

The Effect of Signal-Temporal Uncertainty during Childhood

Angela Yarnell Bonino

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 Speech and Hearing Sciences in the Department of Allied Health Sciences.

Chapel Hill
2012

Approved by,

Lori J. Leibold, Ph.D.

Emily Buss, Ph.D.

John H. Grose, Ph.D.

Katarina L. Haley, Ph.D.

Joseph W. Hall III, Ph.D.

ABSTRACT

ANGELA YARNELL BONINO: The Effect of Signal-Temporal Uncertainty during
Childhood

(Under the direction of Lori J. Leibold, Ph.D.)

Children are more susceptible to interference from competing background sounds than adults. In the laboratory, children's increased susceptibility to interference from competing sounds is evident for both tone detection and speech recognition tasks. For both types of stimuli, this age effect is pronounced when the background sounds are complex and/or unpredictable. In light of children's increased susceptibility to interference from competing sounds, this dissertation tested the hypothesis that performance would improve if children were provided a cue indicating when in time to listen for target signals. This hypothesis was tested for both tone detection and word recognition in the presence of competing tones or speech, respectively. Parallel studies tested the hypothesis that similar factors affect performance for the two different types of stimuli.

For both the tonal and the speech experiments, performance was measured for five to 13 year old children and a group of adult controls (20 to 34 years). In the first experiment, listeners were asked to detect a 1000-Hz pure-tone signal embedded in a continuous background that was a random-frequency, two-tone masker. Two conditions were tested: (1) temporally-defined, with the listening interval defined by a light cue and (2) temporally-uncertain, with no light cue. Thresholds estimated from psychometric functions fitted to the data indicated that both children and adults benefited from access to the light cue (mean =

11.2 dB). Similar benefit was observed across the three listener age groups of children and adults. In the second experiment, listeners were asked to repeat target monosyllabic words presented in two different backgrounds: a two-talker masker or speech-shaped noise. Target words were presented at three different delays relative to the illumination of an interval light: (1) a fixed delay of 0.5 s, (2) a fixed delay of 5.5 s, or (3) a randomly selected delay. In contrast to Experiment 1, a clear benefit of a defined listening interval was not seen for word recognition or phoneme identification scores. However, examination of the types of phoneme errors produced provided insight into the factors that may be responsible for differences in performance across the two backgrounds.

ACKNOWLEDGEMENTS

I would like to express gratitude to my dissertation advisor, Dr. Lori J. Leibold. After my first year in the program, Lori provided me the opportunity to work in her lab for the summer. Now five years later, I am thankful that I had the opportunity to be in such a supportive and rigorous environment during my doctoral training. Throughout my program, Lori has continued to provide high quality mentoring and support that has evolved as my skill set and needs changed. I know that the mentoring I received from Lori will forever shape me both professionally and personally. I would also like to thank Dr. Emily Buss for being a co-mentor of this dissertation. Emily has been extremely generous with her time and knowledge. Emily's support made it possible to run these experiments in MATLAB and was also invaluable during the data analysis phase. In addition to Lori and Emily, Drs. Joseph W. Hall III, John H. Grose and Katarina L. Haley served on my committee. I appreciate their insight and generosity with their time.

I am grateful to all the members of the Human Auditory Development Lab for their help with these projects. In particular, the study presented in Chapter 4 would not have been possible without significant contributions from several lab members, including: Abby Bennett, Bryan Allard, Jack Hitchens, Rachel Fann, and Chas Phillip. Their help was invaluable with every stage of this study, from the development of the target word corpus and editing sound files to data collection and offline coding. Programming assistance from Steve Lockhart made the offline coding a possibility as well as its analysis. I would also like to thank the families who participated in these experiments.

I also thank my family for their unwavering support. My parents, Mark and Vicki Yarnell, fostered a love for learning in me and they have continued to support my educational endeavors for over three decades now. My family's support of this process has been extremely evident over the last couple of months. In order to provide the uninterrupted time that I needed to write this semester, we had nearly two months of continual family support. I will be forever in debt to my mom, Lona and Mike Millar (my grandparents), and Eileen and Sam Bonino (my in-laws), who were willing to sleep on the air mattress on our living room floor to make this a reality.

Lastly, I am thankful for the support and encouragement from my daughter and husband, Eleanor and Christopher Bonino. The decision for Chris and me to simultaneously return to school to pursue doctoral training and then ultimately have a child, while challenging, is a blessing. It has been fantastic to have a spouse who fully understands this process and what is involved. Not to mention, a spouse who can help with formatting and teach you how to use a graphing program. Furthermore, Eleanor provides a balance and joy in our lives that only she can do. I could not imagine going through this process without the two of them.

Support for this research was provided by grants from the National Institutes of Health (NIDCD; F31DC010308 and R01DC011038) and a Dissertation Completion Fellowship from the University of North Carolina at Chapel Hill.

TABLE OF CONTENTS

1. CHAPTER 1: INTRODUCTION	1
2. CHAPTER 2: REVIEW OF THE LITERATURE	5
2.1. Development of the Ability to Listen in Noise	5
2.2. Potential Mechanisms for Children’s Increased Susceptibility to Informational Masking	8
2.2.1. Sound Source Segregation	9
2.2.2. Selective Auditory Attention	10
2.3. The Importance of Knowing When in Time to Listen	11
2.4. Considerations for Applying the Same Theoretical Framework to Tonal and Speech Stimuli	13
2.5. Figures	17
3. CHAPTER 3: EFFECT OF SIGNAL-TEMPORAL UNCERTAINTY DURING CHILDHOOD: DETECTION OF A TONAL SIGNAL IN A RANDOM-FREQUENCY, TWO-TONE MASKER.....	18
3.1. Introduction	18
3.2. Methods	25
3.2.1. Listeners	25
3.2.2. Stimuli.....	26
3.2.3. Conditions and Procedure	27
3.2.4. Procedure for Fitting Psychometric Functions	29

3.3. Results	30
3.3.1. Goodness of Fit	31
3.3.2. Threshold Estimates and the Effect of Signal-Temporal Uncertainty	32
3.3.3. Response Bias	35
3.3.4. Other Estimated Parameters of the Psychometric Function	37
3.4. Discussion	38
3.4.1. Effect of Signal-Temporal Uncertainty	39
3.4.2. Decision Strategy	41
3.4.3. Susceptibility to Masking	43
3.4.4. Methodology Differences	45
3.5. Conclusions	49
3.6. Endnotes	50
3.7. Tables	51
3.8. Figures	52
4. CHAPTER 4: LISTENING AT FIXED OR RANDOM DELAYS: WORD RECOGNITION IN A TWO-TALKER OR A SPEECH-SHAPED NOISE MASKER FOR CHILDREN AND ADULTS	58
4.1. Introduction	58
4.2. Methods	63
4.2.1. Listeners	63
4.2.2. Stimuli and Conditions	63
4.2.3. Testing Procedure	67
4.2.4. Offline Scoring Procedure	69

4.3. Results	71
4.3.1. Adaptive Thresholds	71
4.3.2. Word Recognition	72
4.3.3. Phoneme Identification	75
4.3.4. Error Patterns	78
4.4. Discussion	82
4.4.1. Age Effects in Susceptibility to Masking	82
4.4.2. Benefit of a Cue Indicating When in Time to Listen.....	83
4.4.3. Error Patterns	85
4.5. Conclusions	87
4.6. Endnotes	89
4.7. Figures	90
5. CHAPTER 5: CONCLUSIONS	97
5.1. Summary of Findings	97
5.2. Unresolved Questions for Future Work	99
REFERENCES	102

LIST OF TABLES

Table 3-1. Mean parameter estimates from the fitted psychometric functions by listener age group and temporal condition.....	51
---------------------------------------------------------------------------------------------------------------------------------	----

LIST OF FIGURES

Figure 2-1. Individual and mean effect of signal-temporal uncertainty from Bonino and Leibold (2008)	17
Figure 3-1. Stimuli schematic	52
Figure 3-2. Examples of psychometric function fits.....	53
Figure 3-3. Mean threshold estimates for the temporally-uncertain and temporally-defined conditions	54
Figure 3-4. Individual and mean benefit scores.....	55
Figure 3-5. Threshold estimates in the temporally-uncertain and temporally-defined condition for individual listeners.....	56
Figure 3-6. Individual and mean estimates of criterion	57
Figure 4-1. Schematic of delay conditions	90
Figure 4-2. Individual and mean masked threshold estimates	91
Figure 4-3. Mean word recognition and phoneme identification scores by condition	92
Figure 4-4. Change in individual listener's phoneme identification scores associated with a fixed delay	93
Figure 4-5. Mean error patterns for children and adults by condition	94
Figure 4-6. Individual error patterns for children in the two-talker masker conditions	95
Figure 4-7. Proportion of masker intrusions as a function of signal-to-noise ratio	96

CHAPTER 1

INTRODUCTION

Compared to adults, children are more susceptible to interference from competing background sounds (referred to as auditory masking). In the laboratory, children's increased susceptibility to auditory masking is commonly reported until about eight years of age for both tone detection and speech recognition tasks in steady-state noise maskers (e.g., Allen & Wightman, 1994; Elliott & Katz, 1980; Nishi et al., 2010; Nitttrouer & Boothroyd, 1990). In comparison to steady-state noise maskers, a more prolonged time course of development has been reported for tasks using complex and/or unpredictable maskers (e.g., Lutfi et al., 2003b; Wightman et al., 2003). Moreover, greater child-adult differences have been reported for maskers that are complex and unpredictable compared to noise maskers (e.g., Hall et al., 2002; Leibold & Neff, 2007; Lutfi et al., 2003b; Oh et al., 2001). Children's increased susceptibility to masking appears to reflect immature central auditory processes, rather than immaturities in the peripheral auditory system (reviewed by Werner & Leibold, 2004). Prolonged development of the central auditory system may increase children's difficulty in perceptually parsing the target signal from the background or attending to the target signal while ignore irrelevant sounds.

In light of children's difficulties in noise, studies have attempted to determine if children can use acoustic cues that improve adults' ability to hear and recognize target signals presented in complex backgrounds (e.g., Hall et al., 2005; Wightman et al., 2003;

Wightman & Kistler, 2005). However, the results are mixed in terms of which specific cues children use and the degree to which children benefit compared to adults. Moreover, the relative benefit for some or all cues might change with maturation and listening experience. Understanding the developmental course of susceptibility to auditory masking and determining cues that improve hearing in noise is important because children learn language in natural listening environments that are noisy and unpredictable.

The specific objective of this dissertation was to determine if children benefited from a cue indicating when in time to listen. Additionally, a wide age range of children (5 to 13 years) were tested to examine changes in performance during childhood. These objectives were evaluated for two studies using different stimuli; tones (Chapter 3) and speech (Chapter 4). The central hypothesis for both experiments was that, like adults, children would benefit from a cue that indicated when in time to listen for the signal in the presence of a spectro-temporally complex masker. Furthermore, consistent with other studies (e.g., Leibold & Bonino, 2009; Leibold & Neff, 2007; Lutfi et al., 2003b; Wightman et al., 2010), it was expected that age-related differences in susceptibility to masking would be apparent. It was expected that young children would be more susceptible to masking under complex listening conditions than older children or adults. This dissertation work increases our knowledge of how children listen in natural settings where multiple and unpredictable sound sources are often present, such as classrooms.

The specific aims of this dissertation were:

(1) To establish the degree to which children benefited from a cue indicating when in time to listen for signals embedded in complex maskers. Knowing when in time

to listen for relevant sounds may be an important cue in natural listening environments which contain multiple sound sources that can occur at unpredictable times. Recent evidence suggests that defining the listening interval provides a release from masking for adults when the task is to detect a tone in the presence of a multi-tonal masker (Bonino & Leibold, 2008; for exception see Richards et al., 2011). Although school-age children do not benefit from all of the same cues shown to benefit adults (e.g., Hall et al., 2005; Wightman et al., 2003; Wightman & Kistler, 2005), they are able to use other cues from an early age (e.g., Hall et al., 2005; Leibold & Bonino, 2009; Leibold & Neff, 2007). For complex maskers, it was predicted that providing a cue indicating when in time to listen for a signal would improve children's tone detection (Chapter 3) and word recognition (Chapter 4). Knowing when to listen may be particularly beneficial to children in light of their increased susceptibility to interference from complex maskers.

(2) To characterize susceptibility to interference from complex maskers across childhood. Consistent with previous developmental studies (e.g., Hall et al., 2002; Leibold & Neff, 2007; Lutfi et al., 2003b; Oh et al., 2001), we expected that children would be more susceptible to masking than adults for both tone detection (Chapter 3) and word recognition (Chapter 4) when target signals were presented in complex maskers. Furthermore, it was predicted that younger children would be more susceptible to masking than older children (e.g., Leibold & Bonino, 2009; Leibold & Neff, 2007). For the tone detection task, our approach was to examine changes in masker effectiveness across three age groups of children (5- to 7-year-olds, 8- to 10-year-olds, and 11- to 13-year-olds). Using this approach we were able to delineate the time course of development for tone detection in a continuous, random-frequency, two-tone masker. For the speech study, performance was examined for children

who were 5 to 13 years of age. Performance differences between age groups of children were not evaluated in the second experiment.

CHAPTER 2

REVIEW OF THE LITERATURE

The goal of the following review is to develop the central hypothesis for this research which is that children can benefit from a cue indicating when in time to listen for the signal in complex listening conditions. This literature review is divided into four sections. First, this chapter reviews the findings from previous developmental studies of auditory masking, for both tone detection and speech recognition tasks. The second section attempts to explain children's increased susceptibility to masking as a consequence of immature central auditory processes, including sound source segregation and selective attention. The third section develops the hypothesis that a cue indicating when in time to listen for a signal improves performance for both children and adults. In the final section, the strengths and limitations of using the same theoretical framework for tonal and speech stimuli are elaborated upon. These sections are intended to provide the reader a succinct overview of the literature pertaining to this dissertation. Please note that both studies (Chapter 3 and Chapter 4) also include a literature review, which is more specific to the particular study.

2.1. Development of the Ability to Listen in Noise

A prolonged time course of development has been observed for detecting a pure-tone signal embedded in noise or in other pure-tone complexes. Preschoolers' thresholds are elevated compared to adults' for detecting a pure-tone signal in quiet or in the presence of

wide-band noise maskers (e.g., Allen & Wightman, 1994; Jensen & Neff, 1993). For example, Allen and Wightman (1994) reported that 3- and 4-year-olds' average detection threshold was 13 dB worse than adults' for detecting pure-tone signals in Gaussian noise. Child-adult threshold differences are exacerbated as the masker becomes more complex. For example, an increased child-adult difference was observed for detecting a fixed-frequency pure-tone signal embedded in Gaussian noise with an additional single, fixed-level, random-frequency pure tone that was remote in frequency from the signal (Allen & Wightman, 1995). Allen and Wightman (1995) reported that the average child-adult difference for this task was at least 25 dB for the same group of 3- and 4-year-olds reported in the earlier study (Allen & Wightman, 1994). However, thresholds could not be obtained for half of the children when the random masker tone was present. In addition to larger child-adult differences, complex maskers have been associated with a prolonged course of auditory development (e.g., Lutfi et al., 2003b; Wightman et al., 2003). For example, substantial masking has been observed for detecting pure-tone signals in the presence of a multi-frequency tonal masker, with frequency components that vary randomly on each presentation (e.g., Neff & Green, 1987). This stimulus manipulation often results in substantial masking for all age groups, but child-adult differences can be more than 50 dB (e.g., Oh et al., 2001). Moreover, differences between younger and older children suggest that the ability to perform these tasks is being refined during childhood (e.g., Hall et al., 2005; Leibold & Neff, 2007; Oh et al., 2001; Wightman et al., 2003) and continues to develop until at least 13 to 16 years of age (Lutfi et al., 2003b; Wightman et al., 2003). This type of background creates substantial masking even though experimental controls are employed to minimize the overlap in peripheral excitation between the response to the signal and the response to the masker.

Thus, the term “informational masking” has been used to distinguish these effects from peripheral or “energetic” masking (reviewed by Kidd et al., 2007).

A similar trend has been observed for children’s speech-in-noise performance. Children require a more favorable signal-to-noise ratio for relatively simple tasks in which listeners are asked to recognize words presented in filtered noise (e.g., Elliott et al., 1979; Nitttrouer & Boothroyd, 1990). Compared to filtered noise, larger child-adult differences have been reported for speech maskers that contain a limited number of talkers (e.g., Bonino et al., in press; Hall et al., 2002; Wightman & Kistler, 2005; Wightman et al., 2010). For example, Hall et al. (2002) had 5- to 10-year-olds identify spondee words in the presence of a continuous masker using a four-alternative, forced-choice picture pointing task. In the continuous speech-shaped noise condition, thresholds were approximately 2 dB greater for children than for adults. In contrast, the difference in mean thresholds for children and adults was 7 dB for the continuous two-talker masker condition. This difference indicates that child-adult difference in susceptibility to masking is larger for speech maskers that are perceptually more complex. Additionally, Wightman and Kistler (2005) found that children as old as 13 to 16 years of age performed more poorly than adults. In that closed-set task, listeners identified key items in target sentences in the presence of a contextually-similar distracter sentence simultaneously presented to the same ear. Across studies, speech maskers with a limited number of talkers often produce perceptual masking, in addition to energetic masking, for children and adults. The term perceptual masking was originally used by Carhart et al. (1969) to explain the poorer performance observed for the speech masker condition than for the noise masker condition. Carhart et al. proposed that perceptual masking may be a result of the central auditory system being unable to perceptually parse apart the speech signal from

the competing masker. The effect of perceptual masking appears to be greatest for a two-talker masker and is minimal once the masker contains 10 talkers for adults (Freyman et al., 2004). Recently, it has been suggested that perceptual and informational masking may be closely related or may be the same phenomenon (e.g., Brungart, 2001; Freyman et al., 2004; Wightman & Kistler, 2005).

2.2. Potential Mechanisms for Children's Increased Susceptibility to Informational Masking

The underlying mechanisms responsible for children's increased susceptibility to informational (or perceptual) masking are not clear. However, the mechanisms do not appear to be at the level of the peripheral auditory system. The cochlea and auditory nerve appear to be capable of providing the brain a fairly precise representation of the auditory input by six months of age (reviewed by Werner & Leibold, 2004). Instead, children's increased susceptibility to masking appears to be result of immature central auditory processes (reviewed by Leibold, 2012). Two central auditory processes that appear to experience a prolonged developmental trajectory are sound source segregation and selective auditory attention. Sound source segregation is the ability to perceptually parse the target signal from the competing background sounds. After the incoming auditory information has been segregated, the brain needs to be able to attend to the auditory input of interest and ignore competing sounds. This process is referred to as selective attention. Although presented as separate processes here, it is difficult to determine their individual contributions to children's immature performance on behavioral measures used in the laboratory.

2.2.1. *Sound Source Segregation*

One likely central auditory process that is developing during childhood is the ability to perceptually parse the target signal from the competing background sounds. This dissertation will refer to this process as sound source segregation, but a more generalized definition of this process is auditory scene analysis (Bregman, 1990). One method commonly used to assess adults' auditory scene analysis abilities is to manipulate the stimulus to affect the number of perceived auditory streams. Results from this method indicate that adults are able to use segregation cues such as spatial location, onset/offset timing, timbre, and how sound change over time (reviewed by Bregman, 1993; Yost, 1991). Modified streaming studies suggest that infants are able to use frequency, intensity, timbre and spatial cues (Demany, 1982; Fassbender, 1993; McAdams & Bertoncini, 1997). However, it is not clear how precisely infants use these cues (Demany, 1982; Fassbender, 1993; McAdams & Bertoncini, 1997). Few developmental data are available with this method for school-aged children because young children would have a difficult time reporting the number of perceived streams (for exception see Sussman et al., 2007).

Informational masking paradigms have been used to study sound source segregation, because it is believed that informational masking occurs when the listener is unable to perceptually parse the signal from the masker. This claim has been supported by the observation that introducing cues believed to promote sound source segregation often provide a release from informational masking (or improved performance) (e.g., Kidd et al., 1994; Neff, 1995). Developmental studies using informational masking paradigms indicate that sound source segregation is developing during childhood. To date, researchers have examined whether children are able to benefit from the introduction of the following

segregation cues that have shown to result in a release from masking for adults: spatial separation (Garadat & Litovsky, 2007; Litovsky, 2005), dichotic presentation of the signal and the masker (Hall et al., 2005; Wightman et al., 2003; Wightman & Kistler, 2005; Wightman et al., 2010), asynchronous temporal onset of the signal compared to the masker (Hall et al., 2005; Leibold & Neff, 2007), spectro-temporal coherence (Bonino et al., in press; Hall et al., 2005; Leibold & Bonino, 2009), and visual access to a talker's face (Wightman et al., 2006). Interestingly, findings from these studies indicate that children as young as five years are able to use some cues as effectively as adults (Bonino et al., in press; Garadat & Litovsky, 2007; Hall et al., 2005; Leibold & Bonino, 2009; Leibold & Neff, 2007; Litovsky, 2005). In contrast to those findings, immature performance has been reported for other cues, even into adolescence (Hall et al., 2005; Wightman et al., 2003; Wightman & Kistler, 2005; Wightman et al., 2006; Wightman et al., 2010). Furthermore, differences in the ability to use some of these cues have been reported across childhood (e.g., Wightman & Kistler, 2005). Findings from these studies suggest, that at least for some cues, (1) children are unable to use them as effectively as adults, and (2) that the degree to which a cue improves performance changes over childhood.

2.2.2. *Selective Auditory Attention*

A related mechanism that is likely also developing during childhood is selective attention. To examine infants' ability to selectively listen in the frequency domain, Bargones and Werner (1994) measured infants' and adults' detection of pure-tone signals presented at expected and unexpected frequencies. Infants detected unexpected frequencies with the same consistency as they did expected frequencies. In contrast, adults seldom detected the presence

of unexpected signals. Findings from this study suggest that infants do not listen selectively like adults. Rather it is likely that infants are “broadband listeners,” monitoring many auditory filters. Unselective listening in the frequency domain may explain infants’ and children’s increased susceptibility to auditory masking compared to adults (e.g., Leibold & Neff, 2011; Lutfi et al., 2003b; Werner & Bargones, 1991).

In contrast to evidence of infants’ unselective listening in the frequency domain, infants appear to be able to form temporal expectations for sounds. Werner et al. (2009) showed that like adults, infants had better sensitivity to signals presented at the expected delay than when their presentation time was delayed in the unexpected condition. The discrepancy in results between the frequency and temporal domain may reflect methodology differences or the fact that development trajectory for selective auditory attention is domain specific.

2.3. The Importance of Knowing When in Time to Listen

Of specific interest for this dissertation is the effect of signal-temporal uncertainty, which is operationally defined here as the benefit observed when listeners have a temporally-defined listening interval compared to performance in a temporally-uncertain listening interval. The effect of signal-temporal uncertainty appears to be small (2 to 3 dB) for tone detection in noise (e.g., Egan et al., 1961; Green & Weber, 1980). Compared to these previous studies, a larger effect of signal-temporal uncertainty was observed by Bonino and Leibold (2008).

Bonino and Leibold measured detection thresholds for a 125-ms, 1000-Hz signal embedded in a random-frequency, multi-tonal masker. The maskers were continuous and

were either random-frequency, two-tone complexes or broadband noise samples. Due to a programming error, signals were not synchronized with the onset of one burst of a train of 130-ms masker bursts as was originally reported. Rather, signals occurred 80 ms after the onset of a masker burst. Four trained adults were tested in separate conditions with a temporally-defined or a temporally-uncertain listening interval for each masker. Figure 2-1 shows the average effect of signal-temporal uncertainty (± 1 S.E.) for both maskers. Data from individual listeners are provided as well. Data points above the horizontal dashed line indicate a benefit from the cue. The average effect of signal-temporal uncertainty for the random two-tone masker condition was 9 dB, whereas the effect was 2 dB in broadband noise.

Results from Bonino and Leibold (2008) suggest that effects of signal-temporal uncertainty are larger for conditions that produce substantial informational masking. This interpretation is also supported by findings from Best et al. (2007) and Varghese et al. (2012). In their studies, listeners could more accurately identify a target birdsong embedded in other birdsongs when provided with a cue indicating the time interval that contained the target. In contrast to these studies, Richards et al. (2011) reported no effect of signal-temporal uncertainty for detecting a tone in a random-frequency, multi-tonal masker. In that study, performance was compared between a continuous masker (temporally-uncertain) and a masker gated for each interval (temporally-defined). Thresholds were substantially better in the continuous masker, despite the condition being considered temporally-uncertain. Methodology differences used to create the temporally-defined condition may explain the contrast in results. Richards et al. (2011) manipulated the masker to define the listening

interval, whereas the other studies used a light cue to define the listening interval (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012).

2.4. Considerations for Applying the Same Theoretical Framework to Tonal and Speech Stimuli

In this dissertation, the same theoretical framework has been applied to two different types of auditory stimuli, tones and speech. One of the primary reasons for applying the same framework to this research is that similar developmental trajectories for auditory masking have been reported for both tonal and speech stimuli. As was reviewed above, by seven to eight years of age, children appear to be adult-like for both detecting tones (e.g., Elliott & Katz, 1980) and recognizing speech (e.g., Elliott et al., 1979; Nitttrouer & Boothroyd, 1990) in maskers expected to produce primarily energetic masking. However, increased child-adult differences in performance and a prolonged trajectory of development have been reported for maskers expected to produce substantial informational (or perceptual) masking. Specifically, this trend has been reported for tone detection in random-frequency, multi-tonal maskers (e.g., Hall et al., 2005; Leibold & Neff, 2007; Lutfi et al., 2003b; Oh et al., 2001) and for speech tasks using a masker with a limited number of competing talkers (e.g., Bonino et al., in press; Hall et al., 2002; Wightman & Kistler, 2005; Wightman et al., 2010). Similar developmental patterns for multi-tonal and speech maskers are consistent with the idea that the underlying mechanisms for children's increased susceptibility to masking are either the same or are closely related for the two stimulus sets.

This rationale is further strengthened by the mounting evidence that listeners appear to make masker confusions under complex listening situations for both tonal and speech

stimuli. In the classic simultaneous informational masking paradigm (Neff & Green, 1987), experimental controls are employed to reduce overlap of peripheral excitation patterns for the signal and the masker. Therefore, elevated thresholds in this condition are attributed to listeners failing to either perceptually separate the signal from the masker and/or selectively attend to the signal (reviewed by Kidd et al., 2007). In addition to observed threshold changes for these conditions, listeners' psychometric functions tend to have a reduced upper asymptote (e.g., Lutfi et al., 2003a). Lutfi et al. (2003a) suggested that, even at supra-threshold levels, listeners are confusing the masker for the signal on some proportion of trials. Error patterns for speech tasks also suggest that listeners make masker confusions (e.g., Brungart et al., 2001; Wightman & Kistler, 2005; for exception see Helfer & Freyman, 2009). For example, Wightman and Kistler (2005) asked 4- to 16-year-olds and adults to identify key items in the target message in the presence of a contextually-similar distractor message simultaneously presented to the same ear. For this closed-set task, very few errors were randomly selected responses. Rather, incorrect responses were items presented in the distractor message. The same error pattern was observed for both children and adults. Confusing the masker for the signal suggests that listeners were unable to either perceptually separate or attend to the target word in the presence of the competing speech (e.g., Brungart, 2001; Wightman & Kistler, 2005).

The last line of support for a similar underlying mechanism for susceptibility to masking is that tone detection and speech recognition can be improved with the provision of auditory grouping cues. One such example of successfully applying the same framework to both tonal and speech studies comes from our own recent work. Based on the theoretical framework used for the multi-tonal work (Leibold & Bonino, 2009), a speech study was

designed to examine the improvement for word recognition with a carrier phrase (e.g., “say the word ____”). For this study (Bonino et al., in press), it was predicted that listeners would benefit from a carrier phrase prior to the target word for conditions of high perceptual masking. This prediction was based on the observation that children are better at segregating a tonal signal when multiple bursts of the signal were provided in a random-frequency, multi-tonal masker (Hall et al., 2005; Leibold & Bonino, 2009). This finding is thought to be because the signal is spectrally-fixed across time while the masker is spectrally-uncertain. Results from Bonino et al. (in press) are also consistent with the carrier-phrase promoting spectro-temporal coherence of the target speech stream. In the continuous two-talker masker condition, the average carrier-phrase benefit was 17% for 5- to 10-year-old children. In this line of research, it has been helpful to use the same framework for both tonal and speech stimuli. Similarly, Wightman and his colleagues have used the same framework for dichotic listening with good agreement between the two stimulus sets (Wightman et al., 2003; Wightman & Kistler, 2005; Wightman et al., 2010).

Mounting evidence supports the application of the same theoretical framework to both speech and non-speech stimuli, but there are a couple of caveats to consider. First, studies using tonal stimuli can rigorously control spectral and temporal properties, whereas natural speech is dynamic. These parameter differences may affect the benefit of a particular grouping cue. Secondly, speech maskers will generate energetic masking in addition to perceptual masking. In contrast to the protected region imposed for most multi-tonal experiments, the voices of the masker and signal will have some frequency overlap. (For a speech paradigm that eliminates simultaneous masking see Kidd et al. (2008).)

The last caveat is that additional linguistic-cognitive factors may affect performance on speech recognition tasks. For example, listeners can use contextual information from preceding words in the sentence to improve performance (e.g., Fallon et al., 2002; Nittrouer & Boothroyd, 1990). Also, listeners recognize speech despite wide variation of vocal properties and articulation patterns across speakers. This process is referred to as talker normalization (e.g., Ladefoged & Broadbent, 1957). Moreover, both children and adults have better word recognition when words are presented by the same talker across trials than when the talker is varied (e.g., Mullennix et al., 1989; Ryalls & Pisoni, 1997). These effects tend to be large for conditions in speech-shaped noise; however, there are limited data on these effects from adults in speech maskers (e.g., Brungart et al., 2001; Helfer & Freyman, 2009; Kidd et al., 2008). It would not be surprising if developmental effects for cognitive-linguistic factors were greater for a two-talker masker than for speech-shaped noise.

2.5. Figures

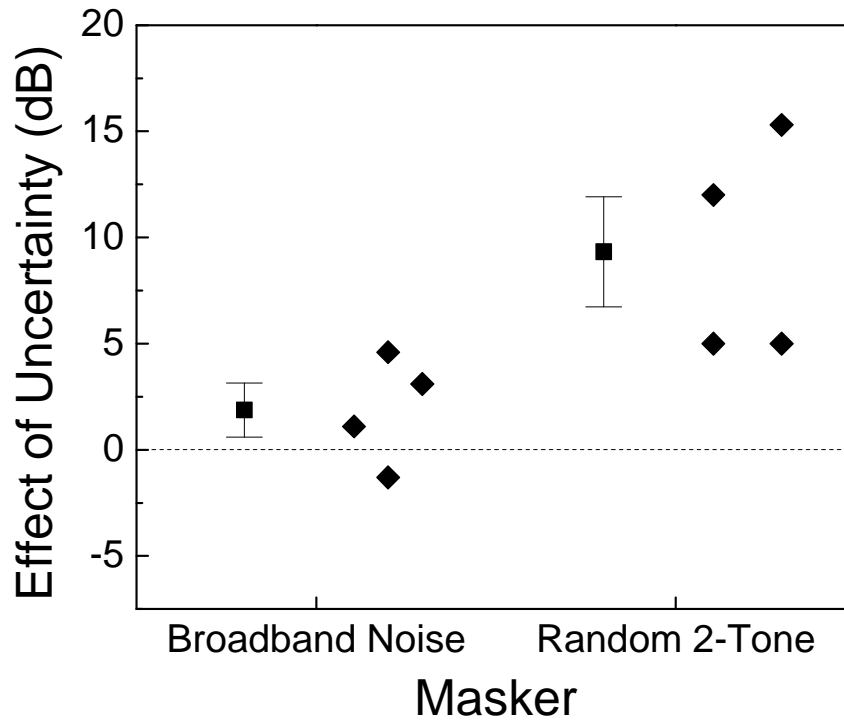


Figure 2-1. Individual and mean effect of signal-temporal uncertainty from Bonino and Leibold (2008)

The mean (± 1 S.E.) and individual benefit associated with having a defined listening interval, or the effect of signal-temporal uncertainty, in a broadband noise or a random, two-tone masker as reported by Bonino and Leibold (2008). Data points above the dash line indicate a benefit from the cue.

CHAPTER 3

EFFECT OF SIGNAL-TEMPORAL UNCERTAINTY DURING CHILDHOOD: DETECTION OF A TONAL SIGNAL IN A RANDOM-FREQUENCY, TWO-TONE MASKER

3.1. Introduction

This study examined the degree to which children benefit from a cue indicating when in time to listen for a tonal signal. This effect of signal-temporal uncertainty was operationally defined as the difference between listeners' thresholds for a 1000-Hz pure-tone signal when provided a temporally-defined listening interval compared to a temporally-uncertain listening interval. In the current study, the signal was embedded in a continuous, random-frequency, two-tone masker. Frequency components for the random-frequency masker were varied on each presentation, a manipulation that often produces substantial masking (e.g., Neff & Green, 1987). In this paradigm, experimental controls are employed to reduce energetic masking, produced as the result of overlapping excitation patterns in the peripheral auditory system (e.g., Fletcher, 1940). Rather, listeners' elevated thresholds in this paradigm appear to be the result of limited or ineffective central auditory processes, including separating the signal from the masker and selectively attending to the signal (e.g., Durlach et al., 2003). This type of masking has been termed "informational" (reviewed by Kidd et al., 2007). Previous studies have found relatively small effects of signal-temporal uncertainty for the detection of a pure tone in a noise masker expected to produce primarily energetic masking (e.g., Egan et al., 1961; Green & Weber, 1980). However, recent studies

suggest a larger effect of signal-temporal uncertainty in the presence of maskers believed to produce informational masking (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012; for exception see Richards et al., 2011). These findings suggest that a cue indicating when in time to listen for a target sound may assist adults in conditions in which they have difficulty segregating and/or attending to the target signal in the presence of complex maskers. Given children's increased susceptibility to informational masking relative to adults' (e.g., Leibold & Neff, 2007; Oh et al., 2001; Wightman et al., 2003), a cue indicating when in time to listen may be particularly beneficial for children.

This work was motivated by multiple observations across laboratories that children are more susceptible to auditory masking than adults (e.g., Allen & Wightman, 1995; Elliott & Katz, 1980; Hall et al., 2005; Werner & Bargones, 1991). Although child-adult differences have been observed for children under about eight years of age for conditions thought to be dominated by energetic masking (e.g., Allen & Wightman, 1994; Elliott & Katz, 1980; Jensen & Neff, 1993), more pronounced and prolonged child-adult differences have been observed for complex maskers believed to produce substantial informational masking (e.g., Allen & Wightman, 1995; Leibold & Neff, 2007; Lutfi et al., 2003b; Oh et al., 2001; Wightman et al., 2003). In the "simultaneous multi-tonal masker paradigm" introduced by Neff and Green (1987), the listener's task is to detect a fixed-frequency pure-tone signal presented simultaneously with a random-frequency multi-tonal masker. Using this paradigm, substantial masking has been observed for both children and adults, with child-adult differences as large as 50 dB (e.g., Oh et al., 2001). Moreover, results from studies employing these types of maskers have demonstrated improved performance with increased age during childhood (Leibold & Bonino, 2009; Leibold & Neff, 2007; Lutfi et al., 2003b;

Wightman et al., 2003). For example, Leibold and Neff reported significantly higher thresholds for 5- to 7-year-old compared to 8- to 10-year-old children for the detection of a 1000-Hz tone in the presence of two-tone maskers with either fixed or random spectra. Despite improvements in performance with increasing age during childhood, significant child-adult differences have been observed in adolescents for detection of a pure-tone signal embedded in a random-frequency, multi-tonal masker (Lutfi et al., 2003b; Wightman et al., 2003).

Mounting evidence suggests that children's greater susceptibility to informational masking compared to adults is due to immature central auditory processes, rather than peripheral immaturity (reviewed by Werner & Leibold, 2004). Although the specific central auditory mechanisms responsible for children's increased susceptibility to informational masking are not fully known, immature sound source segregation and/or selective attention processes may be responsible for children's difficulties (reviewed by Leibold, 2012). Sound source segregation is the process by which a listener perceptually parses apart sounds of interest from the competing background sounds. Features that assist adults in performing sound source segregation include spatial separation, asynchronous onset, incoherence of dynamic stimulus properties, and differences in sound quality or timbre (reviewed by Bregman, 1993). Results are mixed from the limited studies examining children's ability to use sound source segregation cues. Young children appear to be effective in their use of some cues, such as onset asynchrony (Hall et al., 2005; Leibold & Neff, 2007). In contrast, the ability to use other cues appears to follow a slower developmental time course (e.g., Hall et al., 2005; Wightman et al., 2003; Wightman & Kistler, 2005). For example, Hall et al. (2005)

reported that school-aged children (6 to 10 years of age) were not adult-like in their ability to use a lateralization cue based upon interaural intensity differences under headphones.

Related to the ability to segregate sounds, it has also been suggested that infants and children have immature selective auditory attention (e.g., Bargones & Werner, 1994; Wightman et al., 2010). Selective attention is the process of focusing on the auditory information of interest while disregarding irrelevant sounds. Results from studies of adults have been interpreted as indicating that their sensitivity to target signals can be improved by forming expectations about the signal. Stimulus features that adults can use to form expectations include frequency (e.g., Dai et al., 1991; Greenberg & Larkin, 1968), temporal position (e.g., Chang & Viemeister, 1991; Egan et al., 1961), duration (e.g., Dai & Wright, 1995), and spatial position (e.g., Arbogast & Kidd, 2000). Many of these studies used the probe-signal method (Greenberg & Larkin, 1968) to measure a listener's sensitivity. In this method, the listener is "set up" to listen for the primary signal either through training or because it is presented with a greater probability, thus it is thought of as the "expected" signal. In some trials the primary signal is then replaced by a probe signal. Probe signals are considered the "unexpected" signals. For probe-signal studies which have measured selective attention in the frequency domain (e.g., Greenberg & Larkin, 1968), detection thresholds are first measured for each of two signals alone. Then, signal levels are normalized relative to those single-signal thresholds. For these conditions, adults are more likely to detect expected than unexpected signals. Thus, adults are able to use stimulus features to selectively attend to a signal and to form expectations about the features of subsequent stimuli.

Infants' selective attention abilities appear not as refined as those of adults in the frequency domain. Using the probe-signal paradigm, Bargones and Werner (1994) reported

that infants were equally sensitive at detecting signals at expected and unexpected frequencies. In contrast to the data from infants, Bargones and Werner (1994) and others (e.g., Dai et al., 1991) have reported that adults have better sensitivity for the expected frequency tone. It is possible that the ability to listen selectively in the frequency domain continues to develop well into childhood. Greenberg et al. (1970) measured children's abilities to selectively attend based on signal frequency for a group of five children (6 to 8 years) and six adults (20 to 25 years). Mean results were interpreted by Greenberg et al. (1970) to indicate that 6- to 8-year-olds are able to listen selectively like adults. However, individual data suggest that some of the children tested were not as selective in the frequency domain as the adult listeners were.

One consequence of unselective listening in the frequency domain is that children are more susceptible to auditory masking. For example, a remote-frequency broadband noise that produces little or no masking for adults results in elevated threshold for most infants and some 4- to 6-year-olds (Leibold & Neff, 2011; Werner & Bargones, 1991). This finding is consistent with the idea that infants and children monitor a wider range of auditory filters than adults. Monitoring a wide frequency range may also explain why infants and children are often more susceptible than adults to masking produced by random-frequency, multi-tonal maskers (Lutfi et al., 2003b). For example Lutfi et al. (2003b) measured tone detection thresholds for children (4 to 16 years) and adults using a two-interval, force-choice procedure. In this task listeners detected a 1000-Hz tone embedded in a random-frequency masker. The number of masker components was manipulated across conditions, resulting in maskers thought to produce different amounts of informational masking. Results from a principal component analysis of the data suggest that children and adults use the same

listening strategy. Lutfi et al. (2003b) suggested that the variability seen between age groups and within age groups is explained by less selective auditory attention to the signal frequency. Less selective auditory attention may be the result of children monitoring a greater number of independent auditory filters.

In contrast to unselective listening in the frequency domain, limited data suggest that infants can form temporal expectancies under some conditions. Using the probe-signal method, Werner et al. (2009) measured infants' detection thresholds for a 150-ms, 1000-Hz pure-tone signal in a continuous broadband noise. On 75% of the trials the signal was presented at 500 ms following a brief increase in masker level. In the remaining trials the signal was presented either at 200 ms or 800 ms. Both infants and adults showed better sensitivity to the expected delay of 500 ms than for the unexpected delay of 800 ms. However, infants may not be able to use this information to improve tone detection. In an additional experiment presented by Werner et al. (2009), infants' sensitivity was measured in the no cue condition and in two cue conditions, in which the listening interval was temporally-defined with a brief increment or decrement in the noise level. Unlike adults, who showed improved sensitivity in the cue conditions, infants' sensitivity was similar for the no-cue and cue conditions. Thus, although infants are able to form temporal expectation, the most appropriate way to provide the cue and the degree to which they can use this information remains to be clarified.

Of specific interest for this study is to determine if children can use a visual cue indicating when in time to listen for a pure-tone signal in a random-frequency, two-tone masker. Relatively small effects of signal-temporal uncertainty have been reported for tone detection in noise for adults, for which the masker is thought to primarily produce energetic

masking (e.g., Egan et al., 1961; Green & Weber, 1980; Watson & Nichols, 1976). For example, Green and Weber (1980) measured detection thresholds in a two-interval, forced-choice task. The temporal position of the 50-ms, 1000-Hz pure-tone signal was varied within a sequence of 10 bursts of a broadband noise masker. Each masker burst was 50 ms in duration with a silent interval of 100 ms between masker bursts. Listeners were more sensitive to the signal when it always coincided with the fifth masker burst compared to when the temporal position was randomized to coincide with any of the 10 masker bursts. However, the average effect of signal-temporal uncertainty was less than 3 dB. In contrast to these findings, recent studies indicate a larger effect of signal-temporal uncertainty with maskers thought to produce substantial informational masking (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012). For example, Best et al. (2007) found that listeners could more accurately identify a target birdsong embedded in other birdsongs when they were provided with a cue indicating the time interval that contained the target. In a related study, Bonino and Leibold (2008) measured detection thresholds for a 1000-Hz tone embedded in either a random-frequency, two-tone masker or broadband noise. Adults demonstrated an average improvement of 9 dB for the two-tone masker when the listening interval was defined compared to the temporally-uncertain condition. Similar to previous investigations, a small average improvement of 2 dB was observed for the broadband noise condition. These results suggest that the effect of signal-temporal uncertainty is larger for conditions that produce substantial informational masking.

In order to measure the effect of signal-temporal uncertainty during childhood, the current study measured detection thresholds for children who were five to 13 years old. Following Bonino and Leibold (2008), listeners were asked to detect a 120-ms, 1000-Hz

signal presented in a continuous random-frequency, two-tone masker. This masker was selected because it was thought to produce substantial informational masking (e.g., Bonino & Leibold, 2008; Neff & Dethlefs, 1995). Results of this study were examined for both age-related changes in susceptibility to masking and effects of signal-temporal uncertainty. It was predicted that susceptibility to masking would decrease with age. However, that all age groups would have substantial benefit from the visual cue indicating when in time to listen for the signal.

3.2. Methods

3.2.1. *Listeners*

Listeners were 35 children and 10 adults. Children were recruited in three age groups: 5- to 7-year-olds (n=11), 8- to 10-year-olds (n=13), and 11- to 13-year-olds (n=11). The average age was 6.79 years (SD=0.83 years) for the 5- to 7-year-old group, 9.44 years (SD=0.94 years) for the 8- to 10-year-old group, and 12.56 years (SD=0.79 years) for the 11- to 13-year-old group. The rationale for recruiting within these three age ranges was to broadly examine whether the ability to benefit from the temporal cue improved with increasing age throughout childhood. A group of 10 adults was also tested (20 to 34 years; mean=26.68 years; SD=6.2 years). An additional six listeners did not complete testing. Two children (6.7 and 7.3 years) were unable to complete training. A 7.3-year-old's testing session was terminated early because of a consistently high false alarm rate (>75%) during phrase 2 of testing. Finally, three listeners (11.5, 12.2, and 31.4 years) withdrew from the study prior to completion. All listeners had thresholds in quiet of ≤ 20 dB HL bilaterally for octave frequencies from 0.25 to 8 kHz (ANSI, 2004). The sole exception was a 7.7-year-old,

who had two thresholds of 25 dB HL in the non-test ear. Listeners had no known history of chronic ear disease. Listeners were tested individually in a double-walled, sound-attenuated booth (IAC). Adults were tested in a single 2-hour visit. Children were tested over 2 or 3, 1-hour test sessions with regular breaks. This research was approved by the institutional review board at The University of North Carolina at Chapel Hill.

3.2.2. *Stimuli*

Thresholds were measured for a 1000-Hz pure-tone that was 120-ms in duration (including 5-ms, \cos^2 onset/offset ramps). The signal was embedded in a masker train composed of bursts of a random-frequency, two-tone masker. Each two-tone masker burst was 120 ms, including 5-ms, \cos^2 onset/offset ramps. During the plateaus of each masker burst, the overall level was 50 dB SPL (47 dB/component). The train of masker bursts was presented continuously in each block of trials, with no temporal overlap between successive bursts. The two frequency components making up the masker were independently selected on each 120-ms burst, one drawn at random from a uniform distribution with a range of 300 to 920 Hz and the other from a uniform distribution of 1080 to 3000 Hz. A protected region of 920 to 1080 Hz was used to reduce the contribution of energy-based masking within a presumed auditory filter centered at the signal frequency (Moore & Glasberg, 1983). The 1000-Hz signal, when present, was gated simultaneously with a single burst of the two-tone masker.

Stimuli were generated at a 25-kHz sampling rate. The signal and masker were attenuated digitally, played out of a real-time processor (TDT, RP2), routed through separate programmable attenuators that were set to 0 (TDT, PA5), mixed (TDT, SM3), and passed

through a headphone buffer with 3 dB of attenuation (TDT, HB7). Stimuli were presented to the listener's right ear through an insert earphone (Etymotic, ER1).

3.2.3. *Conditions and Procedure*

Listeners completed testing for two separate conditions: (1) temporally-defined and (2) temporally-uncertain. In the temporally-defined condition, a light cue consisting of five LED lights on a handheld unit defined the listening interval. A schematic of the temporally-defined condition is provided in Figure 3-1, with the light cues indicated by the gray boxes. A signal trial (bold dash) is shown in the left-hand box and a no-signal trial is shown in the right-hand box. The onset of the 600-ms light cue was synchronous with the onset of the second masker burst following the initiation of each trial. The signal, when present, always occurred 240 ms after the activation of the light cue. The temporally-uncertain condition was identical to the temporally-defined condition, except that the listener was not provided a light cue. Thus, the listener was not aware of when the trial was initiated by the experimenter in the temporally-uncertain condition.

Listeners sat inside the sound booth and responded with a hand raise. An experimenter sitting outside of the booth initiated each trial with a key press when the listener was judged to be "ready." In the temporally-defined condition, trials were initiated only if the experimenter judged the listener to be looking at the handheld device. When a trial was initiated, listeners had 4 s to indicate that they heard a signal. A failure to respond during that observation window was coded as a "miss" if a signal was present.

For each condition, the testing protocol consisted of one training phase and two testing phases. During training and both phases of testing, the signal/no-signal probability

was 0.5. The signal was presented at a clearly audible level in the training phase, based on results for adults from Bonino and Leibold (2008), and results for both children and adults from extensive pilot testing. The training period concluded when listeners were able to correctly identify 4/5 signals and 3/5 no-signals for a block of 10 training trials for each condition. After finishing training, listeners completed two testing phases.

Because individual variability tends to be large in the presence of multi-tonal maskers (e.g., Neff & Dethlefs, 1995), adaptive thresholds were measured in phase 1 of testing in order to determine the signal levels for each listener in phase 2 of testing. Adaptive thresholds were measured for both temporal conditions using a single-interval, one-up, one-down, adaptive procedure to estimate the 50% point on the psychometric function (Levitt, 1971). An initial step size of 4 dB was used, decreasing to 2 dB after the second reversal. The track only changed direction on signal-present trials. Testing concluded after six reversals, with threshold being the average signal level of the last four reversals. A minimum of three tracks was completed for each condition. For eight listeners whose threshold estimates varied more than 5 dB across tracks, additional tracks were conducted. The three tracks with the closest agreement were used to determine the mean threshold, except for three children. For these three children, only the last two tracks were used because the initial threshold estimates were different than threshold estimates from subsequent tracks by ≥ 10 dB. Testing order for the temporal conditions was randomized across the six threshold tracks.

The average threshold across tracks was used to select the signal levels for phase 2 of testing. Based on this average threshold, signals were presented at five levels during phase 2 of testing. The method of constant stimuli with a single-interval, Yes/No procedure was used. A run consisted of 40 trials across five signal intensities, with a 0.5 *a priori* signal/no-signal

probability, as in phase 1 of testing. Each track contained four signal and four no-signal trials at each of five levels, with trials presented in random order. The mid-point of the five signal levels was the average threshold corresponding to 50% correct signal detection from phase 1. The remaining four signal levels were +8, +4, -4 and -8 dB relative to the mid-point. The mid-point and/or level spacing were adjusted for most listeners across runs. Specifically, the mid-point was adjusted for approximately 80% of functions, either because asymptotic performance was not observed at the +8-dB level in the first run or because a hit rate of $\leq 20\%$ was not observed at the -8-dB level. For approximately 20% of functions, spacing was reduced to collect more data near the middle of the listener's psychometric function. Each listener completed four runs for each temporal condition. The temporal condition was randomized across the eight runs, although it was held constant within a run. If time permitted, one additional run was completed for listeners judged to have variable performance by the experimenter. An additional run was completed for five children, in one condition each, who had one run with a false alarm rate that was higher than their other runs. For these children, the run with the highest false alarm rate, pooled across all five signal levels, was omitted from further analysis.

3.2.4. *Procedure for Fitting Psychometric Functions*

Individual psychometric functions were fitted to listener data from 160 trials collected during phase 2 of testing for each condition. Prior to fitting the data, individual PC_{max} scores were calculated for each signal level. PC_{max} is the percent correct score corrected for bias. Fits were made using the procedure described by Dai (1995), modified to allow upper asymptote to be a free parameter of the fit (Dai & Micheyl, 2011). This procedure is based on

a chi-squared goodness of fit test, weighted by the number of observations at each level.

Individual psychometric functions were assumed to have a form of:

$$PC_{max} = \frac{1}{2}\lambda + (1 - \lambda)\Phi\left(\frac{d'}{\sqrt{2}}\right) \quad (1)$$

where λ relates to upper asymptote $(1 - \frac{\lambda}{2})$, and Φ is the cumulative Gaussian probability function. The first term represents the proportion of trials where the listener's performance was at chance, whereas the second term reflects the listener's performance for the proportion of trials where the listener was attending to the signal. The detectability index d' was defined as:

$$d' = \left(\frac{x}{\alpha}\right)^\beta \quad (2)$$

where x is signal strength, α is threshold, and β is psychometric function slope. Both α and β are free parameters, and α is the signal strength at $d' = 1$. Permissible parameter values were restricted for both slope ($0.05 \leq \beta \leq 3.0$) and upper asymptote ($0.001 \leq \lambda \leq 0.5$). Functions were considered to be poorly fitted if $R^2 < 0.5$.

3.3. Results

Fitting psychometric functions to child data is challenging for a variety of reasons. First, the number of trials is often limited. Second, auditory masking data are often more variable across individual children than across individual adults, especially for younger children (e.g., Allen & Wightman, 1995; Lutfi et al., 2003b; Oh et al., 2001). Third, children may show lower upper asymptote performance than adults (e.g., Allen & Wightman, 1994), necessitating the inclusion of a parameter to estimate upper asymptote (λ) in the fitting

procedure. This reduced upper asymptote indicates some degree of general inattentiveness to the task or that children do not hear the signal on some portion of the trials, even at supra-threshold levels. Asymptotic performance can also be reduced if on some trials the listener “confuses” the masker for the signal (e.g., Lutfi et al., 2003a), which may be the result of a segregation failure. This is particularly evident in data that are collected in maskers expected to produce substantial informational masking. Psychometric functions for multi-tonal masker conditions exhibit increased between-subject variability compared to those collected with broadband noise for both children and adult listeners (e.g., Durlach et al., 2005; Lutfi et al., 2003a).

Mean estimates for the fitting parameters (threshold, slope and upper asymptote ($1 - \frac{\lambda}{2}$)), R^2 and criterion are provided in Table 3-1 for each listener age group based on temporal condition. Examples of individual psychometric function fits are provided in Figure 3-2. This figure includes data from three 5- to 7-year-olds in the temporally-uncertain condition. The observed data are indicated by the squares. The grayscale indicates the number of trials for a particular signal level, with white being the minimum (n=8) and black being the maximum (n=32) for these three listeners. The fitted functions are indicated by the solid black lines, where upper asymptote is a free-parameter of the model. For comparison, fits are also provided with the upper asymptote fixed to 1.0,¹ indicated by the dashed line.

3.3.1. *Goodness of Fit*

Functions were considered to have acceptable fits if $R^2 \geq 0.5$. Of the 45 listeners tested in the current study, 32 met this R^2 criterion in both the temporally-uncertain and temporally-defined conditions. Thirteen listeners (12 children and 1 adult) failed to achieve

this criterion. However, no listener had fits of $R^2 < 0.5$ for both conditions. Of the 12 functions with an $R^2 < 0.5$ for the child data, eight were in the temporally-defined condition and four were in the temporally-uncertain condition. Children with fitted functions with an $R^2 < 0.5$ were not concentrated in one age group: three were 5 to 7 years of age, five were 8 to 10 years of age, and four were 11 to 13 years of age. One adult (31.1 years of age) had a fit with an $R^2 < 0.5$ in the temporally-uncertain condition. Further analyses, reported below, were restricted to data from the 32 listeners who had fits of $R^2 \geq 0.5$ in both conditions.

Psychometric functions were also examined for non-monotonicity. It has been reported in the literature that listeners' performance can be worse than anticipated near a 0 dB signal-to-noise ratio (SNR; e.g., Brungart, 2001; Wightman & Kistler, 2005). Non-monotonic performance for these studies is thought to be the result of listeners' access to a level difference cue between the signal and the masker for levels below and above 0 dB SNR. At a 0-dB SNR this cue is not available to the listener, thus, exacerbating informational masking. For each listener, residuals were calculated by finding the difference between the predicted and observed PC_{max} values. Individual listener's residuals were compiled as a function of signal level. Visual inspection of these data indicated that the mean and the spread of residuals were consistent as a function of signal level. In particular, there was no evidence that thresholds were consistently underestimated at 0-dB SNR or overestimated at other SNRs.

3.3.2. *Threshold Estimates and the Effect of Signal-Temporal Uncertainty*

Group differences. Mean estimates of masked threshold (± 1 SE), based on a $d'=1$, are provided in Figure 3-3 for the four listener age groups. Thresholds are represented by

filled squares for the temporally-uncertain condition, and by open circles for the temporally-defined condition. In the temporally-uncertain condition, mean thresholds were 47.7 dB SPL for 5- to 7-year-olds, 49.4 dB SPL for 8- to 10-year-olds, 49.5 dB SPL for 11- to 13-year-olds, and 48.7 dB SPL for adults. All four listener age groups demonstrated lower mean thresholds in the temporally-defined compared to the temporally-uncertain condition. Mean thresholds in the temporally-defined condition were 38.3 dB SPL for 5-to 7-year-olds, 38.2 dB SPL for 8- to 10-year-olds, 35.4 dB SPL for 11- to 13-year-olds, and 37.9 dB SPL for adults.

In order to examine the benefit of having a defined listening interval (the effect of signal-temporal uncertainty), the threshold for the temporally-defined condition was subtracted from the threshold for the temporally-uncertain condition for each listener. The effect of signal-temporal uncertainty is shown in Figure 3-4 for the group mean (± 1 SE) and individual data, plotted as a function of listener age group. A positive score, above the dotted horizontal line, indicates that the listener benefited from the cue. All listeners appeared to benefit from defining the listening interval, except for one child (7.8 years of age). The average benefit was 9.3 dB for 5- to 7-year-olds, 11.2 dB for 8- to 10-year-olds, 14.1 dB for 11- to 13-year-olds, and 10.8 dB for adults.

A repeated measures analysis of variance (ANOVA) of threshold was performed to assess developmental effects for susceptibility to masking and signal-temporal uncertainty. This analysis included the within-subjects factor of Condition (temporally-defined, temporally-uncertain) and the between-subjects factor of Age Group (5- to 7-year-olds, 8- to 10- year-olds, 11- to 13-year-olds, and adults). The main effect of Condition was significant [$F(1,28)=138.87, p<0.001$]. This result confirms the trend in Figure 3-4, indicating that all

listener age groups benefited from the cue. However, neither the main effect of Age Group [$F(3,28)=0.04$, $p=0.99$] nor the Condition \times Age Group interaction [$F(3,28)=1.02$, $p=0.40$] was significant. The lack of any significant age-related effects, confirms similar performance across the four age groups, both in terms of susceptibility to masking and the effect of signal-temporal uncertainty.

Individual differences. Consistent with other studies of informational masking involving both children and adults (e.g., Lutfi et al., 2003a; Oh et al., 2001), there was considerable between-subjects variability. In Figure 3-5, individual thresholds are plotted as a function of age for the temporally-uncertain (filled squares) and the temporally-defined (open circles) conditions. The black vertical lines indicate the effect of signal-temporal uncertainty for each listener. The gray symbols indicate results for the 13 listeners who had fits of $R^2 \geq 0.5$ in only one of the two conditions. For listeners with fits of $R^2 \geq 0.5$ in both conditions, shown in black, masked thresholds in the temporally-uncertain condition ranged from 25.5 (20.2-year-old) to 59.7 dB SPL (12.9-year-old). Masked thresholds in the temporally-defined condition ranged from 5.7 dB SPL (11.3-year-old) to 48.8 dB SPL (34.3-year-old). Despite large between-subjects variability in susceptibility to masking, nearly all listeners benefited from the listening interval being defined in time. The average effect of signal-temporally uncertainty was 11.2 dB, although seven listeners showed more than a 15-dB effect. Only one listener, a 7.8-year-old, showed less than a 4-dB effect of signal-temporal uncertainty (a 0.3 dB effect).

One possible limitation of the current study is that 13 listeners were excluded because of a poorly fitted psychometric function. It is possible that these listeners performed differently than those with well-fitted functions. Recall that all listeners with fits of $R^2 < 0.5$

in one condition had fits associated with $R^2 \geq 0.5$ in the other condition. All but two of these threshold estimates were within two standard deviations of the range of performance obtained for listeners who had well-fitted functions for both conditions. The two outlier thresholds were lower than the mean data, and both were from children: a 12.3-year-old in the temporally-defined condition and a 5.8-year-old in the temporally-uncertain condition. However, there were two listeners (a 11.3- and a 20.2-year-old), with good fits for both functions, who also had similarly low thresholds.

3.3.3. *Response Bias*

In the single-interval procedure used in the current study, listeners were asked to indicate when they heard a signal, and trials were recorded as no responses when the listener did not respond within the allotted time. Previous findings indicate that adults tend to adopt a conservative (strict) response criterion for indicating that a signal is present in this type of paradigm (e.g., Lucluyse & Meddis, 2009; Marshall & Jesteadt, 1986). In order to examine the effect of bias on threshold, Marshall and Jesteadt (1986) compared adults' thresholds measured in quiet for the standard clinical procedure and for two psychophysical methods: a two-interval, forced-choice (2IFC) adaptive procedure and a Yes/No procedure with undefined observation intervals. Marshall and Jesteadt (1986) reported that the average threshold obtained with the clinical procedure was 6.5 dB greater than for the 2IFC procedure. Response bias had a minimal effect (1.2 dB) on threshold differences between the procedures, despite listeners being more conservative for Yes/No procedures.

In order to examine age-related differences in response bias for the two temporal conditions in the current study, criterion estimates were calculated for the 32 listeners for

whom data fits in both conditions were associated with $R^2 \geq 0.5$. Responses for all signal levels were pooled for each listener by temporal condition (temporally-defined or temporally-uncertain). The criterion (c) was defined as (Snodgrass & Corwin, 1988):

$$c = -\frac{1}{2}[z(H) + z(F)] \quad (3)$$

with $z(H)$ and $z(F)$ being the z -scores associated with the hit rate and false-alarm rate, respectively. For listeners who did not produce any false alarms, a value of 0.5 was used as the number of false alarm trials, resulting in a false alarm rate of 0.6%.² An unbiased listener would have a criterion of 0. For a biased listener, a negative value indicates the listener is more liberal in their decision to indicate a signal is present, and a positive value indicates the listener is more conservative.

Individual estimates of criterion are provided in Figure 3-6, with open circles indicating estimates in the temporally-defined condition and open squares representing the temporally-uncertain condition. Mean bias estimates (± 1 SE) are represented by filled symbols and are provided for each listener age group to the left of the individual data. Listeners were more conservative in the temporally-uncertain listening condition than when the listening interval was defined. Mean data are similar across the four age groups. Across all listener age groups, the mean criterion was 0.91 (SE = 0.06) for the temporally-uncertain condition and 0.07 (SE = 0.006) for the temporally-defined condition. Note that individual differences appeared to be larger in the temporally-uncertain condition (range: 0.16 to 1.31) than in the temporally-defined condition (range: 0.01 to 0.12).

To test these trends, a repeated measures ANOVA was conducted with the within-subjects variable of Condition (temporally-defined, temporally-uncertain) and the between-

subjects variable of Age Group (5- to 7-year-olds, 8- to 10-year-olds, 11- to 13-year-olds, and adults). The main effect of Condition was significant [$F(1,28)=217.37, p<0.001$]. However, neither the main effect of Age Group [$F(3,28)=0.30, p=0.83$] nor the Condition \times Age Group interaction [$F(3,28)=0.33, p=0.81$] was significant. Thus, listeners of all ages were more conservative in their response of “yes” when the listening interval was uncertain compared to when it was defined.

3.3.4. *Other Estimated Parameters of the Psychometric Function*

Although threshold was the primary psychometric function parameter of interest in the current study, upper asymptote and slope were also estimated for each listener during the fitting process. Mean estimates of these parameters are provided in Table 3-1 for each of the listener age groups. However, upper asymptote and slope estimates should be interpreted with care. During data collection, most listeners had a limited number of trials which resulted in $\geq 90\%$ PC_{max} scores. This resulted in upper asymptote estimates of the current study to be reduced compared to those in the published literature (e.g., Lutfi et al., 2003a). Across listeners, the average upper asymptote was 93.1% (SE=0.8%) in the temporally-uncertain condition and 86.8% (SE=1.1%) in the temporally-defined condition. However, the effect of upper asymptote did not appear to have a large effect on threshold estimates in the current study. Fixing the upper asymptote at 100% ($\lambda=0$) still resulted in the same overall pattern of results for this experiment.

3.4. Discussion

This study measured tone detection in a continuous random-frequency, two-tone masker. The signal was presented at either defined or uncertain listening intervals to test the main hypothesis that children would benefit from a visual cue indicating when in time to listen. Previous studies have shown that children's performance can improve when auditory grouping cues are provided. Like adults, children are able to use asynchronous onset of the signal and masker (Hall et al., 2005; Leibold & Neff, 2007) and multiple presentations of the signal burst (Hall et al., 2005; Leibold & Bonino, 2009) to improve thresholds. However, other studies have reported that children are unable to use some auditory grouping cues that adults can use. For example, presenting the masker and signal to separate ears results in substantial improvement for adults, whereas children appear to have minimal benefit (e.g., Hall et al., 2005; Wightman et al., 2003). For the current study, it was predicted that children would benefit from the visual cue. This hypothesis was based on the observation that infants were able to form temporal expectations (Werner et al., 2009) and that school-aged children have a release from informational masking when other temporal cues are provided (e.g., Hall et al., 2005; Leibold & Neff, 2007).

One qualification of this hypothesis, however, was that a robust effect of signal-temporal uncertainty would only be observed for maskers expected to produce substantial informational masking. Findings from this study do indeed suggest that the random-frequency, two-tone masker produced substantial informational masking for listeners of all ages. Specifically, masked thresholds were greater than predicted by overlapping patterns of excitation along the basilar membrane. An excitation-based model of loudness (Moore et al., 1997) has been successfully used to predict masked thresholds in previous studies (e.g., Van

Der Heijden & Kohlrausch, 1994), based on the partial loudness of the signal. Applying this model to the current stimuli, the predicted amount of energetic masking produced varies widely depending on the two frequency components selected for each masker. Using a criterion partial loudness of 2 phons, the model predicts a threshold of 41.6 dB for a masker composed of 920 Hz and 1080 Hz and 6.3 dB for a masker composed of 300 and 3000 Hz. In comparison to these predictions, thresholds in the temporally-uncertain condition ranged from 25.5 dB SPL to 59.7 dB SPL, with a mean threshold of 48.8 dB SPL across all listeners. The observed thresholds are consistent with this masker producing substantial informational masking.

3.4.1. *Effect of Signal-Temporal Uncertainty*

As predicted, all three age groups of children (5 to 7, 8 to 10 and 11 to 13 years of age) and the group of adults benefited from the light cue indicating when in time to listen for the target signal. Moreover, no developmental differences in the effect of signal-temporal uncertainty were observed. The average effect of signal-temporal uncertainty was 11.2 dB across all listeners, and the group average thresholds varied from 9.3 to 14.1 dB across the four age ranges. These results are consistent with the average effect of 9 dB (range of 5 to 15 dB) for four trained adult listeners reported by Bonino and Leibold (2008). Bonino and Leibold (2008) used similar stimuli to the current study with two exceptions: (1) there was a temporal asynchrony between the onset of the signal and the masker burst and (2) the listening interval in that study was defined by having the listener initiate the start of the trial. Defining the listening interval either through a visual cue or allowing the listener to initiate

the trial appears to result in a similar benefit of knowing when to listen for the 1000-Hz signal.

The effect of signal-temporal uncertainty is larger in this study than the 2- to 3-dB effect typically reported for testing in quiet or in noise (e.g., Egan et al., 1961; Green & Weber, 1980; Watson & Nichols, 1976). Additional pilot data collected in the laboratory are consistent with a small effect of signal-temporal uncertainty in broadband noise. Using procedures similar to the current study,³ detection thresholds in broadband noise were measured for an additional 6.7-year-old and five adults (21 to 31 years of age). Comparing thresholds for the temporally-uncertain and the temporally-defined conditions, the 6.7-year-old showed an effect of 3.0 dB. The effect for adults ranged from 1.6 to 3.8 dB, with a mean effect of signal-temporal uncertainty of 2.5 dB (SD=1.0 dB). The relatively large effect (11.2 dB) seen for the current study in the random-frequency, two-tone masker supports the idea that the effect of signal-temporal uncertainty is greater for conditions resulting in high rather than low informational masking.

The current findings are in general agreement with recent studies which have examined the effect of knowing when to listen in conditions believed to produce substantial informational masking (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012). As described above, Bonino and Leibold (2008) observed a larger effect for the same group of adult listeners in a random-frequency, two-tone masker than in a broadband noise. A similar pattern of results has also been reported for bird song stimuli (Best et al., 2007; Varghese et al., 2012). Varghese et al. (2012) asked listeners to detect a target bird song presented in either a chorus of novel bird songs or in a noise masker which had the same long-term spectral characteristics. When the target and competing bird song chorus were perceived as

originating from the same spatial location, listeners were better able to identify the target when a temporal light cue was provided. The temporal cue did not improve performance in noise for this spatial configuration. In contrast to these studies, Richards et al. (2011) reported that listeners did not benefit from a defined listening interval. In that study, adults were asked to detect a signal comprised of series of three 60-ms, 1000-Hz tone bursts. Listeners were tested in two masker conditions, “continuous” and “pulsed.” In the continuous condition, a series of 60-ms random-frequency, masker bursts, with no temporal gap, was continuously played for a run of trials. In the “pulsed” condition, the masker was a series of five bursts, which was only presented during the listening interval. If listeners benefited from a defined listening interval, it was expected that thresholds would be better in the pulsed condition than when the listening interval was not defined in the continuous condition. However, the opposite result was found. Thresholds were 11.7 dB better in the continuous condition than the pulsed condition. One explanation for the lack of an effect of signal-temporal uncertainty compared to the other three studies (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012) is methodology differences. Whereas Richards et al. (2011) gated the masker to define the listening interval, the current study and other three published studies (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012) did not manipulate the masker, but rather provided a light cue in the temporally-defined condition.

3.4.2. *Decision Strategy*

All listener age groups were conservative in the temporally-uncertain condition. The average response bias was 0.91 in the temporally-uncertain condition across all listeners. This finding is consistent with conservative bias estimates, which are commonly seen for

adult listeners tested in the modified observer-based paradigm (e.g., Leibold & Werner, 2006; Werner & Marean, 1991; Werner et al., 2009). In this paradigm, the listening interval is not defined for the listener, but it is defined for the experimenter judging the listener's behavior. Werner et al. (2009) reported that mean bias was 1.15 for a group of adult controls for tone detection in broadband noise. In contrast to adults, data collected from infants in the observer-based paradigm consistently indicate no response bias (e.g., Leibold & Werner, 2006; Werner & Marean, 1991; Werner et al., 2009). Although infants appear to be unbiased listeners, the school-aged children tested in the current study showed a conservative decision strategy similar to adults in the temporally-uncertain condition.

In contrast to the conservative strategy used by listeners in the temporally-uncertain condition, both children and adults adopted an unbiased decision strategy with the provision of the light cue. The average response bias in the temporally-defined condition was 0.07 across all listeners. Werner et al. (2009) also observed that adult listeners adopted a less strict criterion when the listening interval was defined by a level increment cue compared to the temporally-uncertain condition. Thus, in the current study, listeners adopted an unbiased listening strategy in the temporally-defined condition as is thought to be case for the two-interval, forced-choice procedure. Furthermore, it has been suggested that the differences in adults' threshold estimates in quiet between a Yes/No procedure and a two-interval, forced-choice procedure may be a result of signal-temporal uncertainty (Lucluyse & Meddis, 2009; Marshall & Jesteadt, 1986). One explanation for why a defined listening interval results in listeners becoming unbiased is that it allows them to take advantage of their knowledge of the signal/no-signal ratio. In the current study the ratio was 0.5, and listeners were reminded of this probability before the start of each temporally-defined run. Regardless of the

underlying reason, differences in decision strategy are apparent across the two temporal conditions. Although the effect of bias appears to be small (1.2 dB) for conditions in quiet (Marshall & Jesteadt, 1986), it may be larger in the presence of random-frequency, multi-tonal masker. The current findings highlight the importance of controlling for bias, since it may differ across conditions.

3.4.3. *Susceptibility to Masking*

Although masking was observed for both children and adults, developmental differences in susceptibility to masking were not observed in the current study. This result was unexpected. Previous studies have consistently reported increased susceptibility to informational masking for children compared to adults (e.g., Lutfi et al., 2003b; Oh et al., 2001), as well as age-related improvements during childhood (e.g., Leibold & Bonino, 2009; Leibold & Neff, 2007). The surprising result of similar thresholds for children and adults raises the question of whether adults' thresholds in the current study are poorer than in previous studies, whether children's thresholds are better than in previous studies, or whether the data of both children and adults differ from those in previous studies.

The current threshold estimates for adult listeners appear to be generally consistent with other adult data in the literature. In the current study, adults' mean threshold was 49 dB SPL (-1 dB SNR) in the temporally-uncertain condition. The average threshold across four trained adult listeners tested with similar stimuli by Bonino and Leibold (2008) was 28.3 dB SPL (or -21.7 dB SNR) in the temporally-uncertain condition. Thresholds reported by Bonino and Leibold (2008) are likely better than those reported in the current study, because there was a temporal asynchrony of the signal and masker of 80 ms. Recent results in our

laboratory from three different trained adults, indicate that the average thresholds was 49.2 dB SPL (or -0.8 dB SNR) in the temporally-uncertain condition when the onset of the signal and masker were synchronized and all other procedures were identical to that of Bonino and Leibold (2008). Richards et al. (2011), as described earlier, measured adults' detection of a signal composed of series of three 60-ms, 1000-Hz tone bursts presented in a continuous random-frequency masker at an overall level of 63 dB SPL. Thresholds in that study were 55 dB SPL (-8 dB SNR). This mean threshold is 7 dB better in SNR than the current study's adult mean threshold. Leibold and Werner (2006) also tested a control group of adults in a continuous random-frequency, two-tone masker at an overall level of 60 dB SPL. Unlike the current study and Richards et al. (2011), Leibold and Werner inserted a 300-ms silent interval between each of the 300-ms masker bursts. Using a single-interval adaptive procedure, Leibold and Werner (2006) reported that the mean threshold for detecting a 300-ms, 1000-Hz pure-tone signal was approximately 60 dB SPL (0 dB SNR). Although the methodology differences between these studies make a direct comparison difficult, thresholds for adults in the current study appear to be consistent with those of Leibold and Werner (2006) and slightly elevated compared to those of Richards et al. (2011). Furthermore, considering the size of child-adult differences typically reported for random-frequency, multi-tonal maskers, the differences seen in threshold estimates for this group of adults compared to previous studies does not adequately account for the lack of a child-adult difference.

Previous studies have reported substantially higher thresholds for children in the presence of random-frequency, multi-tonal maskers than observed in the current study (e.g., Leibold & Bonino, 2009; Leibold & Neff, 2007; Lutfi et al., 2003b; Oh et al., 2001). For example, in Leibold and Bonino (2009) the average threshold was 84.5 dB SPL (or 24.5 dB

SNR) for 5- to 7-year-olds and 66.4 dB SPL (or 6.4 dB SNR) for 8- to 10-year-olds. In comparison, the average adult threshold was 50.5 dB SPL (or -10 dB SNR). In Leibold and Bonino (2009) the task was to detect a two-burst, 50-ms, 1000-Hz signal in a masker sequence of 10 two-tone, 50-ms bursts at an overall level of 60 dB SPL. In contrast to these thresholds, estimates in the current study are substantially lower for children. For example, 5- to 7-year-olds had an average threshold of 47.7 dB SPL (or -2.3 dB SNR) in the temporally-uncertain condition and 38.3 dB SPL (or -11.7 dB SNR) in the temporally-defined condition. Thus, the similar masked thresholds observed across age in the current study appear to largely reflect relatively better performance for children compared to previous developmental work.

3.4.4. *Methodology Differences*

In light of children's relatively low thresholds in the current study it is worth considering the influence of methodology differences between the current and previous studies. The first difference is that supra-threshold signals were presented frequently throughout testing in the current study. These supra-threshold trials (e.g., those associated with +4 or +8 dB) were distributed uniformly over the trials in phase 2 of testing. In contrast to adaptive procedures used in previous studies, five signal levels were presented during a block of 40 trials, with level being randomly assigned across trials. It is possible that exposure to supra-threshold signal levels may have reminded the listener about the features of the signal. Research with adult listeners for a two-interval task indicate that detection thresholds for a pure tone embedded in a multi-tonal masker can improve by as much as 20 dB when provided a pre-trial "reminder" of the signal in quiet (Richards & Neff, 2003). This

frequent reminder of the signal may have allowed the listener to more selectively monitor auditory filters in the frequency region of the signal. However, there are three considerations that cast doubt on this interpretation for the current study. First, it is not clear how effective this type of cue would be considering the long delay between trials. Secondly, while clearly audible signals may have reminded the listener what the signal sounds like, they might have also increased the listener bias; that is, the inclusion of high-level signals could have reduced the listener's probability of responding to less clearly audible signals. Thirdly, previous developmental studies which used a pre-trial reminder cue have reported large child-adult differences for susceptibility to multi-tonal maskers (Leibold & Neff, 2007; Lutfi et al., 2003b; Oh et al., 2001).

The other primary methodology difference is that this experiment used a continuous, random-frequency, two-tone masker with no temporal offset between masker bursts. In contrast, other developmental studies that used a random-frequency, two-tone masker have typically used either a single brief masker burst (Lutfi et al., 2003b; Oh et al., 2001; Wightman et al., 2003) or a brief masker train (Bonino & Leibold, 2008; Hall et al., 2005) played during each interval. One exception is Leibold and Werner (2006), who used a continuous masker to assess the effects of informational masking for infants. In that study, infants and a control group of adults detected a 300-ms, 1000-Hz pure-tone signal that, when present, was gated with the onset of a 300-ms masker burst. Although the masker was on continuously during a block of trials, there was a 300-ms silent interval between each 300-ms masker burst.

Although highly speculative, three possible explanations are considered for why a continuous masker could result in equivalent thresholds for children and adults. The first

possible explanation is that a continuous masker may provide listeners with a better opportunity to learn the statistics of the masker compared to a gated masker. Recent findings from Richards et al. (2011), as discussed earlier, indicate that playing the masker continuously resulted in better tone detection (of 11 dB on average) and less individual variability than when a gated interval paradigm was used. Results from additional experiments by Richards et al. (2011) indicate that thresholds improved when additional masker bursts were provided before the signal in the pulsed condition. One interpretation of the findings in Richards et al. (2011) is that listeners are better able to detect a signal in the continuous condition because they notice a change in the features (or statistics) of the masker when they are exposed to the masker alone before the signal interval. In the current study, listeners may be able to learn features of the masker, including its level and frequency parameters. However, doubt is cast on this possible explanation because a listener should be able to learn the statistics from a large number of disjointed trials, not necessarily requiring the masker to be on continuously. Furthermore, it is not clear why children would perform like adults in the current study and not for other developmental studies which have presented the same masker frequencies over disjointed trials (e.g., Leibold & Neff, 2007), unless there are developmental differences in the ability either to form or to use the statistics in that condition.

A second possible explanation is that the masker's temporal regularity and spectral restrictions could be weak cues for streaming in the current study. Findings from studies that examined the number of perceived auditory streams suggest that the time period required to build up an auditory stream is at least on the order of several seconds for adult listeners (e.g., Anstis & Saida, 1985; Bregman, 1978; Carlyon et al., 2001). Furthermore, listeners are

biased towards maintaining two separate streams once they are formed; however, introducing several seconds of silence or unpatterned noise can reduce this bias (Bregman, 1978). Thus, one explanation for reduced thresholds in the current study is that children are able to build up a masker stream given the long listening opportunity afforded by the continuous masker. However, if that was the case, adults should show benefit as well, provided they are not at floor already.

The last possible explanation is that children may perform more poorly than adults when the masker is gated because they are more prone to the distracting effects of onsets than adults. Compared to a continuous noise masker, gated maskers are associated with both elevated thresholds (e.g., Green, 1969) and poor frequency-selective listening (e.g., Dai & Buus, 1991). Specifically, Dai and Buus (1991) measured adults' sensitivity for target and probe signals in a band of noise (630 to 1470 Hz) using the probe-signal method (Greenberg & Larkin, 1968). Similar to previous studies of frequency-selective listening (e.g., Greenberg & Larkin, 1968), Dai and Buus (1991) observed that four adults selectively listened for the 1000-Hz target signal when the masker was on continuously. In contrast, when the masker was gated for each interval, listeners detected the probe signals with equal or greater probability compared to the target signals. Thus, a gated masker impaired listeners' abilities to selectively listen in the frequency domain. As proposed by Dai and Buus (1991), these data suggest that a gated masker results in either listeners monitoring several bands or in a widening of the auditory filter (e.g., Green, 1969). It is possible that there would also be differences in listeners' abilities to selectively listen in a gated versus continuous random, two-tone masker. A continuous masker may be particularly beneficial for children, because it

has been suggested that their increased susceptibility to informational masking is a result of monitoring a greater number of independent auditory filters (Lutfi et al., 2003b).

3.5. Conclusions

In the current study, listeners benefited from a visual cue indicating the timing of the listening intervals for a signal presented in a random-frequency masker. Furthermore, the benefit associated with presenting the visual cue was similar across the four listener age groups (5- to 7-year-olds, 8- to 10-year-olds, 11- to 13-year-olds and adults) tested. This average effect was 11.2 dB, which is substantially larger than previous reports of the effect of signal-temporal uncertainty in quiet or broadband noise conditions (e.g., Egan et al., 1961; Green & Weber, 1980). This finding is consistent with the idea that knowing when in time to listen is particularly beneficial under conditions expected to produce substantial informational masking, and that the ability to benefit from temporal cues is adult-like in school-aged children for the conditions tested here.

3.6. Endnotes

¹ Fixing the upper asymptote to 100% ($\lambda=0$) still resulted in the same overall pattern of threshold results for this experiment.

² There were 13 cases in which the total number of false alarms was zero. The 0.6% false alarm rate was calculated by using the 0.5 correction divided by the number of no-signal trials ($n=80$).

³ The masker was a continuous stream of 120-ms broadband noise bursts (300 to 3000 Hz). All other methods were identical to the current study for the 6.7-year-old child. Two modifications were made for adult listeners compared to the current study: only three runs of 40 trials were conducted for each condition, and signal intensities were +4 dB, +2 dB, 0 dB, -2 dB and -4 dB relative to the threshold values obtained during the adaptive threshold procedure (phase 1 of testing).

3.7. Tables

Age Group	Temporally-Defined					Temporally-Uncertain				
	Threshold	Slope	Upper Asymptote	R ²	Bias	Threshold	Slope	Upper Asymptote	R ²	Bias
5-7 yrs	38.34 (2.41)	1.78 (0.38)	0.84 (0.02)	0.79 (0.03)	0.08 (0.01)	47.66 (1.90)	0.94 (0.32)	0.92 (0.01)	0.88 (0.03)	0.99 (0.12)
8-10 yrs	38.19 (1.58)	0.60 (0.11)	0.86 (0.02)	0.76 (0.05)	0.06 (0.01)	49.36 (2.03)	0.67 (0.25)	0.95 (0.02)	0.79 (0.06)	0.91 (0.12)
11-13 yrs	35.38 (5.48)	0.75 (0.23)	0.90 (0.01)	0.78 (0.05)	0.06 (0.01)	49.50 (3.80)	1.04 (0.33)	0.93 (0.02)	0.78 (0.04)	0.90 (0.13)
Adults	37.95 (2.68)	1.38 (0.33)	0.87 (0.02)	0.70 (0.03)	0.07 (0.01)	48.72 (3.15)	1.25 (0.34)	0.93 (0.01)	0.86 (0.03)	0.83 (0.11)

Table 3-1. Mean parameter estimates from the fitted psychometric functions by listener age group and temporal condition

Mean estimates of threshold (dB SPL), slope ($\log d'/\text{dB}$), and upper asymptote ($1 - \frac{\lambda}{2}$) are provided from the fitted psychometric functions in the temporally-defined and temporally-uncertain listening conditions by listener age group. Values of R^2 show the proportion of variance in the data accounted for on average by individuals' fitted functions. The standard error is provided for each estimate in parenthesis. These values are from the 32 listeners who had fits associated with $R^2 \geq 0.5$ for both conditions.

3.8. Figures

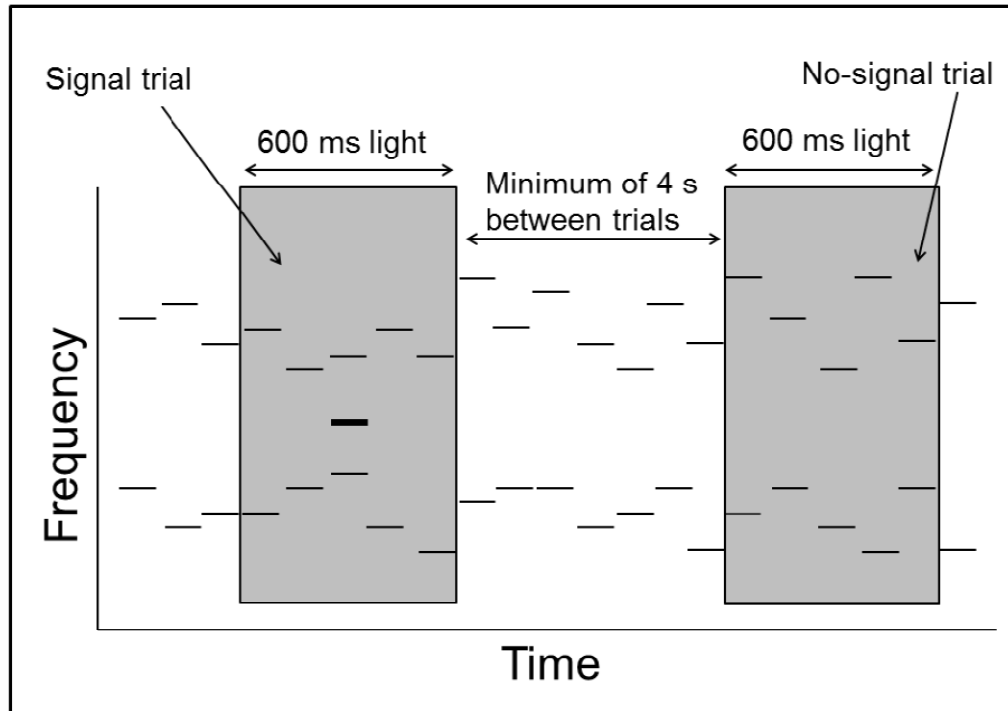


Figure 3-1. Stimuli schematic

Schematic of the temporally-defined condition, in which the signal was embedded in the middle of a 600-ms light cue. The 1000-Hz pure-tone signal (bold, black dash) was presented in a continuous two-tone, random-frequency masker (black dash). The shaded gray box represents the light cue, with the left box being a signal trial and the right box being a no-signal trial. Each run was 40 trials, with a minimum delay of 4 s between trials. The signal/no-signal probability was 0.5. In the temporally-uncertain condition (not shown), all other stimulus parameters were the same except the light cue was not provided.

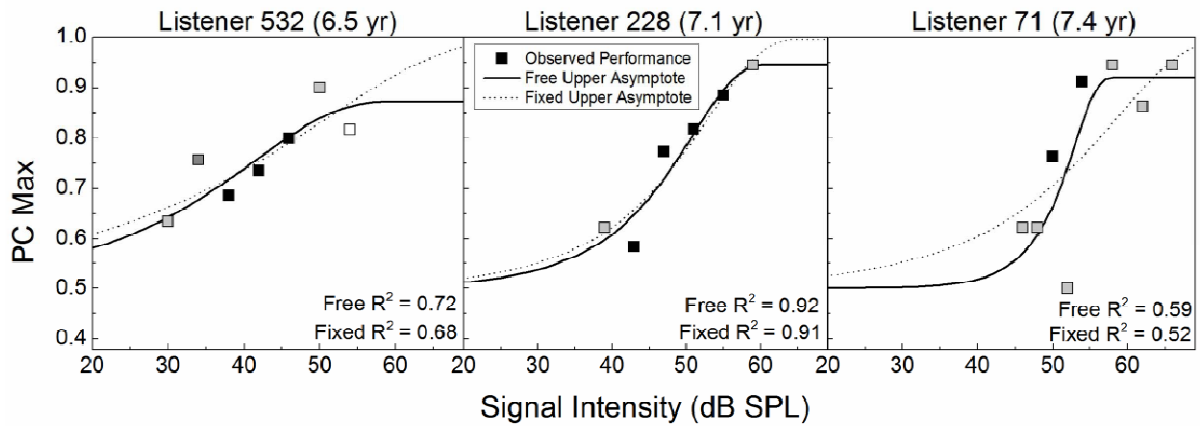


Figure 3-2. Examples of psychometric function fits

Examples of three psychometric function fits are provided with the upper asymptote as either a fixed or free parameter in the fitting model. All fits shown here are in the temporally-uncertain condition, and each panel shows data from a different listener from the 5- to 7-year-old age group. Observed PC_{max} scores are represented by the squares, with the number of trials indicated by the grayscale intensity. For example, signal levels with only eight trials are represented by white-filled squares and levels with 32 trials are black-filled squares. The solid black line represents the fit with lapse rate as a free parameter in the model. The dashed line is the fitted function with the upper asymptote equal to 1.0. The R^2 values for each fit are provided.

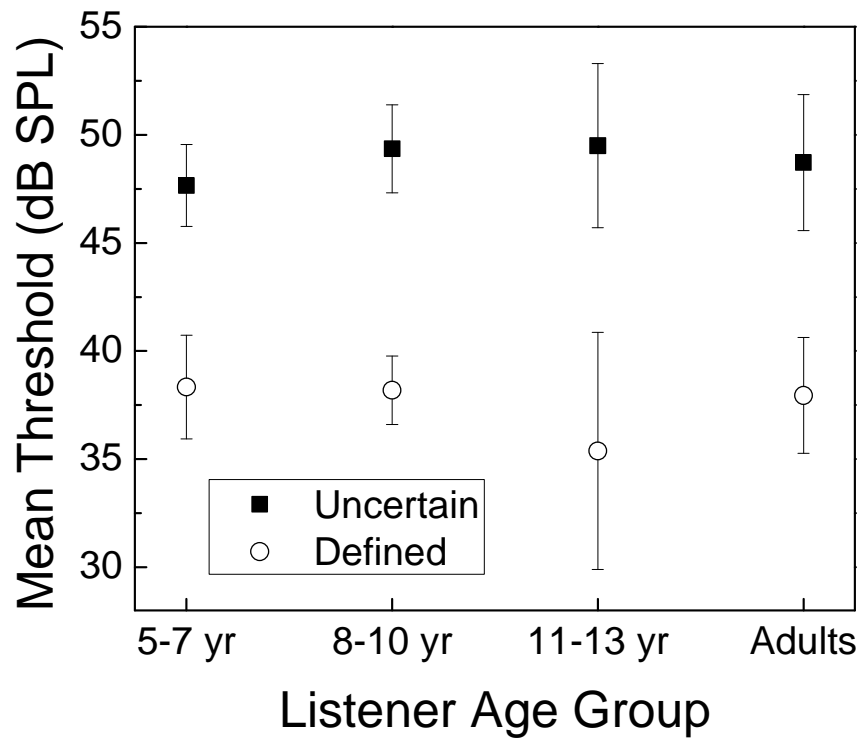


Figure 3-3. Mean threshold estimates for the temporally-uncertain and temporally-defined conditions

Mean threshold estimates (± 1 SE) for the temporally-uncertain (filled square) and the temporally-defined (open circle) listening conditions are provided as a function of listener age group.

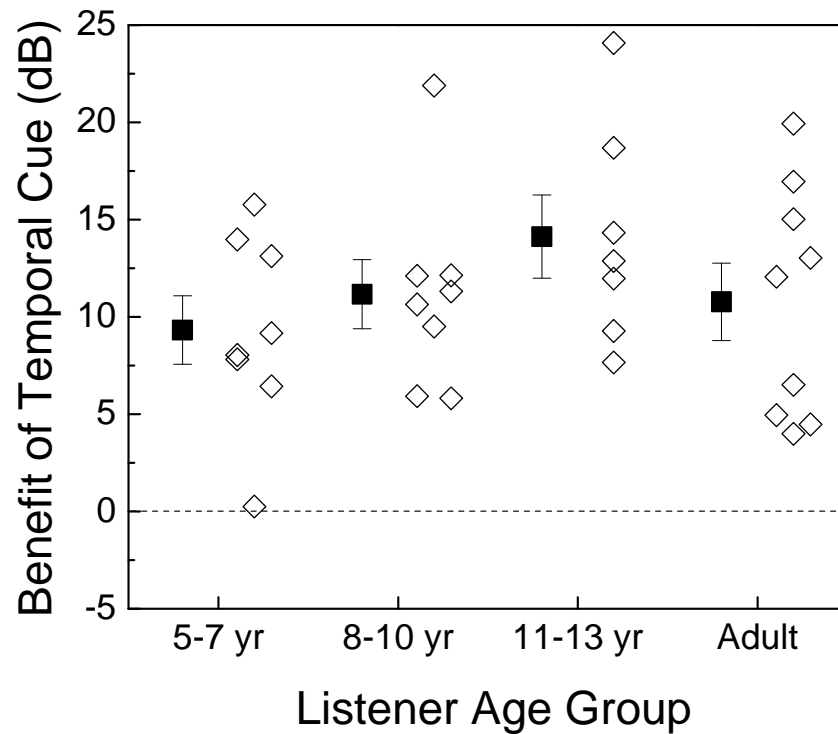


Figure 3-4. Individual and mean benefit scores

Benefit of the temporal cue, or effect of signal-temporal uncertainty, was defined as the difference between threshold in the temporally-uncertain and temporally-defined conditions. Mean benefit scores (± 1 SE) are indicated by the filled square for each listener age group. Data for individual listeners are represented by the diamond symbols. Data points above the horizontal dashed line, or positive scores, indicate that the listeners benefited from a cue indicating when in time to listen for the signal.

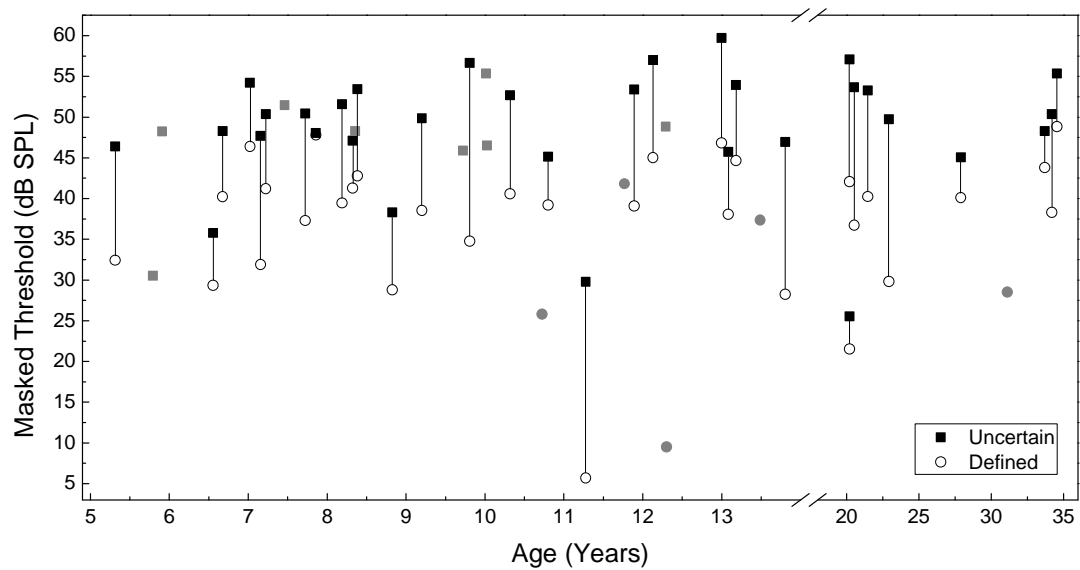


Figure 3-5. Threshold estimates in the temporally-uncertain and temporally-defined condition for individual listeners

Individual performance is shown as a function of listener age in the temporally-uncertain (filled square) and temporally-defined (open circle) listening conditions. The vertical line between each individual's data points indicates the amount of benefit for the cue. The absence of a vertical line indicates no benefit with a defined listening interval or missing data due to failure to meet the R^2 criterion. Data from listeners who were excluded from data analysis, because of a poorly fitted function ($R^2 < 0.5$), are represented by the gray symbols. For these listeners, estimates from their well-fitted functions are shown.

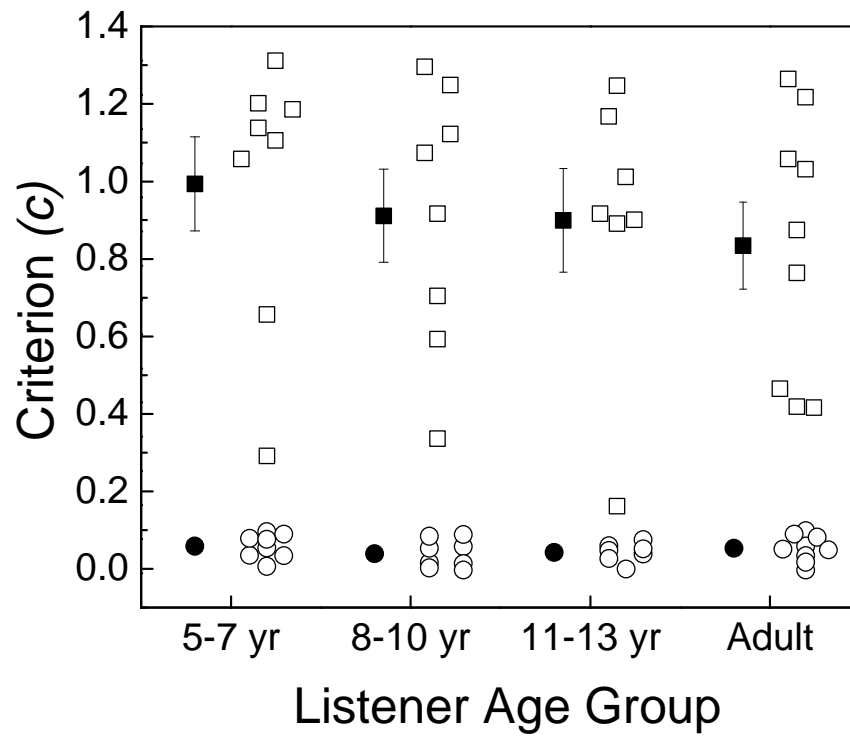


Figure 3-6. Individual and mean estimates of criterion

Individual criterion estimates are provided for the four listener age groups for the temporally-defined (open circles) and temporally-uncertain (open squares) conditions. Mean criterion estimates are indicated by the filled symbol to the left of the individual data; error bars are ± 1 SE.

CHAPTER 4

LISTENING AT FIXED OR RANDOM DELAYS: WORD RECOGNITION IN A TWO-TALKER OR A SPEECH-SHAPED NOISE MASKER FOR CHILDREN AND ADULTS

4.1. Introduction

This study measured monosyllabic word recognition in a two-talker or a speech-shaped noise masker for children and adults. Target words were presented at either a fixed or a random delay in relationship to the interval light, the illumination of a response box. The purpose of this study was to test the hypothesis that a cue indicating when in time to listen for a target word improves children's and adults' masked word recognition. This hypothesis was based on recent findings that a cue indicating when in time to listen can result in substantial benefit for both tone detection and birdsong identification tasks when the signal is embedded in a complex non-speech masker (Best et al., 2007; Bonino & Leibold, 2008; Bonino et al., Chapter 3; Varghese et al., 2012; for exception see Richards et al., 2011). These recent studies have reported larger benefits than previously observed for tone detection tasks in broadband noise (e.g., Egan et al., 1961; Green & Weber, 1980). Differences across these maskers suggest that a cue indicating when in time to listen is particularly beneficial for listening conditions in which it is difficult to perceptually parse the signal from the masker and/or attend to the signal while ignoring the masker. Based on this interpretation, the prediction for the current study was that compared to a random delay, listeners would benefit

from a fixed delay between the onset of the interval light and the onset of the target word. We predicted that this benefit would be larger for the two-talker masker than for speech-shaped noise. Furthermore, it was predicted that school-aged children would be able to use this cue as well as adults.

Recent results demonstrated that 5- to 13-year-old children had a mean tone detection threshold improvement of 11 dB when provided a visual cue indicating when in time to listen for the signal (Bonino et al., Chapter 3). Bonino et al. measured tone detection thresholds for a 120-ms, 1000-Hz pure-tone signal presented in a continuous random-frequency, two-tone masker. The masker was composed of a sequence of 120-ms bursts, with no silent interval between subsequent bursts. Each burst contained a pair of randomly selected tones, with one drawn from 300 to 920 Hz and the other from 1080 to 3000 Hz. The protected region around the signal was used to reduce energetic masking (Moore & Glasberg, 1983), produced as the result of overlapping excitation patterns in the peripheral auditory system (e.g., Fletcher, 1940). Despite these experimental controls to reduce energetic masking, a random-frequency, two-tone masker can substantially elevate thresholds (e.g., Neff & Dethlefs, 1995; Neff & Green, 1987). This type of masking is referred to as informational masking, and it appears to be the result of limited and/or ineffective central auditory processes, including separating the signal from the masker and selectively attending to the signal while disregarding the masker (e.g., Durlach et al., 2003).

The 11-dB effect reported by Bonino et al. (Chapter 3) in the presence of the random-frequency, two-tone masker was larger than for previous studies that examined the benefit of a defined listening interval for tone detection tasks in broadband noise (e.g., Egan et al., 1961; Green & Weber, 1980; Watson & Nichols, 1976). Similar to these older studies, a 2-

dB effect in broadband noise was observed in an earlier study conducted in our lab using a similar approach (Bonino & Leibold, 2008). In contrast, the same group of adults showed a 7-dB benefit of a defined listening interval in the random-frequency, two-tone masker.

Further support for a larger benefit of a temporal cue for conditions that result in informational masking comes from Best et al. (2007) and Varghese et al. (2012). In those studies, listeners were more accurate at identifying a target birdsong embedded in novel competing birdsongs when they were provided a cue indicating the time interval that contained the target. Thus, the benefit of knowing when in time to listen for the signal appears to be large for conditions expected to produce substantial informational masking.

Given the previous results using non-speech stimuli, the prediction for the current study was that presenting target words at a fixed temporal position within a 6.1-s listening window would likewise improve recognition of the word. One qualification of this hypothesis, however, was that a substantial benefit would only be observed for conditions in which the listener had difficulty attending to and/or segregating the signal from the competing background. The approach used in the current study was to compare performance across speech-shaped noise and a two-talker speech masker. Performance is generally poorer for a two-talker masker for both children and adults than for a speech-shaped noise masker with the same long-term average spectrum (e.g., Bonino et al., in press; Brungart et al., 2001; Hall et al., 2002). For example, Brungart (2001) used a closed-set task to measure adults' recognition of two target words embedded in either a same-sex, two-talker masker or a modulated speech-shaped noise masker. At a 0-dB signal-to-noise ratio (SNR), average performance was almost perfect for the speech-shaped noise condition, yet was less than 60% for the two-talker condition. On open-set, word recognition tasks, listeners appear to also

perform more poorly in two-talker masker than for the speech-shaped noise condition (e.g., Bonino et al., in press; Freyman et al., 2004). For example, Bonino et al. (in press) reported that adults' mean word recognition of monosyllabic words in isolation at a 0-dB SNR was 18% worse in the presence of a two-talker masker than in speech-shaped noise. Large variability in susceptibility to perceptual masking has been reported both within and between age groups (e.g., Bonino et al., in press; Brungart et al., 2001; Hall et al., 2002). Furthermore, child-adult differences are more pronounced for a two-talker masker than for speech-shaped noise (e.g., Bonino et al., in press; Hall et al., 2002). Children's and adults' increased difficulty with a two-talker masker appears to be the result of central factors. Carhart et al. (1969) were the first to use the term perceptual masking and proposed that it was the result of limited or ineffective central auditory processing in the ability to perceptually parse the speech target from the competing speech masker. It has recently been suggested that perceptual masking observed for speech stimuli and informational masking observed for tonal stimuli are exactly the same phenomenon or are closely related (e.g., Brungart, 2001; Freyman et al., 2004; Wightman & Kistler, 2005).

Listeners' error patterns also suggest substantial perceptual masking when the masker contains a limited number of talkers. Several studies (e.g., Brungart, 2001; Brungart et al., 2001; Wightman & Kistler, 2005) have reported error patterns collected for the coordinated response measure (CRM; Moore, 1981). For this closed-set task, sentences are presented in the format of "Ready (*call sign*) go to (*color*) (*number*) now." For many experiments using the CRM approach, listeners' ability to correctly identify two key words (color and number) was measured in the presence of a speech masker or speech-shaped noise. Of particular interest for the current paper is the pattern of errors observed for listeners in the monaural

condition, in which a target and masker CRM sentence are simultaneously presented to the same ear by talkers of the same sex. Because the possible responses are limited and are dissimilar words, if a listener hears only a portion of the target word he/she should be able to guess it correctly. Thus, it was expected that when energetic masking resulted in the word being inaudible, listeners would randomly guess from the remaining three colors and seven numbers. However, Brungart (2001) reported that nearly all incorrect responses were coordinates from the masker for adults tested at a 0 dB SNR. Similar error patterns have also been reported for children in the same condition (Wightman & Kistler, 2005). Thus, error patterns for both children and adults suggest that errors are a result of intrusions from the competing speech masker.

Because a two-talker masker appears to produce substantial perceptual masking, it was predicted that a cue indicating when in time to listen for a target word would be particularly beneficial for both children and adults. In order to test this hypothesis, masked word recognition was measured in two fixed delay conditions (short and long) and one random delay condition relative to the activation of the interval light. Word recognition and phoneme identification scores in the fixed delay conditions were compared to performance in the random delay condition. It was expected that performance would be similar for the two fixed delay conditions. However, in the two-talker masker condition, it was predicted that scores from the fixed delay conditions would be higher than those from the random delay condition. In contrast, little or no benefit of a fixed delay was expected in the speech-shaped noise conditions. In addition to analyzing word and phoneme level performance, listeners' error patterns were examined. Consistent with the error patterns reported for the CRM (e.g., Brungart, 2001; Brungart et al., 2001; Wightman & Kistler, 2005), it was expected that

listeners would also have incorrect responses in the two-talker masker conditions which were due to masker intrusions in the current study. However, it was also expected that providing a cue indicating when in time to listen in the fixed delay conditions would reduce listeners' susceptibility to masker intrusions.

4.2. Methods

4.2.1. *Listeners*

A total of 12 children and seven adults participated in this experiment. Children were between five and 13 years of age. The mean age was 9.31 years (range=5.7 to 13.2 years, SD= 2.62 years). The mean age of adults was 25.53 years (range=20.1 to 31.8 years, SD=5.69).¹ All listeners had thresholds in quiet of ≤ 20 dB HL bilaterally for octave frequencies from 0.25 to 8 kHz (ANSI, 2004). All listeners were native speakers of English and had no known history of chronic ear disease. An additional three adults were excluded because of experimenter error, in which the incorrect program was conducted for one run. Adults and a 12.9-year-old were tested individually in a sound-attenuated booth in a single, 2-h visit, including regular breaks. The remaining children were tested in two or three, 1-h sessions, including regular breaks. An experimenter sat inside of the booth with all of the children during testing and entered their responses.

4.2.2. *Stimuli and Conditions*

Speech target words. Word recognition performance was measured for monosyllabic words. A corpus of words familiar to young children was developed based on children's reading lists for kindergarten and first grade (Enchanted Learning, 1998a; 1998b). Five adults

reviewed the initial corpus to identify homophones and confirm that young children would be familiar with the words. All of the adults were hearing or speech pathology graduate students and were native speakers of American English. After omitting homophones and potentially ambiguous or unfamiliar words, there were 842 items in the corpus. These speech targets were recorded by a male native speaker of American English in a double-walled booth (IAC). A condenser microphone (AKG-C1000S) was placed approximately six inches from the speaker's mouth using a microphone stand. The single-channel recordings were amplified (TDT MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (CARDDELUXE). Each target word was recorded twice, with the carrier phrase "say the word" prior to the target word. Audacity sound editing software (v 1.2.6) was used to manually isolate target words, which were then saved as individual wav files. Wav files were scaled to equivalent root mean square (RMS) level and resampled at a rate of 24.414 kHz using MATLAB. The first recording of the target word was used unless undesirable sound quality characteristics (e.g., distortion, peak clipping, or irregular speaking rate) were noted.

To verify the sound quality of the recordings, an open-set word recognition task in quiet was performed. The listeners were the same five adults described above. The 842 target words were presented at 60 dB SPL to the right ear through an ER1 Etymotic insert earphone. No listener missed more than a total of 11 words (>98% correct). A total of 18 target words were re-edited either because they were missed by at least one listener or were noted for having undesirable sound qualities. After re-editing was completed, the listening check was repeated by two adults for the full corpus. Following the same presentation procedure as above, one new adult and one adult from above listened to all of the words. One listener missed two target words and the other listener missed four target words. In addition

to removing one word due to distortion, 145 words were removed from the target word corpus because they were present in the two-talker masker created for this experiment. Any word that appeared in both the target word corpus and in the PBK word lists was removed from the target word corpus for this experiment. The end result was 696 words in the corpus that could be drawn during the experiment as target words.²

Maskers. Listeners were tested in each of two separate continuous backgrounds: a two-talker or a speech-shaped noise masker. The two-talker masker consisted of two streams of concatenated monosyllabic words. Each stream was composed of a different adult male's productions of the 150 Phonetically-Balanced Kindergarten words (PBK; Haskins, 1949). These two talkers were only used to present masker words. The masker words were recorded following the same procedure described above for the target words. Audacity sound editing software (v1.2.6) was used to manually isolate target words, which were then saved as individual wav files. All files were scaled to equal RMS levels. The 150 individual wav files for each talker were concatenated head-to-tail into blocks of 10 to 15 words, without replacement. These blocks were then concatenated head-to tail, with each masker stream following a different order. The stream from one talker was 82 s in duration, and the stream from the other talker was 86 s. The two masker streams were summed and then looped to create a 28-min sample of the two-talker masker. The spectrum of the speech-shaped noise masker was matched to the spectrum of the two-talker masker. This was achieved by calculating the power spectrum from an 86-s sample of the two-talker masker and then applying the same spectral shape to the noise to create an 86-s sample. This sample was then looped to create a 28-min sample. Each masker was recorded on a CD. Throughout testing, the masker was presented continuously at a fixed level of 60 dB SPL.

Stimuli generation. Custom MATLAB software controlled the presentation of target words. The software randomly selected target words from the corpus, without replacement during a run. The masker was played by a CD player, passed into the real-time processor (TDT RP2). The target and masker streams were digitally attenuated separately, mixed, played out of the real-time processor, routed to the headphone buffer (TDT HB7), and presented to the listener's right ear through an insert earphone (Etymotic, ER1). A custom MATLAB script controlled the timing of target word presentation and the illumination of lights on a handheld device, indicating the listening interval.

Delay conditions. Listeners completed three delay conditions for each masker. Delay conditions were defined relative to the illumination of the interval light, three LED lights on a hand-held device. The interval light was activated 0.5 s after the listener initiated the start of a trial by pressing a button on the same device. The interval light remained illuminated for a total duration of 6.1 s, an interval referred to as the listening window. A schematic of the three conditions is depicted in Figure 4-1. In the figure, the illumination of the interval light is represented by the gray boxes. The black boxes indicate the onset time of the target word for each condition. In the two fixed delay conditions the target word was always presented at the same delay after the activation of the interval light, either 0.5 s for the short fixed delay condition or 5.5 s for the long fixed delay condition. In contrast, in the random delay condition, the target word was presented at a randomly selected delay, drawn from a uniform distribution of 0.5 to 5.5 s, after the activation of the interval light. The condition (short fixed delay, long fixed delay, or random delay) was held constant across a run of trials.

4.2.3. *Testing Procedure*

Listeners were tested in a double-walled, sound-attenuating booth (IAC). Listeners initiated all trials by pressing a button. Listeners were instructed to repeat the target word after the interval light turned off. The experimenter, located inside of the booth, entered children's responses via a keyboard. After entering the response, the target word was displayed on the computer monitor, and the experimenter was prompted to score the response as correct or incorrect. Adult listeners entered and scored their responses independently, and the experimenter initiated the start of runs from the adjacent control room.

Listeners completed a familiarization phase followed by two stages of testing. During the familiarization phase, listeners completed 50 target words in quiet. Target words were played at 60 dB SPL. In order to reduce testing time for this phase, the interval light was reduced to 2.2 s and the words were presented randomly between 0.5 and 2 s. All listeners were able to achieve the required criterion of $\geq 80\%$ accuracy in the familiarization phase. Data from this phase were analyzed in the same way as the stage 2 data (discussed below) to evaluate word recognition and phoneme identification in quiet for listeners with this corpus of target words.

After the familiarization phase, each listener completed two stages of testing. Because individual variability tends to be large in the presence of a two-talker masker (e.g., Bonino et al., in press; Hall et al., 2002), adaptive thresholds were measured in stage 1 of testing in order to obtain individual estimates of the SNR used in stage 2 of testing. In stage 1, two adaptive threshold tracks were completed in the random-delay condition for each masker type. A single-interval, adaptive threshold procedure was used. The level of the

masker was fixed at 60 dB SPL, and the level of the target words was adapted using a one-up, one-down rule to estimate a threshold corresponding to 50% on the psychometric function (Levitt, 1971). An initial step size of 4 dB was used, decreasing to 2 dB after the second reversal. Testing concluded after eight reversals, with threshold being computed as the mean signal level at the last six reversals. If threshold estimates for a given masker condition varied more than 5 dB across the two blocks an additional threshold run was completed. Four listeners completed three runs for a condition: two listeners for the two-talker masker and two different listeners for speech-shaped noise. For these listeners, the two runs with the greatest agreement were used to calculate their threshold estimates.

In stage 2, a total of two runs of 25 target words were completed in each delay condition at a fixed presentation level. Specifically, the presentation level of the target words was the average of the thresholds estimated in stage 1. In order to verify that this level resulted in performance that was $<70\%$ correct, the first run in this stage was always the random-delay condition. If performance was $\geq 70\%$ correct, the level of the signal was decreased until performance was $<70\%$ for a run of 25 words. The level adjustment was based on individual performance, but in most situations the signal was decreased in 4 dB steps for the two-talker masker and 2 dB steps for the speech-shaped noise masker. The presentation level of the target words was decreased for three children in the two-talker masker and six children in the speech-shaped noise condition. Additionally, the signal level was decreased for two adults in the speech-shaped noise condition.

After the presentation level was verified for a run in the random delay condition, listeners completed the five remaining runs over two blocks. Block 1 consisted of a run for each of the fixed delay conditions. Block 2 consisted of a run for each of the delay

conditions. Testing order was randomized within each of these blocks. Results were summed across the two runs for each condition, thus, performance was calculated based on 50 trials. Listeners were assigned, in a counterbalanced order, to complete either the speech-shaped noise or two-talker masker condition first. All stages of testing were completed for a given masker before the masker was changed.

The spoken productions of the listeners were recorded throughout testing, allowing responses to be scored offline by two blinded coders. A condenser microphone (AKG-C1000S) was fixed in position approximately six inches from the speaker's mouth. The microphone was connected to a laptop computer. This computer was located inside of the booth during the testing of children, but was external to the booth for adults. Using Audacity (version 1.3 Beta), the recordings were digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz. Recordings were then spliced by hand in Audacity (version 1.3 Beta) and responses were saved as individual WAV files.

4.2.4. *Offline Scoring Procedure*

The spoken responses for all children and adults were phonetically coded offline. Two coders were recruited, both having previous experience with similar transcription using the International Phonetic Alphabet (IPA). Prior to coding responses, the two coders independently transcribed the target words from the corpus using a computer-readable phonemic transcription system (i.e., Klattese). Transcriptions followed the conventions described by Storkel and Hoover (2010). Transcriptions were then compared to those available from a phonotactic probability and neighborhood density calculator developed from child corpora of spoken American English (Storkel & Hoover, 2010). If a target word was

not located in the calculator, even after the removal of affixes, then dictionary transcriptions from Longman (1983) were adopted.

A custom program was developed in MATLAB to assist in the coding of responses. For each coder, a coding order was randomly generated which contained the 247 runs collected from all of the listeners. Coders completed all trials of a given run in the same order that trials occurred during testing. For each response, coders were provided with an oscilloscopic view of the waveform and were able to play the response as many times as desired. Coders were then prompted to enter an orthographic representation of the response. Based on the entry, the program searched through the corpus of both target and masker words to determine if a phonemic transcription was available. If so, the transcription was provided on the screen and the response was scored automatically as correct or incorrect. If a transcription was not found, coders were able to enter the phonemic representation of the response and were prompted to score the response manually. If the listener did not provide a response for a trial, coders could indicate 'no response'. The coders' responses were then saved to a record file for later analysis. By using this coding process, coders were blind to the delay condition, masker, target word played, and the listener's performance in other conditions.

Calculation of reliability. A primary coder scored all runs. In order to calculate reliability, a secondary coder scored 14.6% of the runs (n=36). Runs were randomly selected: two of these runs were in quiet (50 words per run) and 34 runs were masked (25 words per run). Thus, the secondary coder scored a total of 950 trials. A percent agreement score was calculated for two comparisons: correct/incorrect score of response and phonetic transcription. The correct/incorrect reliability score required a perfect match to qualify as an

agreement. To calculate percent agreement for this comparison, the total number of agreements across trials was divided by the total number of trials that were examined by both coders. To determine the coder reliability for the phonetic transcriptions, the number of phonemes in the primary coder's entry that were also present in the secondary coder's entry was computed; the order of the phonemes was not incorporated in this statistic. Thus, comparisons were made for a total of 3,046 phonemes across the 950 trials. In order to calculate percent agreement for the phoneme analysis, the number of matched phonemes was divided by the total number of phonemes entered by the primary coder.

4.3. Results

4.3.1. *Adaptive Thresholds*

Thresholds were measured adaptively in stage 1 of testing, corresponding to the SNR at which listeners achieved 50% correct performance. Figure 4-2 presents the mean thresholds (± 1 SE) by age group for both masker conditions in the random delay condition. Thresholds for individual listeners are provided to the right of the mean data, represented by open symbols. Thresholds in the speech-shaped noise are represented by circles and thresholds in the two-talker masker are represented by triangles. The average threshold in the speech-shaped noise was 62.3 dB SPL for children (range: 56.4 to 67.5 dB SPL) and 57.8 dB SPL for adults (range: 54.3 to 60.5 dB SPL). Elevated thresholds were seen for both age groups in the two-talker compared to the speech-shaped noise masker. Average thresholds in the two-talker masker were 68.5 dB SPL for children (range: 63.0 to 75.5 dB SPL) and 62.2 dB SPL for adults (range: 60.6 to 64.0 dB SPL). To quantify this difference, the amount of perceptual masking was calculated. Perceptual masking was operationally defined as the

difference in threshold for the two-talker masker and speech-shaped noise. The mean perceptual masking effect was 6.2 dB for children and 4.5 dB for adults.

A repeated measures analysis of variance (ANOVA) was used to test the statistical significance of the trends observed in Figure 4-2. The analysis included the within-subjects factor of Masker (two-talker, speech-shaped noise) and the between-subjects factor of Age Group (child, adult). Both of the main effects were significant: Masker [$F(1,17)=30.88$, $p<0.001$] and Age Group [$F(1,17)=23.29$, $p<0.001$]. However, the Masker \times Age Group interaction was not significant [$F(1,17)=0.78$, $p=0.39$]. These results confirm the observation in Figure 4-2, that children were more susceptible to masking than adults in both conditions. Additionally, results confirmed that thresholds were higher in the two-talker masker than the speech-shaped noise for both children and adults. Thus, although children's thresholds were higher than adults for both maskers, the amount of perceptual masking was similar for children and adults, as confirmed by a non-significant Masker \times Age Group interaction.

4.3.2. Word Recognition

For the estimates of word level performance, a response was scored as correct only if it was an identical match to the target word. Percent correct scores were based on 50 trials for each condition. These scores were transformed to rationalized arcsine units for all statistical analyses, to counteract non-uniformity of variance (Studebaker, 1985). Scores are provided in percent correct for both the text and figures.

Coder reliability. Across the 36 runs (950 total words) that both coders examined, average coder agreement was 96.5% for scoring the trials as either correct or incorrect.

Word recognition in quiet. Children's mean word recognition in quiet for the familiarization phase was 91.2% (SE=1.3%). Performance ranged from 84% (6.7-year-old) to 98% (12.6-year-old), with only three children having word recognition scores that were <90%. Mean word recognition in quiet was 97.1% (SE=1.4%) for adults. One adult had a score of 90% correct, whereas all other adults scored $\geq 96\%$. In order to examine age-related differences for word recognition in quiet, a one-way ANOVA was conducted, with Age Group (child, adult) as the between-subjects factor. Results confirmed that adults had better word recognition scores in quiet than children for this corpus [$F(1,17)=11.39, p=0.004$]. Furthermore, results from a Pearson product-moment correlation suggested that word recognition scores improved with age for children ($r(12)=0.59, p=0.04$). However, all children were able to meet the familiarization criterion of $\geq 80\%$ correct in quiet.

Word recognition scores across delay and masker conditions. Mean word recognition scores (± 1 SE) from stage 2 of testing are presented in the top panel of Figure 4-3. The three delay conditions (short, long, and random) are grouped by masker type, with two-talker data on the left and speech-shaped noise data on the right. Mean word recognition scores are represented by open squares for children and filled squares for adults.

As expected based on the individual levels chosen for each listener in stage 1 of testing, performance in stage 2 appears to be equated for the random delay condition across masker type and listener age group. For the random delay condition, adults' mean word recognition score was 53.7% in the speech-shaped noise and 57.7% in the two-talker masker. Similar performance was seen for children, with a score of 55% for both masker conditions. In order to test this trend, a pair of one-way ANOVA was conducted, one for the speech-shaped noise masker and the other for the two-talker masker. Results confirmed that word

recognition scores were equivalent across the two listener age groups in the random delay condition for both speech-shaped noise [$F(1,17)=0.09, p=0.77$] and the two-talker masker [$F(1,17)=0.33, p=0.58$].

The main goal this study was to test the hypothesis that compared to the random delay condition the fixed delay conditions would result in greater word recognition improvement for the two-talker masker than for the speech-shaped noise masker. In order to test this hypothesis, benefit scores were computed for the fixed delay conditions. Because performance was equivalent across listener age groups in the random delay condition, word recognition scores in the random delay condition were subtracted from both the short and long delay conditions. Across all conditions, the benefit score was small (close to 0%) for most children and adults. For the two talker masker conditions, the average benefit scores were 1.3% (short delay) and 4.5% (long delay) for children and 3.1% (short delay) and 2% (long delay) for adults. For the speech-shaped noise conditions, the average benefit scores were -1.7% (short delay) and 2% (long delay) for children and 2.3% (short delay) and 6.9% (long delay) for adults.

A repeated measures ANOVA was conducted in order to examine the benefit of a fixed delay across the two masker types. The benefit score (in RAU) was compared between two within-subjects variables, Delay (short, long) and Masker (speech-shaped noise, two-talker), and the between-subjects variable of Age Group (children, adults). None of the main effects nor the interactions were significant: main effect of Delay [$F(1,17)=2.55, p=0.13$], main effect of Masker [$F(1,17)=0.02, p=0.88$], main effect of Age Group [$F(1,17)=1.83, p=0.19$], Delay \times Masker interaction [$F(1,17)=1.27, p=0.28$], Delay \times Age Group interaction [$F(1,17)=0.32, p=0.58$], Masker \times Age Group interaction [$F(1,17)=0.70, p=0.41$], and Delay

× Masker × Age Group interaction [$F(1,17)=0.91, p=0.35$]. Thus, in contrast to the hypothesis for this experiment, the benefit of providing a fixed delay was not significantly greater in the two-talker compared to the speech-shaped noise masker. Indeed, a visual inspection of the data shows little benefit of a fixed delay for either masker type.

4.3.3. Phoneme Identification

Phoneme level performance was evaluated to examine whether the types of errors listeners make differ across masker type or across delay conditions. The number of phonemes that were in both the target word and the response were determined. This analysis did not control for phoneme order within the response or account for the possibility that a response could include additional phonemes, not contained in the target word. As was done above, scores are presented in percent correct for the text and figures, however all statistical analyses were conducted in RAUs (Studebaker, 1985).

Coder reliability. For the 950 trials that both coders examined, the primary coder transcribed a total of 3,046 phonemes. Ninety-six percent of the total phonemes coded by the primary coder were in agreement with the transcriptions from the secondary coder.

Phoneme identification scores in quiet. The mean phoneme identification score in quiet from the familiarization phase was 96.9% (SE=0.53%) for children. Individual children's scores ranged from 93.8% (8.1-year-old) to 99.4% (12.6-year-old) correct. The mean phoneme identification score in quiet was 99.2% (SE=0.35%) for adults. Adult scores ranged from 97.5% to 100% correct. In order to examine age-related differences in performance in quiet, a one-way ANOVA was conducted with Age Group (child, adult) as

the between-subjects factor. Results confirmed significant child-adult differences for phoneme identification scores in quiet [$F(1,17)=11.29, p=0.004$].

Mean phoneme identification scores across delay and masker conditions. Mean phoneme identification scores (± 1 SE) are provided in the bottom panel of Figure 4-3. For all conditions, phoneme identification scores were greater than those reported for word recognition (top panel of Figure 4-3). In the above word recognition analysis, scores in the random-delay condition were similar for the two age groups across the two maskers. However, for this analysis, scores in the random delay condition appear to be higher in the speech-shaped noise conditions than for the two-talker conditions. In order to confirm the statistical reliability of this trend, a repeated measures ANOVA was conducted with the within-subjects factor of Masker (speech-shaped noise, two-talker masker) and the between-subjects factor of Age Group (child, adult). The main effect of Masker was significant [$F(1,17)=12.76, p=0.002$]. Neither the main effect of Age Group [$F(1,17)=0.10, p=0.75$] nor the Masker \times Age Group interaction [$F(1,17)=0.01, p=0.92$] was significant. Thus, phoneme identification scores in the random delay condition were equivalent across the two listener age groups, but not across the two masker conditions.

The other trend seen in Figure 4-3 is that there appears to be a small benefit for phoneme identification with a fixed delay in the two-talker masker. Across all listeners, the benefit of a fixed delay in the two-talker masker was 5.3% in the short delay condition and 4.9% in the long delay condition. For speech-shaped noise, the benefit was -1.0% in the short delay condition and 1.2% for the long delay condition. Using a repeated measures ANOVA, the phoneme benefit score (in RAU) was compared between two within-subjects variables, Delay (short, long) and Masker (speech-shaped noise, two-talker), and the between-subjects

variable of Age Group (children, adults). None of the main effects nor the interactions were significant: main effect of Delay [$F(1,17)=0.56, p=0.47$], main effect of Masker [$F(1,17)=3.50, p=0.08$], main effect of Age Group [$F(1,17)=4.2, p=0.04$], Delay \times Masker interaction [$F(1,17)=1.42, p=0.28$], Delay \times Age Group interaction [$F(1,17)=0.01, p=0.91$], Masker \times Age Group interaction [$F(1,17)=0.65, p=0.43$], and Delay \times Masker \times Age Group interaction [$F(1,17)=0.31, p=0.58$]. Thus, the differences in the phoneme identification scores between the fixed temporal delay conditions and the random delay condition were not significantly different across the two masker conditions.

Individual differences. The individual data are generally consistent with the group data reported above. In Figure 4-4, performance is plotted as a function of fixed-delay benefit for individual listeners by age. Benefit scores were calculated by subtracting performance in the random delay condition from performance in either the short or the long delay condition. Positive scores, above the horizontal hashed line, indicate a benefit for either the short delay condition (filled circle) or the long delay condition (open squares). Data are separated into two panels by masker type; the top panel shows the benefit scores for the speech-shaped noise condition and the bottom panel shows the benefit scores for the two-talker masker condition. Although most listeners had little or no benefit of a fixed delay, there were a few listeners with benefit scores of approximately 10% in both masker conditions. Furthermore, a 12.9-year-old showed substantial benefit of a fixed delay in the two-talker masker condition: 39% in the short delay condition and 34% in the long delay condition.

4.3.4. *Error Patterns*

Error patterns were compared across both delay condition and masker type for children and adults. Trials scored as incorrect were subsequently coded into one of three mutually-exclusive categories. The three categories were: (1) no response, (2) response unrelated to the target word and (3) response related to the target word. No response trials were indicated by the coder during the offline scoring procedure. To determine if a response was related or unrelated to the target word, the phonemes in the target word and response were compared. If $\geq 50\%$ of phonemes in the target word were contained in the response, the trial was coded as “related.” However, if $< 50\%$ of the target word’s phonemes were provided in the response, the trial was coded as “unrelated.” The proportion of each error category was then determined based on the number of incorrect trials (out of 50 trials) for a given condition for each listener.

Mean error patterns are shown in Figure 4-5. Errors are separated by masker type, with the right panel showing patterns for the speech-shaped noise condition and the left panel showing patterns for the two-talker masker condition. Within a panel, error patterns are shown for the three delay conditions, grouped by listener age group. The three left-most columns are mean data for children and the three right-most columns are mean data for adults. Each column shows the proportion of errors for no response (‘NR,’ bottom, gray), unrelated to the target word (middle, white), and related to the target word (top, hashed). Additionally, the thin black boxes represent the proportion of incorrect trials in which the listeners’ responses were an identical match to one of the words used to create the two-talker masker.

For the speech-shaped noise condition, mean error patterns appear to be similar across both age group and condition. The rate of no responses was small in the speech-shaped noise masker, accounting for 3% or less of errors for both children and adults. The majority of errors were responses that were related to the target word. For example, in the random delay condition, 70.4% (SE = 5.9%) of adult errors were related to the target word.

A different pattern of errors was seen for the two-talker masker conditions. First, the proportions in the ‘no response’ category for children tended to be higher in the two-talker masker than in the speech-shaped noise conditions. The highest rate of no response for children in the two-talker masker was in the random delay condition, accounting for 17.8% (SE = 4.0%) of errors. Another trend seen in Figure 4-5 was that incorrect responses were less likely to be partially correct in the two-talker than the speech-shaped noise masker. However, the probability of producing a related response in the two-talker masker was higher in the fixed delay conditions than in the random delay condition. For example, 31.1% (SE = 3.9%) of adult errors were related to the target word in the random delay, two-talker condition. Adults’ proportion of related errors increased in the fixed delay conditions: 44.4% (SE = 3.6%) for the short delay and 42.7% (SE = 2.8%) for the long delay. For children, the mean proportion of related errors in the two-talker maskers was 56.6% (SE = 5.8%) for the short delay condition and 49.6% (SE = 5.5%) for the long delay condition. In comparison, the average proportion of related errors for children was 41.5% (SE = 4.0%) in the random-delay, two-talker masker condition. This observation suggests that listeners were more likely to hear at least part of the target word presented in the two-talker masker when the delay was fixed compared to when the delay was random.

Errors related to target words. A multivariate analysis of variance (MANOVA) for repeated measures was conducted on the proportion of related errors (in RAUs) to determine the effect of condition, Masker (two-talker, speech-shaped noise) and Delay (random, short, long), across one level, Age Group (children, adults). The assumption that variance-covariance matrices are equal across the cells formed by the between-subjects effects was met [Box's M test; $F(21,577.41)=1.15, p=0.29$]. Additionally, variance ratios were similar for the within-subject variables for each age group. Two significant effects were reported for this analysis: the effect of Masker [$F(1,17)=43.44, p<0.001$] and the Masker \times Delay interaction [Wilk's $\Lambda = 0.0.51, F(2,16)=7.56, p=0.005$]. None of the other effects nor interactions were significant: Delay [Wilk's $\Lambda = 0.91, F(2,16)=0.78, p=0.47$], Age Group [$F(1,17)=2.08, p=0.17$], Delay \times Age Group [Wilk's $\Lambda = 0.96, F(2,16)=0.35, p=0.71$], Masker \times Age Group [$F(1,17)=1.91, p=0.18$], or Masker \times Delay \times Age Group [Wilk's $\Lambda = 0.98, F(2,16)=0.13, p=0.88$].

A pair of repeated measures ANOVA was conducted to follow up the significant Masker \times Delay interaction. With data collapsed across the two age groups, the proportion of related errors significantly differed across delay condition for the two-talker masker [$F(2,36)=8.68, p=0.001$] but not for speech-shaped noise [$F(2,36)=1.31, p=0.28$]. Pairwise comparisons, based on the estimated marginal means with Bonferroni adjustments, were then conducted to examine performance across the delay conditions for the two-talker masker. For the two-talker masker, both the short ($p=0.007$) and the long ($p=0.047$) delay conditions were more likely to have errors related to the target word than the random delay condition. The proportion of related errors was similar across the two fixed delay conditions ($p=0.28$).

Masker intrusions. One explanation for the decrease in errors related to the signal in the two-talker conditions is that listeners' incorrect responses were related to masker words, either in part or in entirety, rather than being related to the target word. In order to examine this possibility, the proportion of masker intrusions was calculated by determining the number of incorrect trials in which the response was an exact match to one of the 150 masker words. For this analysis, the short and long delay conditions were collapsed into one variable, weighted by the number of errors, to represent the fixed conditions. This score, as well as the score for the random delay condition, were then transformed to RAUs (Studebaker, 1985).

A MANOVA for repeated measures was conducted on the proportion of masker intrusions to determine the effect of condition, Masker (two-talker, speech-shaped noise) and Delay (random, fixed), across one level, Age Group (children, adults). The assumption that variance-covariance matrices are equal across the cells formed by the between-subjects effects was met [Box's M test; $F(10, 723.3) = 2.28, p = 0.012$].³ Additionally, variance ratios were similar for the within-subject variables for each age group. For the MANOVA, all three main effects were significant: Delay [$F(1, 17) = 7.38, p = 0.015$], Masker [$F(1, 17) = 92.42, p < 0.001$], and Age Group [$F(1, 17) = 38.85, p < 0.001$]. Additionally, both the Masker \times Age Group [$F(1, 17) = 20.23, p < 0.001$] and the Masker \times Delay [$F(1, 17) = 14.05, p = 0.002$] interactions were significant. The Delay \times Age Group interaction was not significant [$F(1, 17) = 1.45, p = 0.25$], nor was the Masker \times Delay \times Age Group interaction [$F(1, 17) = 2.51, p = 0.13$]. This analysis confirmed the trend seen in Figure 4-6 that masker intrusions were more likely in the two-talker masker conditions than in the speech-shaped noise conditions. For speech-shaped noise, the average proportion of masker intrusions appears to be similar across the three delay conditions, as well as across the two age groups (11.1% to 17.2%). In

addition to a greater occurrence of masker errors, masker intrusions were more likely in the random delay than the fixed delay conditions for the two-talker masker conditions. In the random delay, two-talker masker condition, 66.4% (SE = 3.1%) and 27.6% (SE = 5.2%) of incorrect responses matched a masker word for adults and children, respectively. The last trend confirmed was that children were less likely to have masker intrusions than adults in the two-talker masker conditions. However, children's error patterns were similar to adults' in the speech-shaped noise condition.

Individual differences. Substantial individual variability was seen for error patterns across children in the two-talker masker. Figure 4-6 provides error patterns for individual children, rank ordered by age. Error patterns are shown in three panels based on delay condition: top panel is the short delay, middle panel is the long delay, and the bottom panel is the random delay. Adults showed little between-subjects variability. Adults' mean error patterns are provided on the far right of each panel.

4.4. Discussion

4.4.1. Age Effects in Susceptibility to Masking

In order to account for within and between age group variability in susceptibility to masking, thresholds were measured adaptively in stage 1 of testing for both maskers. Consistent with studies that have examined masked speech recognition (e.g., Allen & Wightman, 1992; Bonino et al., in press; Hall et al., 2002; Nishi et al., 2010; Wightman & Kistler, 2005), significant age effects were observed for both masker conditions. In the current study, the average child-adult difference was 4.5 dB for speech-shaped noise and 6.3 dB for the two-talker masker.

The perceptual masking effect, defined as the difference in threshold between the two masker conditions, was similar for children and adults. The lack of a significant child-adult difference in perceptual masking is inconsistent with results from a spondee recognition task in a continuous masker by Hall et al. (2002). In that study, a perceptual masking effect of 6.7 dB was seen for children and an effect of 2.3 dB was observed for adults. Children's perceptual masking effect of 6.2 dB in the current study is similar to that of Hall et al. (2002). However, the adults in the current study showed a perceptual masking effect of 4.5 dB. This effect is larger than the 2.3 dB effect reported by Hall et al. (2002) but less than the 6.6 dB effect reported by Carhart et al. (1969) for spondee words.

4.4.2. *Benefit of a Cue Indicating When in Time to Listen*

The primary goal of this work was to test the hypothesis that a cue indicating when in time to listen would be associated with greater improvement in speech recognition performance for conditions of high perceptual masking than of low perceptual masking. However, word recognition and phoneme identification analyses indicated little or no benefit associated with presenting words at a fixed delay relative to a random delay in the presence of either masker. The lack of substantial word recognition or phoneme identification score improvement with the provision of the cue in the two-talker masker condition is in contrast to findings from non-speech stimuli (Best et al., 2007; Bonino et al. (Chapter 3); Varghese et al., 2012).

Methodology differences may explain the inconsistent findings between the current study and those of Bonino et al. (Chapter 3), the multi-tonal study which motivated this work. The first difference is that the current study was an open-set word recognition task. In

contrast, Bonino et al. (Chapter 3) used a detection task. There were also differences in how temporal uncertainty was created between the two studies. In the current study, the 6.1-s listening window was defined for all conditions, and the position of the word within that window was manipulated. In contrast, Bonino et al. (Chapter 3) only defined the listening window in the temporally-defined condition. A 600-ms light cue defined the listening interval for each trial in that condition. In the temporally-uncertain condition, listeners were given no indication of when a trial was initiated, apart from hearing the signal, and were not provided trial-by-trial feedback. There was a minimum interval of 4 s between trials, but this delay could be substantially longer for the listener if multiple no-signal trials were presented in a row or he/she did not hear the signal during a trial. Thus, the listening window varied across a run of trials in the temporally-uncertain condition and could be rather long given the briefness of the signal (120 ms). A third related difference is the length of the temporal window. It is possible that a random delay of 5 s, as used in the current study, is not long enough to result in substantial temporal uncertainty for monosyllabic words. If temporal conditions were created which were more similar to those used by Bonino et al. (Chapter 3), the effect of knowing when in time to listen might be more consistent for tone detection and word recognition.

In contrast to the word recognition and phoneme identification scores, error patterns in the two-talker masker were different between the random delay condition and the fixed delay conditions. Compared to the random delay condition, both children and adults were significantly more likely to provide incorrect responses that were classified as related to the target word in the fixed delay conditions. Consistent with this finding, the proportion of incorrect responses which were matches to the masker word also decreased in the fixed delay

conditions. These observations are consistent with the interpretation that the provision of a fixed delay assists the listener in hearing at least part of the target word and thus, reducing masker intrusions. One explanation for why a benefit of the fixed delay was seen for the error pattern data and not for the phoneme level analysis is that different coding schemes were used. Recall that individual phonemes were compared between the target and response to calculate the phoneme identification score, whereas errors were classified by using a 50% phoneme matching criterion in the error pattern analysis. It may be that these analyses would be in better agreement if a more sophisticated coding system was used to calculate phoneme identification scores. For example, a coding scheme that took into consideration the position of the phoneme within the word.

4.4.3. Error Patterns

Differences in error patterns across maskers. Although error pattern data should be interpreted with care, these data suggest that error patterns are different across the two maskers tested. In the speech-shaped noise conditions, nearly 70% of errors were classified as related errors, suggesting that listeners were able to hear at least part of the target word on most trials. In contrast to speech-shaped noise, listeners were less likely to have errors that were related to the target word in the two-talker masker condition. One interpretation of this finding is that listeners were more likely to have intrusions from the masker in the two-talker masker condition. Although it is not possible to know which masker words were played on a given trial, an overall increase in the number of masker words provided as responses was seen for the two-talker masker conditions compared to the speech-shaped noise conditions. This finding is consistent with studies using a closed-set task with the CRM corpus which

have shown that incorrect responses tend to be intrusions from the masker for both children and adults (e.g., Brungart, 2001; Brungart et al., 2001; Wightman & Kistler, 2005; Wightman et al., 2010).

Results from the speech-shaped noise conditions provide an estimate of the baseline probability of producing a response which is a match to a masker word. Across delay conditions in speech-shaped noise, 13% of listener errors were matches to the masker word. This value is a reflection of the similarity between the target and masker words. On average, a masker word shares at least half of the phonemes with 52.5 target words (range: 3 to 199 words; SD = 46.7 words) or 7.5% of the target words in the corpus. One factor that could influence the rate of masker intrusions in the speech-shaped noise condition is previous experience with the two-talker masker. However, the examination of the number of masker intrusions in the speech-shaped noise condition based on masker testing order did not suggest this was a factor for this study [$F(1,17)=0.63$, $p=0.44$]. Thus, regardless of previous experience, on some proportion of incorrect trials, listeners appear to guess a masker word because it is available to them in their lexicon, not necessarily because it is a direct result of a masker intrusion.

Age-related differences in error patterns. The only significant child-adult difference in error patterns was that children were less likely than adults to provide masker words as a response in the two-talker masker conditions. However, examination of the individual error patterns for the child data (Figure 4-6) suggests that there may be differences across childhood. For example, the four youngest children were more likely to provide responses related to the target word than most of the older children and all of the adults for the two-talker conditions. Additionally, these younger children appear to be less likely to provide

masker words as responses compared to older children or adults. One possible explanation is that these younger children required a more advantageous SNR for testing which may have provided a level difference cue between the target word and masker. In order to examine this possibility, individual listener's proportion of masker intrusions in the random delay, two-talker condition was plotted as a function of SNR in Figure 4-7. Adults' data are represented by the filled circles and children's data are the open squares. To the right of children's data points are their ages (in years). Data from four older children are inconsistent with the explanation that the rate of masker intrusions is related to SNR. These children were tested at a comparable SNR as adults (+3 dB SNR) and they all had fewer masker intrusions than the adult listeners. Furthermore, a 12.9-year-old who had a high rate of masker intrusions (53%) was tested at +8 dB SNR, a SNR greater than any of the adult listeners. Regardless, the pattern of performance for the limited data set does not discount the possibility that the higher signal levels used for younger children is related to their decreased likelihood of masker confusions. An alternative explanation is that there may be a relationship between no response errors and masker intrusions errors. It may be that when adults do not hear the target word they guess a masker word that they recently heard, whereas children may be more likely to not provide a response. Thus, future research is needed to examine the rate of masker intrusions and how it is affected by age and SNR.

4.5. Conclusions

The primary purpose of this experiment was to test the hypothesis that listeners would show greater benefit from a cue indicating when in time to listen for a target monosyllabic word embedded in a two-talker masker than in speech-shaped noise. However, word and

phoneme level analyses did not provide evidence of a benefit for either masker condition. An exploratory analysis of error patterns suggests differences in the types of errors made by listeners in the two-talker masker between the random and fixed delay conditions. This discrepancy in error patterns was observed for both children and adults. Future research is needed to determine if the results found here extend to longer windows of temporal uncertainty.

4.6. Endnotes

¹ Two of the adults (31.2 and 31.8 years) had previously participated as trained listeners in studies using multi-tonal stimuli. Their performance in the current study was similar to the naïve adults.

² Using the on-line calculator developed by Storkel and Hoover (2010), word frequency, neighborhood density and phonotactic probability estimates were calculated for each target word. These estimates were based on child corpora. Across the 696 target words, mean estimates were the following: word frequency (in log base 10) was 2.97 (SD=0.97), neighborhood density was 11.4 (SD=6.87), average positional segment frequency was 0.05 and average biphone frequency was 0.005.

³ An α level of 0.001 was used to determine significance for the Box's M test, due to oversensitivity of this analysis, as recommended by Tabachnick and Fidell (2007).

4.7. Figures

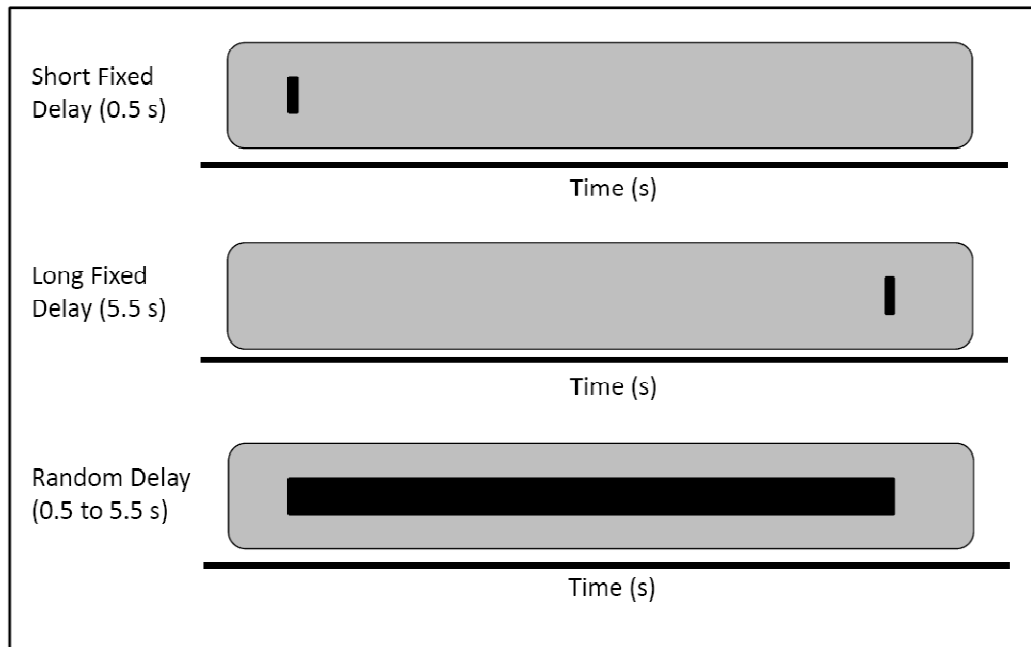


Figure 4-1. Schematic of delay conditions

Schematic of the three delay conditions, in which the target was embedded in a 6.1-s light cue. The length of the shaded gray box represents the duration of the light cue, and the black box represents the relative onset of the target word. The top two panels represent the fixed delay conditions. In these conditions, the onset of the target word was delayed by 0.5 s (short delay, top panel) or 5.5 s (long delay, middle panel) relative to the onset of the light. The bottom panel represents the random delay condition, in which the onset of the target word was delayed by some random period, between 0.5 and 5.5 s, relative to the onset of the light.

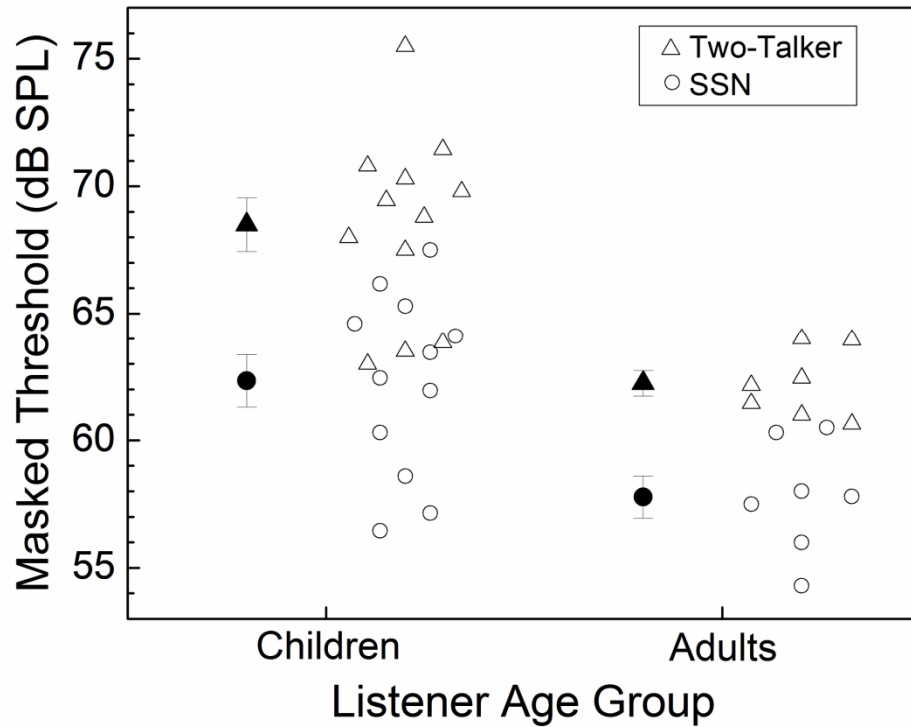


Figure 4-2. Individual and mean masked threshold estimates

Mean (± 1 SE) and individual masked threshold estimates obtained during the adaptive testing in stage 1 are provided for children (left) and adults (right) in the random-delay condition. Threshold estimates for the two-talker masker condition are represented by triangles and for the speech-shaped noise condition by circles. Mean thresholds are indicated by filled symbols, and individual data are indicated by open symbols.

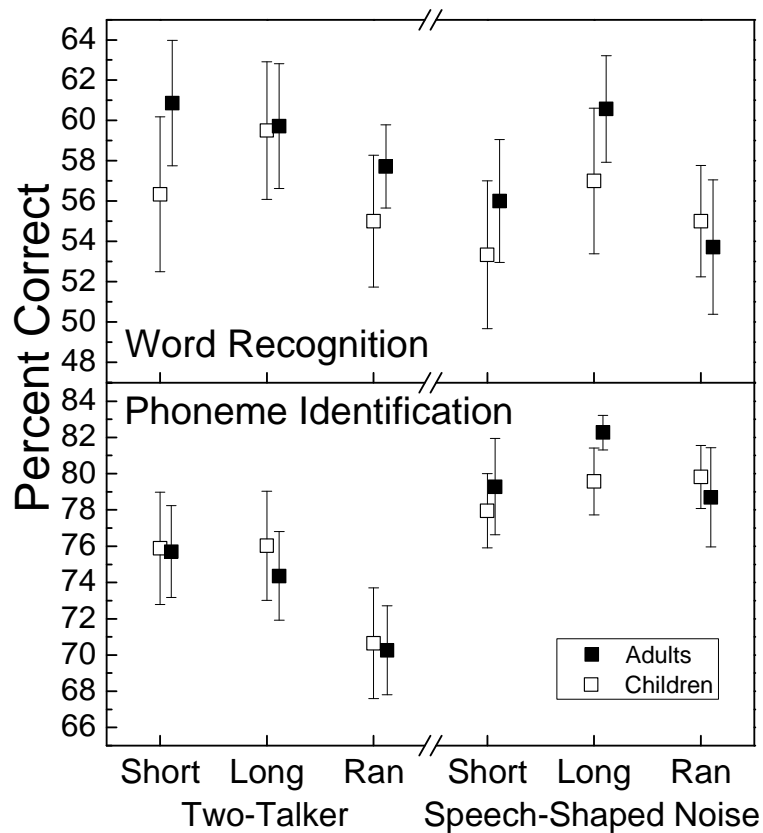


Figure 4-3. Mean word recognition and phoneme identification scores by condition

Mean word recognition scores (± 1 SE) are provided in the top panel and mean phoneme identification scores (± 1 SE) in the bottom panel. Performance is provided by condition for children (open squares) and adults (filled squares). Scores are presented for the three delay conditions by masker type, with two-talker on the left and speech-shaped noise on the right.

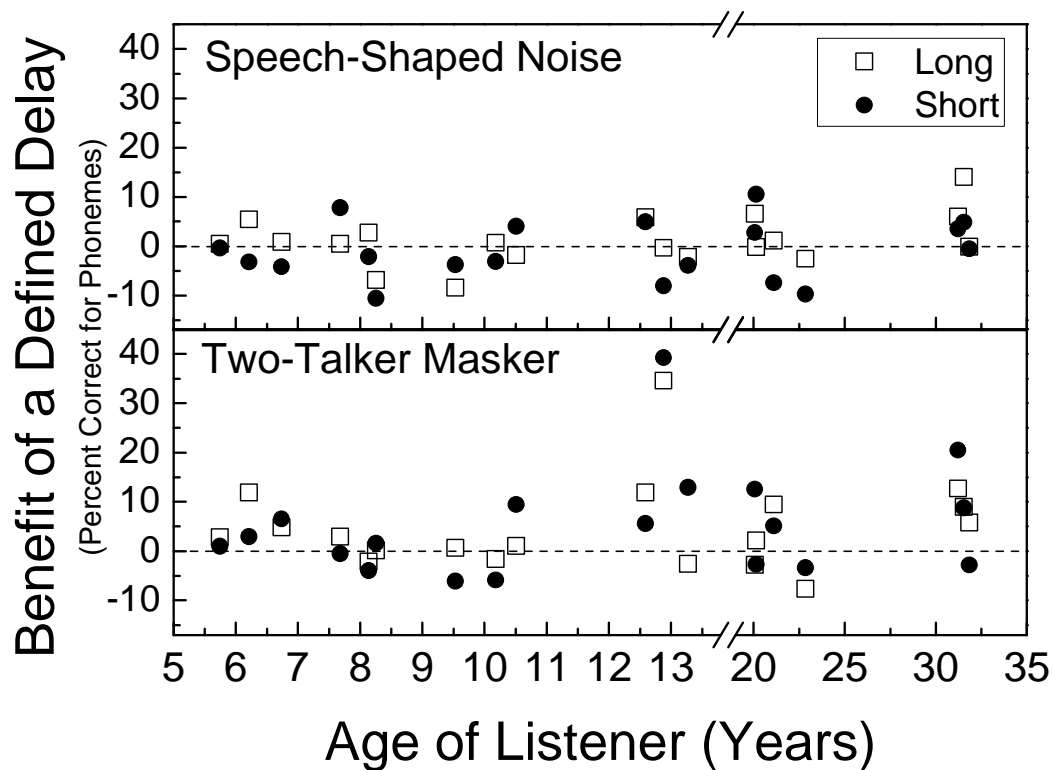


Figure 4-4. Change in individual listener's phoneme identification scores associated with a fixed delay

The difference in phoneme recognition scores between the random delay condition and the fixed delay conditions are provided for individual listeners plotted by age. Benefit scores in the speech-shaped noise are provided in the top panel and benefit scores in the two-talker masker are provided in the bottom panel. Data points above the horizontal dashed line indicate that the listeners had improved scores for the short delay condition (filled circle) or the long delay condition (open square) compared to the random delay condition.

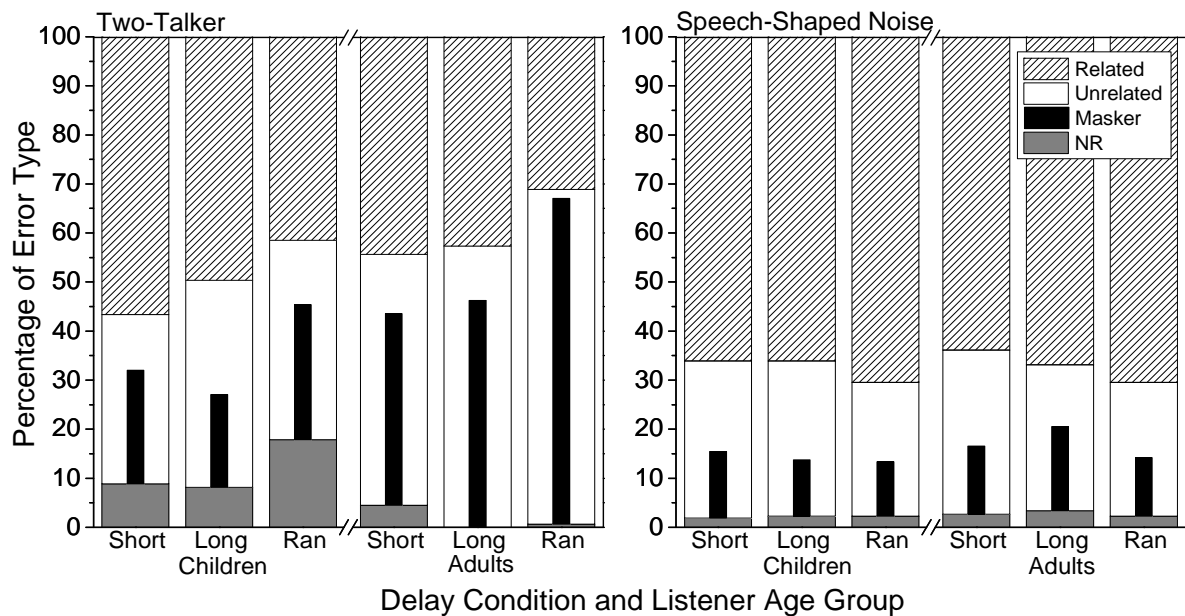


Figure 4-5. Mean error patterns for children and adults by condition

Mean error patterns are provided across conditions for both age groups. Error patterns for the two-talker masker conditions are in the left panel and for the speech-shaped noise conditions in the right panel. Within a panel, delay conditions are grouped by age, with children on the left and adults on the right. For each condition, the proportion of errors was calculated for three categories: no response (“NR;” gray), unrelated to the target word (white), and related to the target word (hashed). Also, the proportion of incorrect trials in which the response was an exact match to a masker word is indicated by the vertical extent of the black box.

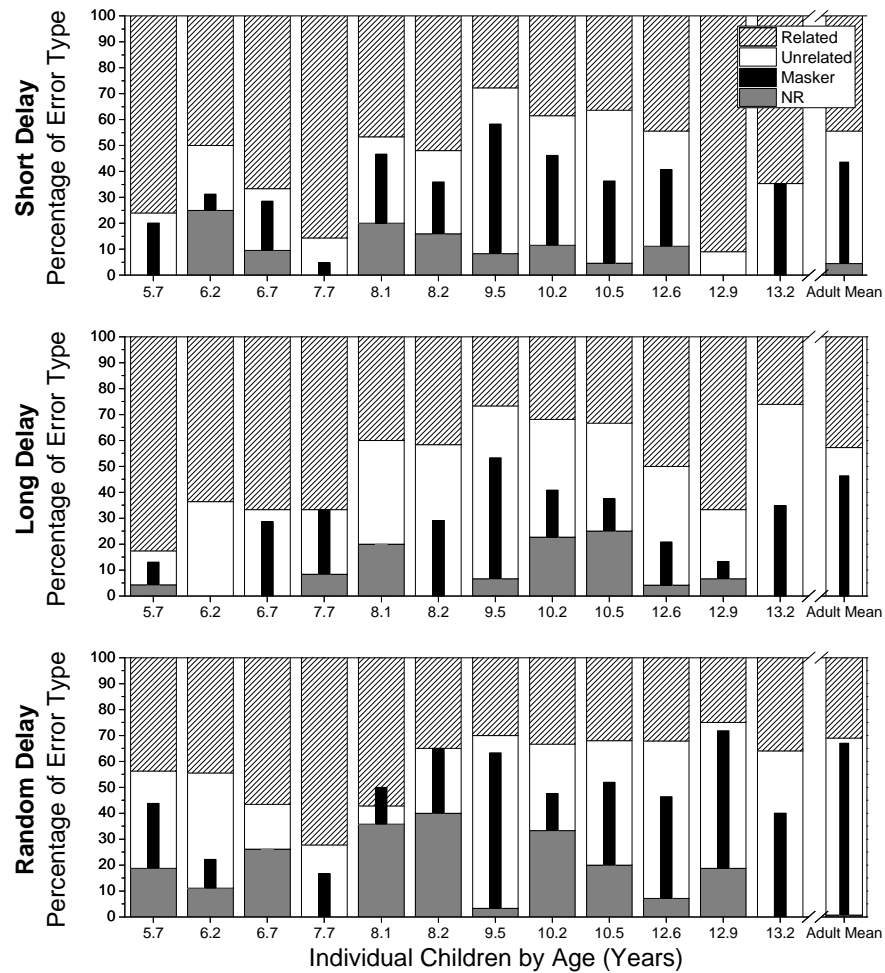


Figure 4-6. Individual error patterns for children in the two-talker masker conditions

Separate panels are provided for each of the three delay conditions. Errors were classified into three categories: no response (“NR;” gray), unrelated to the target word (white), and related to the target word (hashed). The black box represents the proportion of incorrect trials in which the response was a masker word. For comparison, mean adult error patterns are also provided for each condition on the far right-hand side of each panel.

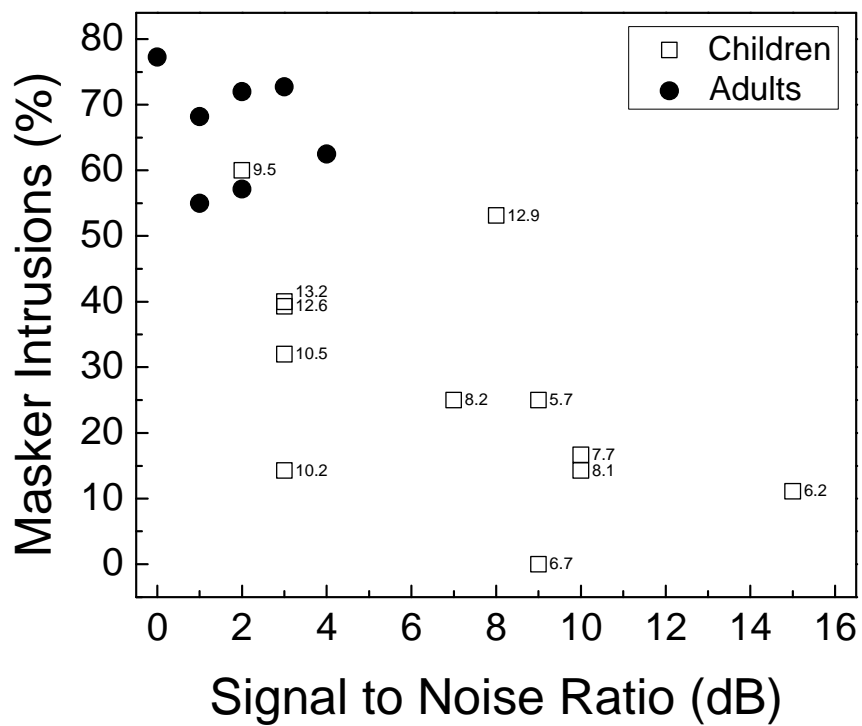


Figure 4-7. Proportion of masker intrusions as a function of signal-to-noise ratio

The proportion of masker intrusions in the random delay, two-talker condition is provided as a function of signal-to-noise ratio (dB SNR). Individual data are provided for children (open squares) and adults (filled circles). For children, each listener's age (in years) is provided to the right of his/her data point.

CHAPTER 5

CONCLUSIONS

5.1. Summary of Findings

The primary purpose of this dissertation was to test the hypothesis that a cue indicating when in time to listen for a signal would be a beneficial cue for children when listening in complex maskers. This hypothesis was based on the observation that larger effects of signal-temporal uncertainty have been reported for maskers expected to produce substantial informational masking (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012) than the 2 to 3 dB effect typically reported for noise maskers (e.g., Egan et al., 1961; Green & Weber, 1980). In order to test this hypothesis, the effect of signal-temporal uncertainty was measured for two different stimuli, tones (Chapter 3) and speech (Chapter 4).

In the first experiment (Chapter 3), based on Bonino and Leibold (2008), 35 children and 10 adults detected a 1000-Hz pure-tone signal presented in a random-frequency, two-tone masker. Thresholds were measured in two conditions, temporally-defined or temporally-uncertain. Across all three age groups of children (5 to 7, 8 to 10, and 11 to 13 years of age) and adults, the average benefit was 11 dB when listeners were provided a defined listening interval. This finding was consistent with the other published studies which have reported substantial benefit of a cue indicating when in time to listen for conditions expected to produce informational masking (Best et al., 2007; Bonino & Leibold, 2008; Varghese et al., 2012).

In the second study (Chapter 4), word recognition was measured for 12 children (5 to 13 years of age) and seven adults in a two-talker masker or speech-shaped noise. In each masker, listeners completed three delay conditions, in which the delay of the target word was either random or controlled relative to the onset of a 6.1-s interval light. Inconsistent with our hypothesis and the results from Chapter 3, neither word recognition or phoneme identification scores improved for the fixed delay conditions compared to the random delay condition for either children or adults. However, a preliminary analysis of error patterns suggests possible differences between the fixed delay conditions and the random delay condition in the two-talker masker. Additionally, differences in error patterns were reported between the speech-shaped noise conditions and the two-talker masker conditions for this study.

As discussed in Chapter 4, methodology differences may explain the inconsistent findings between the two studies reported in this dissertation. One key difference between the two experiments was in how temporal uncertainty was created. In the multi-tonal study (Chapter 3), a light cue indicated the listening window in the temporally-defined condition. In contrast, listeners were only aware of a trial occurring if they heard the signal in the temporally-uncertain condition. Listeners often had a long listening window because a minimum 4-s delay was imposed between trials, plus 50% of trials did not contain a signal. Thresholds in the temporally-uncertain condition were then compared to the temporally-defined condition, in which a light cue defined a 600-ms listening window. In contrast, the 6.1-s listening window was always defined for the speech study (Chapter 4) and the comparison was across conditions in which the target words were presented at a fixed versus a random delay within the defined listening window. It may be that a variable and/or a longer

listening window is required in order to observe a sizeable effect of signal-temporal uncertainty for conditions expected to produce substantial informational/perceptual masking. Another consideration is that the 120-ms tonal signal was brief in duration compared to monosyllabic words. It may be that a larger effect of signal-temporal uncertainty would be observed for shorter speech stimuli, such as a consonant identification task.

5.2. Unresolved Questions for Future Work

Based on the findings in this dissertation, four questions are presented here to be addressed by future research:

Why do children and adults have similar thresholds for a tone detection task in a continuous, random-frequency, two-tone masker? An unexpected finding in Chapter 3 was that children did not show greater susceptibility to the random-frequency, two-tone masker than adults. This result is inconsistent with previously published studies which have used brief stimuli in a two-interval, forced-choice task (e.g., Hall et al., 2005; Leibold & Bonino, 2009; Lutfi et al., 2003b; Oh et al., 2001). The use of a continuous masker in the current study may explain the improved thresholds reported for children in the current study. Future research is needed to test the hypothesis that children are more prone to the distracting effects of onsets. Of particular interest would be to compare tone detection thresholds for a gated versus a continuous multi-tonal masker. It would be expected that both children and adults would have similar thresholds in the continuous masker based on findings in Chapter 3. Furthermore, thresholds should be elevated in the gated condition compared to the continuous condition (Richards et al., 2011). However, a substantial child-adult difference may be observed in the gated masker condition. This pattern of results would be consistent

with previously reported findings (e.g., Lutfi et al., 2003b; Oh et al., 2001). If these hypotheses were confirmed, the next step would be to parametrically manipulate the continuous masker to provide insight into the specific features that are responsible for eliminating the large child-adult difference reported by other studies that used a gated random-frequency, multi-tonal masker.

Would a benefit of knowing when in time to listen for the target word be observed if conditions were more temporally uncertain than those used in Chapter 4? In contrast to the multi-tonal work presented in Chapter 3, no significant improvement in word recognition or phoneme identification scores were reported when words were presented at fixed delays compared to a random delay. One possible explanation for the lack of observed benefit in Chapter 4 was that that temporal window of 6.1 s was not long enough to result in substantial temporal uncertainty for monosyllabic words. Instead of presenting words at either fixed or random delays within a defined temporal window, a future study should evaluate the effect of having a defined versus an uncertain listening window. This manipulation would be more analogous to the approach used for the multi-tonal study (Chapter 3) and should result in greater temporal uncertainty.

Are there developmental differences in the likelihood that listeners have masker intrusions when listening for target words presented in a two-talker masker? Results from the preliminary error pattern analysis (Chapter 4) suggest that children are less likely to have masker intrusions than adult listeners in the two-talker masker conditions. Further examination of data for individual children suggests that younger children may be less likely to make masker intrusions than older children. Additional data need to be collected from children to determine if these developmental trends are reliable. One challenge with

interpreting these effects is that listeners were tested at different signal-to-noise ratios (SNRs). Furthermore, younger children tended to be tested at more advantageous SNRs than older children or adults. In order to better understand the effect of SNR on error patterns, collecting data from adults at a range of SNRs would be helpful.

Do lexical properties of the target words influence word recognition scores in a two-talker masker? Krull et al. (2010) examined five to 12 year old children's word recognition in speech-shaped noise. Word lists varied on two parameters: lexical density (i.e., the number of similar sounding words or "neighbors" to a target word) and word frequency (i.e., how often a word is used in the language). As predicted by the Neighborhood Activation Model (NAM; Luce & Pisoni, 1998), listeners had poorer word recognition for target words that were classified as low-frequency words from dense neighborhoods than if the target words were classified as high-frequency words from sparse neighborhoods. The effect of lexical properties of the signal has not been evaluated for maskers that create substantial perceptual masking. One possible outcome may be that lexical effects are larger for the two-talker masker than for the speech-shaped noise condition. However, it may also be possible that these effects are no longer apparent because more errors are related to masker intrusions than being related to the target word. A preliminary evaluation of these effects will be conducted with the data collected in the fixed delay conditions to determine if future research is warranted.

REFERENCES

- Allen, P., & Wightman, F. L. (1992). Spectral pattern discrimination by children. *Journal of Speech and Hearing Research, 35*, 222-233.
- Allen, P., & Wightman, F. L. (1994). Psychometric functions for children's detection of tones in noise. *Journal of Speech and Hearing Research, 37*, 205-215.
- Allen, P., & Wightman, F. L. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech and Hearing Research, 38*, 503-511.
- ANSI. (2004). Methods for manual pure-tone threshold audiometry (Vol. ANSI S3.21-2004). New York: Author.
- Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*(3), 257-271.
- Arbogast, T. L., & Kidd, G. (2000). Evidence for spatial tuning in informational masking using the probe-signal method. *Journal of the Acoustical Society of America, 108*(4), 1803-1810.
- Bargones, J. Y., & Werner, L. A. (1994). Adults listen selectively; Infants do not. *Psychological Science, 5*(3), 170-174.
- Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology, 8*(2), 294-304.
- Bonino, A. Y., & Leibold, L. J. (2008). The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex. *Journal of the Acoustical Society of America, 124*(5), EL321-EL327.
- Bonino, A. Y., Leibold, L. J., & Buss, E. (in press). Release from perceptual masking for children and adults: Benefit of a carrier phrase. *Ear and Hearing*.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4*(3), 380-387.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Bradford Books, MIT Press.

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition* (5 ed., pp. 10-36). New York: Oxford University Press.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109(3), 1101-1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, 110(5), 2527-2538.

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *Journal of the Acoustical Society of America*, 45(3), 694-703.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 115-127.

Chang, P., & Viemeister, N. F. (1991). Temporal windows for signal presented at uncertain times. *Journal of the Acoustical Society of America*, 90(4), 2248.

Dai, H. (1995). On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods. *Journal of the Acoustical Society of America*, 98(6), 3135-3139.

Dai, H., & Buus, S. (1991). Effect of gating the masker on frequency-selective listening. *Journal of the Acoustical Society of America*, 89(4), 1816-1818.

Dai, H., & Micheyl, C. (2011). Psychometric functions for pure-tone frequency discrimination. *Journal of the Acoustical Society of America*, 130(1), 263-272.

Dai, H., Scharf, B., & Buus, S. (1991). Effective attenuation of signals in noise under focused attention. *Journal of the Acoustical Society of America*, 89(6), 2837-2842.

- Dai, H., & Wright, B. A. (1995). Detecting signals of unexpected or uncertain durations. *Journal of the Acoustical Society of America*, 98(2), 798-806.
- Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development*, 5, 261-276.
- Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B. G., Colburn, H. S., & Kidd, G. (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. *Journal of the Acoustical Society of America*, 118(4), 2482-2497.
- Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *Journal of the Acoustical Society of America*, 113(6), 2984-2987.
- Egan, J. P., Greenberg, G. Z., & Schulman, A. I. (1961). Interval of time uncertainty in auditory detection. *Journal of the Acoustical Society of America*, 33(6), 771-778.
- Elliott, L. L., Conners, S., Kille, E., Levin, S., Ball, K., & Katz, D. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. *Journal of the Acoustical Society of America*, 66(1), 12-21.
- Elliott, L. L., & Katz, D. R. (1980). Children's pure tone detection. *Journal of the Acoustical Society of America*, 67(1), 343-344.
- Enchanted Learning (1998a). First grade vocabulary words. Retrieved August 7, 2008, from <http://www.enchantedlearning.com/Vocabulary/VocabWords1.shtml>
- Enchanted Learning (1998b). Kindergarten vocabulary words. Retrieved August 7, 2008, from <http://www.enchantedlearning.com/Vocabulary/VocabWordsK.shtml>
- Fallon, M., Treheb, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *Journal of the Acoustical Society of America*, 115(5), 2242-2249.
- Fassbender, C. (1993). *Auditory Grouping and Segregation Processes in Infancy*. Kaste Verlag, Norderstedt.

- Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12, 47-65.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America*, 115(5), 2246-2256.
- Garadat, S. N., & Litovsky, R. Y. (2007). Speech intelligibility in free field: Spatial unmasking in preschool children. *Journal of the Acoustical Society of America*, 121(2), 1047-1055.
- Green, D. M. (1969). Masking with continuous and pulsed sinusoids. *Journal of the Acoustical Society of America*, 46(4), 939-946.
- Green, D. M., & Weber, D. L. (1980). Detection of temporally uncertain signals. *Journal of the Acoustical Society of America*, 67(4), 1304-1311.
- Greenberg, G. Z., Bray, N. W., & Beasley, D. S. (1970). Children's frequency-selective detection of signals in noise. *Perception & Psychophysics*, 8(3), 173-175.
- Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method. *Journal of the Acoustical Society of America*, 44(6), 1513-1523.
- Hall, J. W., Buss, E., & Grose, J. H. (2005). Informational masking release in children and adults. *Journal of the Acoustical Society of America*, 118(3), 1605-1613.
- Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, 23, 159-165.
- Haskins, H. (1949). *A phonetically balanced test of speech discrimination for children*. master's thesis, Northwestern University, Evanston, IL.
- Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *Journal of the Acoustical Society of America*, 125(1), 447-456.
- Jensen, J. K., & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science*, 4(2), 104-107.

- Kidd, G., Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *Journal of the Acoustical Society of America*, 124(6), 3793–3802.
- Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America*, 95(6), 3475-3480.
- Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational masking. In W. A. Yost, A. N. Popper & R. R. Fay (Eds.), *Auditory Perception of Sound Sources* (pp. 143-189). New York: Springer.
- Krull, V., Choi, S., Kirk, K. I., Prusick, L., & French, B. (2010). Lexical effects on spoken-word recognition in children with normal hearing. *Ear and Hearing*, 31(1), 102-114.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98-104.
- Leibold, L. J. (2012). Development of auditory scene analysis and auditory attention. In L. A. Werner, R. R. Fay & A. N. Popper (Eds.), *Human Auditory Development* (pp. 137-161). New York: Springer.
- Leibold, L. J., & Bonino, A. Y. (2009). Release from information masking in children: Effect of multiple signal bursts. *Journal of the Acoustical Society of America*, 125(4), 2200-2208.
- Leibold, L. J., & Neff, D. L. (2007). Effects of masker-spectral variability and masker fringes in children and adults. *Journal of the Acoustical Society of America*, 121(6), 3666–3676.
- Leibold, L. J., & Neff, D. L. (2011). Masking by a remote-frequency noise band in children and adults. *Ear and Hearing*, 32(5), 663-666.
- Leibold, L. J., & Werner, L. A. (2006). Effect of masker-frequency variability on the detection performance of infants and adults. *Journal of the Acoustical Society of America*, 119(6), 3960–3970.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467-477.

- Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *Journal of the Acoustical Society of America*, 117(5), 3091–3099.
- Longman. (1983). *Longman Dictionary of American English: A Dictionary for Learners of English*. New York: Longman.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1-36.
- Lucluyse, W., & Meddis, R. (2009). A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners. *Journal of the Acoustical Society of America*, 126(5), 2570-2579.
- Lutfi, R. A., Kistler, D. J., Callahan, M. R., & Wightman, F. L. (2003a). Psychometric functions for informational masking. *Journal of the Acoustical Society of America*, 114(6), 3273-3282.
- Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., & Callahan, M. R. (2003b). One factor underlies individual differences in auditory informational masking within and across age groups. *Perception & Psychophysics*, 65(3), 396-406.
- Marshall, L., & Jesteadt, W. (1986). Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures. *Journal of Speech and Hearing Research*, 29(82-91).
- McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America*, 102(5), 2945-2953.
- Moore, B. C., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74(3), 750-753.
- Moore, B. C., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio Engineering Society*, 45, 224-240.
- Moore, T. (1981). *Voice communication jamming research*. In AGARD Conference Proceeding 331: Aural Communication in Aviation, pp. 2:1-2:6, Neuilly-Sur-Seine, France.

- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85(1), 365-378.
- Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *Journal of the Acoustical Society of America*, 98(4), 1909-1920.
- Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *Journal of the Acoustical Society of America*, 98(1), 125-134.
- Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics*, 41(5), 409-415.
- Nishi, K., Lewis, D. E., Hoover, B. M., Choi, S., & Stelmachowicz, P. G. (2010). Children's recognition of American English consonants in noise. *Journal of the Acoustical Society of America*, 127(5), 3177-3188.
- Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America*, 87(6), 2705-2715.
- Oh, E. L., Wightman, F. L., & Lutfi, R. A. (2001). Children's detection of pure-tone signals with random multitone maskers. *Journal of the Acoustical Society of America*, 109(6), 2888-2895.
- Richards, V. M., & Neff, D. L. (2003). Cuing effects for informational masking. *Journal of the Acoustical Society of America*, 115(1), 289-300.
- Richards, V. M., Shub, D. E., & Carreira, E. M. (2011). The role of masker fringes for the detection of coherent tone pips. *Journal of the Acoustical Society of America*, 130(2), 883-892.
- Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, 33(3), 441-452.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1), 34-50.

- Storkel, H. L., & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods*, 52(2), 497-506.
- Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, 28, 455-462.
- Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5-11 year-old children. *Hearing Research*, 225, 117-127.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). New York: Pearson Edition, Inc.
- Van Der Heijden, M., & Kohlrausch, A. (1994). Using an excitation-pattern model to predict auditory masking. *Hearing Research*, 80, 38-52.
- Varghese, L. A., Ozmeral, E. J., Best, V., & Shinn-Cunningham, B. G. (2012). How visual cues for when to listen aid selective auditory attention. *Journal of the Association for Research in Otolaryngology*, 13(3), 359-368.
- Viemeister, N. F., & Schlauch, R. S. (1992). Issues in infant psychoacoustics. In L. A. Werner & E. W. Rubel (Eds.), *Developmental Psychoacoustics* (pp. 191-210). Washington, D.C.: American Psychological Association.
- Watson, C. S., & Nichols, T. L. (1976). Detectability of auditory signals presented without defined observation intervals. *Journal of Acoustical Society of America*, 59(3), 655-668.
- Werner, L. A., & Bargones, J. Y. (1991). Sources of auditory masking in infants: Distraction effects. *Attention, Perception & Psychophysics*, 50(5), 405-412.
- Werner, L. A., & Leibold, L. J. (2004). Ecological developmental psychoacoustics. In J. G. Neuhoff (Ed.), *Ecological Psychoacoustics* (pp. 191-217). New York: Academic Press.
- Werner, L. A., & Marean, G. C. (1991). Methods for estimating infant thresholds. *Journal of the Acoustical Society of America*, 90(4), 1867-1875.

- Werner, L. A., Parrish, H. K., & Holmer, N. M. (2009). Effects of temporal uncertainty and temporal expectancy on infants' auditory sensitivity. *Journal of the Acoustical Society of America*, 125(2), 1040-1049.
- Wightman, F. L., & Allen, P. A. (1992). Individual differences in auditory capability among preschool children. In L. A. Werner & E. W. Rubel (Eds.), *Developmental Psychoacoustics* (pp. 113-133). Washington, D.C.: American Psychological Association.
- Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., & Oh, E. L. (2003). Children's detection of pure-tone signals: Informational masking with contralateral maskers. *Journal of the Acoustical Society of America*, 113(6), 3297-3305.
- Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *Journal of the Acoustical Society of America*, 118(5), 3164-3176.
- Wightman, F. L., Kistler, D. J., & Brungart, D. S. (2006). Informational masking of speech in children: Auditory-visual integration. *Journal of the Acoustical Society of America*, 119(6), 3940-3949.
- Wightman, F. L., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in dichotic informational masking paradigms. *Journal of the Acoustical Society of America*, 128(1), 270-279.
- Yost, W. A. (1991). Auditory image perception and analysis: The basis for hearing. *Hearing Research*, 56, 8-18.