Selective Auditory Attention in Adults: Effects of Rhythmic Structure of the Competing
Language

by

Leigh Ann Reel, Au.D.

A Dissertation

In

COMMUNICATION SCIENCES AND DISORDERS

Submitted to the Graduate Faculty of
Texas Tech University Health Sciences Center in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

Candace Bourland Hicks, Ph.D.

Reneé Bogenschutz, Ph.D.

Melinda Corwin, Ph.D.

Sylvia Martinez, Ph.D.

Steven Zupancic, Au.D., Ph.D.

Paul P. Brooke, Jr., Ph.D., FACHE
Dean of the School of Allied Health Sciences

August, 2009

Copyright 2009, Leigh Ann Reel

Acknowledgements

I would first like to offer my deepest gratitude to Dr. Candace Hicks, my dissertation committee chair, for the guidance and encouragement she provided, not only during my work on the dissertation, but also throughout my 4 years in the Ph.D. program. She spent countless hours listening to my ideas, questions, and concerns, offering invaluable advice, and providing feedback on each draft of the dissertation. I feel so blessed to have had such a wonderful mentor, and I can only hope to one day set the same example for my own students. I would also like to thank the other members of my dissertation committee: Dr. Steven Zupancic, the member who so patiently advised me regarding the technical details of making the competing speech recordings, editing the recordings, and planning the setup of the test equipment; Dr. Reneé Bogenschutz, the member who made it her mission to help me find the subjects that I needed for my study; Dr. Melinda Corwin, the member who was so helpful when I was having a surprisingly difficult time finding native English speakers to read for the recordings; and Dr. Sylvia Martinez, the member whose Spanish skills were an answer to my prayers when I so desperately needed a native speaker to help edit the Spanish text. I learned so much from working with each of these individuals, and I will be forever grateful for the wise advice and encouraging words that they offered as I planned and conducted my dissertation.

In addition to my committee members, I would also like to thank Mary Perez at Lubbock Christian University for her assistance in my search for English and Spanish text for the competing speech recordings and Karin Senft at the TTU International Student and Scholar Service (ISSS) Office for e-mailing my recording volunteer flyer to the TTU and TTUHSC international students. A special thank you also goes to Dr.

Katsura Aoyama at the TTUHSC Department of Speech, Language, and Hearing Sciences, Aki Shimada at the TTU ISSS Office, and Hilda Martinez and Wiebke Schirrow, two TTU graduate students, for the hours they each devoted to finding native speakers for the recordings and helping me edit the 16 non-English recordings. In addition, I am so grateful to the numerous volunteers who read for a recording in their native language and the subjects who were willing to sit through 2 hours of tedious listening tasks in order to participate in my study. These individuals were answers to my prayers, and I cannot thank them enough for giving up their time to help me, despite the fact that I was a complete stranger to most of them.

Special thanks also goes to Dr. Rajinder Koul, chair of the TTUHSC Department of Speech, Language, and Hearing Sciences, and Dr. Nick Coppola, director of the TTUHSC Clinical Practice Management Program, for the wonderful advice they each provided when I was struggling to determine the best way to account for the “nested” relationship between variables during the statistical analysis of my data.

I would also like to thank all of the family and friends who have supported, encouraged, and prayed for me during my 4 years in the Ph.D. program. I am so thankful for the friendship, encouragement, and “life-line” to the outside world that my sister, Lindsay, provided me through her many, many phone calls. At the most stressful times in working on the dissertation, my conversations with Lindsay gave me a much-needed opportunity to vent my frustrations and enjoy the distraction of hearing about her daily adventures. I am also grateful to my brother, Jodie, and my sister-in-law, Summer, for the weekends they allowed me to spend at their house in the small country community of Girard, Texas. Although much of my time was spent studying, these “mental health

weekends” revived my spirit and gave me the perspective that I needed to face another week of Ph.D. work. In addition, I would like to thank my beautiful nieces, Addison and Coulter, and my adorable nephew, Griffith, for the joy they have brought into my life since the day they each came into this world. Only God could have given me such effective stress relievers as these three little Reels.

I also want to express my appreciation to my parents, Ronnie and Brenda Reel, for always believing in me and giving me the love and support that I needed to face each new challenge in my life, including the decision to pursue the Ph.D. degree. I would like to say a special thank you to my dad for the time and money he has spent to keep my car running so that I could complete the dissertation, get a “real” job, and buy a new car! He has given up countless hours of his time to diagnose and treat my car problems, yet I never once remember hearing him complain. It has meant so much to know that I could always call my dad, and he would do anything to make sure that I could get to the school for my Ph.D. classes and meetings. A special thank you also goes to my mother for the emotional (and at times, financial) support she has given me as I worked to complete the Ph.D. degree. My mother has called me almost every day over the past 4 years and has patiently listened to the details of my life in the Ph.D. program. She has shared my worries and stress over each new challenge, and I am so grateful to her for the hours she has spent listening, empathizing, and encouraging me each and every day.

I would also like to thank my aunt, Linda Sanders, for the very generous “scholarship” she gave me when my financial resources were depleted during the Ph.D. program and for caring enough to make the 7-hour trip to Lubbock with my mother to help me celebrate after my dissertation defense. In addition, I am so grateful to my

friend, Marci Foster, for patiently listening to the never-ending details of my life in the Ph.D. program and offering the encouragement that I needed keep going.

Last, but certainly not least, I want to express my heartfelt appreciation to Dr. Candace Hicks, her husband, Jason, and their sons, Noah and Eli, for all they have done to help me over the last 4 years. They have given me a home away from home in Lubbock and have been with me through each step of this long journey, from helping me celebrate the good times to consoling and encouraging me during the bad times. They have fed me countless meals, and their leftovers have saved me from having to pull out the Honey Nut Cheerios box many nights! Dr. Hicks has frequently given up her evening and weekend time to answer my questions, listen to my concerns, and offer her feedback on my Ph.D. “emergencies.” Jason has devoted countless hours to fixing my seemingly endless car problems, and his incredible automotive skills and generosity have literally saved me thousands of dollars in mechanic bills. Noah and Eli have given me the joy and laughter that I needed to endure the stress of the Ph.D. program, and I feel so blessed that they seem to consider me to be a part of their family.

I cannot imagine how I could have completed the Ph.D. degree without the support that I received from the Hicks family. Dr. Hicks and Jason believed that the Ph.D. was the right choice for me long before I was willing to even consider the possibility, and I cannot thank them enough for steering me along the right path, even though that path turned out to be very stressful much of the time! I now see that the Ph.D. was part of God’s plan, and I thank Him every day for sending the Hicks family to bless my life by giving me the encouragement, moral support, and friendship that I needed to fulfill this part of His plan.

Table of Contents

Acknowledgements	ii
Abstract	viii
List of Tables	x
List of Figures	xi
Chapter	
I. Introduction	1
<i>Statement of the Problem</i>	1
<i>Energetic and Informational Masking</i>	3
<i>Effects of Non-Linguistic Characteristics of Competing Speech</i> <i>Signals</i>	6
<i>Effects of Linguistic Characteristics of Competing Speech</i> <i>Signals</i>	24
<i>Effects of Native Language Rhythmic Structure on Speech</i> <i>Processing</i>	34
<i>Goals and Hypotheses</i>	55
II. Methods	59
<i>Subjects</i>	59
<i>Pre-Test Procedures</i>	60
<i>Test Stimuli</i>	64
<i>Assignment of Target Lists to Test Conditions</i>	75
<i>Mixing of the Target and Competing Speech</i>	76
<i>Order of Test Conditions</i>	77
<i>Individual Playlists</i>	77
<i>Test Environment and Equipment</i>	78
<i>Test Setup and Procedures</i>	79
<i>Scoring and Data Entry</i>	81
<i>Research Design</i>	82
III. Results	87
<i>Descriptive Analysis</i>	87
<i>Repeated Measures Analysis of Variance (ANOVA)</i>	89
IV. Discussion	95
<i>Differences Between Masker A and Masker B for Each</i> <i>Language</i>	96
<i>Relationships Between the Masking Effectiveness of Different</i> <i>Languages</i>	101

<i>Limitations of the Current Study and Implications for Future</i>	
<i>Research</i>	111
<i>Conclusions</i>	121
References.....	125
Appendices	
A. Subject Recruitment Flyer	131
B. Informed Consent Form.....	132
C. Subject Questionnaire	135
D. Hearing Screening Form	138
E. Sample Hearing in Noise Test (HINT) Sentence Lists	139
F. Recruitment Flyer for Language Recording Participants	142
G. Questionnaire for Language Recording Participants	143
H. Sample Scoring Forms.....	145
I. Instructions Given Prior to Testing in the Quiet Conditions	147
J. Instructions Given Prior to Testing in the Competing Conditions	149
K. Results of Additional Statistical Analyses	150

Abstract

Research to date has provided limited evidence related to how selective auditory attention is affected by: 1) differences between the vocal and/or speaking characteristics of different masking talkers and 2) rhythmic properties of the language of the competing (or masking) speech signal. The current study addressed these gaps in existing research by measuring sentence reception thresholds (sSRTs) for 50 monolingual English-speaking adults in conditions where the competing speech was spoken in a language with a stress-based rhythm (i.e., English or German), a syllable-based rhythm (i.e., Spanish or French), or a mora-based rhythm (i.e., Japanese) (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi, Jusczyk, & Johnson, 2000). In all conditions, the target speech signal consisted of meaningful English sentences spoken by a male talker; whereas, the competing speech signal consisted of continuous speech from 1 male and 1 female native speaker of the language assigned to that condition. For each language, 2 different 2-talker masking signals were created (i.e., labeled Masker A and Masker B), with a different male-female pair of native speakers used for each signal. An average sSRT was obtained in 10 types of competing speech conditions, including 1 Masker A condition and 1 Masker B condition for each of the 5 languages.

Analysis of results revealed that for each language, there was a significant difference between subjects' performance in the Masker A and Masker B conditions. These findings suggest that the vocal/speaking characteristics of different mixed-gender 2-talker maskers (i.e., 1 male and 1 female masking talker) can provide different degrees of masking effectiveness, regardless of whether the masking talkers are speaking the listeners' native language or an unfamiliar language. A significant effect of the language

of the competing speech was also seen, with significantly greater listening difficulty associated with conditions where the competing speech was spoken in English, German, or Japanese, as compared to Spanish or French. This pattern of results leads to 2 primary conclusions: 1) the underlying rhythmic structure of the language of the competing speech has a greater impact on masking effectiveness than the meaning of the competing speech, and 2) competing speech spoken in a language from the listener's native rhythmic category (i.e., stress-based) is more difficult to ignore than competing speech spoken in a language from a non-native rhythmic category (i.e., syllable-based), at least for some languages. Together, these results extend the findings of previous studies by providing evidence that exposure to the rhythm of the native language affects not only how a listener processes speech in quiet (e.g., Cutler, Mehler, Norris, & Segui, 1983; Sebastián-Gallés, Dupoux, Costa, & Mehler, 2000), but also how he/she processes a target speech signal in the presence of speech from 2 masking talkers.

List of Tables

1	Random Assignment of English and Spanish Text.....	84
2	Random Assignment of Hearing in Noise Test (HINT) Lists to Two-Talker Masking Pairs	85
3	Pairwise Comparisons Between the Overall Mean Sentence Reception Threshold (sSRT) for Each Language	94

List of Figures

1	Mean sentence reception threshold (sSRT) for Masker A and Masker B for each language	93
2	Overall mean sentence reception threshold (sSRT) for each language (no Masker A – Masker B distinction)	93

Chapter I

Introduction

Statement of the Problem

On a daily basis, humans are faced with the challenge of listening to speech in environments where other sounds are present. These competing sounds may include non-speech noises (e.g., noise generated by equipment), speech from other talkers, and/or a combination of these sounds. In order to accurately perceive speech in competing conditions, listeners must have adequate selective auditory attention skills in order to select and attend to the relevant speech information (i.e., the target signal) while simultaneously ignoring any irrelevant auditory stimuli in the environment (i.e., the masking or competing signals) (Medwetsky, 2002). Without these skills, listeners would be unable to function effectively in the complex acoustical conditions often present in home, work, school, and social environments.

Previous studies have shown that the ability to selectively attend to speech from a target talker while ignoring speech from other talkers is differentially affected by non-linguistic factors, such as differences between the spatial location (e.g., Freyman, Balakrishnan, & Helfer, 2004), intensity (e.g., Cooke, Garcia Lecumberri, & Barker, 2008), number (e.g., Simpson & Cooke, 2005), and gender (e.g., Brungart, 2001) of the target and masking talkers. However, there is not a clear consensus regarding how selective auditory attention performance is affected by: 1) differences between the vocal/speaking characteristics of the masking talkers and 2) the linguistic characteristics of competing (or masking) speech signals. To date, only one previous study has attempted to assess how the vocal/speaking characteristics of masking talkers impact the

extent to which a competing speech signal interferes with selective auditory attention (Freyman, Helfer, & Balakrishnan, 2007). In addition, previous studies related to effects of the linguistic properties of competing speech have primarily focused on listeners' abilities to ignore different forms of English speech, including forward/reversed (e.g., Rhebergen, Versfeld, & Dreschler, 2005), meaningful/non-meaningful (e.g., Chermak, 1975), and grammatical/ungrammatical speech (e.g., Chermak, Vohnhof, & Bendel, 1989). In contrast, the ability to ignore speech spoken in different languages has been investigated by very few studies (e.g., Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). Linguistic research indicates that certain languages can be grouped according to whether they exhibit stress-based, syllable-based, or mora-based rhythm (e.g., Ramus, Nespor, & Mehler, 1999), and these differences in language rhythm have been shown to have a significant impact on how infants and adults process speech in quiet listening conditions (e.g., Cutler et al., 1983; Cutler, Mehler, Norris, & Segui, 1986; Nazzi et al., 2000; Sebastián-Gallés et al., 2000).

In light of these findings, one question of interest relates to whether the rhythmic structure of a language may also affect a listener's ability to process speech in conditions where one or more talkers are speaking simultaneously. Because three primary classes of language rhythm have been proposed (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi et al., 2000), a thorough investigation into effects of the rhythmic structure of competing speech would require comparison between listeners' abilities to ignore speech spoken in languages from each of the rhythmic classes. However, no previous study has used competing speech spoken in more than two languages. Given these limitations of previous research, the primary goals of the current study were to evaluate whether

monolingual English-speaking adults' selective auditory attention was affected by: 1) differences between the vocal/speaking characteristics of different mixed-gender pairs of masking talkers (i.e., one male and one female talker per masking signal) and 2) the rhythmic structure (i.e., stress-based, syllable-based, or mora-based) of the language in which the competing speech was spoken.

Energetic and Informational Masking

To better understand selective auditory attention, studies have compared listeners' performance in different types of competing speech signals. Such studies have shown that non-linguistic (e.g., spatial separation, intensity) and linguistic (e.g., meaning, grammar) characteristics can influence the degree of masking provided by a competing speech signal (e.g., Freyman et al., 2004; Rhebergen et al., 2005). Masking refers to the process by which one sound (i.e., the masker) "obscures or reduces the response to another" sound (i.e., the target) (Pickles, 1988, p. 103). The interference caused by a competing auditory signal can be attributed to two basic forms of masking: energetic and informational masking. Occurring primarily at the level of the cochlea, energetic masking takes place when parts of the target signal are not audible due to overlap between the spectral characteristics of the masker and the target (Brungart, Simpson, Ericson, & Scott, 2001). As a result of energetic masking, the listener can only perceive portions of the target auditory signal and must therefore rely on incomplete information when attempting to determine what he/she heard (Cooke et al., 2008). The amount of energetic masking that occurs is determined by the intensity of the masking signal relative to the intensity of the target signal within different frequency channels (Scott, Rosen, Wickham, & Wise, 2004). As such, noise signals (e.g., white, pink, and speech

noise) primarily interfere with speech perception due to the energetic masking that results from energy that is broadly distributed across the range of audible frequencies. The effects of energetic masking increase as the intensity of the masking signal increases relative to the intensity of the target signal (i.e., as the signal-to-noise ratio decreases) (Brungart, 2001). Speech-in-speech listening conditions also result in some degree of energetic masking when the spectrum of the target speech signal overlaps with that of the competing speech signal, but this effect is minimized by the amplitude fluctuations that occur in natural speech (Scott et al., 2004).

Unlike energetic masking, informational masking occurs at levels of the auditory system higher than the cochlea (Brungart, 2001), but definitions of this form of masking vary widely throughout the literature. More general definitions propose that informational masking may occur for non-speech and speech signals (e.g., Wightman & Kistler, 2005). However, the informational form of masking is most commonly defined as a phenomenon that occurs in listening situations where the target and competing signal(s) both consist of speech. According to one such definition, informational masking is caused by overlap between the non-linguistic and linguistic characteristics of the target and competing speech signal(s) (e.g., Freyman et al., 2004; Scott et al., 2004). Cooke and colleagues (2008) have extended this general definition and proposed that informational masking may be the result of one or more problems that occur as the listener attempts to: 1) segregate simultaneous speech signals, 2) selectively focus attention only on the words spoken by the target talker, and 3) accurately process the components of the target speech signal. For example, according to this view, informational masking may occur because the listener is unable to completely segregate

the target speech from the competing speech signal(s). In this type of situation, the listener may mistakenly assign elements of the masker to the target (e.g., formant transitions, cues as to manner of articulation, phonemes, syllables, or whole words). In addition, the presence of competing speech may result in informational masking because the listener experiences difficulties focusing his/her attention only on the target speech signal. The listener's attention may be drawn to the competing speech, thus causing him/her to miss portions of what the target talker said. Similarly, segregating two or more simultaneous speech signals, processing the two signals, and directing attention only to the information spoken by the target talker places a high cognitive demand on the listener's processing and attention resources. Informational masking may occur if the listener's limited cognitive resources are unable to handle these simultaneous processing demands. Finally, if the listener understands the language of the competing speech signal, informational masking is increased, "perhaps due to the engagement of language-specific decoding processes for both masker and target, which in turn increases cognitive load" (Cooke et al., 2008, p. 415).

In competing speech conditions, it can be difficult to isolate the relative effects of energetic and informational masking, partially due to the ambiguous criteria used to define the informational form of masking. The distinction between energetic and informational masking notwithstanding, numerous studies have shown that the overall degree of masking (i.e., the combined effects of energetic and informational masking) that occurs when two or more speech signals are presented simultaneously depends on various non-linguistic (e.g., Brungart, 2001; Cooke et al., 2008; Freyman et al., 2004) and linguistic factors (e.g., Chermak, 1975; Rhebergen et al., 2005; Van Engen & Bradlow,

2007). In general, such studies have shown that overall masking effectiveness (i.e., the degree of interference provided by a competing speech signal) is reduced and selective auditory attention performance subsequently improves as the characteristics of the competing speech signal become increasingly different from those of the target speech signal. Listeners are able to use such differences to separate the target speech signal from the competing speech signal(s) and then focus their attention selectively on the information spoken by the target talker. For example, differences between the spatial location (e.g., Freyman et al., 2004), intensity (e.g., Cooke et al., 2008), number (e.g., Simpson & Cooke, 2005), and gender (e.g., Brungart, 2001) of the target and masking talker(s) can provide non-linguistic cues that assist listeners in differentiating between the target and competing speech signals(s). In addition, a listener's ability to selectively attend to a target speech signal may also be influenced by linguistic differences between the target and competing speech, such as differences in semantic content (e.g., Chermak, 1975) and differences in the characteristics of the language in which each speech signal is spoken (e.g., Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). The following sections will provide a summary of evidence from previous studies that have investigated how non-linguistic and linguistic factors, such as these, can influence the overall masking effectiveness of a competing speech signal.

Effects of Non-Linguistic Characteristics of Competing Speech Signals

Spatial Separation

One non-linguistic factor known to affect selective auditory attention in competing speech conditions is the spatial separation between the source of the target speech and the source of the competing speech signal(s). Numerous studies have shown

that a significant release from masking (i.e., a reduction in masking) occurs and selective auditory attention performance therefore improves when the target and competing speech signal(s) originate from different locations along the horizontal plane (e.g., Freyman et al., 2004; Freyman et al., 2007; Freyman, Helfer, McCall, & Clifton, 1999; Rakerd, Aaronson, & Hartmann, 2006). This spatial release from masking has been shown to occur across various masking talker conditions, including different signal-to-noise ratios (SNRs), different numbers of masking talkers, and different target-masker gender relationships. However, as will be discussed in the following sections, the degree of benefit provided by target-masker spatial separation varies depending on the specific conditions of the listening environment.

In non-reverberant listening conditions, target-masker spatial separation generally improves recognition of the target speech signal due to two primary factors: the head shadow effect and binaural interaction (Freyman et al., 1999). The head shadow effect occurs when the target and masking speech signals are presented from different locations, such that the listener's head is positioned between the source of the masking speech and the ear that is farthest away from the masking source. In order to reach the far-ear side, the masking speech signal must travel around the listener's head, and in doing so, the intensity of the masking speech is reduced by the time it reaches the far-ear. As a result, the SNR (i.e., the intensity of the target speech relative to the intensity of the masking speech) at the far-ear is improved as compared to the SNR that occurs when the target and masking speech are both presented from directly in front of or directly behind the listener. The head shadow effect occurs primarily for high frequency components of the speech signal.

In contrast, spatial separation mainly improves the detection of low frequency speech components through a phenomenon referred to as binaural interaction (Freyman et al., 1999). Binaural interaction is the process by which the auditory system is able to separate the target and masking speech signals by making use of differences between the time that it takes for each signal to reach the right versus the left ear (i.e., interaural time differences). For example, when a target speech signal is presented from directly in front of the listener and masking speech is presented from 60 degrees azimuth (e.g., 60 degrees to the right of midline), the target speech will reach both ears at approximately the same time; whereas, the masking speech will reach the right ear slightly earlier than the left ear. Because of binaural interaction, the listener's auditory system is able to integrate the auditory information received by each ear and then utilize interaural (i.e., between ears) timing differences to separate the target speech from the masking speech signal.

In a reverberant listening environment, sound reflections mix with the original target and masking speech signals, thus decreasing the advantages provided by the head shadow effect and binaural interaction (Freyman et al., 1999). However, previous research has shown that target-masker spatial separation still results in a release from masking even in rooms that are highly reverberant, although the degree of benefit is reduced relative to what is seen in non-reverberant listening environments. When reverberation is present, listeners are usually still able to accurately localize the sources of the target and masking speech due to a perceptual phenomenon called the precedence effect. This effect occurs when the original sound and its reflections are detected in close succession, and the auditory system subsequently perceives the signals as a single, fused sound originating from the location of the first sound (Moore, 1997). Therefore, in a

reverberant listening environment, the precedence effect allows the listener to perceive the original target signal and its reflections to be one fused signal originating from the location of the original target signal. Likewise, the listener is also able to perceive the original masking signal and its reflections to be a single fused sound originating from the location of the original masking signal. As a result, the benefits of target-masker separation are preserved, at least to some degree, even in reverberant listening environments.

Signal-to-Noise Ratio (SNR)

In addition to target-masker spatial separation, the SNR has also been shown to significantly impact the masking effectiveness of a competing auditory signal. In a listening environment, the SNR is the amount by which the intensity of the target auditory signal exceeds or is surpassed by the intensity of any competing auditory signals (Crandell & Smaldino, 2000). A positive SNR indicates that the intensity of the target auditory signal is higher than that of the competing auditory signal; whereas, a negative SNR indicates that the intensity of the target signal is lower than that of the competing signal. For example, a SNR of +5 dB indicates that the intensity of the target signal is 5 dB higher (i.e., louder) than that of the competing signal. In contrast, a SNR of -5 dB means that the intensity of the target auditory signal is 5 dB lower (i.e., softer) than that of the competing auditory signal.

When a target speech signal is presented in the presence of a masking noise signal, increases in SNR (i.e., a better SNR) result in improvement in speech recognition scores (e.g., Brungart, 2001). This is referred to as a monotonic pattern because increases in the independent variable, the SNR, are never associated with decreases in the

dependent variable, speech recognition scores. For noise maskers, this monotonic pattern occurs because improving the SNR (i.e., by either decreasing the intensity of the masking noise or increasing the intensity of the target auditory signal) results in a decreased level of energetic masking, a change which can only increase the amount of information that the listener is able to detect in the target signal. With further increases in the SNR, speech recognition performance will eventually reach a plateau at or near 100% correct once the SNR is so high that the listener is able to detect all portions of the target auditory signal. Beyond this point, further increases in the SNR will not significantly improve speech recognition because the point of maximum performance has already been reached. On the other hand, the monotonic pattern associated with noise maskers means that decreasing the SNR (i.e., by either increasing the intensity of the masking noise or decreasing the intensity of the target auditory signal) always results in an increased level of energetic masking, a change which can only decrease the amount of information that the listener is able to detect in the target signal. Further decreases in the SNR will eventually render the target signal inaudible to the listener, thus causing speech recognition scores to plateau at or near the level of chance.

In contrast to this pattern of results, changes in SNR do not produce monotonic changes in speech recognition performance when the competing auditory signal consists of speech from one or more competing (or masking) talkers. For competing speech conditions, increases in the SNR may result in little to no change, a noticeable increase, or even a decrease in speech recognition performance, depending on factors such as spatial cues, the number of masking talkers, and the target-masker gender relationship. For example, when the target and masking talker(s) are spatially separated, speech

recognition performance is better across all SNRs as compared to performance when all talkers are presented from the same location (Freyman, Balakrishnan, & Helfer, 2001; Freyman et al., 2004; Freyman et al., 1999; Rakerd et al., 2006). In addition, studies have shown that when spatial cues are available, increases in SNR result in more rapid increases in speech recognition performance across positive and negative SNRs as compared to the more gradual changes in performance that are seen when the target and masking talker(s) are presented from the same location (Freyman et al., 1999; Freyman et al., 2001; Freyman et al., 2004). This general pattern of results has been documented at least for single-talker (Freyman et al., 1999; Freyman et al., 2001) and two-talker (Freyman et al., 2001; Freyman et al., 2004) maskers.

The effects of SNR also differ depending on the number of masking talkers. The interactive effects of these two variables are best demonstrated by first considering results obtained when the target talker and masking talker(s) are presented from the same location and are of the same gender. Under such conditions, Brungart (2001) found that when the masker consisted of a single talker, speech recognition performance changed very little for SNRs from -12 to 0 dB and from +10 to +15 dB, but considerable improvement was seen as the SNR increased from 0 to +10 dB. Other studies using a single competing talker have shown a similar pattern of results. For example, Cooke and colleagues (2008) found that speech recognition performance increased (i.e., improved) rapidly as the SNR increased from 0 to +6 dB, but speech recognition scores actually declined as the SNR increased from -9 to -3 dB. Similarly, speech recognition results reported by Rakerd and colleagues (2006) indicated that for a single masking talker, performance declined slightly as the SNR increased from -12 dB to 0 dB.

The same general pattern of results has been reported for competing speech signals consisting of two or three masking talkers. For example, Brungart et al. (2001) found that when the target and masking talkers were presented from the same location and were of the same gender, changes in SNR had a similar effect on speech recognition performance, regardless of whether a two-talker or a three-talker masker was used. For both masking conditions, speech recognition performance changed very little as the SNR increased from -18 to -9 dB, but performance increased dramatically as the SNR increased from -9 to +12 dB. Rakerd and colleagues (2006) reported similar findings, with performance in both the two-talker and three-talker masking conditions showing only slight improvement with increasing SNR from -8 to 0 dB, followed by a considerable increase in performance as the SNR increased from 0 to +4 dB.

Together, these findings lead to two primary conclusions regarding the effects of SNR on the overall masking effectiveness of speech from one, two, or three masking talkers. First, at positive SNRs, speech recognition performance tends to increase rapidly as the SNR increases (Brungart, 2001; Cooke et al., 2008), although there is some evidence that performance plateaus once a SNR of about +9 to +10 dB is reached (Brungart, 2001). Second, from SNRs in the negative range to approximately 0 dB, speech recognition performance generally plateaus, shows very little improvement, or even decreases as the SNR increases (Cooke et al., 2008; Rakerd et al., 2006). On the surface, this pattern of results seems somewhat counterintuitive, given that as the SNR increases, the target speech signal should become more audible to the listener. However, Brungart (2001) proposed that these findings can be explained by the fact that listeners rely on differences between the target and competing speech signal(s) in order to separate

the signals and selectively focus attention only on the target information. When the target and masking talker(s) are of the same gender and are presented from the same location, as was the case for the previously described results, listeners must use other cues, such as intensity differences, in order to separate the target and masking talker(s). At positive SNRs, listeners are better able to selectively attend to the target talker because his/her voice is louder than that of the masking talker(s), and as the SNR increases, speech recognition performance increases because the target talker's voice becomes increasingly louder than the voice(s) of the masking talker(s).

In contrast, at negative SNRs, the level of the target speech is softer than that of the masking speech (Brungart, 2001). As the SNR decreases, the intensity of the masking speech becomes increasingly louder relative to the target speech, thus increasing the level of energetic masking. Subsequently, the target talker's voice may become inaudible to the listener when the target and masking signal overlap in time and frequency. However, as the SNR decreases below 0 dB, the listener is able to benefit from the increasing difference between the intensity of the target and masking speech, despite the fact that the target speech becomes the softer signal. In fact, the target-masker intensity difference actually provides an advantage that is greater than or at least equal to the deleterious effects of any increases in energetic masking that may occur. Together, these factors help to explain why speech recognition performance has been found to remain relatively stable, decrease very little, or even increase slightly as the SNR decreases from 0 dB to around -12 dB.

Finally, in addition to interacting with spatial separation and the number of masking talkers, the effects of SNR also vary as a function of the gender relationship

between the target and masking talkers. For example, studies have shown that in conditions where one target and one masking talker are present and the two are of a different gender, speech recognition scores remain relatively high (i.e., 70% correct to approximately 100% correct) and show only slight improvements as the SNR increases from -12 dB to +15 dB (Brungart, 2001; Brungart et al., 2001). Similarly, when the target and masking talkers are of the same gender, speech recognition scores remain relatively stable and show little to no improvement as the SNR increases from -12 to 0 dB. However, the scores in the same-gender target-masker condition are poorer than those obtained in the different-gender condition across this range of SNRs. In addition, unlike the pattern seen in the different-gender condition, speech recognition scores in the same-gender condition typically show rapid increases as the SNR increases from 0 dB to approximately +9 dB, at which point a ceiling effect is typically reached.

Together, these findings indicate that when a competing auditory signal consists of speech, the effects of SNR vary depending on other non-linguistic factors, including the spatial location of the target and masking talkers, the number of masking talkers, and the target-masker gender relationship. As previously described, studies have shown that such factors can interact in complex ways. As such, changes in SNR may result in any number of effects on selective auditory attention performance, including no effect, an improvement, or a decline in performance, with the direction and magnitude of these changes varying depending on other non-linguistic characteristics of the masking speech signal.

Number of Talkers

In investigating the effects of non-linguistic characteristics of competing speech signals, studies have also focused on how selective auditory attention performance is affected by the number of masking talkers. For example, Simpson and Cooke (2005) measured changes in adult listeners' abilities to identify consonants as the number of talkers in the competing speech signal increased. In each condition, testing was performed at a SNR of – 6 dB. Results revealed that consonant identification scores decreased as the number of masking talkers increased, up to 6 talkers. However, performance did not change significantly as the number of masking talkers increased from 8 to 128. To explain these findings, the investigators proposed that energetic masking increased as the number of simultaneous talkers in the competing speech signal increased, presumably due to greater overlap between the spectrum of the target and masking signals. In contrast, the effects of informational masking were hypothesized to have reached a peak with smaller numbers of masking talkers, presumably because listeners were able to detect more of the content of the competing signal when fewer masking talkers were present. Other studies have reported a similar pattern of findings, such as poorer speech recognition performance in the presence of 2 masking talkers as compared to a single masking talker (Freyman et al., 2001) and greater difficulty recognizing speech in the presence of a 6-talker masker as compared to a 2-talker masker (Van Engen & Bradlow, 2007).

In addition to these findings, evidence also indicates that interactive effects occur between the number of masking talkers and the SNR in the listening environment, such that increasing the number of masking talkers may have a different effect on selective

auditory attention performance depending on the SNR. Results from previous studies indicate that the interaction between number of talkers and SNR primarily occurs when the number of masking talkers is increased from one talker to two talkers. For example, Brungart and colleagues (2001) found that when no spatial or gender cues were available, speech recognition performance declined dramatically when the number of masking talkers increased from one talker to two talkers, but this effect was primarily seen at SNRs from -12 dB to 0 dB. Across this range of negative SNRs, speech recognition performance for the two-masking-talker condition improved as the SNR improved; whereas, performance in the single-masking-talker condition remained stable across the negative SNRs. As such, speech recognition performance became more similar between the one-masking-talker and the two-masking-talker condition as the SNR improved up to 0 dB SNR. However, at SNRs from 0 dB to +12 dB, increasing the number of masking talkers from one talker to two talkers resulted in only very small changes in speech recognition performance. In contrast, increasing the number of masking talkers from two talkers to three talkers had little to no effect on speech recognition performance, regardless of the SNR. In other words, performance was similar in the two- and three-masking-talker conditions across the tested range of negative and positive SNRs (i.e., -18 dB to +12 dB).

Together, these results indicate that the addition of a second masking talker (i.e., increasing the number of masking talkers from one talker to two talkers) causes a dramatic decline in the listener's ability to selectively attend to a target speech signal, but only when the SNR is negative. As described in the previous section, Brungart (2001) proposed that when a single masking talker is present, speech recognition performance

changes very little as the SNR decreases below 0 dB because intensity differences between the target speech and the masking speech provide an advantage that balances or even outweighs the increase in energetic masking that occurs as the SNR declines. However, the results reported by Brungart and colleagues (2001) suggest that the addition of a second masking talker interferes with a listener's ability to make use of these target-masker intensity differences at negative SNRs. One possible explanation for this finding is that at negative SNRs, the presence of a second masking talker may result in an increase in energetic masking that is sufficient to prevent the listener from catching enough "glimpses" of the softer target speech signal in order to be able to differentiate it from the louder masking speech signal. However, if the performance differences between the one- and two-masking-talker conditions were due primarily to energetic masking, then one would also expect that speech recognition performance would decline noticeably when the number of masking talkers increased from two talkers to three talkers. Yet, in contrast to this prediction, Brungart and colleagues (2001) found that increasing the number of masking talkers from two talkers to three talkers had little to no effect on speech recognition performance at negative or positive SNRs. Therefore, it seems that energetic masking alone cannot account for the dramatic performance decrement that occurs at negative SNRs when a second masking talker is added.

One alternative explanation is that speech from two masking talkers may interfere more with performance than speech from one masking talker because of the increased cognitive demands associated with: 1) separating the target speech signal from not one but two masking speech signals and 2) simultaneously focusing attention only on the information spoken by the target talker. This type of "higher level" interference could be

the result of an increase in informational masking (Brungart et al., 2001). However, additional research is needed to definitively determine which factors are responsible for the dramatic decline in selective auditory attention performance that occurs at negative SNRs when a second masking talker is added but not when the number of masking talkers is increased from two talkers to three talkers.

Effects of the number of masking talkers can also vary as a function of the spatial separation between the target and masking signals. For instance, Freyman and colleagues (2001) found that adult listeners' sentence recognition performance was poorer when the masker consisted of 2 talkers as compared to when the masker consisted of a single talker. This pattern of results was observed across the tested SNRs of -12 dB to 0 dB and in two spatial conditions: 1) target and masking sentences presented from the same speaker positioned in front of the listener (F-F condition) and 2) target sentences presented from a front speaker and masking sentences presented from both the front speaker and the right speaker (F-RF condition). However, the effect of the number of masking talkers was greatest when the target and masking signals originated from the same location (F-F condition). Freyman and colleagues (2004) later extended these findings using the same test stimuli and test setup, with the exception that testing was performed using maskers consisting of speech from 2, 3, 4, 6, and 10 talkers. Results revealed that sentence recognition scores were poorer in the F-F condition (i.e., no spatial separation) as compared to the F-RF condition across all tested SNRs and all numbers of masking talkers, but as the number of masking talkers increased from 2 to 10, sentence recognition scores became increasingly similar between the two spatial conditions.

Therefore, it seems that the benefits of target-masker spatial separation decline as the number of masking talkers increases, at least from 2 to 10 talkers.

Gender of the Masking Talker(s)

The masking effectiveness of a competing speech signal can also vary as a function of the gender of the target and masking talker(s). For example, Brungart (2001) compared adult listeners' abilities to recognize English phrases spoken by a target talker while attempting to ignore English phrases spoken by one masking talker. The effect of gender was investigated using conditions where the target and masking talker were of a different gender, of the same gender, or were the same talker. In each type of condition, testing was performed at SNRs from -12 dB to +15 dB. The results revealed that listeners performed similarly whether the target talker was a male or a female. However, speech recognition scores differed considerably depending on the relationship between the gender of the target talker and the gender of the masking talker. For example, across the tested SNRs, speech recognition performance was best when the target and masking talker were of a different gender and poorest when the target and masking phrases were spoken by the same talker. Other studies using one masking talker have reported a similar pattern of results (e.g., Cooke et al., 2008). One explanation for these findings is that when a single masking talker is present, speech recognition is easiest when the target and masking talker are of a different gender because the vocal characteristics of the two talkers are considerably different. On the other hand, performance is poorest when the target and masking speech are spoken by the same talker because there are no differences between the vocal characteristics of the "two" talkers. Together, these findings provide another example of how selective auditory attention performance tends to improve as

differences increase between the non-linguistic characteristics of the target and masking speech signal(s).

As the number of masking talkers increases from one to three, the effects of the target-masker gender relationship become somewhat more complex. However, similar to the results obtained for one masking talker, Brungart and colleagues (2001) found that when two or three masking talkers were used, speech recognition performance was still best when all masking talkers were of a different gender than the target talker and poorest when the target and masking speech signals were all spoken by the same talker. For the two- and three-masking-talker conditions, differences between the speech recognition scores obtained for the mixed-gender (i.e., masking talkers consisted of at least one male and one female) and same-gender (i.e., all masking talkers were of the same gender as the target talker) conditions were somewhat small.

Other evidence suggests that the magnitude of the target-masker gender effect varies as a function of the SNR in the listening environment. For example, Brungart (2001) found that when a single masking talker was used, the difference between performance in the different-gender, same-gender, and same-talker conditions was greatest at negative SNRs and changed very little as the SNR increased from -12 dB to -3 dB. However, as the SNR increased from 0 to +15 dB, the difference between the three conditions decreased, indicating that the target-masker relationship had less of an effect on speech recognition performance as the SNR improved. One explanation for these findings is that at negative SNRs, the masking signal is louder than the target signal; therefore, the listener must rely more on differences between the target and masking voices in order to separate the competing signals and selectively focus attention on the

target signal. As such, the cues provided by differences between the voice characteristics of the target and masking talker have a greater impact on speech recognition at negative SNRs. In contrast, at positive SNRs, the target signal is louder than the masking signal; therefore, the listener can more easily attend to the target signal and is less reliant on differences between the target and masker voice characteristics. As a result, the relationship between the gender of the target and masking talkers has a much smaller effect on speech recognition performance at positive SNRs as compared to the noticeable gender effect that occurs when testing is performed at negative SNRs.

Masking Talker Vocal/Speaking Characteristics

With regard to the masking effectiveness of competing speech, a final non-linguistic factor to be considered is the effect of differences between the vocal or speaking characteristics of individual masking talkers. Individual talkers may vary in areas such as their fundamental frequency, vocal quality, intonation patterns, and speaking rate, and such differences could potentially affect the degree to which speech from a particular masking talker interferes with recognition of a target speech signal (Freyman et al., 2007). Evidence of differences between the masking effectiveness of different masking talkers was recently reported by Freyman and colleagues (2007) based on an experiment that compared listeners' abilities to recognize a target nonsense sentence in the presence of two competing nonsense sentences. In each of the five test conditions, the target sentence was spoken by the same female talker, and the competing sentences were spoken by a different pair of female talkers. For each of the five two-talker maskers, testing was performed at SNRs ranging from -12 dB to 0 dB in two spatial conditions: 1) with the target and masking sentences presented from the same

loudspeaker positioned directly in front of the listener (F-F) and 2) with the target sentences presented from the front loudspeaker and the masking sentences presented from the front and right loudspeakers (F-RF).

Analysis of the results revealed that the five two-female-talker maskers differed in their masking effectiveness, but the degree of the differences varied depending on the SNR and the spatial separation between the target and masking talkers (Freyman et al., 2007). In the F-F condition (i.e., no spatial separation), percent correct scores varied widely between the five two-talker maskers in each SNR condition, with the greatest variability seen at a SNR of -4 dB, where a difference of around 50 percentage points was found between the scores for the least and the most effective two-talker masker. Considerably less variability (approximately 25 percentage points) was noted when testing was performed in the F-F condition (i.e., no spatial separation) at SNRs of -12 dB and 0 dB. In the F-RF condition (i.e., spatial separation), much less variability was seen between the five pairs of female-talker maskers, and these differences were less affected by changes in SNR. For example, comparison between the percent correct scores obtained in the F-RF condition for the least and most effective pairs of masking talkers revealed a difference of approximately 15 percentage points at a SNR of -12 dB and a difference of less than 10 percentage points when testing was performed at a SNR of 0 dB.

Together, these findings suggest that variability between vocal/speaking characteristics may lead to considerable differences in the masking effectiveness of individual pairs of masking talkers, at least when both masking talkers are of the same gender as the target talker (i.e., same-gender two-talker maskers). However, additional

research is needed to determine the degree of individual variability that occurs for other numbers of masking talkers (e.g., single-talker maskers, three-talker maskers) and other target-masker gender configurations (e.g., mixed-gender maskers), as well as to determine if and how factors such as fundamental frequency, prosody, vocal quality, and speaking rate may contribute to this variability.

Collectively, these studies of non-linguistic factors indicate that when a target speech signal and one or more masking speech signals are presented simultaneously, the overall degree of masking (i.e., the combined effects of energetic and informational masking) varies depending on the non-linguistic characteristics of the masking speech signal, including the target-masker spatial separation (e.g., Freyman et al., 2004), the SNR (e.g., Cooke et al., 2008), the number of masking talkers (e.g., Simpson & Cooke, 2005), the target-masker gender relationship (e.g., Brungart, 2001), and variability between the vocal/speaking characteristics of different masking talkers (e.g., Freyman et al., 2007). However, it is very difficult to isolate the individual effects of these factors because each non-linguistic characteristic of a masking speech signal has been shown to interact with the signal's other non-linguistic characteristics. Despite these complex interactions, a common finding across the majority of studies is that the degree of interference (i.e., masking effectiveness) provided by a competing speech signal declines and selective auditory attention performance subsequently improves as the differences increase between the non-linguistic characteristics of the target speech signal and the masking speech signal. This pattern presumably occurs because differences between the spatial location, intensity, number, and/or gender of the target and masking talker(s) provide cues that help the listener to separate the target and masking signal(s) and

selectively attend to the target speech. However, listeners do not rely only on non-linguistic differences in order to separate two or more simultaneous speech signals. Instead, differences between the linguistic characteristics (e.g., meaning, grammar, rhythmic structure) of the target speech and the masking speech can also provide important cues that listeners are able to use when faced with the challenge of selectively attending to speech from a target talker in the presence of speech from other talkers (e.g., Chermak, 1975; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). However, as will be discussed in the following sections, many questions remain unanswered regarding which linguistic factors have the greatest impact on the masking effectiveness of a competing speech signal.

Effects of Linguistic Characteristics of Competing Speech Signals

Speech in the Native Language

Forward speech. When considering how listeners attend to a target talker in the presence of one or more competing talkers, another question of interest relates to whether performance is influenced by the linguistic structure/characteristics of the competing speech signal(s). In other words, do the specific sounds, syllables, words, phrases, and/or sentences of the competing speech signal have an effect on the listener's ability to selectively attend to a target speech signal and ignore speech spoken by one or more masking talkers? To date, a relatively small number of studies have addressed this question, the majority of which have focused on how selective auditory attention performance is affected by the semantic and syntactic structure of competing speech spoken in the listeners' native language. For example, two studies (Chermak, 1975; Chermak et al., 1989) investigated these questions by comparing adult listeners' abilities

to recognize monosyllabic words when the masking speech consisted of one of the following types of English sentences: 1) meaningful and grammatically correct, 2) non-meaningful but grammatically correct (i.e., semantically anomalous), and 3) non-meaningful and grammatically incorrect (i.e., strings of unrelated words). Using a +8 dB SNR, Chermak (1975) found that adults experienced significantly greater difficulty recognizing monosyllabic words when the competing speech signal consisted of semantically anomalous sentences as compared to sentences that were meaningful and grammatically correct or sentences that lacked both meaning and grammar. To explain these findings, Chermak (1975) proposed that very familiar competing messages, such as meaningful/grammatically correct sentences, are quickly recognized by the listener as being irrelevant and are disregarded. Similarly, competing messages that are totally unfamiliar, such as sentences that lack meaning and grammar, are quickly recognized as violating the rules of English and are also disregarded. In contrast, semantically anomalous sentences may attract the listener's attention more because they are not too familiar and yet they are also not totally unfamiliar. These findings suggest that the semantic structure (i.e., the meaning) of the masking speech signal may significantly influence adult listeners' abilities to selectively attend to a target speech signal.

However, a later study by Chermak and colleagues (1989) failed to replicate this pattern of results, despite using the same three types of masking sentences as those used in the Chermak (1975) study. In contrast to the earlier study, Chermak and colleagues (1989) found that at a SNR of +12.5 dB, no significant differences occurred between the listeners' recognition of monosyllabic words in the three types of masking sentences. The reason for the inconsistency of these results remains unclear, although interactive

effects between the linguistic structure of the masking speech and the SNR cannot be ruled out. Although both studies used monosyllabic words as the target speech signal and the same three types of English sentences as the masking speech signal, testing was performed at a +8 dB SNR in the Chermak (1975) study and at a +12.5 dB SNR in the Chermak et al. (1989) study. Therefore, the lack of a significant difference between the three types of English competing sentences in the Chermak et al. (1989) study may indicate that the linguistic structure of the masking speech signal has less impact on selective auditory attention performance at higher SNRs. However, additional research is needed to specifically investigate potential interactions between these and other linguistic (e.g., semantic structure, syntactic structure) and non-linguistic characteristics (e.g., spatial separation, SNR, number of masking talkers, and target-masker gender) of masking speech signals.

One limitation of the two previous studies was their failure to take into account that different levels of semantic structure, and therefore different levels of meaning, can exist within a speech signal. For example, meaningful sentences convey meaning at both the sentence and the word level. The individual words within a meaningful sentence each carry their own meaning, and those words then combine to express a broader meaning at the sentence level. Therefore, even nonsense or semantically anomalous sentences are not completely devoid of meaning. Although the words in such sentences do not combine to express meaning at the sentence level, the individual words within semantically anomalous sentences still carry their own semantic content or meaning. The Chermak (1975) and Chermak et al. (1989) studies viewed their non-meaningful/

grammatical and non-meaningful/ungrammatical masking sentences as examples of speech that lacked meaning or semantic content, and they failed to clarify that meaning was in fact still present at the word level in both conditions. This distinction may help explain why Chermak and colleagues (1989) failed to find any significant differences between the masking effectiveness of the meaningful/grammatically correct, non-meaningful/grammatically correct, and non-meaningful/grammatically incorrect sentences. Perhaps the listeners' abilities to recognize meaningful words within each type of masking sentence resulted in an equivalent degree of difficulty in ignoring the three types of sentences. However, this explanation fails to account for the Chermak (1975) results, which showed that non-meaningful/grammatically correct sentences were significantly more difficult to ignore than the other two types of masking sentences. As such, additional research is needed to further investigate if and how masking effectiveness may be affected by whether a masking speech signal carries meaning at only the word level or at both the sentence and word levels.

Reversed speech. To create a masking speech signal that is truly devoid of meaning at any level, other studies have used reversed recordings of speech spoken in the listeners' native language. Reversal of a speech signal removes the intelligibility and semantic content of the signal but preserves the spectral information (Rhebergen et al., 2005). Previous studies have shown that under most test conditions, reversed competing speech is a less effective masker than meaningful, forward competing speech, at least when the competing speech is spoken in the listeners' native language. For example, Rhebergen and colleagues (2005) found that adult listeners' sentence reception thresholds (sSRTs) were significantly lower (i.e., better) when the masking signal consisted of

reversed speech in the listeners' native language as compared to when the masking signal consisted of forward, meaningful speech. Similar results were obtained by Sperry, Wiley, and Chial (1997) using monosyllabic target words presented with a masking signal consisting of forward six-talker babble or reversed six-talker babble. Word recognition performance was significantly better in the reversed six-talker condition than in the forward six-talker condition across the tested SNR range of -8 dB to +8 dB. In addition, Freyman and colleagues (2001) found that adult listeners' recognition of target nonsense sentences was considerably better when the masker was a "reversed" playback of nonsense sentences spoken by two masking talkers as compared to a "forward" playback of nonsense sentences spoken by two masking talkers.

Together, the results of these studies suggest that selective auditory attention tasks are generally easier when the masking signal consists of reversed speech as compared to when the masking signal consists of forward speech. However, the reasons that underlie this pattern of results remain unclear at this time. One possibility is that reversed speech is a less effective masker because it lacks meaning at both the sentence and word level; whereas, forward speech in the native language can carry meaning at both the sentence and word level (i.e., as in the case with forward, meaningful sentences) or only at the word level (i.e., as in the case of forward nonsense sentences). However, reversal of speech also changes the temporal envelope of the signal such that the rapid onsets and slow offsets typical of plosive consonants in forward speech become slow onsets and rapid offsets (Rhebergen et al., 2005). Therefore, it is possible that the rapid offsets that occur in reversed speech, but not in forward speech, may help the listener to differentiate

between the target and masking speech signals, thus reducing the masking effectiveness of competing speech played in reverse (Sperry et al., 1997).

In contrast to these findings, other studies have reported no significant differences between the masking effectiveness of forward competing speech signals and reversed competing speech signals. For example, Johnstone and Litovsky (2006) found that adult listeners' speech reception thresholds were essentially the same (i.e., no significant differences were noted between conditions) whether the competing signal consisted of forward, meaningful English sentences or reversed English sentences. The reason for the inconsistency between these results and the results of other studies remains unclear at this time, but differences between the target stimuli, competing stimuli, and test procedures used in each individual study may have been contributing factors. Therefore, additional research is needed to identify the conditions in which the meaning of the competing speech signal significantly affects selective auditory attention performance, as opposed to the conditions in which no effect of meaning occurs.

Speech in an Unfamiliar Language

To further investigate the effects of linguistic characteristics of masking speech signals, a small number of studies have compared listeners' selective auditory attention performance in conditions where the competing speech is spoken in the listeners' native language versus a language that is unfamiliar to the listeners. Speech in an unfamiliar language provides a masking signal that has the same basic time envelope as speech in the native language (i.e., with rapid onsets and slow offsets) but is lacking in semantic content, both at the sentence and word level. Therefore, use of speech in an unfamiliar language may provide a better control condition than nonsense sentences or reversed

speech when testing the effects of the semantic content of the masking speech signal. In addition, comparison of results obtained when competing speech is spoken in the native language versus an unfamiliar language may also reveal linguistic characteristics other than meaning that can influence the masking effectiveness of a competing speech signal.

At the present time, only a small number of studies have evaluated listeners' abilities to ignore speech spoken in different languages, and based on these results, the available studies can be divided into two groups. First, the results of two studies have shown that listeners experience greater difficulty selectively attending to target speech in their native language when the competing speech signal is also spoken in the native language as compared to when the competing speech is spoken in an unfamiliar language (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). Together with the evidence from studies using reversed masking speech, the findings of these two studies also seem to support the contention that the semantic content of the masking speech signal has a significant influence on selective auditory attention, such that competing speech that carries meaning is more difficult to ignore than competing speech that lacks meaning. However, this explanation may be overly simplistic, given that two other studies have reported no significant differences between listeners' abilities to recognize a target speech signal when the masking speech was spoken in the listeners' native language versus when the masking speech was spoken in an unfamiliar language (Freyman et al., 2001; Tun, O'Kane, & Wingfield, 2002).

The inconsistency between these results suggests that the masking effectiveness of a competing speech signal is not determined solely by its meaning (or semantic content). Instead, masking effectiveness may also be influenced by other linguistic

characteristics of the languages under comparison. For example, languages vary in the degree to which they either share similarities or differ in areas such as their lexical roots, consonant inventory, vowel inventory, syllable structure, stress patterns, and rhythmic structure (McLeod, 2007). The greater the similarities between the linguistic characteristics of two languages, the more likely it may be that both languages will provide a similar degree of masking effectiveness, even if one language is familiar (and therefore meaningful) to the listener and one language is unfamiliar (and therefore not meaningful) to the listener. Such similarities may explain why the studies by Freyman et al. (2001) and Tun et al. (2002) both showed no significant differences between English listeners' abilities to recognize a target speech signal in conditions where the masking speech was spoken in English, the listeners' native language, versus when the masking speech was spoken in Dutch, a language unfamiliar to the monolingual English-speaking listeners in each study. Previous research indicates that English and Dutch share certain linguistic characteristics. For example, both languages have variable stress patterns, Germanic lexical roots, a similar number of vowels (i.e., 15 for English and 16 for Dutch), and a stress-based rhythmic structure (Sebastián-Gallés et al., 2000). Therefore, in the studies by Freyman et al. (2001) and Tun et al. (2002), the similarities between English and Dutch may have caused the listeners to experience an equal level of difficulty in separating the target English speech from the masking speech signal, regardless of the language in which the masking speech was spoken.

In contrast, differences between the linguistic characteristics of languages may explain why the studies by Garcia Lecumberri and Cooke (2006) and Van Engen and Bradlow (2007) both found that monolingual English-speaking listeners experienced

greater difficulty in ignoring competing speech spoken in their native language (i.e., English) as compared to competing speech spoken in an unfamiliar language. Unlike the two studies that found no significant effect of the language of the masking speech signal, the studies by Garcia Lecumberri and Cooke (2006) and Van Engen and Bradlow (2007) each used competing speech spoken in an unfamiliar language that differed considerably from the listeners' native language. For example, Garcia Lecumberri and Cooke (2006) used speech spoken in English and Spanish as the two masking speech signals. Evidence suggests that these two languages differ in important linguistic areas, such as their lexical roots, number of vowels, and the rhythmic structure of the language. More specifically, English has Germanic lexical roots (Sebastián-Gallés et al., 2000), 18-19 vowels (Smit, 2007), and stress-based rhythm, whereas Spanish has Romance lexical roots (Sebastián-Gallés et al., 2000), only 5 vowels (Goldstein, 2007), and syllable-based rhythm. Similarly, English and Mandarin Chinese, the languages used as the competing speech in the study by Van Engen and Bradlow (2007), also exhibit different linguistic characteristics. For example, unlike English, Mandarin Chinese has 22 vowels (Hua, 2007), uses tones to indicate changes in meaning, and exhibits some characteristics consistent with syllable-timed rhythm (Grabe & Low, 2002).

Therefore, in both of these studies (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007), differences between the linguistic characteristics of the native language and the unfamiliar language (i.e., Spanish or Mandarin Chinese) may have made it easier for the listeners to separate the target native language speech (i.e., English) from the competing speech spoken in the unfamiliar language. Furthermore, linguistic differences between the language of the target speech (i.e., English) and the competing

speech (i.e., Spanish or Mandarin Chinese) may have also caused the competing speech spoken in the unfamiliar language to be more easily recognized as being “irrelevant,” thus making it easier for listeners to ignore. In contrast, when the target and competing speech signals were both spoken in English, overlap between the linguistic characteristics of each signal may have increased the difficulty of the listening task, either due to difficulty in separating the target and competing signals and/or difficulty in ignoring the competing speech.

Together, the results of these four studies suggest that linguistic factors other than meaning may influence the masking effectiveness of competing speech signals. More specifically, these studies indicate that the masking effectiveness of speech spoken in an unfamiliar language is greater when the unfamiliar language has linguistic characteristics similar to those of the listeners’ native language. On the other hand, the masking effectiveness of speech spoken in an unfamiliar language is weaker when the unfamiliar language has very different linguistic characteristics than those of the listeners’ native language. However, these conclusions are based solely on studies that compared listeners’ speech recognition performance in conditions where the competing speech was spoken in two languages: 1) the listeners’ native language and 2) one unfamiliar language. Therefore, it remains to be seen whether this pattern of results would occur if performance was compared in conditions where two or more unfamiliar languages were used for the competing speech signals. For example, would English-speaking listeners perform similarly in conditions where the competing speech was spoken in two unfamiliar languages that share linguistic similarities, such as Spanish and French? Comparing listeners’ abilities to ignore different types of unfamiliar languages could help

to identify which linguistic factors other than meaning may contribute to the masking effectiveness of competing speech signals. As such, additional research in this area could have important implications for the overall understanding of the cues that listeners use when selectively attending to speech in competing speech conditions.

Effects of Native Language Rhythmic Structure on Speech Processing

Results from the previously described studies provide important evidence regarding the cues that listeners use to separate and selectively attend to a target speech signal in the presence of speech from one or more competing talkers. In general, these findings indicate that the difficulty of a selective auditory attention task increases as the non-linguistic and/or linguistic similarities increase between the target and masking speech signal (e.g., Brungart, 2001; Freyman et al., 2004; Garcia Lecumberri & Cooke, 2006; Rhebergen et al., 2005). Together, studies of this nature help to answer the question of how the characteristics of the competing speech signal influence selective auditory attention performance. However, a listener's ability to process a target speech signal in competing speech conditions may be affected not only by the characteristics of the competing speech signal, but also by the listener's own characteristics, such as his/her age and language background. With regard to language background, one question of interest relates to how the linguistic characteristics of a listener's native language may influence the ways in which he/she processes a target speech signal in the presence of different types of competing speech signals. To date, few studies have addressed this question (e.g., Van Engen & Bradlow, 2007). However, there is ample evidence that linguistic characteristics of the native language affect the processing strategies that infant and adult listeners use when listening to speech in quiet. One factor that has been shown

to be particularly important for speech processing in quiet is the rhythmic structure of the listener's native language. Therefore, it is possible that native language rhythmic structure may also shape the ways in which listeners process a target speech signal in competing speech conditions. As such, it is important to consider the available evidence related to the rhythmic structure of different languages and the effects of native language rhythm on speech processing in quiet.

Early Views of Language Rhythm

For over 60 years, researchers in various fields have attempted to answer questions related to whether languages exhibit different rhythmic characteristics and if so, whether native language rhythm shapes the strategies that listeners use to process speech signals. With respect to spoken language, rhythm is a term used to refer to “the regularities which govern the grouping of elements in a language’s phonological structure” (Murty, Otake, & Cutler, 2007, p. 78). However, the issue of which phonological elements contribute to language rhythm remains a matter of much debate. Early researchers in this area (Abercrombie, 1967; Bloch, 1942; Han, 1962; Pike, 1946; all as cited in Grabe & Low, 2002) proposed that rhythm was determined by the type of speech units that were isochronous (i.e., of equal duration) within a particular language. Based on this definition, it was hypothesized that all languages of the world could be divided into three rhythmic categories based on whether the rhythm of the language was stress-timed, syllable-timed, or mora-timed. According to this “rhythmic class hypothesis,” the stress foot was the isochronous interval in stress-timed languages, meaning that stresses or rhythmic feet were separated by intervals of almost equivalent duration. In contrast, syllable-timed languages were thought to have isochronous

syllables, meaning that back-to-back syllables were of approximately equal duration. Finally, the isochronous units in mora-timed languages were said to be morae, which are “sub-units of syllables consisting of one short vowel and any preceding onset consonants” (Grabe & Low, 2002, p. 1). Therefore, mora-timed languages were thought to have successive morae that were of almost equal duration. According to these early views of rhythm, languages such as English, Dutch, German, and Russian were classified as stress-based languages; whereas, Spanish, French, and Italian were said to be examples of syllable-based languages (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi et al., 2000). Japanese, on the other hand, was considered to be the only member of the mora-based rhythmic category. Prompted by these early views of language rhythm, researchers have since conducted behavioral and acoustic-phonetic investigations to determine if such rhythmic classes of languages do in fact exist.

Behavioral Evidence

Discrimination between languages. Overall, results obtained through behavioral studies of infants and adults indicate that at least some languages can be classified according to their rhythmic structure and that exposure to the rhythmic properties of one’s native language influences how a listener processes speech. A key example of such evidence comes from behavioral investigations, which show that newborns from monolingual families can discriminate between languages that belong to different rhythmic classes (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998), but they fail to make the distinction between languages that fall within the same rhythmic category (Nazzi et al., 1998). For example, Mehler et al. (1988) reported that 4-day-old French newborns were able to discriminate between utterances spoken in French (a syllable-

based language) and those spoken in Russian (a stress-based language). To determine which acoustic information newborns use to make this distinction, Nazzi et al. (1998) presented French newborns with sentences that were low-pass filtered to reduce the available segmental information while preserving the rhythmic cues. Responses to a high-amplitude sucking technique suggested that the French newborns were able to make the between-class discrimination between sentences spoken in English (a stress-based language) and sentences spoken in Japanese (a mora-based language). However, the French newborns could not discriminate when the stimuli consisted of utterances from two stress-timed languages (English and Dutch). Thus, the newborns seemed to use “rhythmic information to classify utterances into broad language classes defined according to global rhythmic properties” (Nazzi et al., 1998, p. 756). These findings suggest that early sensitivities allow newborns to discriminate between languages from different rhythmic categories but not between those sharing the same rhythmic structure.

Behavioral evidence also indicates that from birth to 6 months of age, monolingual infants retain the ability to discriminate between languages belonging to different rhythmic categories. This pattern of results has been reported across infants from different linguistic backgrounds, different combinations of language stimuli, and different behavioral test methods. For example, 4-month-old Catalan (syllable-based) monolingual and Spanish (syllable-based) monolingual infants have shown significantly shorter visual orientation latencies when presented with passages spoken in their native syllable-based language as compared to the same passages spoken in English, a stress-based language (Bosch & Sebastián-Gallés, 1997). This pattern of findings was replicated by Nazzi and colleagues (2000) using an adapted version of the head-turn

preference procedure with 5-month-old American English-learning infants. In this study, the infants were successful in discriminating between pairs of languages from different rhythmic classes, including British English (stress-based) versus Japanese (mora-based) and Italian (syllable-based) versus Japanese (mora-based).

While these between-category discrimination abilities are maintained, exposure to the native language rhythm during the first 6 months of life results in developmental changes that allow infants to detect more subtle differences between languages, including some within-category distinctions. Overall, previous studies indicate that infants 4-5 months of age can make within-category language discriminations when the native language is contrasted with an unfamiliar language from the same rhythmic class (Bosch & Sebastián-Gallés, 1997, 2001; Nazzi et al., 2000). However, there is evidence that young infants fail to discriminate between two languages within the same rhythmic category when the maternal language is not included in the language pair (Nazzi et al., 2000). For example, 5-month-old American English-learning infants failed to discriminate between Italian and Spanish, two languages within the syllable-based rhythmic category. It could be argued that the infants were being raised in a monolingual environment consisting of stress-based English and were therefore unlikely to have been exposed to any syllable-timed languages. However, the infants also failed to discriminate between two languages from the native stress-based rhythmic class when both languages were unfamiliar (Dutch vs. German). Together, these findings indicate that the ability to discriminate between languages within the same rhythmic class develops as a result of infants' learning of the unique rhythmic features of their native language. As discussed in the following sections, these early sensitivities to language rhythm may result in the

development of speech processing strategies that are specific to the rhythmic structure of the listener's native language (Cutler, Mehler, Norris, & Segui, 1992).

Speech segmentation strategies. One body of evidence that supports the connection between native language rhythm and speech processing comes from studies that have compared the speech segmentation strategies used by infant and adult listeners from different language backgrounds. To understand speech in their native language, infants from all language backgrounds must first learn to segment continuous speech into individual words (Cutler & Otake, 1994). This task seems almost insurmountable, given that unlike adults, young infants do not have an existing store of familiar words to guide them in identifying the meaningful components of a speech signal. However, despite the daunting nature of the speech segmentation task, studies have shown that by approximately 7.5 months of age, infants are able to detect familiar words within continuous speech spoken in their native language (e.g., Jusczyk & Aslin, 1995). Based on the results of other experiments, it seems that infants may develop this ability to segment speech by using cues provided by the rhythmic structure of the language they are learning (i.e., the native language). For example, Houston, Jusczyk, Kuijpers, Coolen, and Cutler (2000) found that Dutch-learning and English-learning 9-month-old infants were able to segment continuous speech spoken in Dutch, as evidenced by the ability of both groups to detect Dutch words with a strong/weak stress pattern (i.e., stress on the first syllable). The authors hypothesized that the similar pattern of performance noted for Dutch-learning and English-learning infants was likely related to the stress-based rhythmic structure exhibited by both languages. Therefore, it may be that infants learning languages from the same rhythmic class use similar strategies to detect

individual words within a continuous speech stream, regardless of whether they are listening to their native language (i.e., as was the case for the Dutch infants) or to an unfamiliar language from the native language rhythmic class (i.e., as was the case for the English infants).

On the other hand, at least one study has shown that infants are not able to segment words from continuous speech spoken in an unfamiliar language belonging to a non-native rhythmic class. For example, Newman and Jusczyk (1998; as cited in Houston et al., 2000) familiarized English-learning 7.5-month-old infants with target words spoken in Mandarin Chinese and found that the infants failed to detect the familiarized words when they were embedded in continuous Mandarin Chinese speech. The stress-based rhythmic structure of English (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi et al. 2000), the infants' native language, differed from the syllable-based rhythmic structure of Mandarin Chinese (Grabe & Low, 2002), the language in which the speech stimuli were spoken. Therefore, the unfamiliar rhythmic structure of Mandarin Chinese may have interfered with the English-learning infants' abilities to detect individual words within the Mandarin Chinese speech signal.

Together, these two studies provide preliminary evidence that infants' exposure to the rhythmic properties of their native language may influence the type of speech segmentation strategies that they develop. As such, infants learning different languages from the same rhythmic class may develop similar speech segmentation strategies, which they can use to successfully detect words within speech spoken in their native language or within speech spoken in an unfamiliar language from the native rhythmic class (Houston et al., 2000). In contrast, the influence of the native language rhythmic

structure may prevent infants from being able to segment speech spoken in an unfamiliar language from a non-native rhythmic category (Newman & Jusczyk, 1998; as cited in Houston et al., 2000). However, additional research is needed to determine whether the same patterns of results would be found for infants from other language backgrounds and for speech stimuli spoken in other languages. More specifically, studies are needed which systematically compare infants' speech segmentation abilities as a function the rhythmic structure of their native language and the rhythmic structure of the language in which the target speech stimuli are spoken. Such studies could provide additional evidence regarding the role of rhythmic structure in infants' development of speech segmentation strategies.

In contrast to the limited number of studies involving infants, numerous studies have provided evidence suggesting that adult listeners' speech segmentation strategies are shaped by the rhythmic characteristics of their native language (e.g., Cutler & Otake, 1994; Cutler et al., 1983; 1986; Vroomen, van Zon, & de Gelder, 1996). Overall, such studies indicate that for languages from the same rhythmic class, native listeners use similar strategies for segmenting speech; whereas, native listeners of languages from different rhythmic classes employ different speech segmentation strategies. This general pattern of results has been found for adult listeners from different language backgrounds and for speech stimuli spoken in a variety of languages. For example, Cutler and colleagues (1983; 1986) found that native French listeners segment speech using syllabic strategies, meaning that they process continuous speech syllable-by-syllable. French listeners applied this syllable segmentation strategy to different types of speech signals, including speech spoken in French and speech spoken in English. In contrast, English

listeners participating in the study failed to show evidence of segmenting speech by syllables, regardless of whether the speech signal consisted of English words or French words. Instead, other studies have shown that English listeners segment speech based on stress patterns, such that the onset of word boundaries are perceived as occurring at the onset of strong (or stressed) syllables (e.g., Cutler & Butterfield, 1992; Cutler & Norris, 1988).

To explain these findings, Cutler and colleagues (1992) proposed that the differences in the processing strategies used by French and English listeners are a product of differences between the rhythmic characteristics of the two languages. For example, in the French language, boundaries between syllables are relatively clear, and the underlying rhythm of the language is believed to be syllable-based (Cutler, Mehler, Norris, & Segui, 1989). Therefore, the syllabic segmentation strategy used by French listeners corresponds to the underlying rhythmic properties of their native language and thus serves as the most effective method of processing speech spoken in French (Cutler et al., 1992). In contrast, English is characterized by somewhat ambiguous boundaries between syllables; therefore, speech spoken in English would not be efficiently processed using a syllabic segmentation strategy (Cutler et al., 1989). Instead, the strong/weak stress pattern typical of most lexical words in English and the overall stress-based rhythm of the language suggest that the stress-based segmentation strategies used by English listeners are a product of the underlying phonological properties of the English language (Cutler et al., 1992).

Evidence from studies of native listeners of other languages also supports the contention that the rhythmic structure of the native language shapes the ways in which

adult listeners segment continuous speech, both in their native language and in unfamiliar languages. For example, Cutler and Otake (1994) reported that Japanese listeners showed evidence of segmenting speech according to morae, the subsyllabic units that form the basis for the rhythmic structure of speech in Japanese. In fact, the Japanese listeners used mora-based segmentation, regardless of whether the speech signal consisted of their native language (Japanese) or an unfamiliar language from a non-native rhythmic category (English). In contrast, the English listeners participating in the study did not utilize mora-based segmentation when processing speech spoken in Japanese, an unfamiliar language from a non-native rhythmic category. Similarly, Otake, Hatano, Cutler, and Mehler (1993) found that French listeners also failed to segment Japanese speech according to morae but instead showed evidence of using syllable segmentation. In addition to these findings, studies have also shown that native Dutch listeners use speech segmentation strategies similar to those reported for native English listeners, with speech in the native language being segmented using a stress-based strategy in which strong syllables are viewed as markers for the onset of lexical words (e.g., Vroomen et al., 1996).

Together, these studies provide evidence that exposure to the rhythmic structure of the native language shapes the way in which listeners segment continuous speech. More specifically, native listeners of languages from the same rhythmic class (i.e., stress-based, syllable-based, or mora-based) tend to use the same types of speech segmentation strategies; whereas, native listeners of languages from different rhythmic classes differ in the ways in which they process continuous speech into individual words (e.g., Cutler & Butterfield, 1992; Cutler et al., 1983; 1986; Cutler & Otake, 1994; Otake et al., 1993;

Vroomen et al., 1996). For example, native listeners of English and Dutch, two languages with stress-based rhythmic structure, have been shown to segment speech based on the predominant stress pattern found in both languages: the presence of strong syllables at the onset of lexical words (e.g., Cutler & Butterfield, 1992; Cutler & Norris, 1988; Vroomen et al., 1996). In contrast, native listeners of French, a language with syllable-based rhythm, segment speech input syllable-by-syllable (e.g., Cutler et al., 1983; 1986); whereas, native listeners of Japanese employ mora-based segmentation, a pattern that correlates with the underlying mora-based rhythmic structure of the Japanese language (e.g., Cutler & Otake, 1994). Interestingly, evidence reveals that listeners segment continuous speech based on the rhythmic characteristics of their native language, regardless of whether the speech input is spoken in the native language or in an unfamiliar language from a non-native rhythmic class (e.g., Cutler & Otake, 1994). Therefore, results from speech segmentation studies provide further evidence that the linguistic properties of the native language, including its rhythmic structure, significantly shape how a listener processes different types of speech signals (e.g., speech spoken in the native language versus in unfamiliar language) in quiet listening conditions.

Adaptation to degraded speech. In addition to the effects on language discrimination and speech segmentation strategies, evidence also suggests that the rhythmic structure of a listener's native language can influence how speech is processed in degraded listening conditions. Over the course of a typical day, listeners are often faced with the challenge of understanding speech that has been degraded in some way. For example, a speech signal may be degraded by background noise, reverberation, distortion created by a microphone/amplifier system, or the process of speech

synthetization (e.g., the voices used by some automated phone services). Speech degradation may also result from talker-related characteristics such as the talker's accent, articulation errors, and his/her rate of speaking. Initially, such degradation may render the speech signal unintelligible to the listener. However, with continued exposure to the degraded signal, the auditory perceptual system adapts, and the degraded speech becomes more intelligible to the listener (Sebastián-Gallés et al., 2000). This type of perceptual adaptation effect has been demonstrated by studies using speech materials that have been time-compressed (i.e., artificially altered to increase the presentation rate). For example, studies in this area have shown that after listening to a short number of time-compressed sentences during a pre-test phase, the perceptual system adapts, and listeners are able to better understand time-compressed sentences presented during the test phase (Mehler et al., 1993; as cited in Sebastián-Gallés et al., 2000). In contrast, this perceptual adaptation does not occur when listeners hear uncompressed sentences or no sentences during the pre-test phase. Results from another study (Altmann & Young, 1993; as cited in Sebastián-Gallés et al., 2000) revealed that the advantage provided by prior exposure to time-compressed speech also occurred when the pre-test time-compressed signal consisted of nonsense words, thus suggesting that perceptual adaptation is not dependent upon the meaning of the pre-test speech stimuli.

In light of these findings and the hypothesis that some languages share similar rhythmic properties, one might predict that perceptual adaptation would also be more likely to occur when the pre-test degraded speech is spoken in a language belonging to the listeners' native language rhythmic category as opposed to a language from a non-native rhythmic category. To further investigate this prediction, Sebastián-Gallés and

colleagues (2000) compared native Spanish-speaking adults' abilities to understand 5 time-compressed test sentences in their native language after they had been exposed to 10 time-compressed pre-test sentences in one of six languages: Spanish, Italian, French, English, Japanese, and Greek. In each of the six experimental conditions, the listeners' responses were scored in terms of the percentage of syllables correctly recognized from the five test-phase sentences. Analysis of the results revealed that the listeners' recognition of the time-compressed Spanish test sentences was significantly better when the pre-test consisted of time-compressed sentences in Spanish, Italian, or Greek as compared to when the pre-test consisted of no sentences (i.e., the control condition) or time-compressed sentences in the other three languages (i.e., French, English, and Japanese).

These findings lead to several important conclusions. First, it seems that the linguistic characteristics of the adapting stimuli (i.e., the pre-test time-compressed sentences) have a significant effect on the extent to which listeners are able to adapt to and recognize time-compressed speech (Sebastián-Gallés et al., 2000). However, the adaptation effect does not appear to be dependent upon the meaning of the adapting stimuli, considering that the Spanish-speaking adults received a significant adaptation benefit not only from exposure to pre-test time-compressed sentences in Spanish, their native language, but also from exposure to pre-test time-compressed sentences in Italian and Greek, two languages that were not meaningful to the Spanish-speaking listeners. Instead, the results of the study suggest that sublexical factors such as the rhythmic structure and phonological characteristics of the pre-test sentences may significantly affect the degree of adaptation that occurs for the time-compressed test sentences. For

example, significant adaptation only occurred when the Spanish-speaking listeners were exposed to pre-test sentences in Spanish, Italian, and Greek, three languages that have been hypothesized to have a syllable-based rhythmic structure. In contrast, significant adaptation did not occur when the Spanish-speaking listeners were exposed to pre-test sentences spoken in English, a language with stress-based rhythm, or Japanese, a language with mora-based rhythm. Together, these results suggest that exposure to languages within the native language rhythmic class facilitates adaptation to time-compressed speech; whereas, exposure to languages from non-native rhythmic classes does not facilitate adaptation. Finally, in a more general sense, these findings may indicate that the degree of adaptation increases as the similarities increase between the linguistic characteristics of the pre-test and test speech signals.

However, unlike the findings for the other five pre-test languages, the results obtained in the French pre-test condition are not consistent with the idea that adaptation to degraded speech is influenced only by the rhythmic structure of the language in which the pre-test stimuli is spoken (Sebastián-Gallés et al., 2000). Although French is classified as a syllable-based language, like Spanish, Italian, and Greek, exposure to time-compressed pre-test sentences in French did not result in significant improvement in the listeners' abilities to understand time-compressed Spanish sentences during the test phase. Therefore, it seems that rhythmic structure alone cannot account for why adaptation to time-compressed native language speech is facilitated by exposure to some unfamiliar languages but not others. Sebastián-Gallés and colleagues (2000) proposed that the lack of an adaptation effect for French may be related to differences between the phonological characteristics of French and the other syllable-based languages used in the

study. For example, Spanish, Italian, and Greek each have 5 – 7 vowels and use a variable stress pattern, meaning that “prominence can be given to one of the last three syllables of content words” (Sebastián-Gallés et al., 2000, p. 841). In contrast, French contains 14 vowels and uses a fixed, word-final stress pattern, meaning that the stress is placed on the last syllable of all content words. Such differences may have contributed to the lack of adaptation observed for the French pre-test condition, but additional research is needed to determine the extent to which factors such as language rhythm, vowel repertoire, and stress pattern (e.g., variable versus fixed) contribute to a listener’s ability to adapt to a degraded speech signal.

Although these questions remain to be answered, the results obtained by Sebastián-Gallés and colleagues (2000) lead to the general conclusion that pre-test exposure to languages that are phonologically similar to the native language facilitates processing of degraded speech spoken in the native language, but this effect declines as the differences increase between the phonological characteristics of the pre-test language and the listeners’ native language. Therefore, it seems that listeners are sensitive to the linguistic properties of speech spoken in the native language and speech spoken in unfamiliar languages, and these sensitivities affect how listeners process speech, even when they are not listening to the language for content (as is the case for speech spoken in an unfamiliar language). As such, studies of listeners’ adaptation to time-compressed speech provide additional behavioral evidence related to how phonological characteristics of the native language, including rhythmic structure, can affect listeners’ overall speech processing in quiet listening conditions.

Acoustic-Phonetic Evidence

Overall, the previously described behavioral evidence suggests that languages within each of the three rhythmic classes (e.g., stress-based, syllable-based, and mora-based) share certain underlying linguistic properties and that these properties differ between languages from different rhythmic classes. Furthermore, exposure to the rhythmic structure of one's native language has a significant influence on speech processing, to the extent that native listeners of languages from the same rhythmic class process speech in similar ways; whereas, native listeners of languages from different rhythmic categories use different types of speech processing strategies. Thus, behavioral evidence seems to support the contention that at least some languages can be categorized as stress-based, syllable-based, or mora-based. However, questions remain regarding which linguistic elements actually differentiate the three rhythmic classes of languages.

In an attempt to answer these questions, investigators have compared the acoustic-phonetic properties of languages that have historically been classified as stress-based, syllable-based, or mora-based. As previously mentioned, the rhythmic class hypothesis was originally based on the idea that language rhythm was determined by the type of linguistic unit that occurs at regular intervals (i.e., isochronous units) within a particular language (Ramus et al., 1999). According to early views (Abercrombie, 1967; Pike, 1945; both as cited in Grabe & Low, 2002), the rhythm of stress-based languages was believed to be based on the approximately equal intervals between stresses; whereas, rhythm in syllable-based languages was thought to be derived from the approximately equal length of successive syllables. Similarly, the rhythm of mora-based languages was hypothesized to result from successive morae having approximately equal duration.

Although this speech isochrony proposal remained popular among linguists for many years, evidence from acoustic-phonetic studies fail to support the idea that language rhythm is derived from the regularity of linguistic timing units. For example, studies have shown that stress-based languages do not have equal or more regularly patterned intervals between stresses as compared to languages typically classified as having syllable-based rhythm (Dauer, 1983, 1987; Eriksson, 1991; Manrique & Signorini, 1983; Nakatani, O'Connor, & Aston, 1981; Pointon, 1980; Roach, 1982; Wenk & Wioland, 1982; all as cited in Grabe & Low, 2002). Furthermore, syllables and morae are not of equal duration in syllable-based and mora-based languages, respectively (Dauer, 1983, 1987; Pointon, 1980; Roach, 1982; Wenk & Wioland, 1982; all as cited in Grabe & Low, 2002).

Given the lack of evidence to support an isochronous basis for speech rhythm, alternative explanations have been proposed for the rhythmic classification of spoken languages. For example, Dauer (1983, 1987; both as cited in Grabe & Low, 2002) suggested that rhythm is based on the phonological properties of a language, including syllable structure, word stress, and whether or not vowel reduction is used. According to Dauer's account (1983, 1987; both as cited in Grabe & Low, 2002), stress-timed languages have more varied syllable types than syllable-based languages. Furthermore, vowel reduction (i.e., reduction of vowels within unstressed syllables) is found in stress-based languages but not in syllable-based languages. According to this view, rhythmic differences between languages occur because "these features combine with one another to give the impression that some syllables are far more salient than others in stress-timed languages, and that all syllables tend to be equally salient in syllable-timed languages"

(Ramus et al., 1999, p. 268). Dauer (1987; as cited in Ramus et al., 1999) also hypothesized that all languages fall along a continuum depending on the degree to which they exhibit the properties associated with stress-based versus syllable-based languages. Therefore, the more stress-based a language is, the fewer syllable-based qualities that language will possess. According to this continuum model, languages may be very stress-based, very syllable-based, or intermediate, with both stress-based and syllable-based qualities represented within the same language.

Most recently, two new acoustic-phonetic measurement techniques have been used in an attempt to identify reliable acoustic features that differentiate rhythmic classes of languages. For example, Ramus and colleagues (1999) analyzed speech samples from native speakers of eight languages (i.e., English, Polish, Dutch, French, Spanish, Italian, Catalan, and Japanese) to determine the duration of two types of intervals within each language: vocalic and consonantal intervals. A vocalic interval was defined as the length of time that elapsed between the beginning and the ending of a single vowel or a group of vowels; whereas, a consonantal interval was defined as the time that elapsed between the beginning and ending of a single consonant or a group of consonants. Results of these measurements were used to calculate three values for each sentence: the percentage of the total sentence duration that involved vocalic intervals (i.e., % V), the standard deviation of the duration measured for intervals of vowels/clusters of vowels (i.e., ΔV), and the standard deviation of the duration measured for intervals of consonants/clusters of consonants (i.e., ΔC).

Analysis of the results revealed a significant difference between the proportion of vocalic intervals (%V) measured for languages typically categorized as belonging to

different rhythmic classes (Ramus et al., 1999). Each rhythmic class of languages also differed significantly in the variability of consonantal intervals (ΔC). These results indicated a negative correlation between %V and ΔC values, with greater %V values associated with lower ΔC values and vice versa. No significant differences were noted between the ΔV values measured for each language class. These findings were interpreted as evidence that rhythmic differences between languages are related to differences in the diversity of syllable structures. Therefore, languages with more diverse syllable structures may also be more likely to have “greater...variability in the number of consonants and in their overall duration in the syllable, resulting in a higher ΔC ” (i.e., the standard deviation of the consonantal interval durations) (Ramus et al., 1999, p. 273). Furthermore, as the diversity of syllables increases, the proportion of consonants usually increases, thus resulting in a decreased % V (i.e., the percentage of the total sentence duration that involved vocalic intervals). For example, English uses full vowels and reduced vowels which results in a smaller % V than what is seen in French, a language that does not use vowel reduction (Grabe & Low, 2002). In contrast, English exhibits a higher ΔC than French due to the more diverse and complex syllable structures found in the English language.

Together, the findings reported by Ramus and colleagues (1999) provide evidence of acoustic properties that differentiate languages belonging to the three standard rhythmic classes: stress-based, syllable-based, and mora-based. However, it is important to keep in mind that these results were based on measurement of the vocalic and consonantal intervals found in only eight languages. Therefore, it is currently unknown

whether other languages can be reliably categorized into one of the three standard rhythmic classes.

Acoustic evidence for the rhythmic categorization of languages has also been provided by Grabe and Low (2002). Similar to Ramus and colleagues (1999), Grabe and Low (2002) analyzed speech samples from different languages by measuring the duration of vocalic intervals and intervocalic intervals (i.e., essentially the same as the consonantal intervals measured by Ramus et al., 1999). Speech samples from 18 different languages were analyzed, including languages typically categorized as stress-based (e.g., British English, German, Dutch, and Thai), syllable-based (e.g., Tamil, Spanish, French, and Singapore English), mora-based (e.g., Japanese), and languages that have not been previously classified (e.g., Greek, Mandarin, and Romanian). For each language, a Pairwise Variability Index (PVI) was calculated separately for vocalic intervals (i.e., the duration of vowels) and for intervocalic intervals (i.e., the duration of intervals occurring between successive vowels) to determine the variability in duration between adjacent pairs of each type of interval.

Analysis of the results revealed that high vocalic nPVI values (i.e., indicating greater variability in the duration of adjacent pairs of vocalic intervals) were associated with English, Dutch, and German, three languages typically classified as exhibiting stress-based rhythm (Grabe & Low, 2002). Low vocalic nPVI values (i.e., indicating less variability between the duration of adjacent pairs of vocalic intervals) were found for French and Spanish, two languages typically placed in the syllable-based rhythmic category. However, low vocalic nPVI values were also noted for Japanese, a language typically categorized as having mora-based rhythmic structure. Although differences in

vocalic nPVI values differentiated between some stress-based and syllable-based languages, no consistent differences were noted between intervocalic PVI values (i.e., the variability between adjacent pairs of intervocalic intervals) for languages belonging to these two rhythmic categories.

Together, the results reported by Grabe and Low (2002) provide acoustic evidence of the distinction between German, English, and Dutch as stress-based languages and French and Spanish as syllable-based languages. However, the results obtained for speech spoken in Japanese fail to support the idea that Japanese exhibits its own unique rhythmic structure, separate from that of stress-based and syllable-based languages. In addition, the acoustic-phonetic results obtained for the other 12 tested languages did not neatly align with either the stress-based or the syllable-based languages. Instead, results for these languages were spread across the continuums for both vocalic and intervocalic PVI values. This lack of a clear delineation between some languages provides further evidence that languages can vary in the degree to which they exhibit classic stress-based or syllable-based rhythm.

Despite questions as to whether all languages fall into the three rhythmic categories, existing behavioral and acoustic-phonetic evidence indicates that at least some languages have rhythmic properties consistent with specific classes (e.g., English, German, Spanish, French), and as early as infancy, listeners seem to detect these rhythmic differences between languages (e.g., Grabe & Low, 2002; Nazzi et al., 1998; Ramus et al., 1999). Behavioral studies of infants and adults further suggest that these sensitivities to language rhythm play a critical role in how listeners process speech in quiet listening conditions, such that native listeners of languages from the same rhythmic

class tend to process speech in similar ways; whereas, native listeners of languages from different rhythmic classes generally use different types of speech processing strategies (e.g., Cutler et al., 1983; 1986; Cutler & Otake, 1994; Houston et al., 2000; Otake et al., 1993; Vroomen et al., 1996). Together, these findings lead to the general conclusion that the rhythmic structure of the native language serves as a type of filter through which listeners process all incoming speech signals, regardless of the language in which the speech input is spoken (Sebastián-Gallés et al., 2000). This conclusion is largely based on results from studies of how listeners process speech in quiet. Therefore, additional research is needed to investigate whether the relationship between native language rhythmic structure and speech processing extends to conditions in which listeners must process speech from two or more simultaneous talkers.

Goals and Hypotheses

Overall, results from previous studies of selective auditory attention lead to the conclusion that a competing speech signal generally provides a greater degree of masking as the similarities increase between characteristics of the target and competing speech signals (e.g., Brungart, 2001; Freyman et al., 2004; Garcia Lecumberri & Cooke, 2006; Rhebergen et al., 2005). However, this conclusion has been based primarily on evidence related to effects of non-linguistic characteristics of target and competing speech signals, such as differences between the spatial location (e.g., Freyman et al., 2004), intensity (e.g., Cooke et al., 2008), number (e.g., Simpson & Cooke, 2005), and gender (e.g., Brungart, 2001) of the target and masking talkers. In contrast, few studies have investigated how selective auditory attention performance is affected by: 1) differences between the vocal/speaking characteristics of the masking talkers and 2) the language in

which the competing speech is spoken. For example, only one previous study has evaluated how selective auditory attention is affected by individual variability in the vocal and/or speaking characteristics of masking talkers (Freyman et al., 2007). Similarly, only a small number of studies have specifically investigated the masking effectiveness of speech spoken in different languages (e.g., Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007), and to date, no previous study has compared listeners' speech recognition performance in conditions where the competing speech was spoken in more than two languages. Previous behavioral and acoustic-phonetic studies have shown that some languages differ in their rhythmic structure and that these differences have a significant effect on how a listener processes speech in quiet listening conditions (e.g., Cutler & Otake, 1994; Cutler et al., 1983; 1986; Grabe & Low, 2002; Nazzi et al., 2000; Ramus et al., 1999; Sebastián-Gallés et al., 2000). As such, a question of primary interest relates to if and how the rhythmic structure of the language of a competing speech signal may affect a listener's ability to selectively attend to and process a target speech signal. In an attempt to answer this question, the current study compared monolingual English-speaking adult listeners' abilities to selectively attend to meaningful English sentences in conditions where the competing speech was spoken in a stress-based language (English or German), a syllable-based language (Spanish or French), or a mora-based language (Japanese) (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi et al., 2000). The primary goals of the study were to determine how monolingual adult listeners' selective auditory attention performance was affected by the:

- Semantic content (i.e., meaning) of the competing speech,

- Rhythmic structure of the language in which the competing speech was spoken, and
- Individual differences between the vocal/speaking characteristics of different pairs of masking talkers of the same language.

Based on previous research findings, the following hypotheses were proposed prior to initiation of the current study:

- Within each of the five languages, no significant difference would be seen between the masking effectiveness of two different pairs of masking talkers (e.g., English two-talker Masker A versus English two-talker Masker B, German two-talker Masker A versus German two-talker Masker B, etc.).
- Listeners would experience greater difficulty attending to the target English sentences in conditions where the competing speech was spoken in English as compared to conditions where the four unfamiliar languages were used as the competing speech (i.e., German, Spanish, French, and Japanese).
- Listeners would experience greater difficulty attending to the target sentences in conditions where the competing speech was spoken in an unfamiliar language from the native stress-based rhythmic category (i.e., German) as compared to conditions where the competing speech was spoken in one of the three unfamiliar languages from non-native rhythmic categories (i.e., Spanish, French, and Japanese).
- Listeners' selective auditory attention performance would be similar between the conditions where the competing speech was spoken in an unfamiliar

language from a non-native rhythmic class (i.e., Spanish, French, and Japanese).

Chapter II

Methods

Subjects

Results from 50 monolingual English-speaking adult females were analyzed in the current study. Subjects ranged from 20 to 44 years of age, with a mean age of 23.72 years. All subjects were recruited for participation using direct person-to-person solicitation and posted notices of solicitation (see Appendix A). To participate in the study, each subject was required to meet all of the following criteria:

- Sign an informed consent form (see Appendix B),
- Be 18 to 55 years of age,
- Have English as his/her first language, with little to no familiarity with any languages other than English,
- Be at least a high school graduate,
- Have no history of a diagnosed permanent hearing loss, auditory processing disorder, language disorder, learning disability, and/or dyslexia, and
- Pass a hearing screening at 15 dB HL in each ear at 250-8000 Hz.

A total of 12 subjects were excluded from the study, including 8 subjects who failed to meet one or more of the inclusion/exclusion criteria. Results from 4 additional subjects were also excluded, including 3 subjects who were unable to complete the testing during a single session due to scheduling conflicts and 1 subject who failed to comply with the test protocol (i.e., was text messaging during the testing). Results from the remaining 50 subjects were included in the data analysis. Previous studies investigating monolingual adult listeners' selective auditory attention abilities have included sample sizes ranging

from 8 (Rhebergen et al., 2005) to 66 subjects (Van Engen & Bradlow, 2007). Therefore, the current study's sample size of 50 approximates the maximum number of subjects included in previous studies in this area.

Each subject received a small decorative sign as a token of appreciation for his/her participation in the study. All signs were handmade by the investigator. A decorative sign was given to each of the 62 subjects that enrolled in the study, regardless of whether they completed the testing. Subjects did not receive any additional compensation for participating in the study.

Pre-Test Procedures

Prior to enrollment in the study, each subject signed an informed consent form (see Appendix B). All subjects included in the study then completed the questionnaire, underwent a hearing screening, and were tested in each of the 22 conditions described in the following sections.

Questionnaire

After signing the informed consent form, each subject was required to complete a brief questionnaire prior to any testing (see Appendix C). The investigator read Questions 1 – 9 aloud to the subject and recorded his/her responses on the questionnaire. The subject was then given the questionnaire and asked to read and answer Questions 10 and 11. The questionnaire addressed topics such as the subject's age, educational degree(s) obtained, exposure to/experience with languages other than English, and history of any disorders that might affect his/her speech recognition performance. With regard to language experience, subjects were asked to: 1) list any languages other than English that they could speak and/or understand and 2) specify when/where these non-English

languages were learned. If an individual had experience with another language, he/she was then asked to respond to the following two statements included on the questionnaire:

***10. Please circle the number of the sentence that best describes your fluency in speaking a language other than English.*

- 1 My speech is so halting and fragmentary that conversation is virtually impossible.*
- 2 My speech is usually hesitant, and I often give up due to language limitations. I mainly give one or two word utterances.*
- 3 My speech in conversation is frequently disrupted by my need to search for the correct manner of expression. I use short phrases and sentences.*
- 4 My speech in conversation is generally fluent, with occasional lapses while I search for the correct manner of expression.*
- 5 My speech in conversation is fluent and effortless, approximately like that of a native speaker.*

***11. Please circle the number of the sentence that best describes your ability to understand a language other than English.*

- 1 I understand little or no conversation.*
- 2 I have great difficulty following what is said. I can comprehend only conversation spoken slowly and with frequent repetition and rephrasing.*
- 3 I understand most of what is said in conversations at slower than normal speed with some repetition and rephrasing.*
- 4 I understand conversation, although occasional repetition and rephrasing may be necessary.*
- 5 I understand conversation in a variety of settings.*

Questions 10 and 11 of the subject questionnaire were adapted from the Minnesota Modified Student Oral Language Observation Matrix (MN-SOLOM) (Minnesota State Department of Education [MSDE], 2003; as cited in MSDE, n.d.) which is an adaptation of the Student Oral Language Observation Matrix (SOLOM) (Ortiz,

1995; as cited in von Hapsburg & Peña, 2002). The SOLOM and the MN-SOLOM are both informal rating scales designed to allow teachers to evaluate the English oral language proficiency of students for whom English is a second language. Although not a standardized measure, the SOLOM has been widely used in states such as California and Minnesota to supplement the information gained through standardized tests of language proficiency (MSDE, n.d.).

Each subject's responses to Questions 10 and 11 were used to determine whether or not the subject met the inclusion criterion of "little to no familiarity with any languages other than English." Only native English-speakers who scored a 1 or a 2 on Questions 10 and 11 of the questionnaire were included in the study. Individuals who scored a 3 or greater on Question 10 and/or Question 11 were excluded from participating. This criterion helped ensure that the monolingual English-speaking adults participating in the study would only be able to understand the competing speech spoken in English and not the competing speech spoken in the other four languages (i.e., German, Spanish, French, and Japanese). As such, the meaning of the competing speech could be ruled out as a contributing factor underlying the subjects' performance in the German, Spanish, French, and Japanese competing speech conditions.

In addition to language experience, the questionnaire also addressed whether the subject had ever been diagnosed with a permanent hearing loss, auditory processing disorder, language disorder, learning disability, and/or dyslexia. The sentence reception threshold (sSRT) task used in the current study required subjects to attend to target sentences spoken in English while simultaneously ignoring continuous speech playing in the background. As such, successful performance on the task required a combination of

auditory and language related skills. Previous studies have shown that individuals with hearing loss, some forms of auditory processing disorder, language, and/or learning problems perform more poorly on listening-in-noise tasks than listeners with normal skills in these areas (e.g., Crandell, Henoch, & Dunkerson, 1991; as cited in Crandell & Smaldino, 2002; Crandell, Smaldino, & Flexer, 1995). Therefore, individuals were excluded from the current study if they had ever been diagnosed with a disorder in one or more of these areas.

Hearing Screening

To further ensure that hearing sensitivity did not affect the study's results, all subjects were required to pass a pure-tone hearing screening prior to being tested in the experimental conditions. Otoscopy was performed using a Welch Allyn Pocketscope otoscope to identify any abnormalities of the ear canal or eardrum which might affect the subject's performance on the target sentences. For the pure-tone hearing screening, the subject was seated in a sound treated booth wearing TDH 50P Telephonics headphones or E-A-R Tone 3A insert earphones. A GSI 61 audiometer was used to present pure-tone signals at 15 dB HL at the following frequencies: 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz. To pass the hearing screening, the subject was required to respond consistently to the tones presented at 15 dB HL at each of the target frequencies (i.e., 250 – 8000 Hz) in each ear. Hearing screening results for each subject were recorded on the screening form shown in Appendix D. Subjects who did not pass the hearing screening were counseled on the results and given information on where to receive a complete audiological evaluation. These subjects were not required to complete the remainder of the test session and were therefore excluded from the data analysis.

Test Stimuli

Target Sentences

Subjects who met all inclusion criteria were then tested in the 22 conditions used in the current study. In each test condition, the target speech signal consisted of meaningful English sentences from the Hearing in Noise Test (HINT) (Nilsson, Soli, & Sullivan, 1994; see Appendix E for sample HINT sentence lists). Lists of single words are often used in clinical practice for speech recognition and speech reception threshold (SRT) testing. However, lists of meaningful sentences were selected as the target stimuli for the current study in order to more closely approximate the “normal spectral weighting, level fluctuations, intonations, [and] pauses...associated with conversational speech” (Nilsson et al., 1994, p. 1086).

Of the commercially available sentence lists used for audiological testing, the HINT best met the needs of the current study for several reasons. First, the HINT sentences are meaningful, contain six to seven syllables, are written in General American English, and have been judged to represent approximately a first grade reading level (Nilsson et al., 1994). As such, the HINT sentences were deemed appropriate for the monolingual American English-speaking adults targeted for the study. In addition, the HINT consists of 250 different sentences divided into 25 lists of 10 sentences, with all lists reportedly equated for “naturalness, length, and intelligibility” (Nilsson et al., 1994, p. 1095). The phoneme distribution is balanced across all 25 HINT sentence lists, and on the commercially available compact disc (CD) recording of the test, all sentences are spoken by the same male professional voice actor. The availability of 25 equivalent sentence lists, each consisting of 10 different sentences, was sufficient for a different

sentence list to be used in each of the study's 22 test conditions. Use of a different sentence list in each condition reduced the likelihood of practice effects which could occur if the same lists were repeated between conditions. Finally, the HINT offered the advantage of having been specifically developed for adaptive sSRT measurement, the same assessment technique used in each condition of the current study. Adaptive sSRT procedures are used to determine the lowest (i.e., softest) intensity at which a listener can correctly recognize sentences approximately 50% of the time (Nilsson et al., 1994). Like other speech recognition tests, adaptive sSRT measurement can be performed in quiet or in background noise. However, the sSRT is unique in that it is not limited by the floor and/or ceiling effects which can occur when speech recognition testing is administered at a single intensity (i.e., to obtain a percent words/sentences correct score).

A closer look at data obtained during development of the HINT provides additional evidence regarding the equivalency and sensitivity of the HINT sentence lists. For the current study, such information was used to select the HINT sentence lists that would be most likely to produce reliable, equivalent results and to determine the number of HINT sentence list measurements that would be necessary to achieve adequate test sensitivity. With regard to list equivalency and reliability, the developers of the HINT reported that the 25 sentence lists were carefully selected and evaluated to ensure that all were "normed for naturalness, difficulty, and reliability" (Nilsson et al., 1994, p. 1086). For example, the reported test development data indicates that no significant differences were found between listeners' performance on the 25 lists, and with the exception of List 4, the means for all lists were found to be within 1 dB of the overall mean. Based on these findings, List 4 was excluded from use in the current study. In addition, within

subject comparisons of repeated measurements of HINT sSRTs revealed that use of one 10-sentence HINT list for each sSRT measurement allowed for the detection of threshold differences of ± 2.98 dB in quiet and ± 2.41 dB in noise. However, by averaging the results obtained for two 10-sentence HINT lists, the sensitivity of the test improved such that threshold differences of ± 1.94 dB were detected in quiet, and differences of ± 1.49 dB were detected in noise. Use of a third HINT sentence list for each measurement only resulted in a minimal increase in sensitivity (i.e., ± 1.41 dB and ± 1.20 dB in quiet and in noise, respectively). Based on these findings, performance on two 10-sentence HINT lists were averaged for the quiet condition and for each of the 10 types of competing conditions (i.e., Masker A and Masker B conditions for each of the five languages) in order achieve the best balance between test sensitivity and efficiency of testing in the current study.

Cool Edit Pro Version 2 (2002) audio editing software was used to adapt the original HINT recordings for the purposes of the current study. Each list was first edited to remove the speech noise saved to the second channel of the original HINT list recordings. The recording of each HINT list in quiet (i.e., without the speech noise) was then saved to a Dell Optiplex 745 desktop computer as a one-channel *.wav file. By listening to each of the HINT list recordings, it was discovered that List 6 was missing from the original HINT CD, and significant distortion could be heard in the original recording of List 21. As such, List 6 and List 21 were excluded from use in the current study. As previously mentioned, List 4 was also excluded based on data reported by the developers of the HINT (Nilsson et al., 1994).

The remaining 22 HINT list recordings were then edited to increase the length of the silence that occurred between each sentence and to equalize the overall intensity of the sentence lists. The original HINT recordings included 4.5 seconds of silence between each sentence. However, the user manual of the HINT (Nilsson, McCaw, & Soli, 1996) recommends that the examiner manually pause the CD after each sentence in the list. Therefore, the actual length of time between the presentation of each sentence could vary. To achieve adequate control in the current study, pilot testing was conducted to determine the length of time between sentences that would be sufficient for subjects to respond to the target stimuli and for the investigator to record the response and appropriately adjust the presentation level of the target sentences. Based on results of the pilot testing, the Cool Edit Pro Version 2 (2002) audio editing software was used to insert a 20-second segment of silence between each sentence. Twenty seconds of silence was also added to the beginning of each HINT list, so that the first sentence was always preceded by an equal amount of silence. As such, in the current study, the same period of time elapsed between all sentences within each HINT list, for all conditions and across all subjects. Therefore, the subjects' performance should not have been affected by variations in the amount of time that passed between presentation of the target sentences. After increasing the period of silence between each sentence, the Cool Edit Pro Version 2 (2002) audio editing software was then used to normalize the recording of each edited HINT list to 96% of peak value in order to ensure that all lists had the same overall root mean square (RMS) values.

Competing Speech

The current study compared subjects' selective auditory attention performance in conditions where the competing speech was spoken in one of five languages, including English, German, Spanish, French, and Japanese. These languages were selected based on previous behavioral and acoustic-phonetic evidence that supports the categorization of English and German as stress-based languages, Spanish and French as syllable-based languages, and Japanese as a mora-based language (e.g., Culter & Otake, 1994; Grabe & Low, 2002; Nazzi et al., 1998; Nazzi et al., 2000; Otake et al., 1993; Ramus et al., 1999). By using competing speech spoken in languages from different rhythmic categories, the current study was able to evaluate effects of the rhythmic structure of the competing speech on selective auditory attention performance.

In each competing speech test condition, the masking signal consisted of speech from two native speakers (i.e., one male and one female) of the masking language assigned to that condition. To more closely approximate real-world listening environments, all masking signals used in the current study consisted of natural, continuous speech, as opposed to the lists of unrelated meaningful or nonsense sentences used in some previous studies (e.g., Freyman et al., 2001; Johnstone & Litovsky, 2006; Rhebergen et al., 2005; Van Engen & Bradlow, 2007). The competing speech signals used in the current study were created using speech from adult native speakers of English, German, Spanish, French, and Japanese. For each of the five languages, male and female native speakers were recruited from the Lubbock area using flyers (see Appendix F), e-mails, and face-to-face solicitation. A person was considered to be a "native" speaker if he/she had spoken the target language from birth. Individuals who

learned the target language as their second language were excluded from participating. All talkers for the language recordings were 18 years of age or older (age range: 23 to 56 years; mean age: 32.45 years) and had no history of a speech or language disorder. Each participant completed a brief questionnaire, which included questions related his/her language background (see Appendix G). Native speakers were not paid for making the recordings, but each speaker was given a gift bag that included homemade snacks and other small tokens of appreciation (e.g., note cards, pens, pencils).

All recordings used as the competing speech were made in a sound treated booth in the TTUHSC Department of Speech, Language, and Hearing Sciences using an Isomax headset microphone (i.e., a headset/boom microphone combination) and a TASCAM DA-P1 digital audio tape (DAT) recorder. For each recording, the participant was seated in the sound treated booth wearing the boom microphone positioned approximately 2- 4 inches from his/her mouth. The investigator adjusted the recording sensitivity to be within the target range of +20 to +12 dB throughout each recording in order to compensate for any slight differences between the microphone positioning for different talkers. All recordings were made using a sampling rate of 44.1 K Hz. Each recording was later uploaded and saved to a desktop computer as a one-channel *.wav file using the Cool Edit Pro Version 2 (2002) audio editing software.

For the recording, each talker read for approximately 20 minutes. Talkers were instructed to read the text as smoothly as possible, using a natural speaking rate, and with minimal pausing between sentences. Talkers were also instructed to speak using an even, natural intonation pattern, avoiding the use of monotonous or overly dramatic intonation. All text used for the recordings met the following criteria:

- Was written in the native language at approximately a seventh grade reading level or higher,
- Included content that would be appropriate for child or adult listeners (i.e., no mature content),
- Contained no dialogue between characters,
- Contained no references to well-known American proper names, and
- Was of sufficient length for the talker to read for approximately 20 minutes without repeating any text.

Ideally, translations of the same text would have been used for the recordings for all five of the competing languages. However, at the time of the current study, no text could be located which: 1) met the previously described criteria and 2) was available in English, German, Spanish, French, and Japanese translations. As a result, for the German, French, and Japanese recordings, each talker read continuously from a different passage written in their native language, but all passages of text met the previously described general criteria related to reading level and content. Therefore, the passages used for the German, French, and Japanese recordings were not translations of the text used for the English and Spanish recordings. The types of text used for the recordings in German, French, and Japanese included passages from novels (without dialogue), biographies, self-help books, encyclopedias, newspapers, and non-professional journals (e.g., topics related to travel, culture, and politics in the native country).

Unlike the other three languages, a direct translation of the same text was used for the recordings made in English and Spanish. Although subjects in the current study were limited to English-speaking adults with little to no experience with any other languages, it

is anticipated that a group of Spanish-English bilingual adults will be tested in a follow-up study. As such, equivalency between the content and sentence structure of the English and Spanish text was deemed essential since these two competing languages would be meaningful to the monolingual subjects participating in the current study (i.e., only the English was meaningful) and/or the bilingual subjects that will participate in the anticipated follow-up study (i.e., both the English and Spanish will be meaningful). To ensure that the content and sentence length/structure of the English and Spanish competing speech signals were approximately equivalent, corresponding passages were selected from the Spanish translation and the English translation of the same bilingual education textbook. The native English speakers each read a different passage from the English translation of the textbook: *Teaching Reading and Writing in Spanish and English in Bilingual and Dual Language Classrooms* (Freeman & Freeman, 2006). Each native Spanish speaker then read the corresponding passage from *La Enseñanza de la Lectura y la Escritura en Español e Inglés en Salones de Clases Bilingües y de Doble Inmersión* (Freeman & Freeman, 2007), the Spanish translation of the same bilingual education textbook.

The investigator reviewed the English text and selected passages that were conducive to oral reading. Each passage was then edited by the investigator (a native speaker of English) to remove or change any wording which might cause difficulties during oral reading, including:

- References,
- Lists,
- Individual phonemes or syllables,

- Proper names of people and places,
- Terms that might be difficult for the average person to read aloud smoothly (e.g., sociopsycholinguistic view of reading), and
- Abbreviations or acronyms (e.g., SSR).

A Spanish-English bilingual individual reviewed the Spanish version of the selected passages and ensured that the changes made to the English text were also made to the Spanish text. As such, the passages read for the Spanish recordings were a direct translation of the passages read for the English recordings.

The edited English text was then divided into four sections so that no two English masking talkers read the same text. In the same way, the edited Spanish text was divided into the same four sections so that each Spanish masking talker read a different passage of text. The passages of text were divided among the masking talkers so that each of the four English masking talkers read the English translation of the Spanish text read by their Spanish counterpart. For example, the English Masker A female read the English translation of the text read by the Spanish Masker A female. Text assignments for the English and Spanish masking talkers are shown in Table 1.

Recordings of 52 native speakers were made for the current study. For each of the five languages, the two male and two female recordings that best met the study's criteria were selected (e.g., speakers who read most fluently, read with relatively even intonation and even speaking rate, and who read text that met the criteria for reading level and content). As such, recordings from a total of 20 different talkers were selected for use in the study (i.e., 2 male and 2 female talkers for each of the five languages).

Information obtained from the language questionnaire shown in Appendix G was used to

ensure that for each language, the 4 native speakers selected for the study used the same general dialect of their native language when reading the text for their recording. With assistance from a fluent speaker of the target language, each selected recording was then edited using the Cool Edit Pro Version 2 (2002) audio editing software in order to:

- Remove sentences where the talker produced a marked dysfluency, such as stumbling, clearing his/her throat, laughing, or coughing,
- Remove any sentences where the talker used a recognizable American proper name,
- Remove any sentences from the non-English recordings where the talker used a recognizable English word, and
- Select the best continuous section of the recording (e.g., sections where the talker read most fluently with the fewest errors).

As a native speaker of English, the investigator ensured that the editing procedures did not violate the boundaries of surrounding sentences within each of the four selected English recordings. To maintain the same control within the German, Spanish, French, and Japanese recordings, all editing of the non-English recordings was performed with assistance from a fluent speaker of the language being edited. For each non-English recording, the fluent speaker listened to the entire recording and used the hard copy of the text to identify the exact beginning and ending of all sentences that contained a marked dysfluency, a recognizable American proper name, and/or a recognizable English word. The investigator then used the Cool Edit Pro Version 2 (2002) audio editing software to remove these sentences from each recording. For the English and Spanish recordings, all sentences removed from the recording in one

language were also removed from the recording in the other language to ensure that the content remained the same between the recordings.

After selecting the best continuous section, the investigator edited the 20 recordings to ensure that all pauses were approximately equivalent within and across recordings. For the current study, the criterion for the maximum allowable pause time was determined based on previous evidence reported by Smith (2005) related to the duration of pauses that typically occur between sentences when text is read aloud. In a study of three English speakers and two French speakers, Smith (2005) measured the duration of the pauses that each talker made between sentences when reading text aloud in their native language. For consecutive sentences that continued or elaborated on the same topic, mean pause times between sentences were 551.5, 703.5, and 940.5 milliseconds (ms) for the three English speakers, as compared to 738 and 957 ms for the two French speakers (C. Smith, personal communication, July 21, 2008). For the purpose of the current study, shorter pauses between sentences were desirable in order to minimize the amount of silence that occurred within the competing speech recordings. Therefore, the criterion for the maximum allowable pause duration within the competing speech recordings was set at 600 to 700 ms, based on the shorter mean pause times reported for the speakers in the Smith (2005) study. Using this criterion, the investigator used the Cool Edit Pro Version 2 (2002) audio editing software to measure the pauses within each of the 20 recordings. All pauses that exceeded 700 ms were shortened to 600 to 700 ms in length. After editing the dysfluencies and pauses, each recording was then normalized to 96% of peak value to ensure that all 20 recordings had equivalent RMS values.

The normalized language recordings were then used to create two different male-female babble recordings for each of the five languages, with each two-talker babble recording consisting of a different male and a different female masking talker. For each language, one male recording (Male 1) was paired with one female recording (Female 1). Using the Cool Edit Pro Version 2 (2002) audio editing software, the Male 1 and Female 1 recordings were then digitally mixed to create the first two-talker babble signal for each language (i.e., English A, German A, Spanish A, French A, and Japanese A). The remaining male recording (Male 2) and the remaining female recording (Female 2) in each language were digitally mixed to create the second two-talker babble signal for each of the five languages (i.e., English B, German B, Spanish B, French B, and Japanese B). Table 2 shows which individual masking talkers were assigned to Masker A and Masker B for each of the five languages. Each two-talker mixed recording was normalized to 96% of peak to ensure that the overall RMS value of each mixed signal was equivalent. Each normalized two-talker recording was then saved as a single-channel *.wav file. To control for potential individual variability between vocal/speaking characteristics, each subject was tested in both Masker A and Masker B competing conditions for each language.

Assignment of Target Lists to Test Conditions

In the current study, each subject was tested twice in a quiet condition (i.e., no competing speech) and twice in each of the 10 types of competing speech conditions (i.e., one Masker A and one Masker B condition for each of the five languages). Therefore, testing was performed in a total of 22 different conditions for each subject. A different HINT sentence list was randomly assigned to each of the 22 test conditions, such that the

same HINT list was always presented with the same randomly selected test condition.

Table 2 shows the specific HINT list assigned to each of the 22 test conditions.

Mixing of the Target and Competing Speech

For the 20 competing test conditions, the Cool Edit Pro Version 2 (2002) audio editing software was used to digitally mix the recording of each HINT sentence list (i.e., the target speech stimuli) with the two-talker babble recording randomly assigned to it. For each competing speech condition, the recording of the masking speech began 20 seconds before the start of the first HINT sentence. Previous research has shown that in at least some test conditions, use of a continuous masking signal provides greater masking effectiveness than use of a gated masking signal (i.e., a masker that is only presented when the target speech is presented) (e.g., Hall, Grose, Buss, & Dev, 2002). Therefore, the masking speech used in the current study continued uninterrupted for the duration of the 10 HINT sentences used in each competing condition. The masking speech ended after the final word of the last HINT sentence on the list. Digital mixing of the target speech recordings and the masking speech recordings ensured that the start and end times for the masking speech and target speech recordings were consistently staggered across all competing test conditions (i.e., as compared to the variability that would occur if the investigator were to manually start and stop the target and masking recordings for each condition). Each mixed recording of the target and competing speech was saved as a two-channel *.wav file, with the HINT list saved to the right channel and the two-talker babble recording saved to the left channel. This allowed the investigator to separately adjust the intensity of the target speech recording and the intensity of the masking speech recording via a separate channel of the audiometer.

Order of Test Conditions

For each subject, testing was performed first in the two quiet test conditions, with the order of testing for the two “quiet” HINT lists being randomly selected. The order of the 20 competing test conditions was randomized using a partial counterbalancing procedure in which 62 different condition orders were created using an online random order generator (Carter, 2008). In order to do this, the 20 competing conditions were first labeled using the numbers 1-20 (see Table 2). The online program was then used to generate 62 random orderings of this set of numbers. This information was used to create a different random ordering of the 20 competing test conditions for each subject enrolled in the study.

Individual Playlists

For ease of testing, the iTunes Version 8.1 (Apple, 2008) digital music software program was downloaded to a Dell Latitude D810 laptop computer and used to create a different digital playlist for each of the 62 subjects enrolled in the study. The two “quiet” HINT files (i.e., the HINT sentence lists without the competing two-talker speech) and the 20 HINT/two-talker babble mixed files were then saved to the same laptop. For every subject, the playlist began with a 1000 Hz calibration tone lasting 15 seconds, followed by the two HINT lists randomly assigned to be tested in quiet for all subjects (i.e., List 12 and List 23). The 1000 Hz calibration tone was created using the Cool Edit Pro 2 (2002) audio editing software. After normalizing the tone to 96% of peak value, the intensity of the calibration tone was adjusted so that the minimum RMS amplitude was equivalent to the minimum RMS amplitude seen among the two-talker mixed recordings. For each playlist, the 20 HINT/two-talker mixed recordings were organized according to the

random ordering of conditions generated for that subject. Each playlist was saved to the laptop under the code name assigned to the associated subject (e.g., Subject 1, Subject 2, etc.).

Test Environment and Equipment

All subjects were tested at the Texas Tech University Health Sciences Center (TTUHSC) Department of Speech, Language, and Hearing Sciences in one of two sound treated booths. Prior to initiation of the study, all audiometric equipment was calibrated to ANSI 2004 standards. For all testing, the subject was seated in the sound treated booth wearing TDH 50P Telephonics headphones or E-A-R Tone 3A insert earphones. Insert earphones were used unless otoscopy revealed significant cerumen in one or both ears. Individuals with significant cerumen were tested using headphones. Using this criterion, 39 subjects were tested using insert earphones, and 11 subjects were tested using headphones. All testing was administered using a GSI 61 two-channel audiometer and the Dell Latitude D810 laptop. For all testing, the laptop and the iTunes (Apple, 2008) volume controls were set to the maximum settings.

The same investigator administered the testing to all subjects. During the testing, the investigator was seated at the audiometer in front of the sound booth window. As such, the investigator was visible to the subject during each test condition. However, the audiometer controls and the scoring forms were not visible to the subject. A different scoring form was created for each of the 22 test conditions (see Appendix H for two examples). Prior to each test session, the investigator organized the score sheets to match the random ordering of conditions generated for that subject.

Test Setup and Procedures

Immediately before each subject arrived, the investigator confirmed that all equipment settings were appropriate. The investigator then presented the 1000 Hz calibration tone and adjusted the audiometer's External A and External B input levels so that the VU meter for each channel read 0 dB. After calibrating the external inputs, the investigator performed a listening check (using headphones and insert earphones) by listening to the first competing test condition of the subject's assigned playlist. During the listening check, the investigator ensured that the target and competing speech signals could be heard simultaneously through the right and left headphone/insert earphone.

Each subject was tested during one 2-hour test session. Due to the length of the testing, all subjects were told that they had the option to take a brief break at any time during the testing. During the test session, each subject underwent sSRT testing in a total of 22 conditions, including 2 tests administered in quiet and 20 tests administered with competing speech (i.e., 2 tests with Masker A and 2 tests with Masker B for each of the five languages). For each condition, a different sentence list from the HINT (Nilsson et al., 1994) served as the target stimuli. For all test conditions, the target sentences and the competing speech (if applicable) were both presented simultaneously to the subject's right and left ear (i.e., diotic presentation). In each test condition, the subject's sSRT was determined using the adaptive procedures described in the user manual of the HINT (Nilsson et al., 1996), with the exception that the recordings were not paused between the presentation of each HINT sentence. As previously described, the need to pause the recording was eliminated by inserting a 20-second period of silence between each sentence for all 22 HINT lists used in the current study.

For all subjects, testing was performed first in the two quiet conditions in order to familiarize each subject with the voice of the male talker who would be presenting the target HINT sentences. Prior to testing in the quiet condition, each subject was given the same set of test instructions (see Appendix I). To begin the testing, the first sentence of the target HINT list was presented at -10 dB HL (i.e., a level likely to be below each subject's threshold). The presentation level was increased by 4 dB until the subject was able to repeat the first sentence of the target HINT list with 100% accuracy. This intensity level was then recorded on the score sheet.

From this point on, a new sentence was presented at each intensity level tested. The presentation level was then decreased by 4 dB, and the second sentence on the list was presented. The investigator recorded the presentation intensity used for each subsequent sentence on the scoring form. If the sentence was repeated correctly, a plus sign (+) was marked on the score sheet, and the presentation level was decreased before playing the next sentence. For sentences 1 – 5, a decrease of 4 dB was used. For sentences 5 – 10, a decrease of 2 dB was used. If any word in the sentence was repeated incorrectly, a minus sign (-) was marked on the score sheet, and the investigator recorded the subject's response on the score form. The presentation level was then increased before playing the next sentence. For sentences 1 – 5, an increase of 4 dB was used. For sentences 5 – 10, an increase of 2 dB was used. Based on the subject's response to sentence 10, the presentation level that would be used for sentence 11 was calculated and then recorded on the score sheet.

Prior to the competing speech test conditions, each subject was given the same set of instructions regarding the task (see Appendix J). During each test condition, the

competing speech signal was presented continuously at 65 dB HL. To begin the testing, the first sentence of the target HINT list was presented at 10 dB below the noise level (i.e., at 55 dB HL). In each competing condition, the remaining test procedures were the same as those used in the quiet condition, with the intensity of the target sentences being increased or decreased based on the subject's responses.

Scoring and Data Entry

For all 50 subjects, the sSRT score for each of the 20 competing test conditions was entered into an electronic spreadsheet file. The results obtained in the two quiet test conditions were not included because testing in quiet was only performed to familiarize subjects with the target talker's voice and to ensure that subjects could perform the sSRT task. Following the procedures recommended in the HINT manual (Nilsson et al., 1996), the sSRT for each HINT list was determined by calculating the mean presentation level used for sentences 5 – 11. To increase sensitivity, testing was performed twice in each of the 10 types of competing conditions (i.e., Masker A and Masker B for each of the five languages), resulting in a total of 20 sSRT scores for each subject (i.e., 2 scores for English Masker A, 2 scores for English Masker B, etc.). For each of the 10 types of competing conditions, the mean sSRT was determined by averaging the score obtained on the two HINT lists administered in that condition. For example, the 2 sSRT scores obtained in the two English Masker A conditions were averaged together to determine the subject's mean sSRT for English Masker A. This resulted in 10 mean sSRT scores for each subject, with 1 mean sSRT score for each of the 10 types of competing conditions. These 10 mean sSRT scores were then analyzed across the 50 subjects to investigate potential effects of: 1) variability between the masking effectiveness of different two-

talker maskers within each language (i.e., English Masker A versus English Masker B, German Masker A versus German Masker B, etc.) and 2) differences in the overall masking effectiveness of competing speech spoken in the five languages (i.e., English versus German, English versus Spanish, etc.).

Research Design

The current study used a completely within-subjects design with two independent variables (i.e., pairs of masking talkers and language of the competing speech) and one dependent variable (i.e., sSRT). The pairs of masking talkers factor consisted of 10 levels because a different male-female pair of masking talkers was used in each of the 10 competing conditions. The language factor consisted of 5 levels, including English, German, Spanish, French, and Japanese. As such, the current study's research design did not meet the criteria for a true factorial design because the test conditions did not include "all possible combinations of the levels of the independent variables" (Keppel & Wickens, 2004, p. 550). For example, there was no condition in which English Male 1 and English Female 1 provided the competing speech spoken for the German, Spanish, French, or Japanese conditions. Instead of a true factorial design, the current study involved a nested within-subjects design in which each level of the pairs of masking talkers factor only appeared at one level of the language factor. In other words, a different pair of masking talkers was used for each of the 10 competing speech conditions (i.e., 2 conditions for each of the five languages). For each language, one pair of masking talkers was randomly labeled "Masker A," and the other pair of masking talkers was randomly labeled "Masker B." As such, there was no common factor among the Masking A pairs of talkers or among the Masking B pairs of talkers. The pairs of

masking talkers factor was simply nested within the language factor (Keppel & Wickens, 2004). Because of this nested relationship, tests of the overall relationship between Masker A and Masker B across the five languages would not provide meaningful information and were therefore not analyzed in the current study.

Table 1

Random Assignment of English and Spanish Text

Two-Talker Masking Pair	Individual Talkers	Text Read for the Recording
English A	English Male 1 (44) English Female 1 (35)	Text 3 (English translation) Text 1 (English translation)
Spanish A	Spanish Male 1 (51) Spanish Female 1 (52)	Text 3 (Spanish translation) Text 1 (Spanish translation)
English B	English Male 2 (40) English Female 2 (43)	Text 4 (English translation) Text 2 (English translation)
Spanish B	Spanish Male 2 (49) Spanish Female 2 (47)	Text 4 (Spanish translation) Text 2 (Spanish translation)

Table 2

Random Assignment of Hearing in Noise Test (HINT) Lists to Two-Talker Masking Pairs

Test Condition #	Two-Talker Masking Pair	Individual Talkers	HINT List
Quiet 1			12
Quiet 2			23
1	English A	English Male 44 English Female 35	16
2	English A	English Male 44 English Female 35	19
3	German A	German Male 29 German Female 30	8
4	German A	German Male 29 German Female 30	24
5	Spanish A	Spanish Male 51 Spanish Female 52	14
6	Spanish A	Spanish Male 51 Spanish Female 52	1
7	French A	French Male 18 French Female 11	5
8	French A	French Male 18 French Female 11	18
9	Japanese A	Japanese Male 26 Japanese Female 2A	13
10	Japanese A	Japanese Male 26 Japanese Female 2A	11
11	English B	English Male 40 English Female 43	22

Table 2. Continued

Random Assignment of Hearing in Noise Test (HINT) Lists to Two-Talker Masking Pairs

Test Condition #	Two-Talker Masking Pair	Individual Talkers	HINT List
12	English B	English Male 40 English Female 43	20
13	German B	German Male 33 German Female 27	9
14	German B	German Male 33 German Female 27	7
15	Spanish B	Spanish Male 49 Spanish Female 47	17
16	Spanish B	Spanish Male 49 Spanish Female 47	15
17	French B	French Male 22 French Female 20	10
18	French B	French Male 22 French Female 20	3
19	Japanese B	Japanese Male 32 Japanese Female 31	25
20	Japanese B	Japanese Male 32 Japanese Female 31	2

Chapter III

Results

Descriptive Analysis

All subjects were able to complete the testing in the two quiet conditions with no difficulties, thus indicating that their understanding of the sSRT task was sufficient to proceed with the 20 competing speech test conditions. The mean sSRT scores for each of the 10 types of competing speech conditions (i.e., one Masker A mean and one Masker B mean for each of the five languages) were converted into bar graphs to assist in identifying trends in performance within and between the conditions tested for each of the five languages. In each condition, the competing speech was held constant at 65 dB HL, while the presentation level of the target HINT sentences was increased and decreased based on the subject's performance. Therefore, in the figures, higher mean sSRT scores indicate that the intensity of the target sentences had to be higher in order for subjects to repeat the sentences correctly. In other words, higher mean sSRTs suggest that subjects experienced greater difficulty with the task. In contrast, lower mean sSRTs suggest that subjects experienced less difficulty in selectively attending to the target sentences while ignoring the competing speech.

Figure 1 displays the mean sSRTs for Masker A and Masker B for each of the five language conditions. Within each language, there is a noticeable difference between subjects' performance in the Masker A and Masker B condition. This difference was greatest when the competing speech was spoken in English, with a difference of 4.42 dB between the mean sSRT for English Masker A and English Masker B. Differences between Masker A and Masker B within the other four languages were 1.93 dB for

German, 2.05 dB for Spanish, 2.73 dB for French, and 1.86 dB for Japanese. These differences suggest that for each of the five languages, individual differences between the two pairs of masking talkers affected the listeners' abilities to ignore the competing speech signal, despite the fact that both pairs of talkers included one male and one female native speaker of the target language.

Figure 2 illustrates the collapsed mean sSRT obtained when subjects' scores for the Masker A and Masker B condition were averaged within each of the five languages to create one overall mean for each language. Comparison between the mean sSRTs for the five languages revealed several interesting trends. First, mean sSRTs were similar for conditions where the competing speech was spoken in languages from the same rhythmic category. For example, mean sSRTs of 58.30 dB and 57.82 dB were obtained when the competing speech was spoken in English and German (respectively), two languages belonging to the stress-based rhythmic category. Similar performance was also seen between conditions when the competing speech was spoken in languages belonging to the syllable-based rhythmic category, as evidenced by the mean sSRTs obtained in the Spanish (56.11 dB) condition and the French (55.77 dB) condition. However, mean sSRTs were noticeably higher when the competing speech was spoken in English or in German (two stress-based languages), as compared to when the competing speech was spoken in Spanish or French (two syllable-based languages). Given the subjects' English-only language background, these results suggest that subjects experienced greater difficulty in ignoring speech spoken in a language from their native rhythmic class as compared to languages from a non-native rhythmic class. This increased level of difficulty associated with competing speech from the native rhythmic class occurred,

regardless of whether the speech was meaningful (i.e., as in the English condition) or not meaningful (i.e., as in the German condition) to the listeners. Unlike the other two languages from a non-native rhythmic class, the mean sSRT for the Japanese condition (58.31 dB) was similar to the scores obtained in the English and German conditions, suggesting that subjects also experienced a greater degree of difficulty in ignoring competing speech spoken in Japanese.

The mean sSRT for each language can also be viewed as a SNR by subtracting the mean sSRT from the intensity of the competing speech. Given that the competing speech was presented at 65 dB HL in all competing conditions, the mean sSRT values shown in Figure 2 are equivalent to SNRs of -6.70 dB for English, -7.18 dB for German, -8.89 for Spanish, -9.23 dB for French, and -6.69 dB for Japanese. For each of the five languages, the negative mean SNR value reveals that the intensity of the competing speech exceeded the mean intensity of the target sentences. For example, the mean SNR of -6.70 dB for the English condition indicates that the intensity of the competing speech (65 dB HL) was 6.70 dB higher than the mean intensity of the target speech (58.30 dB HL). Like the sSRT means, the SNR values also indicated that the listeners could understand in more detrimental conditions (i.e., could understand at a more negative SNR) when the competing speech was spoken in Spanish or French, as compared to when the competing speech was spoken in English, German, or Japanese.

Repeated Measures Analysis of Variance (ANOVA)

A series of statistical analyses was performed to investigate differences between the mean sSRT scores for Masker A and Masker B for each language and differences between subjects' performance across the five language conditions. All data were

analyzed using SPSS 17.0, and statistical significance was determined using a criterion of $p \leq .05$. A repeated measures ANOVA was first conducted with pair of masking talkers (Masker A or Masker B) and language (English, German, Spanish, French, or Japanese) as the within-subjects factors. Because the study used a completely within-subjects design (i.e., every subject was tested in every condition), a between-subjects factor was not specified. The main effect of masker, $F(1, 49) = 53.27, p < .01$, and the main effect of language, $F(4, 196) = 61.73, p < .01$, were both significant. To further investigate the significant main effect of masker, a series of paired t-tests was performed to compare the mean sSRT for Masker A and Masker B within each language. Results revealed a significant difference between the mean scores for Masker A and Masker B for English, $t(49) = 15.59, p < .01$, German, $t(49) = -4.95, p < .01$, Spanish, $t(49) = -5.75, p < .01$, French, $t(49) = 10.30, p < .01$, and Japanese, $t(49) = 5.90, p < .01$.

Post hoc comparisons were then performed using the Bonferroni adjustment to identify any significant differences between the overall mean (i.e., the collapsed mean for Masker A and Masker B) for each of the five languages (see Table 3). Results revealed no significant difference between the overall mean for the Spanish condition ($M = 56.11, SE = 0.20$) and the French condition ($M = 55.77, SE = 0.18, p = .77$). These results suggest that the listeners' selective auditory attention performance was the same when the competing speech was spoken in two unfamiliar languages belonging to the syllable-based rhythmic category (i.e., a category which was not native to the listeners). In addition, no significant difference was observed between the overall mean for the English ($M = 58.30, SE = 0.21$) and German ($M = 57.82, SE = 0.20, p = .43$) conditions, indicating that listeners experienced the same degree of difficulty in ignoring competing

speech spoken in their native language and in an unfamiliar language belonging to their native stress-based rhythmic category. However, the overall means for the English condition and the German condition were each significantly greater than the overall means for the Spanish condition (English-Spanish: $p < .01$, German-Spanish: $p < .01$) and the French condition (English-French: $p < .01$, German-French: $p < .01$). Therefore, listeners experienced significantly greater difficulty in ignoring competing speech spoken in a language from their native rhythmic category as compared to languages from a non-native rhythmic category, regardless of whether or not the language of the competing speech was meaningful (as in the English condition) or not meaningful (as in the German condition).

No significant difference was seen between the overall mean for the Japanese ($M = 58.31$, $SE = 0.20$) condition and the English condition ($M = 58.30$, $SE = 0.21$, $p = 1.00$) or between the Japanese condition and the German condition ($M = 57.82$, $SE = 0.20$, $p = .51$). However, the overall mean for the Japanese condition ($M = 58.31$) was significantly greater than the overall mean for the Spanish condition ($M = 56.11$, $SE = 0.20$, $p < .01$) and the overall mean for the French condition ($M = 55.77$, $SE = 0.18$, $p < .01$). Together, these results indicate that listeners treated the competing speech spoken in Japanese the same as competing speech spoken in English or German, despite the fact that Japanese belongs to a rhythmic category that was unfamiliar to the listeners (i.e., mora-based).

Because of the nested relationship between the pairs of masking talkers factor and the language factor, the question remained as to whether a repeated measures ANOVA was the most appropriate way to statistically analyze the results of the current study. In light of this concern, the results were also analyzed using two additional types of tests.

First, a one-way ANOVA was performed without the distinction between Masker A and Masker B to confirm the significant main effect of language which was revealed by the repeated measures ANOVA. In addition, due to the significant differences noted between the Masker A and Masker B means for each of the five languages, the results were also reanalyzed using a univariate ANOVA with language as the within-subjects factor and the difference between the Masker A mean and the Masker B mean as the covariate. Both types of analyses revealed the same pattern of results as the repeated measures ANOVA, thus confirming the validity of the original findings. Appendix K provides a more detailed description of the results revealed by the one-way ANOVA without the Masker A-Masker B distinction and the univariate ANOVA with the Masker A-Masker B difference as the covariate.

Figure 1. Mean sentence reception threshold (sSRT) for Masker A and Masker B for each language.

Figure 2. Overall mean sentence reception threshold (sSRT) for each language (no Masker A – Masker B distinction).

Table 3

Pairwise Comparisons Between the Overall Mean Sentence Reception Threshold (sSRT) for Each Language

Language (1)	Language (2)	Mean Difference ^a (1 – 2)	Standard Error
English	German	0.496	0.234
	Spanish	2.193*	0.247
	French	2.531*	0.197
	Japanese	-0.004	0.242
German	English	-0.486	0.234
	Spanish	1.707*	0.218
	French	2.046*	0.205
	Japanese	-0.490	0.245
Spanish	English	-2.193*	0.247
	German	-1.707*	0.218
	French	0.339	0.188
	Japanese	-2.197*	0.223
French	English	-2.531*	0.197
	German	-2.046*	0.205
	Spanish	-0.339	0.188
	Japanese	-2.536*	0.202
Japanese	English	0.004	0.242
	German	0.490	0.245
	Spanish	2.197*	0.223
	French	2.536*	0.202

Note. Results were obtained using the Bonferroni adjustment for multiple comparisons.

^aMean difference scores represent the difference (in dB) between the overall mean sSRT for each pair of languages.

* $p < .01$.

Chapter IV

Discussion

The current study provides unique insight into how adult selective auditory attention is affected by: 1) variability between the vocal/speaking characteristics of different pairs of masking talkers and 2) the rhythmic structure of the language of the competing speech. Such findings are particularly important given that these two aspects of selective auditory attention have been investigated by very few previous studies. For example, only one other study has explored how differences between the vocal/speaking characteristics of masking talkers may impact the masking effectiveness of a competing speech signal (Freyman et al., 2007). In addition, few studies have compared listeners' abilities to ignore speech spoken in different languages (e.g., Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). Results from previous studies have revealed that some languages differ in their rhythmic structure (e.g., Grabe & Low, 2002; Ramus et al., 1999), and these differences have been shown to have a significant impact on how infants and adults process speech in quiet listening conditions (e.g., Cutler et al., 1983; 1986; Nazzi et al., 2000; Sebastián-Gallés et al., 2000). In light of these findings, one possibility is that the rhythmic properties of a language may also affect a listener's ability to process speech in conditions where one or more talkers are speaking simultaneously. Because three primary categories of language rhythm have been proposed (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi et al., 2000), a test of the relationship between language rhythm and selective auditory attention requires comparison of listeners' speech recognition performance in conditions where the competing speech is spoken in multiple languages, including languages from each of the

three proposed rhythmic classes. However, prior to the current investigation, no study had investigated listeners' selective auditory attention performance in more than two language conditions.

The current study addressed these gaps in existing research by assessing monolingual English-speaking adult listeners' selective auditory attention performance in conditions where the competing speech was spoken in a stress-based language (e.g., English or German), a syllable-based language (e.g., Spanish or French), or a mora-based language (e.g., Japanese) (Abercrombie, 1967; Pike, 1945; both as cited in Nazzi et al., 2000). To evaluate effects of the vocal/speaking characteristics of different masking talkers, two different two-talker competing speech signals (i.e., labeled Masker A and Masker B) were created for each language, with a different male-female pair of native speakers used for each competing speech signal. As such, 10 different types of competing speech signals were created, including one Masker A competing speech signal and one Masker B competing speech signal for each of the five languages. Results from the 10 types of competing conditions were compared to determine how monolingual adult listeners' selective auditory attention performance was affected by the:

- Semantic content of the competing speech,
- Rhythmic structure of the language in which the competing speech was spoken, and
- Individual differences between pairs of masking talkers of the same language.

Differences Between Masker A and Masker B For Each Language

In planning the current study, it was first hypothesized that no significant differences would be seen between the masking effectiveness of the two pairs of masking

talkers within each language (e.g., English Masker A versus English Masker B). Prior to the current investigation, no study had specifically compared the masking effectiveness of different mixed-gender two-talker maskers. One previous study had investigated listeners' selective auditory attention performance in conditions where the competing speech was spoken by different two-talker maskers, but unlike the current study, all conditions were in English and included target and competing speech spoken by female talkers (i.e., a same-gender target-masker configuration) (Freyman et al., 2007). Results of that study revealed that different same-gender two-talker maskers varied considerably in their masking effectiveness, but the study only included descriptive data for each test condition and failed to report whether the differences between conditions reached the level of statistical significance. Therefore, at the time of planning the current study, it remained unknown whether similar differences in masking effectiveness might be seen between different mixed-gender two-talker maskers, such as those used in the current study. Furthermore, there was no evidence as to whether effects of differences between the vocal/speaking characteristics of masking talkers might vary depending on the language of the competing speech.

In light of the limited research in this area, the current study included the null hypothesis that no significant differences would be seen between the two pairs of masking talkers within each language (i.e., English Masker A versus English Masker B). However, the current results failed to support this hypothesis. For each of the five languages used as the competing speech, a significant difference was seen between the listeners' mean sSRT in the Masker A and the Masker B condition. In order to interpret these findings, it is important to first consider the various controls that were used to

ensure that the Masker A and Masker B competing speech signals within each language were as similar as possible. For example, all text read for the English Masker A and English Masker B recordings was taken from the same bilingual education textbook and was therefore written at the same reading level and covered the same general topic areas. As such, differences between the text used for the recordings should not have contributed to the significant difference noted between the English Masker A and English Masker B condition. The text read for the Spanish Masker A and Spanish Masker B recordings was also taken from the same bilingual education textbook, thus ensuring that all competing speech spoken in Spanish was consistent with regard to reading level and content. For German, French, and Japanese, all text used for the Masker A and Masker B competing speech was written at approximately a seventh grade reading level, but the content of the text varied between the conditions because the masking talkers for these languages did not read text from the same source. However, differences between the content of the Masker A and Masker B competing speech were unlikely to have affected selective auditory attention performance in the Spanish, German, French, and Japanese conditions because listeners were unable to understand the competing speech spoken in these languages.

In addition to these controls, other efforts were made to equate the Masker A and Masker B conditions for each language. For example, the HINT lists used as the target speech signal in the test conditions were of equivalent difficulty based on data reported by the test developers (Nilsson et al., 1994). All HINT lists were also normalized to be of the same overall RMS amplitude, thus eliminating any significant differences in the overall intensity of the target sentences used in each competing condition. All individual

recordings of masking talkers and all two-talker mixed recordings were equated for RMS amplitude so that the overall intensity of each two-talker masking signal would be equivalent within and across the five languages. In addition, the 20 competing test conditions (i.e., 2 conditions using Masker A and 2 conditions using Masker B for each of the five languages) were administered to each subject in a different randomly selected order. As such, the order of testing for the Masker A and Masker B conditions for each language did not follow any consistent pattern. Use of a different random order of testing for each subject greatly reduced the possibility that order effects contributed to the significant difference that was seen between the mean sSRT in the Masker A and Masker B condition for each language. In addition, testing in all Masker A and Masker B conditions was administered by the same investigator using the same test environment, equipment, and procedures.

In light of these controls, the target speech, masking speech, and test setup used in the current study should not have contributed to the significant effect of masking talker pair that occurred for each of the five languages. Therefore, differences between the masking effectiveness of the Masker A and Masker B competing speech signals for each language are most likely to have resulted from differences between the vocal and/or speaking characteristics of the individual masking talkers. For example, factors such as the fundamental frequency, vocal quality, prosody, and speaking rate of the masking talkers may have affected the listeners' abilities to separate the target sentences from the competing speech (Freyman et al., 2007). However, additional research is needed to further explore the extent to which each of these factors may contribute to the masking effectiveness of a competing speech signal.

Although further investigation is needed, the current results alone make an important contribution to existing research related to how selective auditory attention is affected by characteristics of the competing speech. To date, the current study is the first to compare the masking effectiveness of different pairs of mixed-gender two-talker maskers (i.e., one male and one female masking talker), as well as the first study to statistically analyze any observed differences in masking talker variability. In addition, results of the current study are the first to demonstrate that differences between the vocal/speaking characteristics of the masking talkers may significantly affect selective auditory attention performance even when the competing speech is not meaningful to the listeners at the sentence or the word level, as was the case for the four unfamiliar languages used in the current study (i.e., German, Spanish, French, and Japanese). However, the current results also provide initial evidence that listeners may be more sensitive to differences between the vocal/speaking characteristics of the masking talkers when the competing speech is spoken in their native language. This conclusion is demonstrated by the fact that in the current study, the difference between the mean sSRT for the Masker A and Masker B conditions was greatest for the English competing speech, with a difference of 4.42 dB, as compared to a difference of 1.93 dB for German, 2.05 dB for Spanish, 2.73 for French, and 1.86 dB for Japanese.

Together, the findings reported by Freyman and colleagues (2007) and those reported in the current study have important implications for the validity of selective auditory attention studies in which the masking talkers differ between conditions. Without adequate control for the variability between the vocal/speaking characteristics of different masking talkers, differences between competing conditions cannot be validly

interpreted. Therefore, future studies should attempt to equate the vocal/speaking characteristics of the masking talkers to the greatest extent possible and/or structure the study's design such that effects of masking talker variability can be measured and used as a covariate in the analysis of overall differences between different competing conditions, as was done in the current study.

Relationships Between the Masking Effectiveness of Different Languages

With the exception of the comparison between Masker A and Masker B within each language, all other hypotheses dealt with the proposed relationships between listeners' selective auditory attention performance in the five language conditions. More specifically, prior to initiation of the current study, it was hypothesized that:

- Listeners would experience greater difficulty attending to the target English sentences when the competing speech was spoken in English as compared to conditions where the competing speech was spoken in one of the four unfamiliar languages (i.e., German, Spanish, French, and Japanese).
- No significant differences in performance would be seen between the three conditions where the competing speech was spoken in an unfamiliar language from a non-native rhythmic class (i.e., Spanish, French, and Japanese).
- Listeners would experience greater difficulty attending to the target sentences when the competing speech was spoken in an unfamiliar language from the native language rhythmic category (i.e., German) as compared to performance in the three conditions where the unfamiliar language belonged to a non-native rhythmic category (i.e., Spanish, French, and Japanese).

Effects of the Meaning of Competing Speech

Results of the current study support some aspects of these hypotheses, but unexpected relationships were also observed between listeners' performance in certain competing language conditions. For example, contrary to the hypothesized outcome, competing speech spoken in English was not significantly more difficult for listeners to ignore than competing speech in all of the other languages. Instead, analysis of the results revealed that the English competing speech conditions were only significantly more difficult than conditions where the competing speech was spoken in Spanish or French, two languages in which the words and underlying rhythm of the language (i.e., syllable-based) were both unfamiliar to the listeners. In contrast, no significant differences were seen between listeners' performance in conditions where the competing speech was spoken in English, German, or Japanese, despite the fact that listeners could not understand the competing speech in the German or Japanese conditions. These results indicate that meaning alone does not determine the masking effectiveness of a competing speech signal. Such findings are consistent with results from previous studies which have shown that at least in some conditions, competing speech spoken in the native language is not significantly more difficult to ignore than competing speech spoken in an unfamiliar language (Freyman et al., 2001; Tun et al., 2002). Together with results from the current study, these findings support the conclusion that linguistic factors other than meaning can influence the masking effectiveness of competing speech.

Competing Speech in English, German, Spanish, and French

As hypothesized, evidence from the current study suggests that adult selective auditory attention performance is significantly affected by the rhythm of the language in

which the competing speech is spoken. The influence of language rhythm in the current study is first illustrated by the fact that no significant difference was seen between listeners' performance in conditions where the competing speech was spoken in languages within the same rhythmic category. For example, listeners experienced the same level of difficulty in conditions where the competing speech was spoken in English or German, two languages that belong to the listeners' native stress-based rhythmic category. Similarly, no significant difference was seen between listeners' performance in conditions where the competing speech was spoken in Spanish or French, two unfamiliar languages that belong to the same non-native rhythmic category (i.e., syllable-based). These findings are consistent with results from two previous studies that compared listeners' selective auditory attention performance in conditions where the competing speech was spoken in the native language and one unfamiliar language from the native rhythmic category (Freyman et al., 2001; Tun et al., 2002). In both studies, no significant differences were seen between listeners' abilities to recognize a target speech signal when the competing speech was spoken in English, the listeners' native stress-based language, versus when the competing speech was spoken in Dutch, an unfamiliar language belonging to the native stress-based rhythmic category.

Collectively, results from these two studies (Freyman et al., 2001; Tun et al., 2002) and those of the current study suggest that listeners are sensitive to the underlying rhythm of the language of the competing speech, such that languages within the native rhythmic category provide the same degree of masking effectiveness, regardless of whether the competing language is meaningful or not meaningful to the listeners. However, it should be noted that the current study is the first to compare listeners'

selective auditory attention performance between two conditions where the competing speech is spoken in two unfamiliar languages belonging to the same non-native rhythmic category (i.e., two syllable-based languages). As such, the current results extend previous evidence to show that languages within a particular rhythmic category provide the same degree of masking effectiveness, regardless of whether the two languages being compared are both from within the native rhythmic category (i.e., English versus German) or both from within a non-native rhythmic category (i.e., Spanish versus French).

In addition to these findings, results of the current study also suggest that competing speech spoken in a language from the listeners' native rhythmic category is more difficult for listeners to ignore than competing speech spoken in a language from a non-native rhythmic category, at least for some languages. For example, as hypothesized, listeners in the current study experienced significantly greater difficulty in selectively attending to the target speech when the competing speech was spoken in English or German, two languages from the native stress-based rhythmic category, as compared to when the competing speech was spoken in Spanish or French, two languages from the non-native syllable-based rhythmic category. These findings are consistent with two previous studies which found that competing speech spoken in the native language was more difficult for listeners to ignore than competing speech spoken in an unfamiliar language from a non-native rhythmic category (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). However, the current study is the first to demonstrate that speech in an unfamiliar language from the native rhythmic category (i.e., German) is significantly more difficult for listeners to ignore than an unfamiliar language from a

non-native rhythmic category (i.e., Spanish and French). Such findings further strengthen the contention that the rhythm of the language of a competing speech signal has a greater impact on masking effectiveness than the meaning of the competing speech.

Competing Speech in Japanese

Contrary to the hypothesized outcome, listeners in the current study did not treat competing speech spoken in Japanese the same as competing speech spoken in Spanish or French, even though all three languages were unfamiliar to the listeners and were from non-native rhythmic categories. Instead, no significant differences were seen between listeners' performance in the English, German, and Japanese competing speech conditions. Therefore, competing speech spoken in Japanese had the same effect on listeners' performance as competing speech spoken in English, the listeners' native language, and German, an unfamiliar language from the native stress-based rhythmic category. Like English and German, competing speech spoken in Japanese was significantly more difficult for listeners to ignore than competing speech spoken in Spanish or French. Together, these results suggest that competing speech spoken in Japanese provides the same degree of masking effectiveness as competing speech spoken in English or German, despite the fact that Japanese is typically considered to have mora-based rhythm, a type of rhythmic structure that should have been unfamiliar to the monolingual English-speaking listeners in the current study.

At the present time, it remains unclear why this unexpected pattern of results occurred for the Japanese competing speech condition. However, a number of extraneous factors can most likely be ruled out due to the controls used in the current study. For example, as previously described, careful procedures were used to ensure that the

competing conditions for all five languages were equivalent with regard to factors such as: the reading level of the text read by each masking talker, the difficulty of the HINT lists used as the target speech signal, the overall intensity of each HINT list as measured by the RMS amplitude, the RMS amplitude of each individual masking talker, the RMS amplitude of each two-talker mixed recording, the test environment, test equipment, and test procedures. As such, extraneous factors related to the test stimuli and/or test conditions should not have differentially affected listeners' performance in the Japanese competing speech conditions.

In addition, the vocal/speaking characteristics of the Japanese masking talkers were unlikely to have caused the Japanese competing speech to be any more difficult to ignore than the competing speech spoken in Spanish or French, the other two unfamiliar languages from non-native rhythmic categories. For each of the five languages, a significant difference was seen between the mean sSRT obtained for the Masker A and Masker B conditions, thus suggesting that differences between the vocal/speaking characteristics of the masking talkers had a significant effect on the results obtained within all five languages. However, the difference between the mean sSRT for the Masker A and Masker B conditions was actually smallest for the Japanese condition, with a difference of 1.86 dB, as compared to a difference of 1.93 dB for German, 2.05 dB for Spanish, 2.73 dB for French, and 4.42 dB for English. In light of these findings, variability between the vocal/speaking characteristics of different masking talkers did not have a greater impact on listeners' performance in the Japanese condition as compared to the other four languages used as the competing speech. Furthermore, when the overall mean sSRT (i.e., the collapsed mean for the Masker A and Masker B conditions) for each

language was compared, the same pattern of relationships between languages was seen, regardless of whether or not the Masker A – Masker B difference for each language was factored into the analysis as a covariate. Therefore, the same overall pattern of results was seen for the Japanese competing speech, even when differences between the vocal/speaking characteristics of the masking talkers were taken into account when comparing the results for each language.

Given these findings, it seems that the linguistic characteristics of the Japanese language itself were most likely responsible for the pattern of results seen for the Japanese competing speech. However, the question remains as to why, in the current study, competing speech spoken in Japanese, a language typically classified as having mora-based rhythm, provided the same degree of masking effectiveness as competing speech spoken in English or German, two languages from the listeners' native stress-based rhythmic class. One possible explanation for these findings is that Japanese may not actually exhibit its own unique mora-based rhythm as classically believed but may instead share at least some similarities with the stress-based rhythm associated with languages such as English and German. The classification of Japanese as a mora-based language has largely been based on behavioral evidence related to the speech processing strategies used by infant and adult listeners in quiet listening conditions. For example, previous studies have demonstrated that from birth to 6 months of age, babies from monolingual families can discriminate between Japanese and languages from the stress-based rhythmic class (i.e., English) (Nazzi et al., 1998; Nazzi et al., 2000) and languages from the syllable-based rhythmic class (i.e., Italian) (Nazzi et al., 2000). However, the same infants have been shown to experience difficulties in discriminating between two

unfamiliar languages from the within a particular rhythmic category (e.g., Nazzi et al., 1998; Nazzi et al., 2000). These results have been viewed as evidence that the underlying rhythmic structure of Japanese is different from that of stress-based and syllable-based languages. In addition, other studies have shown that adult native listeners of Japanese segment continuous speech according to morae, the subsyllabic units believed to underlie the rhythmic structure of Japanese speech (Cutler & Otake, 1994). Specifically, results of such studies have revealed that native Japanese listeners use a mora-based strategy to segment speech, regardless of whether speech is spoken in their native language (Japanese) or in an unfamiliar language from the stress-based rhythmic category (English). In contrast, findings from other studies have shown that native listeners of other languages fail to segment Japanese speech according to morae but instead tend to segment speech by the unit that underlies the rhythm of their native language (e.g., Otake et al., 1993). Together, these behavioral findings have supported the contention that the underlying rhythm of Japanese is distinct from that of languages belonging to the stress-based and syllable-based rhythmic classes.

However, despite this pattern of behavioral results, acoustic-phonetic studies of language rhythm have reported conflicting findings regarding whether the rhythmic structure of Japanese is in fact significantly different from that of stress-based and syllable-based languages. For example, Ramus and colleagues (1999) found that Japanese speech was significantly different from speech spoken in stress-based languages (i.e., English, Dutch, and Polish) and speech spoken in syllable-based languages (e.g., Spanish, French, Catalan, and Italian) when the rhythm of each language was quantified based on the duration of vocalic and consonantal intervals, two properties that are directly

related to the syllable structure of a language. In this particular study, a vocalic interval was defined as the length of time that elapsed between the beginning and end of a single vowel or group of vowels; whereas, a consonantal interval was defined as the time that elapsed between the beginning and end of a single consonant or group of consonants. Results of the study revealed that Japanese differed significantly from the other languages on two measures: the proportion of vocalic intervals within each sentence and the standard deviation of the duration of consonantal intervals within each sentence. Stress-based and syllable-based languages also differed significantly on these two measures. Therefore, the acoustic-phonetic evidence reported by Ramus and colleagues (1999) further supported the idea that Japanese exhibits a rhythmic structure that differs significantly from the rhythm associated with stress-based and syllable-based languages.

In contrast to these findings, evidence from another acoustic-phonetic study indicated that Japanese may not have its own unique rhythm but may instead exhibit rhythmic properties similar to syllable-based and stress-based languages. For example, Grabe and Low (2002) compared the rhythmic structure of different languages by calculating a Pairwise Variability Index (PVI) for vocalic intervals (i.e., the duration of vowels) and intervocalic intervals (i.e., the duration of intervals occurring between successive vowels) to determine the variability in duration between adjacent pairs of each type of interval. Analysis of the results revealed that the PVI for vocalic intervals differentiated stress-based languages, such as English, Dutch, and German, from syllable-based languages, such as French and Spanish. However, for Japanese, the PVI for vocalic intervals was comparable to that of French, a syllable-based language. Comparison between the PVI for intervocalic intervals associated with each language revealed no

consistent differences between stress-based, syllable-based, and mora-based languages. Instead, intervocalic PVI values for Japanese were similar to those of English, a stress-based language.

In addition to these calculations, Grabe and Low (2002) reanalyzed their data using the same techniques as Ramus and colleagues (1999) to measure the proportion of vocalic intervals in each sentence and the standard deviation of the duration of consonant intervals in each sentence. Results of this measurement technique also differentiated between some stress-based languages (e.g., English, German, and Dutch) and syllable-based languages (e.g., French and Spanish). However, the proportion of vocalic intervals within the Japanese sentences was found to be comparable to the values seen for Dutch and German, two stress-based languages. Therefore, the collective results of the study by Grabe and Low (2002) suggested that the rhythmic structure of Japanese may share similarities with stress-based and syllable-based languages, but identification of such similarities appeared to depend on the specific acoustic-phonetic measurement technique used to quantify language rhythm.

Together, results of the current study and the acoustic-phonetic evidence reported by previous studies suggest that Japanese may not belong to its own unique rhythmic class. Instead, the acoustic-phonetic properties of Japanese seem to overlap with those of stress-based and syllable-based languages. In the current study, listeners treated competing speech in Japanese the same as competing speech in English and German, two languages from the stress-based rhythmic category. In contrast, significant differences were seen between how listeners in the current study performed when the competing speech was spoken in Japanese as compared to Spanish or French, two syllable-based

languages. In light of the findings reported by Grabe and Low (2002), the pattern of results seen in the current study appears to indicate that for monolingual English-speaking adults, the acoustic-phonetic properties shared between Japanese and stress-based languages have a greater impact on selective auditory attention performance in comparison to the acoustic-phonetic similarities between Japanese and syllable-based languages. However, additional research is needed to further explore the potential relationship between the underlying rhythm of Japanese and stress-based languages, such as English and German. For example, future studies should investigate how listeners from other language backgrounds, including native Japanese listeners, selectively attend to speech in their native language when the competing speech is spoken in Japanese, stress-based languages, and syllable-based languages. Results of such studies could assist in determining if there is a consistent relationship between the masking effectiveness of competing speech in Japanese and languages typically classified as having stress-based rhythm.

Limitations of the Current Study and Implications for Future Research

The current study offers unique insight into how adult listeners' selective auditory attention performance is affected by variability between the vocal/speaking characteristics of different masking talkers and the underlying rhythm of the language in which the competing speech is spoken. To date, very few studies have investigated either of these factors. Results from the current study address this gap in previous research and subsequently contribute to the overall understanding of how selective auditory attention performance is affected by characteristics of the competing speech signal. However, despite these strengths, certain limitations should be considered before generalizing the

current results beyond the context of the current study. These limitations have important implications for the subject population, test stimuli, and test setup/procedures used in future studies of this nature.

Subjects

The first limitation to be considered relates to the restricted age range of the subjects who participated in the current study. Although the study was open to adults between 18 and 55 years of age, all subjects were undergraduate or graduate students in the TTUHSC Department of Speech, Language, and Hearing Sciences. For the 50 subjects included in the data analysis, subjects' ages ranged from 20 to 44 years of age, with a mean age of 23.72 years. However, only 5 subjects were over 25 years of age. As such, it remains unknown whether the pattern of results observed in the current study can be generalized to other age groups. Previous studies have shown that selective auditory attention performance changes with age, both during childhood and during late adulthood. In general, such evidence has suggested that most aspects of selective auditory attention reach adult maturity by the teenage years (e.g., Johnson, 2000; Neuman & Hochberg, 1983). However, other research has indicated that older adults experience greater difficulties than younger adults in selectively attending to target speech in the presence of competing speech, even when hearing sensitivity is normal or near normal for at least the speech frequencies (e.g., Tun et al., 2002). Furthermore, the meaning of the competing speech signal has been shown to have a greater effect on the selective auditory attention performance of older adults as compared younger adults, at least when the competing speech is spoken in the native language (i.e., meaningful English sentences versus strings of random English words) (e.g., Tun et al., 2002). In light of these

findings, additional research is needed to determine whether the pattern of results observed in the current study can be generalized to other age groups, including children and older adults. Results of such studies could provide important information regarding how a listener's age affects his/her susceptibility to effects of different characteristics of competing speech signals, including variability between the vocal/speaking characteristics of masking talkers and the language of the competing speech.

The language background of subjects in the current study is another factor to be considered in generalizing the current results to other populations of listeners. Previous research has shown that a listener's language background can affect how he/she processes speech, both in quiet conditions (e.g., Cutler & Otake, 1994; Sebastián-Gallés et al., 2000) and in conditions where competing sounds are present (e.g., Garcia Lecumberri & Cooke, 2006). Based on these findings, the subject sample for the current study was limited to monolingual American English-speaking adults. Controlling for language background ensured that listeners' performance on the experimental tasks would not be influenced by differences in their exposure to languages other than English. As such, this control strengthened the internal validity of the current study's results. However, additional research is now needed to determine if a similar pattern of results would be seen for listeners from other language backgrounds. For example, future studies should test monolingual speakers of stress-based languages other than English (e.g., German, Dutch, Russian, etc.), syllable-based languages (e.g., Spanish, French, Italian, etc.), and mora-based languages (e.g., Japanese). Results from these studies will assist in determining if greater masking effectiveness is always associated with competing speech spoken in languages from a listener's native rhythmic category, even for monolingual

speakers of languages other than English. In addition, future studies should also investigate how the language of the competing speech affects the selective auditory attention performance of different groups of bilingual listeners. For example, performance should be compared between bilingual listeners for whom both native languages are from the same rhythmic class (e.g., German-English bilinguals) and those for whom the two native languages are from different rhythmic classes (e.g., Spanish-English bilinguals). Comparing bilingual and monolingual listeners' performance in conditions where the competing speech is spoken in languages from the three proposed rhythmic classes will extend the findings of the current study and provide additional insight into how exposure to the native language(s) affects listeners' abilities to selectively attend to a target speech signal in the presence of competing speech.

Competing Speech

In addition to the subjects' age range and language background, other limitations of the current study relate to characteristics of the competing speech used during each selective auditory attention test condition. Some characteristics of the competing speech signals merit further consideration due to their influence on how results of the current study are interpreted; whereas, other characteristics limit the extent to which the current pattern of results can be generalized to other selective auditory attention conditions. For example, in the current study, the masking talkers of the five languages did not all read the same text for the recordings used to create the competing speech signals for each language. Instead, translations of the same text were only used for the English and Spanish competing speech signals; whereas, the German, French, and Japanese masking talkers each read text from unrelated sources. Ideally, translations of the same text would

have been used for the competing speech in all five languages. However, at the time of the current study, no single source of text could be located which was appropriate for the current study and was available in English, German, Spanish, French, and Japanese translations. Therefore, different passages of text were used for the German, Spanish, French, and Japanese recordings, but the text used for all five languages was still required to meet the following criteria:

- Was written in the native language at approximately a seventh grade reading level or higher,
- Included content that would be appropriate for child or adult listeners (i.e., no mature content),
- Contained no dialogue between characters,
- Contained no references to well-known American proper names, and
- Was of sufficient length for the talker to read for approximately 20 minutes without repeating any text.

These controls helped to minimize the possibility that differences between the text read for each recording would significantly impact the listeners' selective auditory attention performance in the competing speech conditions. However, even without such controls, differences between the content of the text read by the masking talkers should not have affected listeners' performance in the German, Spanish, French, or Japanese competing speech conditions, given that the monolingual English-speaking subjects in the current study were unable to understand any of these languages. Furthermore, differences in the text used as the competing speech seem unlikely to have significantly impacted the current pattern of results, given that a significant difference was seen

between listeners' performance in the English and Spanish competing conditions, despite the fact that translations of the same text were used for the English and Spanish competing speech signals. However, if possible, future studies should attempt to use translations of the same text for the competing speech in each language to rule out the possibility of extraneous effects of text differences on the masking effectiveness of different competing speech signals.

Other aspects of the competing speech used in the current study limit the extent to which the current pattern of results can be generalized to other types of competing speech conditions. For example, each competing signal used in the current study consisted of speech from two native speakers of the language assigned to that condition. However, previous studies have shown that listeners' selective auditory attention performance varies significantly depending on the number of masking talkers used in the competing speech signal (e.g., Simpson & Cooke, 2005). Such evidence has suggested that overall, the masking effectiveness of competing speech increases (i.e., creates a more difficult listening situation) as the number of masking talkers increases up to six talkers (Simpson & Cooke, 2005). However, other evidence has indicated that the effect of the content of the competing speech signal is reduced or eliminated as the number of masking talkers increases. For example, Van Engen and Bradlow (2007) found that the sentence recognition performance of monolingual English-speaking adults was significantly poorer when the competing speech signal consisted of English two-talker babble (i.e., a language familiar to the listeners) as compared to Mandarin Chinese two-talker babble (i.e., a language that was unfamiliar to the listeners). In contrast, the language of the competing speech signal had no significant effect when sentence recognition performance was

measured in six-talker English babble and six-talker Mandarin babble. Based on these findings, two-talker babble was chosen as the masking speech signal for the current study in order to achieve the most effective balance between maximizing the masking effectiveness of the competing speech signal (i.e., by using more than one masking talker) and ensuring that any effects of the linguistic content of the competing signal (e.g., rhythmic structure of the language) could still be detected. However, additional research is needed to determine if listeners' selective auditory attention performance will still be significantly impacted by the language of the competing speech if a different number of masking talkers is used (e.g., a single masking talker for each condition, three masking talkers for each condition, etc.).

In addition to the number of masking talkers, the target-masker gender configuration used for the competing speech must also be considered before generalizing the results of the current study. For all competing conditions, the competing speech signal consisted of continuous speech from one male and one female native speaker of the language assigned to that condition; whereas, the target sentences used in each condition were always spoken by the same male talker (i.e., a different male than the male masking talker). As such, the current study used a mixed-gender two-talker target-masker gender configuration for all competing conditions. This choice was based on previous evidence reported by Brungart and colleagues (2001) which revealed that listeners experienced significantly less difficulty in selectively attending to a target speech signal when the competing speech consisted of two masking talkers of a different gender than the target talker (i.e., a different-gender masking configuration) as compared to conditions where both masking talkers were of the same gender as the target talker

(i.e., a same-gender masking configuration) or when one masking talker of each gender was used (i.e., a mixed-gender masking configuration). No significant difference was found between listeners' selective auditory attention performance in the same-gender and the mixed-gender masking conditions.

Considering these findings, the same-gender and mixed-gender masking configurations were determined to both be appropriate options for the current study; whereas, use of a different-gender masking configuration was ruled out due to the risk of a ceiling effect (i.e., because the listening task might be too easy). Because the target HINT sentence lists were recorded by a male talker, use of a same-gender two-talker masker configuration would have required a minimum of four different male recording volunteers for each of the five languages (i.e., to create a Masker A and a Masker B same-gender two-talker masking signal for each language). However, due to the limited number of native speakers who volunteered for the recordings, there were not a sufficient number of male recording volunteers to meet this requirement. As such, it was necessary to use a mixed-gender two-talker competing speech signal (i.e., one male and one female masking talker) for each of the competing conditions in the current study. Although the evidence reported by Brungart and colleagues (2001) indicated that same-gender and mixed-gender two-talker maskers produced the same degree of masking effectiveness, these findings were obtained using competing speech spoken only in English. Therefore, additional research is now needed to determine whether the language of the competing speech will also significantly affect listeners' selective auditory attention performance if the competing speech signal consists of two masking talkers of the same gender as the target talker (i.e., a same-gender target-masker configuration).

Although the significant main effect of masking talker pair was taken into account during the between language comparisons, one additional limitation of the current study is that the results failed to identify the specific factors that caused the significant difference between performance in the Masker A and Masker B condition for each of the five languages. As previously described, numerous controls were used to ensure that the target speech, competing speech, and test setup used for the Masker A and Masker B test conditions were as similar as possible for all five languages. As such, differences between the masking effectiveness of the Masker A and Masker B competing speech were most likely the result of differences between the vocal/speaking characteristics of the individual masking talkers. However, the methods used in the current study did not allow for investigation into which specific aspects of the masking talkers' vocal and/or speaking characteristics provided the greatest contribution to the overall masking effectiveness of the competing speech signal. In light of this limitation, additional research is now needed to measure the extent to which the masking effectiveness of competing speech is influenced by specific vocal/speaking properties, such as a masking talker's vocal quality, fundamental frequency, prosody, and speaking rate.

Test Setup and Procedures

The test setup and procedures used in the current study also have implications for future research. For example, the current study used an adaptive sSRT procedure to assess listeners' selective auditory attention performance in each competing speech condition. Previous studies have shown that complex interactions occur between the non-linguistic characteristics of competing speech signals. For example, the effects of SNR on selective auditory attention performance have been reported to vary depending

on other non-linguistic aspects of the competing speech signal, including the spatial location of the target and masking talkers (e.g., Freyman et al., 2001; Freyman, 2004), the number of masking talkers (e.g., Brungart, 2001; Brungart et al., 2001; Cooke et al., 2008), and the target-masker gender relationship (e.g., Brungart, 2001). However, much remains unknown regarding how effects of the SNR may interact with the linguistic characteristics of competing speech, including the rhythm of the language spoken by the masking talkers. In light of these uncertainties, an adaptive sSRT procedure was chosen for use in the current study, instead of administering sentence recognition testing at a single SNR in each test condition. Use of the sSRT procedure subsequently revealed important information regarding how selective auditory attention is influenced by linguistic characteristics of competing speech, including the rhythm of the language spoken by the masking talkers. However, additional research is needed to determine if the same effects of the language of the competing speech will be seen if selective auditory attention is measured using other techniques, such as a speech recognition task in which a percent words or percent sentences correct score is derived or a reaction time task in which listening effort is measured for each condition. Both a percent correct speech recognition task and a reaction time task require that a specific SNR be used for each test condition. Therefore, future studies will also need to specifically investigate whether effects of the language of the competing speech may vary depending on the SNR at which the target and competing speech signals are presented.

A final limitation of the current study relates to the length of time required to administer the 22 separate tests included in the protocol (i.e., 2 tests in quiet, 2 tests using Masker A as the competing speech, and 2 tests using Masker B as the competing speech,

for each of the five languages). For each subject, all testing was performed during a single 2-hour test session. As such, the possibility existed that practice effects and/or fatigue could potentially alter the listeners' performance in the different conditions. However, order effects were controlled for in the current study by using a different random ordering of the 20 competing test conditions for each of the 50 subjects included in the study. Therefore, any effects of the testing order should have influenced all of the competing test conditions equally. Although order effects were controlled for by the design of the current study, future studies should consider the length of time in testing, especially if younger or older subject populations are used.

Conclusions

Despite these limitations, the current study makes an important contribution to existing research by providing new evidence related to aspects of selective auditory attention that have not been investigated by other studies. Specifically, the current study is the first to:

- Compare the masking effectiveness of different pairs of mixed-gender two-talker maskers,
- Evaluate how the vocal/speaking characteristics of masking talkers affect selective auditory attention performance in conditions where the competing speech is spoken in an unfamiliar language and is therefore not meaningful to the listeners at the sentence or word level,
- Assess effects of the language in which a competing speech signal is spoken by comparing selective auditory attention performance in more than two language conditions,

- Compare the masking effectiveness of two unfamiliar languages from the same non-native rhythmic category, and
- Compare the masking effectiveness of an unfamiliar language from the listeners' native rhythmic category (i.e., German) to the masking effectiveness of unfamiliar languages from non-native rhythmic categories (i.e., Spanish, French, and Japanese).

By investigating these previously unexplored aspects of selective auditory attention, the current study provides unique insight into how adult selective auditory attention performance is affected by differences between the vocal/speaking characteristics of the masking talkers and the underlying rhythmic structure of the language in which the competing speech is spoken. For example, for each of the five languages used in the current study, listeners' selective auditory attention performance was significantly affected by which male-female pair of masking talkers provided the competing speech signal. This finding suggests that the vocal/speaking characteristics of different pairs of male-female masking talkers can provide different degrees of masking effectiveness, regardless of whether the masking talkers are speaking in the listeners' native language or an unfamiliar language. In addition to the significant effect of masking talker pair (i.e., Masker A versus Masker B for each language), the current results revealed that the language of the competing speech also significantly influences listeners' selective auditory attention performance. For example, no significant differences were seen between listeners' performance in conditions where the competing speech was spoken in English, German, or Japanese, despite the fact that listeners could not understand the competing speech in the German or Japanese conditions. These results

indicate that meaning alone does not determine the masking effectiveness of a competing speech signal. Instead, the current results suggest that the underlying rhythm of the language of the competing speech has a greater impact on masking effectiveness than its meaning. This conclusion is based on the finding that listeners in the current study experienced significantly greater difficulty in ignoring competing speech spoken in languages from their native rhythmic category as compared to languages from the non-native syllable-based rhythmic category, regardless of whether the language from the native rhythmic category was meaningful (i.e., English) or not meaningful (i.e., German) to the listeners.

However, in contrast to this pattern, listeners in the current study treated the competing speech spoken in Japanese the same as competing speech spoken in English or German, despite the fact that Japanese has not historically been considered to be a stress-based language. Although Japanese has long been classified as having a mora-based rhythmic structure, results of the current study provide behavioral evidence that Japanese may not actually exhibit its own unique rhythm. This finding is consistent with evidence from one previous study which found that the acoustic-phonetic properties of Japanese overlapped with those of stress-based and syllable-based languages, depending on the type of phonetic measurements used to quantify language rhythm (Grabe & Low, 2002). However, additional behavioral and acoustic-phonetic research is now needed to further explore the similarities and differences between the rhythmic structure of Japanese and languages from the stress-based and syllable-based rhythmic categories.

Collectively, results of the current study extend evidence from previous studies by showing that the rhythm of the native language shapes not only how a listener processes

speech in quiet (e.g., Cutler et al., 1983; Cutler & Otake, 1994; Sebastián-Gallés et al., 2000), but also how he/she processes a target speech signal in conditions where other talkers are speaking simultaneously in the background. Additional research is now needed to explore how the language of the competing speech affects selective auditory attention performance for monolingual listeners from different age groups and different language backgrounds. The current study can also serve as a basis for future investigations into how the rhythmic structure of the native language(s) affects selective auditory attention among bilingual listeners. Together with the current results, evidence from such studies will lead to a better understanding of how exposure to the rhythmic structure of the native language(s) influences how speech is processed in competing conditions.

References

- Apple, Inc. (2008). iTunes 8.1 [Computer software]. Retrieved from <http://www.apple.com/itunes/download/>
- Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65, 33-69.
- Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2(1), 29-49.
- 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.
- Carter, W. (2008). *Department of psychology random order generator*. Retrieved December 14, 2008, from Oxford Brookes University, School of Social Sciences and Law Website: <http://www.brookes.ac.uk/schools/social/psych/order.html>
- Chermak, G. D. (1975). Intelligibility of monosyllabic words masked by linguistic strings. *Perceptual and Motor Skills*, 41, 587-593.
- Chermak, G. D., Vonnhoff, M. R., & Bendel, R. B. (1989). Word identification performance in the presence of competing speech and noise in learning disabled adults. *Ear and Hearing*, 10(2), 90-93.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, 123(1), 414-427.
- Cool Edit Pro* (Version 2) [Computer software]. (2002). Phoenix, AZ: Syntrillium Software Corporation.
- Crandell, C., & Smaldino, J. (2000). Room acoustics for listeners with normal-hearing and hearing impairment. In M. Valente, H. Hosford-Dunn, & R. J. Roeser (Eds.), *Audiology treatment* (pp. 601-637). New York: Thieme Medical.
- Crandell, C., & Smaldino, J. (2002). Room acoustics and auditory rehabilitation technology. In J. Katz (Ed.), *Handbook of clinical audiology* (5th ed., pp. 607-630). Philadelphia: Lippincott, Williams, & Wilkins.

- Crandell, C., Smaldino, J., & Flexer, C. (Eds.) (1995). *Sound-field FM amplification: Theory and practical applications*. San Diego, CA: Singular Publishing.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.
- Cutler, A. Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature*, 304, 159-160.
- Cutler, A. Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385-400.
- Cutler, A. Mehler, J., Norris, D., & Segui, J. (1989). Limits on bilingualism. *Nature*, 340, 229-230.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24, 381-410.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113-121.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824-844.
- Freeman, Y. S., & Freeman, D. E. (2006). *Teaching reading and writing in Spanish and English in bilingual and dual language classrooms* (2nd ed.). Portsmouth, NH: Heinemann.
- Freeman, Y. S., & Freeman, D. E. (2007). *La enseñanza de la lectura y la escritura en español e inglés en salones de clases bilingües y de doble inmersión* (2nd ed.). Portsmouth, NH: Heinemann.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America*, 109(5), 2112-2122.
- 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.
- Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *Journal of the Acoustical Society of America*, 121(2), 1040-1046.

- Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America*, 106(6), 3578-3588.
- Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*, 119(4), 2445-2454.
- Goldstein, B. A. (2007). Spanish speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 539-553). Clifton Park, NY: Thomson Delmar Learning.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythmic class hypothesis. *Papers in Laboratory Phonology*, 7, 1 – 16.
- 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 & Hearing*, 23(2), 159-165.
- Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, 7(3), 504-509.
- Hua, Z. (2007). Putonghua (modern standard Chinese) speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 516-527). Clifton Park, NY: Thomson Delmar Learning.
- Johnson, C. E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, 43(1), 144-157.
- Johnstone, P. M., & Litovsky, R. Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *Journal of the Acoustical Society of America*, 120(4), 2177-2189.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1-23.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- McLeod, S. (Ed.). (2007). *The international guide to speech acquisition*. Clifton Park, NY: Thomson Delmar Learning.

- Medwetsky, L. (2002). Central auditory processing. In J. Katz (Ed.), *Handbook of clinical audiology* (5th ed., pp. 495-509). Philadelphia: Lippincott, Williams, & Wilkins.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Minnesota State Department of Education. (n.d.). Chapter 7: Language assessment tools. Retrieved July 5, 2007, from Minnesota Department of Education Website: <http://education.state.mn.us/mdeprod/groups/SpecialEd/documents/Manual/003106.pdf>
- Moore, B. C. (1997). *An introduction to the psychology of hearing* (4th ed.). San Diego, CA: Academic Press.
- Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech*, 50(11), 77-99.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology*, 24(3), 756-766.
- Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43, 1-19.
- Neuman, A. C., & Hochberg, I. (1983). Children's perception of speech in reverberation. *Journal of the Acoustical Society of America*, 73(6), 2145-2149.
- Nilsson, M., McCaw, V., & Soli, S. (1996). *Minimum speech test battery for adult cochlear implant patients: User manual*. Los Angeles: House Ear Institute.
- Nilsson, M., Soli, S., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95(2), 1085 – 1099.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 258-278.
- Pickles, J. O. (1988). *An introduction to the physiology of hearing* (2nd ed.). San Diego, CA: Academic Press.
- Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *Journal of the Acoustical Society of America*, 119(3), 1597 – 1605.

- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265 – 292.
- Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech (L). *Journal of the Acoustical Society of America*, 118(3), 1274-1277.
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America*, 115(2), 813-821.
- Sebastián-Gallés, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, 62(4), 834-842.
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N (L). *Journal of the Acoustical Society of America*, 118(5), 2775-2778.
- Smit, A. B. (2007). General American English speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 128-147). Clifton Park, NY: Thomson Delmar Learning.
- Smith, C. L. (2005, September). *Durational prosody and topic organization: Differences between English and French*. Paper presented at the IDP05 International Symposium on Discourse-Prosody Interfaces. Retrieved May 20, 2008, from <http://aune.lpl.univ-aix.fr/~prodige/idp05/actes/smith.pdf>
- Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology*, 8, 71-80.
- Tun, P. A., O’Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, 17(3), 453-467.
- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, 121(1), 519-526.
- von Hapsburg, D., & Peña, E. D. (2002). Understanding bilingualism and its impact on speech audiometry. *Journal of Speech, Language, and Hearing Research*, 45, 202-213.

- Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, 24(6), 744-755.
- 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.

Appendix A

Subject Recruitment Flyer

TEXAS TECH UNIVERSITY
HEALTH SCIENCES CENTER
School of Allied Health Sciences
Department of Speech, Language and Hearing Sciences
Accredited by Council on Academic Accreditation
American Speech-Language-Hearing Association

SUBJECTS NEEDED: Adults Who Only Speak English

Subjects are needed for a research study that will see how adult listeners are able to understand English when there is noise made up of different languages playing in the background. To take part, individuals must:

- Be 18 to 55 years of age,
- Have English as their first language, with little or no familiarity with other languages,
- Be at least a high school graduate, and
- Have no history of a permanent hearing loss, auditory processing disorder, language disorder, learning disability, and/or dyslexia.

All testing will take place in a quiet room in the TTUHSC Department of Speech, Language, and Hearing Sciences. The testing will only take 2 hours to complete. Subjects will receive a free hearing screening as part of the study.

If you might be interested in helping with this research project, please contact the following individual for more information:

Leigh Ann Reel, Au.D., A.B.D.
Ph.D. Candidate
TTUHSC Department of Speech, Language, and Hearing Sciences
leigh.reel@ttuhsc.edu
(806) 743-5678

Texas Tech University HSC
IRB NUMBER: L08-079
IRB APPROVAL DATE: 10/1/2008

3601 4th Street, Stop 6073 | Lubbock, Texas 79430 | T 806.743.5678 | F 806.743.5670

An EEO/Affirmative Action Institution

Appendix B

Informed Consent Form

CONSENT TO TAKE PART IN A RESEARCH STUDY
Signed copy to be provided to subject or authorized representative

This form is not valid without the TTUHSC IRB seal on the bottom of each page.

This is a research study for subjects who voluntarily choose to take part. Please take your time to make a decision, and discuss the study with your family and friends if you wish.

STUDY TITLE: Selective Auditory Attention in Adults: Effects of Rhythmic Structure of the Competing Language

INVESTIGATOR(S) and CONTACT TELEPHONE NUMBER(S):

Candace Bourland Hicks, Ph.D., Audiology Program Director/Assistant Professor
Leigh Ann Reel, Au.D., Ph.D. Candidate
(806) 743-5678

INSTITUTION: Texas Tech University Health Sciences Center

1. Why is this study being done?

The goal of this study is to see how well adult listeners can pay attention to English sentences when there are different languages playing in the background.

2. What will happen during this study?

For this study, you will be asked to come to the TTUHSC Department of Speech, Language, and Hearing Sciences. The following will take place during the test session:

- 1) You will be asked to complete a brief questionnaire that will include questions about things such as your age, education, hearing and speech.
- 2) The researcher will look in both of your ears using an otoscope, a special light used to see the ear canal and eardrum clearly.
- 3) During each test:
 - a. You will be seated in a chair in a sound booth, a special room designed to be very quiet.
 - b. You will wear headphones or small foam earphones that will fit comfortably inside your ears.
 - c. The researcher will present sounds (beeps or sentences) through your headphones/insert earphones.
 - d. None of the sounds will be loud enough to hurt your ears/hearing.
- 4) The researcher will first screen your hearing in each ear using tones or "beeps" presented at different pitches. You will be asked to raise your hand, push a button, or say "yes" every time that you hear a "beep," even if it is so soft that you can barely hear it. If you do not pass the hearing screening, we will give you information on where you can have a full hearing test. You will not have to complete the rest of the test session.
- 5) The researcher will then give you a series of twenty-two short tests in which you will listen for English sentences and repeat the sentences that you hear.
- 6) On two tests, you will listen for the English sentences in quiet (i.e., there will not be any speech playing in the background). On the other 20 tests, you will listen for the English sentences and ignore the speech that you hear playing in the background:

Subject/ AR Initials _____

Texas Tech University HSC
IRB NUMBER: L08-079
IRB APPROVAL DATE: 10/1/2008
IRB EXPIRATION DATE: 3/4/2009

Page 1 of 3
Version 1

- a. On four tests, the speech playing in the background will be spoken in English.
 - b. On four tests, the speech playing in the background will be spoken in German.
 - c. On four tests, the speech playing in the background will be spoken in Spanish.
 - d. On four tests, the speech playing in the background will be spoken in French.
 - e. On four tests, the speech playing in the background will be spoken in Japanese.
- On each of these tests, the speech playing in the background will not be loud enough to be uncomfortable to you or to damage your hearing.
- 7) You can tell the researcher if you need to take a break.

3. How much time will this study take?

You will only be asked to attend one 2-hour test session. If the testing cannot be completed in one session for some reason (for example, you can only come for 1 hour at a time or you get too tired to finish in one session), a second test session will be scheduled.

4. What are the risks or discomforts to me if I join this study?

You may find listening in the background noise conditions to be a little annoying.

5. Are there any benefits to me if I take part in this study?

You will receive a free hearing screening if you take part in this study.

6. What other choices do I have if I don't take part in the research study?

Taking part in this study is your choice. You do not have to take part in this study. If at any time you decide not to be in the study, it will not affect any benefits or rights to which you are entitled. If you are currently a student, your grades will not be affected by whether or not you choose to take part in this study.

7. What about confidentiality and the privacy of my records?

We will keep your involvement in this research study confidential to the extent permitted by law. In addition to the staff carrying out this study, others may learn that you are in the study. This might include federal regulatory agencies, Texas Tech University Health Sciences Center (TTUHSC) representatives, and the TTUHSC Institutional Review Board (a committee that reviews and approves research). These people may review and copy records involving this research.

Study results that are used in publications or presentations will not use your name.

8. Will I receive anything for taking part in this research study?

You will receive a small gift bag with homemade snacks and/or other small tokens of appreciation (e.g., pens, pencils, notecards, etc.).

9. Does anyone on the research staff have a personal financial interest in this study?

No

10. What if I am hurt by participating in this study?

It is very unlikely that you will be hurt by participating in this study, but if so, Texas Tech University Health Sciences Center and its affiliates do not offer to pay for or cover the cost of medical treatment for research related illness or injury. No funds have been set aside to pay or reimburse you in the event of such injury or illness.

Subject/ AR Initials _____

Page 2 of 3

Version 1

Texas Tech University HSC
IRB NUMBER: L08-079
IRB APPROVAL DATE: 10/1/2008
IRB EXPIRATION DATE: 3/4/2009

11. What if I have questions?

For questions about this study, contact the Investigator, CANDACE HICKS, at (806) 743-5678.

If you would like to speak to someone who is not involved in the study about your rights as a participant, research-related injuries, or any other matter related to the study, you can call the TTUHSC Research Protection Hotline: 1-800-396-0918.

Your signature indicates that this research study has been explained to you, that you've been given the opportunity to ask questions, and that you agree to take part in this study. You will be given a signed copy of this form.

Printed Name of Subject

Signature of Subject

Date

Time

Signature of Parent/Guardian
or Authorized Representative

Date

Time

_____ Subject was unable to read and understand the written consent.

Signature of Witness to Oral Presentation

Date

Time

I have discussed this research study with the subject and his or her authorized representative, using language that is understandable and appropriate. I believe I have fully informed the subject of the possible risks and benefits, and I believe the subject understands this explanation. I have given a copy of this form to the subject.

Signature of authorized research personnel who
conducted the informed consent discussion

Date

Time

Subject/ AR Initials _____

Page 3 of 3

Version 1

Texas Tech University HSC
IRB NUMBER: L08-079
IRB APPROVAL DATE: 10/1/2008
IRB EXPIRATION DATE: 3/4/2009

Appendix C

Subject Questionnaire

SUBJECT QUESTIONNAIRE

STUDY TITLE: Selective Auditory Attention in Adults: Effects of Rhythmic Structure of the Competing Language

SUBJECT CODE: _____

The following questions ask you to provide general information about yourself. This information will be used to make sure that individuals participating in the study are similar. Please answer all of the following questions.

If you have any questions about the study or the questionnaire, you can contact Candace Hicks or Leigh Ann Reel at (806) 743-5678.

1. How old are you?

2. What is your gender? Please check the appropriate box.

- ☐ Male
- ☐ Female

3. What is the highest educational degree that you have obtained? Please check the appropriate box.

- ☐ High school graduate
- ☐ Undergraduate/baccalaureate degree
- ☐ Master's degree
- ☐ Doctoral degree
- ☐ Other:

4. Please list all countries in which you have lived.

**5. Have you ever been diagnosed with any of the following disorders?
Please check all boxes that apply.**

- ☐ Permanent hearing loss
- ☐ Auditory processing disorder (APD)
- ☐ Language disorder
- ☐ Learning disability
- ☐ Dyslexia
- ☐ No, I have never been diagnosed with any of these disorders.

6. Is English your first language?

- ☐ Yes
- ☐ No

7. Can you speak or understand any languages other than English?

- ☐ Yes
- ☐ No

If you answered "Yes" to Question 7, please answer the remaining questions (Questions 8-11). If you answered "No" to Question 7, you do not have to answer the remaining questions (Questions 8-11).

8. List all languages other than English that you are able to speak or understand.

9. When and where did you learn to speak a language (or languages) other than English (e.g., foreign language classes, time spent living in another country, etc.)?

****10. Please circle the number of the sentence that best describes your fluency in speaking a language other than English.**

- 1 My speech is so halting and fragmentary that conversation is virtually impossible.
- 2 My speech is usually hesitant, and I often give up due to language limitations. I mainly give one or two word utterances.
- 3 My speech in conversation is frequently disrupted by my need to search for the correct manner of expression. I use short phrases and sentences.
- 4 My speech in conversation is generally fluent, with occasional lapses while I search for the correct manner of expression.
- 5 My speech in conversation is fluent and effortless, approximately like that of a native speaker.

****11. Please circle the number of the sentence that best describes your ability to understand a language other than English.**

- 1 I understand little or no conversation.
- 2 I have great difficulty following what is said. I can comprehend only conversation spoken slowly and with frequent repetition and rephrasing.
- 3 I understand most of what is said in conversations at slower than normal speed with some repetition and rephrasing.
- 4 I understand conversation, although occasional repetition and rephrasing may be necessary.
- 5 I understand conversation in a variety of settings.

****Questions 10 and 11 were adapted from the Minnesota Modified Student Oral Language Observation Matrix (MN-SOLOM) (Minnesota State Department of Education, 2003). The MN-SOLOM (2003) is an adaptation of the Student Oral Language Observation Matrix (SOLOM) (Ortiz, 1995).**

Minnesota State Department of Education (2003). *Minnesota Modified Student Oral Language Observation Matrix*. Retrieved February 10, 2008, from <http://education.state.mn.us/mdeprod/groups/Assessment/documents/Manual/000424.pdf>

Ortiz, A. (1995). *Teacher observation student oral language observation matrix: Classroom supplement*. Sacramento, CA: California State Department of Education.

Appendix D

Hearing Screening Form

Selective Auditory Attention in Adults:
Effects of Rhythmic Structure of the Competing Language

Pure-Tone Hearing Screening (15 dB HL)

Subject Code: _____

Date: _____

Ear	250 Hz	500 Hz	1000 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz	8000 Hz
Right								
Left								

Otoscopic Results:

Right: _____

Pass _____

Left: _____

Fail _____

Appendix E

Sample Hearing in Noise Test (HINT) Sentence Lists

Name _____ Processor _____ Date _____

HEARING IN NOISE TEST

NOTE: Scoring is based on the number of words repeated exactly correct. Words in parentheses are to be scored as correct if *either* word is given in the response.

List 1	Words correct
1. (A/The) boy fell from (a/the) window.	_____/6
2. (A/The) wife helped her husband.	_____/5
3. Big dogs can be dangerous.	_____/5
4. Her shoes (are/were) very dirty.	_____/5
5. (A/The) player lost (a/the) shoe.	_____/5
6. Somebody stole the money.	_____/4
7. (A/The) fire (is/was) very hot.	_____/5
8. She's drinking from her own cup.	_____/6
9. (A/The) picture came from (a/the) book.	_____/6
10. (A/The) car (is/was) going too fast.	_____/6

WORDS correct: _____/53

Percent correct: _____%

Sentence level _____ Noise level _____ S/N ratio _____

Name _____ Processor _____ Date _____

HEARING IN NOISE TEST

NOTE: Scoring is based on the number of words repeated exactly correct. Words in parentheses are to be scored as correct if *either* word is given in the response.

List 2	Words correct
1. (A/The) boy ran down (a/the) path.	_____/6
2. Flowers grow in (a/the) garden.	_____/5
3. Strawberry jam (is/was) sweet.	_____/4
4. (A/The) shop closes for lunch.	_____/5
5. The police helped (a/the) driver.	_____/5
6. She looked in her mirror.	_____/5
7. (A/The) match fell on (a/the) floor.	_____/6
8. (A/The) fruit came in (a/the) box.	_____/6
9. He really scared his sister.	_____/5
10. (A/The) tub faucet (is/was) leaking.	_____/5
WORDS correct: _____/52	
Percent correct: _____%	

Sentence level _____ Noise level _____ S/N ratio _____
--

Name _____ Processor _____ Date _____

HEARING IN NOISE TEST

NOTE: Scoring is based on the number of words repeated exactly correct. Words in parentheses are to be scored as correct if *either* word is given in the response.

List 3	Words correct
1. They heard (a/the) funny noise.	_____/5
2. He found his brother hiding.	_____/5
3. (A/The) dog played with (a/the) stick.	_____/6
4. (A/The) book tells (a/the) story.	_____/5
5. The matches (are/were) on (a/the) shelf.	_____/6
6. The milk (is/was) by (a/the) front door.	_____/7
7. (A/The) broom (is/was) in (a/the) corner.	_____/6
8. (A/The) new road (is/was) on (a/the) map.	_____/7
9. She lost her credit card.	_____/5
10. (A/The) team (is/was) playing well.	_____/5
WORDS correct: _____/57	
Percent correct: _____%	

Sentence level _____ Noise level _____ S/N ratio _____
--

Appendix F

Recruitment Flyer for Language Recording Participants

TEXAS TECH UNIVERSITY
HEALTH SCIENCES CENTER
School of Allied Health Sciences
Department of Speech, Language and Hearing Sciences
Accredited by Council on Academic Accreditation
American Speech-Language-Hearing Association

VOLUNTEERS NEEDED: Native Speakers of Languages Other Than English

The TTUHSC Department of Speech, Language, and Hearing Sciences is needing audio recordings of native speakers of different languages (e.g., French, German, Japanese, Italian, Spanish). Audio recordings will be made as talkers read a short passage in their native language. Each talker will only be asked to attend one 60-minute recording session.

If you are a native speaker of a language other than English and might be interested in helping with this project, please contact the following individual for more information:

Leigh Ann Reel, Au.D.
Ph.D. Candidate
TTUHSC Department of Speech, Language, and Hearing Sciences
leigh.reel@ttuhsc.edu
(806) 743-5678

The TTUHSC Department of Speech, Language, and Hearing Sciences would greatly appreciate any assistance you can offer with this project.

3501 4th Street, Stop 6073 Lubbock, Texas 79430 | T 806.743.3678 | F 806.743.5675

An Equal Opportunity Institution

Appendix G

Questionnaire for Language Recording Participants

**Selective Auditory Attention Performance in Competing Speech Spoken
in Different Languages:**
Questionnaire for Language Recording Participants

Code Number: _____

Age: _____

Native Language: _____

1. Where were you born? Please list the city, state, and country as appropriate.

2. Do you speak a particular dialect of your native language? If so, please specify the dialect below. If you are a native speaker of more than one dialect, please list both dialects, and specify which one you used when reading for the language recording.

3. What did you read for the language recording?

4. How long have you lived in the United States?

5. What is the highest educational degree that you have obtained?

6. When and where did you learn to speak English?

7. How many years have you spoken English?

8. Please estimate the percentage of time that you speak your native language each day.

9. Please estimate the percentage of time that you speak English each day.

10. Would you characterize yourself as being a “fluent/proficient” speaker of your native language? (This question is particularly important for those individuals who primarily speak English now and have not used their native language on a regular basis for many years.)

11. Did you have any speech-language problems growing up?

Appendix H

Sample Scoring Forms

Condition Order: 1

Selective Auditory Attention in Adults: Effects of Rhythmic Structure of the Competing Language

Subject Code: _____ Condition #: Quiet 1 Date: _____

HINT List 12

Presentation Level (in dB HL)

4 dB
Change	1. (A/The) boy went to bed early.	_____
	2. (A/The) woman cleaned her house.	_____
	3. (An/The) sharp knife (is/was) dangerous.	_____
	4. (A/The) child ripped open (a/the) bag.	_____
	5. They had some cold cuts for lunch.	_____

2 dB
Change	6. She's helping her friend move.	_____
	7. They ate (a/the) lemon pie.	_____
	8. They (are/were) crossing (a/the) street.	_____
	9. The sun melted the snow.	_____
	10. (A/The) little girl (is/was) happy.	_____
	11.	_____

Total for Sentences 5 – 11 = _____

Mean for Sentences 5 – 11 = _____

SNR (Mean – 65 dB) = _____

Condition Order: 3

Selective Auditory Attention in Adults: Effects of Rhythmic Structure of the Competing Language

Subject Code: _____ Condition #: 13 Date: _____

HINT List 9

Presentation Level (in dB HL)

4 dB Change	1. School got out early today.	_____
	2. (A/The) football hit (a/the) goalpost.	_____
	3. (A/The) boy ran away from school.	_____
	4. Sugar (is/was) very sweet.	_____
	5. The two children (are/were) laughing.	_____
<hr/>		
2 dB Change	6. (A/The) fire truck (is/was) coming.	_____
	7. Mother got (a/the) sauce pan.	_____
	8. (A/The) baby wants his bottle.	_____
	9. (A/The) ball broke (a/the) window.	_____
	10. There (is/was) a bad train wreck.	_____
	11.	_____
<hr/>		
Total for Sentences 5 – 11 =		_____
Mean for Sentences 5 – 11 =		_____
SNR (Mean – 65 dB) =		_____

Appendix I

Instructions Given Prior to Testing in the Quiet Conditions

Okay, now we're going to do a series of short tests that will measure how well you can understand English sentences when they are spoken very softly in quiet and in background noise. On all of the tests, you will hear the same man reading one sentence at a time. The loudness of the man's voice will change during the testing, and sometimes it will be very soft. On all of the tests, your job is to repeat everything you hear the man say, even if you can only hear part of the sentence. It's okay to guess if you're not exactly sure what he said. Just be sure that you speak clearly and loud enough so that I can hear what you said.

There will be a significant pause between each sentence that the man says, so don't worry if you don't hear him say anything for a while. I just need you to listen closely throughout each pause so that you'll be ready when the man says the next sentence. Most of the sentences will only be presented one time, but at the beginning of each test, the first sentence will be repeated, a little bit louder each time, until you're able to repeat that sentence correctly. This is the only time that the same sentence will be repeated during the testing.

On all of the tests, the loudness levels of the sentences are chosen so that you will not be able to understand 100% of every sentence, so don't be discouraged if you can't repeat the entire sentence every time. No one is able to repeat all of the sentences. On these first two tests, you will hear the man reading one sentence at a time in quiet. There will not be any noise playing in the background during these two tests, so all you have to

do is repeat every sentence that you hear the man say. Do you have any questions?

Okay, then we'll go ahead and do the first two tests.

Appendix J

Instructions Given Prior to Testing in the Competing Conditions

Okay, now we're going to continue with the rest of the testing. On all of the remaining tests, you will hear the same man reading one sentence at a time, but at the same time, you will also hear another male and a female talking continuously in the background. On some tests, these background talkers will be speaking English, but on other tests, the background talkers will be speaking another language. During each test, the loudness of the background talkers will stay the same, but the loudness of the man who is reading the sentences will change.

On each of the tests, your job is to try and ignore what the two background talkers are saying, and listen for the man who is reading the sentences one at a time. When you hear the man say a sentence, I want you to repeat what you heard, even if you were only able to understand part of the sentence. Remember that there will be a significant pause between each sentence that the man says, so don't worry if you only hear the background talkers for a while. Just keep listening closely throughout each pause so that you'll be ready when the man says the next sentence.

Each test will only last about 4 minutes, but all together the tests will take about 1.5 hours to complete. We'll move continuously from one test to the next without pausing, so please let me know if you need to stop the testing and take a break at any time. There's no penalty for taking a break! However, if we're in the middle of a test when you tell me that you need a break, I'll wait until we finish that test, and then, I'll stop the recording. Do you have any questions about anything before we begin? Okay, then we'll get started.

Appendix K

Results of Additional Statistical Analyses

One-Way ANOVA

A one-way ANOVA was performed without the distinction between Masker A and Masker B to confirm the significant main effect of language which was revealed by the repeated measures ANOVA. Masker A and Masker B sSRT scores for each language were included in the analysis, but no distinction was made between the two types of scores. As such, a total of 500 mean sSRT scores were analyzed, including the Masker A mean and the Masker B mean for each of the 50 subjects in each of the five language conditions. Results of the one-way ANOVA revealed a significant relationship between the language of the competing speech and the listeners' mean sSRT scores, $F(4, 495) = 29.13, p < .01$. Means for each language were compared using the Tukey HSD test. Results revealed the same pattern of results that was seen with the repeated measures ANOVA, including:

- No significant difference between the mean scores for the Spanish and French conditions ($p = .83$),
- No significant difference between the mean scores for the English, German, and Japanese conditions (English-German, $p = .56$, English-Japanese, $p = 1.00$, German-Japanese, $p = .55$), and
- Significant differences between all other pairs of languages, including English and Spanish ($p < .01$), English and French ($p < .01$), German and Spanish ($p < .01$), German and French ($p < .01$), Japanese and Spanish ($p < .01$), and Japanese and French ($p < .01$).

Univariate ANOVA (One Covariate)

Due to the significant differences noted between the Masker A and Masker B means for each of the five languages, a univariate ANOVA was also performed with language as the within-subjects factor and the difference between the Masker A mean and the Masker B mean as the covariate. Results revealed a significant main effect of language, $F(4, 244) = 38.47, p < .01$, indicating that a significant relationship remained between the language of the competing speech and the mean sSRT scores, even when the significant differences between Masker A and Masker B were taken into account for each language. Post hoc comparisons were performed using the Bonferroni adjustment to identify any significant differences between the overall mean sSRT for each of the five languages when the Masker A-Masker B difference was used as the covariate. Results revealed the same pattern of results that was seen with the repeated measures ANOVA and the one-way ANOVA (without the distinction between Masker A and Masker B), including:

- No significant difference between the mean score for the Spanish and French conditions ($p = 1.00$),
- No significant difference between the mean score for the English, German, and Japanese conditions (English-German, $p = 0.74$, English-Japanese, $p = 1.00$, German-Japanese, $p = 0.86$), and
- Significant differences between all other pairs of languages, including English and Spanish ($p < .01$), English and French ($p < .01$), German and Spanish ($p < .01$), German and French ($p < .01$), Japanese and Spanish ($p < .01$), and Japanese and French ($p < .01$).