narrow bands. We conclude that along with audibility, frequency resolution plays a significant role in the ability to glimpse speech in spectro-temporally fluctuating noise, and strategies that limit the effect of spread of masking, such as dichotic listening, can improve speech recognition in a complex auditory scene.
REFERENCES


