The Dissertation Committee for Sangeeta C. Kamdar certifies that this is the approved version of the following dissertation:

Sensory and Cognitive Contributions to Speech Perception in Background Noise in Young Adults

Committee:

Craig Champlin, Supervisor

Chang Liu

Julia Campbell

Joan Balash

Rajka Smiljanic

Sensory and Cognitive Contributions to Speech Perception in Background Noise in Young Adults

by

Sangeeta C. Kamdar

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin
December 2018

Dedication

This work is dedicated to…

My husband, Chris. The best human in all the world. The completion of this work would not be possible without his unconditional love and unwavering support. I won the lottery in the life partner department.

My son, Max. You have made me stronger, better, and more fulfilled than I could have ever imagined. Truly, I wrote this story for you, but when it began, I had not realized just how important you (and your future sibling) would be to its completion.

Acknowledgements

There are many people who deserve thanks for the completion of this work. They have helped guide, challenge, and motivate me during my journey.

I am extremely fortunate to have Dr. Craig Champlin as a mentor and advisor. Dr. Champlin is the best advisor and one of the smartest people I know. I am eternally grateful for his scientific advice and knowledge. He is the best role model for a researcher, mentor, and teacher and I would not be where I am today without his support and encouragement. I am also grateful for Joan Balash. I think fondly of my time in her office as an early graduate student. Joan always provided insightful discussions and suggestions about the research and reminded me at times when I needed it the most, “to keep my eye on the prize.” I would like to acknowledge the statistics department, and specifically Erika Hale, who very patiently guided me through the statistics of my project. I would also like to thank the members of my Ph.D committee, Doctors Chang Liu, Julia Campbell, and Rajka Smiljanic for their encouragement and insight.

A good support system is essential to thriving in a graduate program and I am forever grateful to Dr. Amanda Zappler and Dr. Angela Carey for their endless support and positivity. I am lucky to have them as mentors, colleagues, and friends. To my wonderful and loving friends and family, my parents and my in-laws, I do not know how any of this would have been possible with out you.

Sensory and Cognitive Contributions to Speech Perception in Background Noise in Young Adults

Sangeeta C. Kamdar, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Craig A. Champlin

Abstract: In the present study, speech perception in background noise was investigated by asking participants to perform listening tasks in two target speech contexts. Each target speech context was presented in four listening environments. External factors within each listening environment were manipulated, and various internal factors of the listener were subsequently measured. The primary aim of this study was to determine how well sensory and cognitive functions predict speech perception in noise outcomes for various target speech contexts presented in different types and levels of background noise. While this study did not reveal a significant effect for seven of the nine sensory and cognitive variables, this does not suggest that speech perception in noise does not involve a broad range of sensory and cognitive abilities. It's possible that even when individual difference among listeners exist, the sensory and cognitive abilities of most young normal hearing individuals are suitable for the purpose of speech perception in noise.

Table of Contents

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

Speech perception is a complex process involving multiple factors (Anderson, Parbery-Clark, Yi, & Kraus, 2011, Edwards, 2016). Some of these factors are external to the listener and include such things as speaker traits and environmental characteristics (Mattys, Davis, Bradlow, & Scott, 2012). Other factors are internal to the listener and include sensory and cognitive abilities. Further, these external and internal factors interact differentially according to the demands of the listening environment (Edwards, 2016). As a result, listener 'success' (i.e., accurately perceiving the message) depends on both the factors themselves and their interactions (Schneider & Pichora-Fuller 2000; Wingfield, 2016).

In the present study, speech perception in background noise was investigated by asking participants to perform listening tasks in two target speech contexts. Each target speech context was presented in four listening environments. External factors within each listening environment were manipulated, and various internal factors of the listener were subsequently measured. The external factors within the environment were selected because they have been shown to affect speech perception in noise. These factors include target speech context, background noise type and level. Target speech context varied from single words to sentences, background noise type ranged from speech noise to two-talker babble, and background noise level varied from low to high.

The internal factors that may influence speech perception in noise fall into two categories – sensory and cognitive (Edwards, 2016). Sensory abilities are related to the senses and, in this case, specifically to the ears and the sense of hearing (e.g., Edwards, 2016; Schneider & Pichora-Fuller, 2000). These functions are thought to be properties of the auditory periphery (cochlea and sub-cortical pathways) and thus are often referred to as 'bottom-up' processes (e.g., Edwards, 2016).

A listener's sensory abilities that were measured in this study include sound detection and discrimination. Signal detection is essential as a stimulus must be audible before it can be further analyzed and perceived. Discrimination reflects a listener's ability to break down sound into its constituent elements. The primary types of auditory discrimination are associated with the physical features of sound. Critical analytical skills for speech perception are frequency and temporal discrimination. Frequency discrimination requires the listener to reliably distinguish sounds that differ by small amounts in frequency, while temporal discrimination or acuity involves the differentiation of small deviations in duration.

Cognitive abilities are the mental activities that relate to how a listener acquires information (Edwards, 2016) and originate in cortical regions of the brain. Cognitive processing can be referred to as 'top-down' processing, which initiates with our thoughts, and flows down to lower-level functions, such as sensations. A listener's cognitive abilities that are thought to influence speech perception in noise include attention, working memory, language and processing speed (e.g., Dryden, Allen, Henshaw, & Heinrich, 2017; Edwards, 2016). Working memory is responsible for temporarily storing information before it is either dismissed or transferred to long-term memory (e.g., Dryden et al., 2017; Baddeley, 1992). Attention is the process of selectively concentrating on key aspects of incoming information while ignoring other features (Dryden, et al., 2017). Language refers to a shared set of rules related to making up meaningful words that allow an individual to express thoughts (Dryden et al., 2017), and finally, processing speed reflect to the amount of time it takes an individual to do a mental task (Dryden, 2017).

In summary, each person possesses a unique set of sensory and cognitive abilities that enable the perception of speech in noise. However, it also seems likely that a given listener's ability to process speech in noise is not static. The pattern could change gradually over time brought about by developmental factors or aging. Additionally, the pattern could shift rapidly depending on the demands of the listening task.

Clinical tests of speech perception in noise measure a listener's ability in a limited range of conditions and target recognition abilities only. These evaluations provide a partial view into a listener's speech in noise performance and highlight the need to evaluate ecologically valid test environments that reflect more accurately the reality of everyday listening experiences. A better understanding of the relationship between behavioral measures of listening and cognitive abilities will enable clinicians to be more informed in their choice of clinical speech in noise measures and the sensory and cognitive listener abilities that modulate speech in noise performance.

Thus, the primary aim of this study was to investigate a listener's sensory and cognitive abilities that may contribute to speech perception in noise. I was interested in determining what sensory and cognitive factors best predict speech perception in noise and whether the set changes with variations to the listening environment. This study was the first to our knowledge to measure speech perception in noise by systematically varying the target speech context, noise type, and noise level, while also measuring a range of sensory (sound detection and discrimination) and cognitive (attention, working memory, language, and processing speed) abilities to examine individual differences in speech in noise performance. The external factors related to the speaker and the background were manipulated in each listening environment to reflect a range of frequently encountered listening environments.

CHAPTER 2. LITERATURE REVIEW

A listener calls upon a unique set of abilities to perceive speech in noise. These abilities are composed of both sensory and cognitive factors that are internal to the listener (Edwards, 2016). Additionally, a listener perceives speech in various listening environments. The characteristics of these listening environments are external to the listener (Mattys, Davis, Bradlow, & Scott, 2012).

2.1 External Factors

External factors consist of variations to the type and level of speech (Tamati, Gilbert, & Pisoni, 2013). They also include speaker characteristics that result in reduced intelligibility, such as accented speech and speech disfluencies (Tamati et al., 2013; Mattys et al., 2012). Additionally, external factors include degradations to the speech signal caused by the environment, such as the introduction of background sounds and competing talkers (Mattys et al., 2012). The discussion below will review these factors by dividing them into two categories, speaker effects and environment effects, and focus on sub-factors within each category that are most pertinent to audiology.

2.1.1 Speaker effects

Speaker effects include variations to speech context, along with differences in speech quality (Mattys, et al., 2012). Speech context refers to the type of speech and can vary along a continuum, while the quality of speech can be disrupted by a speaker's accent or by disfluent speech (Tamati et al., 2013; Mattys et al., 2012).

Target Speech Context. The speech signal can vary in complexity from an individual phoneme to a single word or a complete sentence (Tamati et al., 2013). The present study focuses on words and sentences as these speech signals convey the meaning of language and are typically used as target speech stimuli in clinical measures of speech perception in noise (Fullgrabe and Rosen, 2016).

In a laboratory environment, one way to investigate how single words and sentences are understood is to have a group of subjects, such as young normal hearing listeners, listen to word and sentence lists at different intensity levels (MacPherson & Akeroyd, 2014). The average percent correct score of these lists is plotted as a function of intensity, and the graph is called a performance-intensity (PI) function (MacPherson & Akeroyd, 2014). The PI function for speech recognition illustrates a listener's ability to identify speech as a function of its intensity and demonstrates two important parameters of the stimulus: the threshold (which is the stimulus level required for a particular level of performance, e.g., 50%), and the slope (which is the rate at which a listener's performance increases with changes in the stimulus level) (MacPherson & Akeroyd, 2014). For normal hearing listeners, the slope of the function shows that the amplitude of speech varies over a range of about 30 to 40 dB SPL (MacPherson & Akeroyd, 2014). As the speech level is raised above a listener's threshold, their recognition performance rises above zero (MacPherson & Akeroyd, 2014). For normal hearing listeners, the PI function shows that at about 40 dB SPL above a listener's threshold, maximum performance is reached (e.g., 100%) (MacPherson & Akeroyd, 2014).

Words and sentences impact the slope of a PI function differently. A sentence is composed of a series of words. The increased segment length of sentences provides the listener with greater acoustic information to identify the speech signal with (Tamati et al., 2013; Firszt, Reeder, & Skinner, 2008). Thus, in comparing the slope of a PI function for sentences to that of words, the function for sentences is steeper (e.g., a listener can achieve a particular level of performance with increasing stimulus level faster) than for words, and, words produced and heard in a sentence are typically identifiable at a lower intensity than words in isolation (Firszt et. al., 2008). PI functions provide Audiologists with information about how well different types of speech can be correctly identified as a function of intensity for different groups of listeners, and their relationship with word and sentence recognition serve as a foundation for clinical measures of speech audiometry.

Target speech context is important to consider when investigating speech perception in noise performance because single words and sentences have been shown to engage cognitive processes differently (Dryden et al., 2017; Heinrich, Henshaw, & Ferguson, 2015). For example, Heinrich and colleagues (2015) examined the relationship between cognition and speech perception in noise when the speech perception task ranged in complexity. The researchers varied target speech context (phonemes, words, sentences), background noise type (quiet, speech noise, 8-Hz modulated noise), and background noise or target speech level and found that the extent to which a listener's cognitive abilities correlate with speech perception measures was dependent on the complexity of the target speech context. The results of Heinrich et al. (2015) indicate that cognition is a predictor for speech in noise performance when the speech context is sentences, but not phonemes or words. In the present study, speech context varied from single words to sentences to explore how external and internal factors discussed in the next sections differentially effect word and sentence recognition.

Other Factors. Speech production converts a person's thoughts into speech. Variations in speech production include differences in speech quality and rate (Tamati et al., 2013; Mattys et al., 2012). Differences in speech quality include accented speech, disordered speech, and speech disfluencies (Tamati et al., 2013; Mattys et al., 2012). Accented speech results in deviations to the segmental and supra-segmental characteristics of the signal, whereas disordered and disfluent speech affects the rhythm, rate, and voice quality of speech (Mattys et al., 2012). While these effects contribute to overall speech intelligibility for a listener, variations in speech production are not included in typical measures of speech perception in noise and therefore beyond the scope of this review.

2.1.2 Environmental effects

As with speaker effects, environmental effects are external to the listener. This category refers to background sounds that are simultaneously present with the speech signal and contribute to difficulty perceiving speech (Tamati et al., 2013; Mattys et al., 2012). Types of background sounds include speech noise and babble, along with distortions caused by the environment, such as reverberation (Mattys et al., 2012). The discussion below will consider the type and level of background sounds as the focus of the present study is on clinical measures of speech perception in noise.

Background Type. The type of background sounds used in clinical speech in noise measures include speech noise or sounds that fluctuate over time, such as babble. Speech noise is white noise that has been filtered to match the long-term average speech spectrum and is frequently used as a source of masking in clinical environments (Tamati et al., 2013; Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014; Lidestam, Holgersson, & Moradi, 2014). In speech noise, spectral and temporal dips are filled in and the majority of the target speech is overlapped by the background sound (Lidestam, et al., 2014). Speech noise results in energetic masking. Energetic masking occurs in the cochlea (Van Engen, et al., 2014) and highlights the energetic effects of the masker over the signal (Hornsby, Ricketts, & Johnson, 2006; Lidestam, et al., 2014). In energetic masking, the background sound physically overlaps the speech signal in the environment, making the signal inaudible to the listener, thus, opportunities for detecting the target speech are reduced, impacting speech in noise performance.

Babble results in background sounds that are made by several speakers talking at the same time. Similar to speech noise, babble competes with target speech and masks the signal of interest (Hornsby, et al., 2006). However, babble also results in informational masking. Informational masking occurs when background sounds perceptually interfere with the speech signal (Hornsby, et al., 2006; Lidestam et al., 2014), for example, when a listener is trying to extract the target speech from several simultaneous competing speech signals at a noisy restaurant. Thus, informational masking highlights the interference cause by the informational component of this masker (Hornsby, et al., 2006). Distinguishing between energetic and information masking can be important because both exert different effects on sensory and cognitive systems (Zekveld, Rudner, Johnsrude, & Ronnberg, 2013; Mattys et al., 2012). For example, background sounds composed of competing talkers can introduce a significant amount of higher-level lexical interference, placing greater emphasis on a listener's cognitive load (Mattys et al., 2012). Additionally, Carhart and colleagues (1975) found competing talker backgrounds that are composed of two to three speakers are often more difficult for listeners than when the masker is stationary, due to the effects of information masking (Carhart, Johnson, & Goodman, 1975).

Unlike speech noise, babble produces temporal fluctuations within the waveform. Temporal fluctuations allow opportunities where the target speech overlaps with a dip in the level of the background sound (Lidestam et al., 2014;

Miller & Licklider, 1950). As a result, babble may allow a listener to acoustically glimpse the speech signal, thus positively impacting speech in noise performance (Lidestam et al., 2014).

As the number of talkers in babble noise increases beyond two or three, the amount of energetic masking produced by babble also increases (Van Engen, et al., 2014). This is because the temporal fluctuations within the background get filled in (Van Engen, et al., 2014). Therefore, the informational masking in babble is greatest when it is composed of relatively few speakers (Van Engen, et al., 2014). In this study, background type was varied from speech noise to two-talker babble to examine the effects of both energetic and information masking and their interactions with external and internal factors.

Background Level. The signal-to-noise ratio (SNR) describes the level of a target signal to the level of a background sound (McShefferty, Whitmer, & Akeroyd, 2016). The more favorable the SNR, the greater the level difference between the target speech signal and background sound, increasing speech perception in noise for a listener (McShefferty, et. al., 2016). If speech is masked by competing background sounds, such as when the level difference between the signal and background sound is low, then it will be more challenging for the listener to perceive the target signal due to the unfavorable SNR (Tamati et al., 2013; Mattys et al., 2012). Speech perception in noise tasks can adjust the level of the background sound or the level of the target signal to achieve favorable or unfavorable SNRs, and both approaches to setting SNR are used in research (Heinrich et al., 2015). Heinrich and colleagues (2015) examined the protocol difference in setting SNR and its relationship with cognition using the Digit Triple Test (DTT) as the DTT can be measured using both methods to establish SNR. Both measures of DTT (varying the noise level or varying the target speech signal) showed similar cognitive profiles. Therefore, Heinrich et al., (2015) concluded that both protocols of determining SNR place similar cognitive demands on a listener and are equally suitable for studying variations in SNR when cognitive load is of interest (Heinrich et al., 2015). In this study, background noise level was varied to observe its interaction with external and internal factors.

2.2 Internal Factors

The internal factors that affect speech perception can be divided into two categories – sensory and cognitive. Within the sensory domain, sound detection, and discrimination will be discussed. Cognitive abilities will include a discussion on working memory, attention, language, and processing speed.

2.2.1 Sensory Abilities

Sensory abilities are internal characteristics of a listener that are related to the sense of hearing (Edwards, 2016, Schneider & Pichora-Fuller, 2000) and reside in the auditory periphery (cochlea and sub-cortical pathways). Sensory abilities are related to bottom-up processing, which refers to the processing of sensory information as it is encoded by the auditory system.

Hearing Sensitivity. Hearing sensitivity is evaluated through a measure of hearing thresholds, called pure tone audiometry (Martin & Clark, 1997). During pure tone audiometry, an audiometer is used to produce sounds at various volumes and frequencies. The listener hears the sounds through earphones and responds by pressing a button. This test evaluates the softest sounds an individual can hear at frequencies most important for speech perception (Martin & Clark, 1997). In the present study, a listener's hearing sensitivity was measured to explore its effect on the external and internal factors.

Frequency Selectivity. Frequency selectivity refers to a listener's the ability to discriminate sounds by frequency (e.g., Oxenham, 2012; Katz, Medwetsky, Burkard, & Hood, 2009). When sound enters the cochlea, different frequencies within the sound stimulate different regions along the cochlea (Oxenham, 2012). This frequency-to-place mapping is referred to as the tonotopic arrangement of the cochlea and is maintained throughout the auditory pathway (Oxenham, 2012).

In addition to tonotopy, a healthy cochlea contains a series of overlapping band-pass filters, called auditory filters (Baer & Moore, 1993). Auditory filters are associated with spectral locations along the basilar membrane and determine the frequency selectivity of the cochlea (Oxenham, 2012; Baer & Moore, 1993). The shape of auditory filters is non-linear and is dependent on the level of the stimulus (Baer and Moore, 1993). As a result, an auditory filter's width along the length of the cochlea increases with increasing frequency and becomes more asymmetrical with increasing level (Baer & Moore, 1993). Auditory filters work by enhancing certain frequencies while attenuating others (Baer & Moore, 1993). For example, auditory band-pass filters allow a range of frequencies within the bandwidth of the filter to pass through, while preventing frequencies outside the cut-off frequencies of the filter.

The narrower the width of a listener's auditory filter, the better their ability to detect vocal tract resonances called formants (Liu & Kewley-Port, 2007). Formants are concentrations of spectral energy around a specific frequency in the speech spectrum (Fant, 1960). Detection of formant frequencies in speech perception help listeners identify vowels, and the acoustic details that are found between formant frequencies, called formant transitions, help listeners identify consonants (Winn, Edwards, & Litovsky, 2015). Successfully perceiving the spectral peaks of formants and the transitions between formants assists a listener in identifying acoustically similar phoneme pairs such as /b/-/d/ and /t/-/k/ (Winn, et al., 2015; Liu & Kewley-Port, 2007).

Whereas narrow auditory filter shapes are associated with better frequency selectivity and speech perception in noise, broadened auditory filter widths negatively impact the contrast between formant peaks and valleys (Baer & Moore, 1993; Winn et al., 2015; Liu & Kewley-Port, 2007). For example, broad filter widths allow closely spaced formants to fall within the same auditory filter so it becomes more difficulty for a listener to discriminate between vowels with similar formant frequencies (Liu & Kewley-Port, 2007). Additionally, even if formant frequencies are categorized into separate auditory filters, broadened filter widths smear spectral information that is important for consonant identification (Liu & Kewley-Port, 2007). Thus, wider auditory filters reduce the signal-to-noise ratio at the output of the filter, making formant peaks and transitions less clear to a listener, negatively impacting speech perception in noise (Liu & Kewley-Port, 2007). In the present study, a listener's frequency selectivity was measured to explore its effect on the external and internal factors.

Temporal Acuity. Temporal acuity refers to the ability to discriminate changes in stimuli over time, for example, detecting a brief gap between two stimuli or identifying that a sound is modulated in some way (Moore, 2008). A listener's temporal acuity may be influenced by their ability resolve both temporal envelope and temporal fine structure information (Moore, 2008). The temporal envelope is conveyed by the slow variations in amplitude of the speech signal over time, while temporal fine structure information is related to more rapid oscillations within the temporal waveform (Moore, 2008).

Several researchers have investigated the role of temporal envelope and fine structure information on speech perception by processing speech sounds so that they contain mainly envelope or fine structure information (Moore, 2008; Moore & Hopkins, 2008). Experiments using envelope speech have found that only a few bands are necessary to provide good speech perception in quiet, however more bands are required when speech is presented in background sounds (Moore, 2008).

While envelope cues are adequate to provide good speech intelligibility in quiet, they do not provide enough information in the presence of background sounds (Moore, 2008). Several investigators have suggested that temporal fine structure information is essential for speech perception in noise, especially when the noise is fluctuating, such as babble (Moore, 2008; Hopkins et al, 2008). For example, Hopkins and Moore (2008) examined speech recognition thresholds in babble in normal hearing listeners. Speech stimuli were filtered into 32 bands and noise-vocoded above the cut-off band, which varied in frequency. Noise-vocoded speech is both spectrally and temporally degraded and conveys speech information through envelope cues (Hopkins & Moore, 2008), such that filtered sound below the cut-off band contains intact temporal fine structure and envelope information, while only envelope information is provided in bands above the cut-off band (Moore, 2008). Normal-hearing listeners showed an improvement in speech recognition thresholds in noise as more temporal fine structure information was added into the signal, which is consistent with the notion that temporal fine structure information plays a significant role in the ability to identify speech in competing talker background (Hopkins and Moore, 2008). In the present study, a listener's temporal acuity was measured to evaluate its effect on the external and internal factors.

Measures of Frequency Resolution and Temporal Acuity. Methods to evaluate frequency resolution include classical psychophysical approaches such as tuning curves, filter shape, and critical band measurements. Similarly, psychophysical procedures to measure temporal acuity include gap detection, temporal modulation transfer functions, and temporal window measures. Standard psychophysical procedures that are used to measure frequency resolution and temporal acuity are traditionally used in laboratory settings and because of the extensive time requirements necessary to administer these measures are unsuitable for clinical use.

It is possible to measure frequency resolution and temporal acuity clinically by estimating the shape of a listener's auditory filter or temporal window (Unoki, Ito, Ishimoto, & Tan, 2006; Charaziak, Souza, & Siegel, 2012). For these measurements, a listener's threshold is obtained in noise and several notched noise and amplitude modulated (AM) noise conditions and the slope of the function for each measure is an estimate of a listener's auditory filter shape or temporal window shape, respectively (Unoki et al., 2006; Charaziak et al., 2012). These measures utilize equipment already found in audiology clinics, measurement techniques familiar to an Audiologist, and can be completed within a reasonable amount of time, which makes them clinically practical. In the present study, clinical approaches to measuring a listener's frequency resolution and temporal acuity were used rather than traditional psychophysical methods.

2.2.2 Cognitive Abilities

Cognitive abilities are the mental activities that relate to how a listener acquires information (Edwards, 2016). These functions reside in the brain. A listener's cognitive abilities that are thought to influence speech perception in noise include attention, working memory, language and processing speed (Dryden, et al., Edwards, 2016). Each of these functions is complex and considerable research has been devoted to understanding each one. Moreover, a comprehensive discussion of these abilities is beyond the scope of this review. A brief overview is given here.

Attention. Attention refers to an individual's ability to focus their attention on a particular stimulus while ignoring a range of other stimuli (Strait & Kraus, 2011). The human nervous system is constantly met with a wide variety of sensory input (Strait & Kraus, 2011). Fortunately, the brain has developed to allow for the modulation of neural activity based on competing demands, which helps an individual select the appropriate behavioral response to stimuli of greatest interest (Strait & Kraus, 2011; Janse, 2012). A subset of attention, called selective attention, is suggested to be responsible for this modulation and allows listeners to tune into a single voice and tune out other competing voices (Strait & Kraus, 2011).

Models of selective attention include filter models (McLeod, 2018). Filter models vary in their assumptions as to how far an unattended unit of information is processed and can be divided into early and late filter models (McLeod, 2018). The early filter model of attention first proposed by Broadbent (1958) suggests that stimuli are filtered at an early stage of processing, before processing the meaning of a word (McLeod, 2018). Stimuli with similar characteristics pass through a filter and are attended to while unrelated information is filtered out (McLeod, 2018). Because a listener has a limited ability to process information, the filter keeps the system from becoming overloaded (Broadbent, 1958; McLeod, 2018). The input that is not initially selected by the filter briefly remains in a sensory buffer, and if it is not processed, it decays rapidly (Broadbent, 1958; McLeod, 2018). Broadbent's model assumes that the filter rejects the unattended message at an early stage of processing (Broadbent, 1958).

Late selection models of attention suggest that information is selected after processing for word meaning. In these models, the monitor attends to all stimuli and acts like a filter that amplifies important information and attenuates unimportant information (McLeod, 2018). Attenuation models of attention like Treisman's (1964) model include an early selection filter that is similar to Broadbent's model; however, the filter in Treisman's model attenuates stimuli that are considered unimportant (Treisman, 1964, McLeod, 2018), but does not completely reject unimportant stimuli. The attenuation theory uses an idea of threshold: if the stimulus exceeds a threshold, it will pass through the filter and become attended to (McLeod, 2018). Thus, as an unattended channel includes weakly attended to information, to gain awareness to this information, it must pass a threshold, which is determined by the word's meaning (McLeod, 2018). Thus, in late filter models, the threshold for each word acts as a filter, and the filter is relying on the meaning of the word (McLeod, 2018).

A cognitive process within the framework of selective attention is inhibitory control (Neil, Valdes, & Terry, 1995). Inhibitory control allows an individual to ignore competing stimuli in the environment in order to select a more appropriate behavior that is directed towards completing a goal (Dryden et al., 2017; Janse 2012). Inhibitory control is thought to play a role in speech perception in noise (Dryden et al., 2017; Janse, 2012). For example, Janse (2012) investigated hearing loss and a measure of inhibitory control to predict listening performance for speech in quiet and in a competing talker background. Janse (2012) found the individual Stroop measure was significantly associated with speech perception performance in the two-talker babble condition. Thus, Janse (2012) concluded that poor inhibition increases a listener's susceptibility to competing background sounds. Additionally, during lexical access, poor inhibitory control makes it harder for a listener to select the appropriate target from their mental lexicon (Dryden et al., 2017).

In a review of the literature, Dryden and colleagues (2017) found a significant correlation between inhibitory control and speech perception in noise of .34. This association was pooled across all listeners (normal hearing and hearing impaired listeners) with the majority of studies using sentences as the target speech. Dryden and colleagues (2017) were not able to assess associations between inhibitory control and speech in noise on different masker types (i.e., unmodulated, modulated, and babble), as there was insufficient data ($n \leq 5$). Inhibitory control is measured through tests that require individuals to focus attention on a given stimulus while ignoring neighboring stimuli. In the present study, a listener's inhibitory control was measured to explore its effect on the external and internal factors.

Working Memory. Working memory is a cognitive process that stores and manipulates a limited amount of information over a brief period of time to complete a complex task (Baddeley, 1992). Working memory is responsible for: processing information across tasks and modalities, holding information in a short-term storage, manipulating information, and holding the products of that manipulation in the same short-term storage (Dryden et al., 2017).

A popular theory of working memory includes the multicomponent model proposed by Baddeley (Baddeley 1992; Dryden et al., 2017). This model of working memory is made up of a central executive system along with three sub-systems: the episodic buffer, the phonological loop, and the visuospatial sketchpad (Baddeley, 1992). The central executive system is suggested to have no retention capacity but controls all working memory processes through its attentional capacity (Baddeley, 1992), and is therefore thought to be responsible for the selection, initiation, and termination of processing routines within working memory (Baddeley, 1992). Additionally, the central executive system is also responsible for focusing, dividing, and switching attention. Following the central executive system, the episodic buffer is a limited capacity system that provides storage of information (Baddeley, 1992). The episodic buffer is also the interface between the information from perception and long-term memory (Baddeley, 1992). The visuospatial sketchpad and phonological loop are thought to be responsible for the temporary storage of information. The phonological loop is where verbal material is managed and the visuospatial sketchpad is where spatial material is managed (Baddeley, 1992).

The psychological construct of working memory is prominent in speech perception literature and has been incorporated into models of speech perception and cognition, such as the Ease of Language Understanding model (ELU) (Dryden et al., 2017). For example, when listening to a conversation, individuals continuously store and update auditory information that is spoken in real time, thus an intact working memory is critical for understanding speech (Olson & Campbell, 2012). ELU suggests that, in favorable listening conditions, language input is complete and matches phonological representations in the listener's mental lexicon. However, when language input is weak or distorted, such as in adverse listening environments, a mismatch may arise (Dryden et al., 2017). In this situation, explicit processing mediated by a listener's working memory capacity is required to match the distorted language input with that stored in the metal lexicon (Dryden et al., 2017). However it is debatable in ELU as to whether working memory is equally important in all listeners or only for those with degraded input (i.e., hearing impaired listeners) (Fullgrabe, 2016).

Dryden and colleagues (2017) reported a moderate association between working memory and speech perception in noise across all listeners ($r = .28$), regardless of hearing status. When the type of background masker was considered separately, significant correlations ranging between .26 and .39 were found for unmodulated noise, modulated noise, and babble (Dryden et al., 2017). Additionally, the review noted that the strength of the relationship between working memory and speech perception in noise increased as the amount of information masking produced by the background sound increased. This is inline with current evidence that suggests information masking may increase a listener's cognitive load (Mattys et al., 2012). However, in the review by Dryden and colleagues (2017), age and hearing status of the listener were not included as subdomains, and listeners with normal hearing and hearing impairment along with younger and older listeners were pooled together.

Working memory is measured through complex span tests (Dryden et al., 2017). These tests require participants to immediately store, recall, and/or sequence different visually and orally presented stimuli. For example, in these measures, pictures of different foods and animals are displayed on a screen with an audio recording of the object and the participant is asked to say the items back in order from smallest to largest. To make a task more challenging, the participant may also be asked to repeat back items in size order in addition to grouping items by category. In the present study, a listener's working memory was measured to explore its effect on the external and internal factors.

Linguistic ability. Gardner first proposed the theory of multiple intelligences in 1983 (Gardner & Hatch, 1989). The theory categorizes human intelligence into different modalities (Gardner & Hatch, 1989). Within Gardner's multiple intelligence theory is linguistic intelligence, which refers to an individual's ability to understand spoken and written language, as well as their ability to speak and write (Gardner & Hatch, 1989). There are four parts of Gardner's linguistic intelligence theory: phonology, syntax, semantics, and pragmatics (Gardner & Hatch, 1989). Phonology refers to how sounds are organized and used in language (Gardner & Hatch, 1989). Syntax refers to the grammatical structure of words and phrases to make sentences (Gardner & Hatch, 1989). Semantics refers to the meaning of words and pragmatics is the ability to interpret the intended meaning of words (Gardner & Hatch, 1989).

Semantic knowledge contributes to individual differences in vocabulary and can explain variability in speech recognition performance (Kaandorp, De Groot, Festen, Smits, & Goverts, 2015). Vocabulary conveys much of the meaning of language, and when an individual listens to speech, they are constantly matching the acoustic information they hear to internal representations of what they have heard before (Kaandorp et al., 2015; Dryden, et al., 2017). In this process, the number of words stored in a listener's mental lexicon, which is like a mental dictionary containing the meaning behind words, could be important. Thus, a listener's vocabulary knowledge can impact a their speech in noise performance. For example, in a study that explored the effect of language proficiency on speech perception in noise, Van Wijngaarden and colleagues (2002) found that non-native listeners need a 1- to 7 dB more favorable signal-to-noise ratio than native listeners to achieve 50% correct sentence recognition on a speech in noise task. Van Wijngaarden and colleagues (2002) concluded that the reduced ability of non-native speakers on speech in noise performance can be explained by less effective use of context clues such as semantic redundancy in this population because of their reduced language proficiency. Language knowledge is measured through tests of receptive vocabulary. In these tests, an individual listens to an audio recording of a word and then selects one of four pictures on a screen that best describe the word's meaning. In the present study, a listener's language knowledge was measured to explore its effect on the external and internal factors.

Processing Speed. Processing speed refers to the amount of time it takes an individual to do a mental task; and is related to how fast a person can understand and react to the information they receive (Dryden et al., 2017). Thus, processing speed can be thought of as the time between receiving a stimulus and responding to that stimulus. Processing speed is associated with speech perception due to the sequential nature of speech (Dryden et al., 2017) and it is suggested that a listener's processing speed becomes even more critical when the speech stimuli are linguistically complex or degraded in some way, such as when listening to speech in noise. In a review of the literature, Dryden and colleagues (2017) found a significant association between processing speed and speech perception in noise (r = .39). When examining target speech, the association between processing speed and speech in noise increased (r = .43) when sentences were used. The researchers noted there was insufficient data to compare processing speed and background type on speech in noise performance (n ≤ 5). Processing speed can be measured on individuals through tests that ask listeners to identify whether two visual patterns are the same or different. In the present study, a listener's processing speed was measured to explore its effect on the external and internal factors and speech in noise outcomes.

2.3 Purpose and Research Question

As mentioned previously, clinical measures of speech perception in noise assess a listener's recognition ability in a limited range of conditions using a single target (words vs. sentences), background type (speech shaped noise vs. babble), and background level (low vs. high). These assessments provide a partial view of a listener's speech in noise performance and emphasize the need to evaluate ecologically valid test environments that reflect more accurately the reality of everyday listening. Heinrich et al., (2015) examined cognition and speech perception in noise when the listening environment was systematically manipulated. However, their selection of background sounds did not include babble, which has been shown to engage cognitive processes differently when compared to speech noise (Mattys et al., 2012). Additionally, babble is a common source of interference in everyday listening. Moreover, Heinrich et al., (2015) did not include variations in background level, which is also a typical occurrence in everyday listening. Thus, the current study investigated speech perception in noise by systematically manipulating the listening environment with variations not only to the target speech context, but also the background type (speech noise vs. babble) and level (low vs. high) while also measuring a range of sensory (sound detection and discrimination) and cognitive (attention, working memory, language, and processing speed) abilities.

The primary aim of this study was to examine sensory and cognitive abilities that may contribute to individual differences in speech perception performance in four representative listening environments per target speech context (single words and sentences). To reflect ecologically valid listening experiences, I manipulated the noise type and level in each listening environment. Increasing the level of the background was expected to increase speech perception difficulty (i.e., reduced word recognition scores) in a straightforward manner. Specifically, low-level noise would yield higher speech perception scores than a high-level noise. Varying the type of background should also affect speech perception performance. Babble noise contains both energetic and linguistic characteristics and thus should cause greater perceptual difficulty than comparable level speech noise, which only has energetic features (Rosen, Souza, Ekelund, & Majeed, 2013). Thus, it was expected that speech noise would produce higher speech perception scores than babble noise when controlling for the overall energy level.

I was interested in determining what sensory and cognitive factors best predict speech perception in noise in the listening environments established in this study and whether these factors changed with variations to the listening environment. Therefore, the research question for this study was: What sensory and cognitive factors best predict speech perception in noise and does the set change with manipulations to the listening environment? Heinrich et al. (2015) reported that cognition (specifically: reasoning, working memory, and attention) is a predictor for speech in noise performance when the speech context is sentences. Therefore, I predicted a greater number of cognitive abilities would emerge as significant in sentence recognition relative to word recognition.

CHAPTER 3. METHODS

3.1 Participants

Participants included 32 monolingual adults (22 females, 10 males) with normal hearing sensitivity between the ages of 18 and 29 years recruited from the University of Texas at Austin, who reported difficulty-understanding speech in noise. Speech in noise ability was assessed using the Speech, Spatial, and Qualities of Hearing Scale (Gatehouse & Noble, 2004). The questionnaire was designed to measure a range of hearing disabilities across several domains with specific attention placed on hearing speech in a variety of competing contexts. Two subscales within the questionnaire were used on the participants: the speech hearing and qualities of hearing subscales. Each subscale included about 15 questions and participants rated their hearing ability with a score out of 10 with higher scores reflecting greater ability. For the speech hearing subscale (SHS), the average score was a 6.4 and for the qualities of hearing subscale (QHS) the average score was a 7.5. It appears Gatehouse & Noble (2004) published some results for listeners with slight hearing impairment. The mean for the SHS was 4.4 (SD 1.4) and for the QHS was 6.3 (SD 1.3).

Participants were not included in the study if a history of neurological, hearing loss, or auditory processing disorder was reported. Prior to participation in the study, informed consent was obtained and each participant received a hearing evaluation and tympanometry to confirm normal auditory function.

3.2 Instrumentation & General Procedures

Data collection was completed in a single session. The session consisted of standard audiometric testing, followed by speech perception, and sensory and cognitive measurements.

3.2.1 Measurement Overview

Standard Audiometric Testing. Assessment of auditory status was undertaken prior to administering the experimental conditions to confirm normal auditory function. This evaluation consisted of an otoscopic exam, pure-tone audiometry, and tympanometry. Pure-tone audiometry was completed on a GSI 61 audiometer, which was calibrated to the ANSI S3.7, 2016 standard. Visual inspection of the calibration sheet showed no more than a .5 dB error at each test frequency. Measures of hearing sensitivity were completed with TDH 39 supra-aural earphones in a sound proof booth which met ANSI standard S3.1, 2013 for the octave frequencies between 250 Hz and 8000 Hz for the right and left ear. Tympanometry was completed using a TympStar Immittance bridge calibrated according to the ANSI S3.39, 1987 standard for the right and left ear.

Speech Perception Measurement. The experimental conditions included measures of speech perception in noise, and sensory and cognitive function. Each participant's data collection began with speech perception in noise measurements. Speech perception in noise testing was completed in a sound treated booth with a TDH 39 earphone. Speech targets and noise were delivered via a GSI 61 audiometer and a media player (FuBar and VLC media player), which were connected to a Dell desktop computer. Within speech perception in noise measurements, isolated word and sentence recognition alternated across subjects. After speech perception in noise data collection was completed, sensory and cognitive data collection began.

Sensory Measurement. Sensory measurements were completed in a sound treated booth with an Asus computer routed to an audiometer and TDH 39 earphone. Sensory stimuli were delivered via a GSI 61 audiometer and media player (FuBar and VLC media player) on the computer. Speech perception in noise and sensory measurement conditions included calibration prior to taking these measurements on each participant.

Cognitive Measurement. Cognitive measures were completed on an Apple iPad with iOS 10.3.3 and the NIH Cognitive Toolbox application with circumaural headphones. These tests were administered in a quiet room.

Sensory and cognitive data collection alternated per participant with further randomization within individual assessments. Data collection took about one hour per participant. Table 1 summarized the testing phases, which will be considered in greater detail in the next section.

Table 1. List of Conditions

Ability	Property
Standard Audiometry	Hearing Function Middle Ear Status
Speech Perception	Single Words Sentences
Sensory	Hearing Sensitivity Frequency Resolution Temporal Acuity
Cognitive	Attention Working Memory Vocabulary Processing Speed

Note. Following standard audiometric testing, data collection began with speech perception in noise measures on each participant. Following speech perception in noise data collection, sensory and cognitive data were collected.

3.2.2 Speech Perception in Noise

Speech stimuli included Arizona Bioindustry Association (AZBio) sentences and Consonant-Nucleus-Consonant (CNC) words obtained from the Minimum Test Speech Battery (MTSB). AZBio sentences were developed at Arizona State University (Dorman, Loizou, Fitzke, & Tu, 2004). The corpus consists of 8 lists of 20 sentences that range in length from 4 to 12 words. Each list of 20 sentences include 10 sentences spoken by two male talkers and 10 sentences spoken by two female talkers (5 sentences per talker) using a conversational speaking style (Dorman et al., 2004). In addition, the sentences had limited contextual cues that make it difficult for a listener to fill in unintelligible words (Dorman et al., 2004). Thus, the AZBio sentences were everyday sentences, with less predictability and greater difficulty than HINT sentences (Dorman, et al., 2004). Sample sentences include:

> A mother always has something better to do.
> You should be used to taking money from ladies.
> Who would lie about cancer for attention?
> Hang the air freshener from your rearview mirror.

The CNC word test consists of lists of monosyllabic words spoken by a male speaker. There was equal phonemic distribution across lists. Each list consisted of 500 test words organized into ten 50-word lists (Lehiste & Peterson, 1959). Background sounds included speech noise (SN) and two-talker babble. SN consisted of noise with a long-term average spectrum similar to that of speech. Two-talker babble consisted of a male and female speaker simultaneously reading a passage from a book named "The New Children's Encyclopedia."

AZBio sentence and CNC word recognition was performed at two noise levels to achieve approximately 25% and 75% correct responses on word recognition. These levels were established via pilot testing. AzBio sentences and CNC words were presented at 65 dB SPL and the level of the noise was varied. AZBio sentences were presented at a signal-to-noise ratio of -10 dB (75 dB SPL) and +0 dB (65 dB SPL) in both the speech noise and two-talker babble conditions. CNC words were presented at a signal-to-noise ratio of -5 dB (70 dB SPL) and +5 dB (60 dB SPL) for both the speech noise and two-talker babble conditions. Table 2 provides a list of speech perception in noise conditions.

To ensure appropriate output from the audiometer, measurements were made with a sound level meter using the A scale and a 6cc coupler. When the dial of the audiometer was set to 50 dB HL, the output on the sound level meter was 65 dB SPL. Additionally, prior to each participant's speech perception data collection, to ensure the output was consistent across all participants, calibration tones for both speech and noise stimuli were played through the media player on the computer routed to the audiometer and the VU meter on the audiometer was adjusted to +2.

Table 2. List of Speech Perception in Noise Conditions

Speech Perception Conditions			
Speech Context	Noise Type	Noise Level	SNR
Words	SN	45 dB HL	5
Words	SN	55 dB HL	-5
Words	BN	45 dB HL	+5
Words	BN	55 dB HL	-5
Sentences	SN	40 dB HL	-10
Sentences	SN	50 dB HL	+0
Sentences	BN	40 dB HL	-10
Sentences	BN	50 dB HL	+0

Note. Speech perception in noise data collection was randomized across participants. Words and Sentence stimuli were presented at 50 dB HL and the level of the noise was varied. Words and Sentences were presented at two noise levels and noise types for a total of eight conditions. SRN = signal-to-noise ratio, SN = speech noise, BN = babble noise, CNC = consonant nucleus consonant, AZBio = Arizona Bioindustry Association.

The participants were asked to repeat back sentences and words presented through an earphone to their right ear in quiet and in the presence of background sounds presented to the same ear. The measurement variable for both sets of stimuli was a percent correct score on word recognition, with each word scored as completely correct or incorrect for words and sentences; no partial credit was given.

3.2.3 Sensory Measurement

Pure Tone Average. Hearing sensitivity was established on each participant from pure tone thresholds obtained during inclusion testing. To measure hearing sensitivity, pure-tone detection thresholds for octave frequencies from 250 Hz – 8000-Hz were evaluated for both ears using the modified Hughson-Westlake threshold search paradigm. In this procedure, the intensity of the tone is first presented at a level the listener can hear clearly. Then the intensity is reduced in 10 dB steps until the listener no longer responded. The intensity of the signal is then increased in 5-dB steps until the listener responds again. From this point on, whenever the listener responds, the signal is decreased-10 dB and whenever the listener fails to respond the signal is increased 5-dB. The intensity, when the signal is being increased, to which the listener responds two out of three times is recorded as threshold. Once threshold is established, a down 4-dB up 2-dB threshold search paradigm was used to refine the threshold measurement. From pure-tone audiometry measurements, the pure-tone average (PTA) for each participant was measured. The PTA refers to the average of hearing threshold levels at a set of specified frequencies, 500, 1000, and 2000-Hz. The PTA was established on each participant as a measure of hearing sensitivity.

Auditory Filter Slope. To estimate frequency resolution (selectivity) the slope of the auditory filter was measured. Auditory Filter Slope (AFS) was established on each participant as a threshold measurement obtained in several notched noise conditions centered at 500 Hz and 4000 Hz. Notched-noise maskers were created by generating two band-pass noises in the frequency domain. Each noise band was 300-Hz wide. By positioning the upper edge of the low-frequency band and the lower edge of the high-frequency band, a spectral notch of a desired bandwidth was established. For the 500-Hz signal, the notch widths included: 0, 100, 200 and 300-Hz. For the 4000-Hz signal, the notch widths included: 0, 800, 1600, and 3200-Hz. The notched-noise masker was presented continuously at 65 dB SPL.

Pure-tone stimuli and notched noise maskers were delivered monaurally to the participant's right ear only. Using the same threshold seeking procedure as in hearing sensitivity measurements, the score for auditory filter shape was a threshold measurement obtained for each condition in dB HL. These thresholds were plotted against notch noise width conditions and the slope of the function was an estimate of the participant's AFS. AFS was established on each participant as a measure auditory filter shape. To confirm the output of the noise, measurements were made with a sound level meter using the A scale and a 6cc artificial ear. When the dial of the audiometer was set to 50 dB HL, the output on the sound level meter was 65 dB SPL. Refer to Appendix A for further details regarding AFS measurement.

Temporal Window Slope. To estimate temporal acuity, the slope of the temporal window was measured. Each participant's Temporal Window Slope (TWS) was measured using a procedure similar to the auditory filter measurement. The slope of the temporal window was measured by determining the participant's threshold at 500 Hz and 4000 Hz while amplitude modulated (AM) noise as delivered to the same ear. Modulating the amplitude of Gaussian noise with a sinusoid created the amplitude modulated-noise masker. The modulation depth of the amplitude-modulated noise was 100%. The modulation frequencies were 0, 3.123, 6.25, and 12.5-Hz. The modulated-noise masker was presented continuously at 65 dB SPL.

The same threshold seeking procedure as in pure-tone sensitivity measurements was sued. Thresholds at 500 Hz and 4000-Hz were obtained in continuous noise (no AM modulation), 12.5 Hz modulation (word), 6.25 Hz modulation (syllable), and 3.125-Hz modulation (phoneme). The slope of temporal window was a threshold measurement obtained for each condition in dB HL. These thresholds were plotted against amplitude modulated noise conditions and the slope of the function was the estimate of the participant's TWS. TWS was established on each participant as a measure of temporal window shape. To confirm the output of the noise, measurements were made with a sound level meter using the A scale and a 6cc artificial ear. When the dial of the audiometer was set to 50 dB HL, the output on the sound level meter was 65 dB SPL. Refer to Appendix B for further details regarding TWS measurement.

3.2.4 Cognitive Assessment

Cognitive assessment was undertaken in the areas of working memory, language, attention, and processing speed using subtests from the NIH Cognitive Toolbox (Weintraub, Dikmen, Heaton, Tulsk, Zelazo, Bauer, & Gershon et al., 2013). The NIH Cognitive Toolbox provides investigators with well-established standardized tests that enable comparisons across studies and populations (Weintraub et al., 2013). The test battery is computerized and includes automated scoring. Prior to each task, the experimenter provided oral instructions, and compliance was monitored to ensure valid results. In addition to oral instructions, instructions were presented visually on the iPad screen. Before each session, the iPad volume was adjusted to a comfortable listening level for each participant. The comfortable listening level was established based on feedback from the participant. All NIH Toolbox measures were administered under circumaural headphones.

Flanker Inhibitory Control and Attention Test. The Flanker Inhibitory Control and Attention test measured each participant's attention and inhibitory control. The test required the participant to focus on a stimulus while ignoring neighboring stimuli. On each trial, the word "middle" appeared on the screen to remind the participant to attend to the middle stimulus (Weintraub et al., 2013). Three arrow stimuli were presented on a screen and participants are instructed to press one of two arrow buttons on a keyboard that corresponded to the direction the middle arrow (Weintraub et al., 2013). Higher scores indicate greater ability to attend to relevant stimuli and ignore attention from neighboring stimuli (Weintraub et al., 2013). In addition to the age-corrected and uncorrected standard scores, and fully corrected t-scores, the Flanker computed score provides examiners with a way to measure improvement or decline from subsequent assessments (Weintraub et al., 2013). The uncorrected standard score was used in this study as it provided an index of the participant's overall performance when compared with the general U.S. population.

List Sorting Working Memory Test. The List Sorting Working Memory Test was used as a measure of working memory for each participant. The test is a sequencing task that required participants to sort and sequence information. A series of pictures were presented visually and auditorily to participants on an iPad screen. Participants were instructed to remember the stimuli in the series, but to repeat the items back to the examiner from smallest to largest verbally (Weintraub et al., 2013). The number of items in a series increased on each trial, increasing a participants working memory load. The test started with a 1-list version where the participant had to sequence one type of stimuli (for example: animals) based on size and then switched to a 2-list version where two types of stimuli had to be sequenced, each in size order (Weintraub et al., 2013). In the 2-list version, the working memory load was taxed further as the stimuli were presented from two categories (animals and food) and the participant had to track and organize stimuli from both categories and repeat items back by size from one category (i.e., animals) followed by the other category (i.e., food) (Weintraub et al., 2013). List Sorting scores are based the number of total correct responses from both lists which make up the List Sorting Total Score (Weintraub et al., 2013). Age-corrected and uncorrected standard scores, and fully corrected t-scores are provided for the List Sorting Working Memory Test. The uncorrected standard score was used in this study as it provided an index of the participant's overall performance when compared with the general U.S. population.

Toolbox Picture Vocabulary Test. Language was measured on each participant using the Toolbox Picture Vocabulary Test (TPVT). The TPVT is a test of receptive language. In this test, the participant was presented with an audio recording of a word and four photographs on an iPad screen and instructed to select the photograph that matched the meaning of the word. After the participant responded, the computer selected the next item based on whether or not the first was answered correctly (Weintraub et al., 2013). Test difficulty was controlled so that the participant had a 50% likelihood of answering each item correctly (Weintraub et al., 2013). A maximum of 25 items are administered. A theta score, similar to a z-score, was calculated for each participant which represents the overall ability of the participant (Weintraub et al., 2013). Age-corrected and uncorrected standard scores, and fully corrected t-scores are provided for the TPVT. The uncorrected standard score was used in this study as it provided an index of the participant's overall performance when compared with the general U.S. population.

Pattern Comparison Processing Speed Test. The Pattern Comparison Processing Speed Test was used to measure processing speed of each participant. In this test, participants were asked to identify whether two visual patterns are the same (by selecting a yes button) or different (by selecting a no button) on an iPad screen. Scores reflected the number of correct items completed in 3 min (Weintraub et al., 2013). Patterns are either identical or varied on one of two dimensions: color or adding/taking something away (Weintraub et al., 2013). Age-corrected and uncorrected standard scores, and fully corrected t-scores are provided for the pattern comparison processing speed test. The uncorrected standard score was used in this study as it provided an index of the participant's overall performance when compared with the general U.S. population.

3.3 Overview of Data Analysis

Data analysis first confirmed all participant scores by checking for outliers. All scores exceeding three standard deviations of the mean were double checked to ensure data integrity. Following the preliminary analysis, descriptive statistics were computed for the covariates and independent variables. The effects of the independent variables on speech perception in noise were analyzed next, followed by correlational analysis to assess relationships among the covariates only and among the covariates and independent variables together. As single words and sentences came from two different corpuses and percent correct was calculated from a different number of total words for each measure, the final analysis included separate analysis for single words and sentences.

CHAPTER 4. RESULTS

This study examined the role of sensory and cognitive properties and their contribution to a listener's speech perception in noise abilities. The sensory properties included hearing sensitivity, frequency selectivity, and temporal acuity. The cognitive properties included attention, working memory, language, and processing speed. Collectively, the sensory and cognitive measures are referred to as covariates. Covariate variables are continuous and represent characteristics of participants. In the subsequent analyses, the covariates were used as predictors of a listener's speech in noise abilities. The contribution of each covariate to speech perception in noise was measured in two types of target speech contexts (single words and sentences) in four representative listening environments. The different listening environments served as the independent variables. The dependent variable measured in each listening context was a percent correct score on word recognition.

4.1 Independent Variables – Listening Conditions

The independent variables in this study included four listening environments per target speech context. The goal was to assess comprehensively a listener's speech in noise ability. In each listening environment, the target speech, background noise level, or background noise type was different. Target speech varied from single words to sentences, increasing the noise level varied signal-to-noise ratio, and background type varied from speech noise to babble. The target speech for single words was Consonant Nucleus Consonant (CNC) words and for sentences was Arizona Bioindustry Sentences. Both sets of speech stimuli were from the Minimum Test Speech Battery (MTSB). The scoring method differed for words and sentences. For single words, percent correct was determined out of 50 words, whereas for sentences, percent correct was determined out of about 140 words. An arcsine transform was not undertaken for the dependent variable as percent correct scores are treated as a continuous variable and are normally distributed. The background type included speech noise, which followed the long-term average of speech and two-talker babble with male and female talkers. Words and sentences were presented at 50 dB HL and the level of the background varied from low to high creating favorable and less favorable signal-to-noise ratios**.** The listening environments are summarized in Table 3.

Table 3. Summary of Listening Environments

Abbreviation	Target Context	Noise Type	Noise Level
wd_sn_lo	Words	Speech	Low (45 dB HL)
wd_sn_hi	Words	Speech	High (55 dB HL)
wd_bn_lo	Words	Babble	Low (45 dB HL)
wd_bn_hi	Words	Babble	High (55 dB HL)
st_sn_lo	Sentences	Speech	Low (50 dB HL)
st_sn_hi	Sentences	Speech	High (60 dB HL)
st_bn_lo	Sentences	Babble	Low (50 dB HL)
st_bn_hi	Sentences	Babble	High (60 dB HL)

Note. Table 3 summarizes the eight listening environments. In each environment, the target speech context, noise type, or noise level was varied. Target speech was presented at 50 dB HL. The abbreviation for each listening environment is provided in the first column. The last column includes the level of the noise. wd = words, st = sentences, sn = speech noise, bn = babble noise, lo = low noise level, hi = high noise level.

` ### 4.1.1 Descriptive Statistics for the Independent Variables

The following section provides descriptive statistics for the independent variables.

Table 4. Descriptive Statistics for the Independent Variables

Abbreviation	N	Minimum	Maximum	Mean	Standard Deviation
wd_sn_lo	32	54%	98%	74%	9.4%
wd_sn_hi	32	12%	36%	22%	6.7%
wd_bn_lo	32	56%	96%	70%	15.2%
wd_bn_hi	32	10%	60%	32%	11.2%
st_sn_lo	32	47%	92%	70%	11.3%
st_sn_hi	32	2%	44%	19%	10.8%
st_bn_lo	32	19%	96%	69%	15.5%
st_bn_hi	32	8%	62%	28%	13.2%

Note. Table 4 provides descriptive statistics for the independent variables. The abbreviation for each listening environment is provided in the first column. N = number of participants, wd = words, st = sentences, sn = speech noise, bn = babble noise, lo = low noise level, hi = high noise level.

For listening environments that included single words in speech noise at high noise level, the mean was 22% (SD = 6.7), babble at a high noise level, the mean was 32% (SD = 11.2), babble at a low noise level, the mean was 70% (SD = 15.2), and speech noise at a low noise level, the mean was 74% (SD = 9.4). For listening environments that include sentences in speech noise at a high noise level, the mean was 19% (SD = 10.5), babble at high noise level, the mean was 28% (SD = 13.2), babble at a low noise level, the mean was 69% (SD = 15.5), and speech noise at a low noise level, the mean was 70% (SD = 11.3). Descriptive statistics for the independent variables are listed in Table 4.

4.2 Language Context Condition

As mentioned previously, stimuli for single words and sentences came from different corpuses. Additionally, percent correct for single words and sentences was calculated from a different number of total words (50 for single words and about 140 for sentences). The results for the two target speech contexts could not be compared directly, and therefore all subsequent analyses were completed separately for single words and sentences.

4.2.1 Effects of the Independent Variables on Speech Perception in Noise

Single Words. A two-factor within group analysis of variance was performed to study the effect of noise level (low and high) and noise type (speech and babble) on the perception of single words as measured by percent correct recognition. A Bonferroni adjustment for multiple comparisons was performed. The results of the ANOVA are summarized in Table 5.

Table 5. Two-factor ANOVA: Single Words

Source	Numerator df	Denominator df	F	Significance
Intercept	1	22	1963.8	0.000
sn_bn*lo_hi	1	93	11.1	**0.001**
sn_bn	1	93	2.6	0.89
lo_hi	1	93	562.5	**0.000**

Note. Table 5 provides the results of the two-factor ANOVA for single words. The analysis revealed a significant main effect for noise level and the interaction between noise type and noise level. sn = speech noise, bn = babble, lo = low noise level, hi = high noise level. Significant effects in bold.

The interaction between noise level and noise type was statistically significant $F(1,93) = 11.1$, $p = .001$. A post-hoc pairwise comparison revealed no significant difference between noise types at the low noise level, $F(1,93) = 1.5$, $p > .05$ but a significant difference was observed at the high-level noise $F(1,93) = 12.2$, $p < .01$. Additionally there was a significant main effect for noise level $F(1,93) = 562.5$, $p < .001$. The main effect for noise type did not reach statistical significance $F(1,93) = 2.6$, $p > .05$. The interaction is shown in Figure 1.

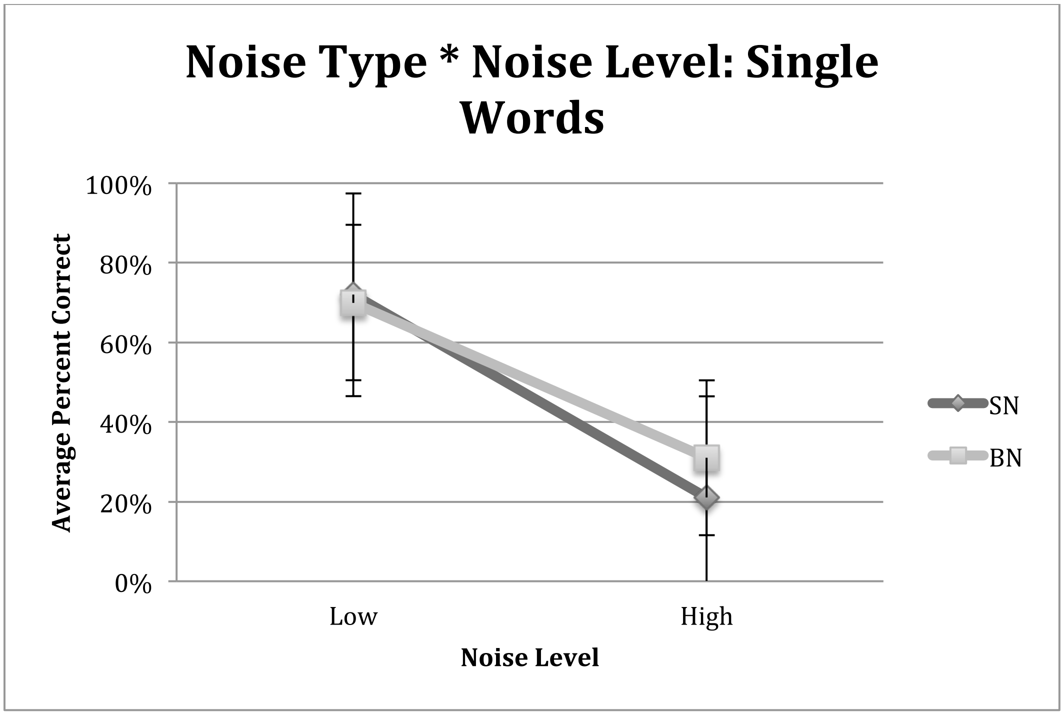


Figure 1 shows the noise type-by-noise level interaction for single words. At the high noise level, listeners performed better when the background was two-talker babble vs. speech noise. SN = speech noise, BN = babble. Error bars represent one standard error of the mean.

Sentences. A two-factor within group analysis of variance was performed to study the effect of noise level (low and high) and noise type (speech and babble) on the perception of sentences as measured by percent correct recognition. A Bonferroni adjustment for multiple comparisons was performed. The results of the ANOVA are summarized in Table 6.

Table 6. Two-factor ANOVA: Sentences

Source	Numerator df	Denominator df	F	Significance
Intercept	1	22	479	0.476
sn_bn*lo_hi	1	93	4.6	**0.001**
sn_bn	1	93	483.5	**0.022**
lo_hi	1	93	7.4	**0.000**

Note. Table 6 provides the two-factor ANOVA results for sentences. The analysis revealed a significant main effect for noise type and noise level and the interaction between noise type and noise level. sn = speech noise, bn = babble, lo = low noise level, hi = high noise level. Significant effects in bold.

The interaction between noise level and noise type was statistically significant $F(1,83.3) = 7.4$, $p < .003$. A post-hoc pairwise comparison revealed no significant difference between noise types at the low noise level, $F(1,101.3) = .167$, $p > .05$ but a significant difference was observed at the high-level noise $F(1,101.3) = 11.8$, $p < .01$. The interaction is shown in Figure 2. Additionally there was a significant main effects for noise level $F(1,87.61) = 483.5$, $p < .0001$ and noise type $F(1,83.39) = 4.6$, $p < .05$.

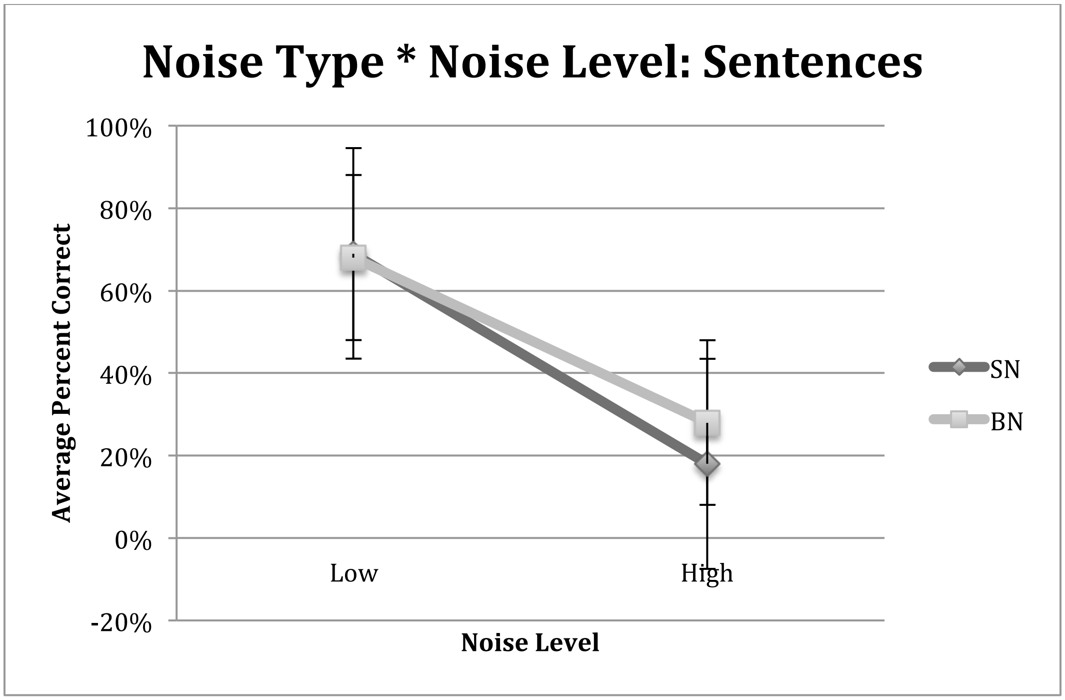


Figure 2 shows the noise type-by-noise level interaction for sentences. At the high noise level, listeners performed better when the background sound was two-talker babble vs. speech noise. SN = speech noise, BN = babble. Error bars represent one standard error of the mean.

4.3 Descriptive Statistics for the Covariates

The following section provides descriptive statistics for the covariates, beginning with the sensory factors followed by the cognitive factors. Each covariate refers to the specific measure of sensory or cognitive function. A summary of the covariates and their means (M) and standard deviations (SD) are summarized in Table 7.

Table 7. Summary of the Covariate Measures

Abbreviation	Measure	N	Minimum	Maximum	Mean	Standard Deviation
PTA	Pure Tone Average	32	-5.0	13.3	4.6	4.6
AFS_500	Auditory Filter Slope at 500 Hz	32	-7.0	-3.4	-5.6	.81
AFS_4000	Auditory Filter Slope at 4000 Hz	32	-14.8	-6.6	-9.7	2.0
TWS_500	Temporal Window Slope at 500 Hz	32	-3.4	-0.32	-2.3	.98
TWS_4000	Temporal Window Slope at 4000 Hz	32	-4.5	-1.4	-2.7	.91
ATN	Attention	32	86	124	110.8	8.8
LNG	Language	32	95	122	108.9	6.5
WM	Working Memory	32	95	117	106.7	5.8
PS	Processing Speed	32	82	143	113.2	20.3

Note. Table 7 provides a summary of the covariates. Higher scores for ATN, LNG, WM, PS reflect better performance. Measure refers to the sensory or cognitive property that was measured on each participant. N = number of participants, PTA = pure tone average, AFS = auditory filter slope, TWS = temporal window slope, ATN = attention, WM = working memory, LNG = language, PS = processing speed.

Pure tone average (PTA) refers to a listener's auditory sensitivity. Normal hearing sensitivity is identified with a PTA equal to or better than 20dB HL. The mean PTA for the participants is 4.6 dB HL (SD = 4.6 dB HL). The mean indicates that the participants in this study have normal hearing sensitivity.

Auditory filter slope (AFS) estimates reflect a listener's frequency selectivity. Sharp frequency resolution enables a listener to distinguish one sound from another on the basis of its frequency content. The mean auditory filter slope for individuals between 18 and 29 years was -4.4 dB HL (SD = 3.3) at 500Hz and -5.6 dB HL (SD = 3.4) at 4000Hz. The mean auditory filter slope for the participants in this study is -5.6 dB HL (SD = .81) at 500Hz and -9.7 dB HL (SD = 2.0) at 4000Hz. These scores are considered typical performance for young adults. For additional information regarding this measurement, refer to Appendix A.

Temporal window slope (TWS) estimates reflect a listener's ability to detect changes in stimuli over time. The mean temporal window slope for individuals between 18 and 29 years at 500 Hz was -2.5 dB HL (SD = 2.6) and at 4000Hz is -2.1 dB HL (SD = 2.3). The mean temporal window slope for the participants in this study at 500Hz is -2.3 dB HL (SD = .98) and at 4000 Hz is -2.7 (SD = .91). These scores are considered typical performance for young adults. For additional information regarding this measurement, refer to Appendix B.

AFS and TWS measurements were both taken at two frequencies: 500 Hz and 4000 Hz. Independent t-tests were performed to determine the presence of a frequency effect for the 500 Hz and 4000 Hz conditions. There was a significant difference in scores for auditory filter slope at 500 Hz (M = -5.7, SD = 1.4) and auditory filter slope at 4000 Hz (M = -9.8 , SD = 2.0) conditions; $t(31) = 10.8$, $p < 0.05$. Similarly, there was a significant difference in scores for the temporal window slope at 500 Hz (M = -9.8, SD = 2.0) and temporal window slope at 4000 Hz (M = -3.5, SD = 2.4) conditions, $t(31) = -13.6$, $p < .05$. For both AFS and TWS, p-values were adjusted for multiple comparisons. Based on these results, all subsequent analyses included the auditory filter slope and temporal window slope measurements at 500 Hz and 4000 Hz.

Cognitive measurements were administered using the NIH Cognitive Toolbox battery for the iPad. Measurements were taken for attention, working memory, language, and processing speed. The following section summarizes the results for the cognitive factors.

Attention assists a listener in focusing on a particular stimulus while filtering out a range of other stimuli (Strait & Kraus, 2011). Participant's attention was measured using the Flanker Inhibitory Control and Attention test. The mean for the Flanker Inhibitory Control and Attention test for individuals between 18 and 29 years is 113.21 (SD = 23.14). The scores for participants in this study fell within one standard deviation of the mean (M = 110.8, SD = 8.8), and thus are considered typical performance for young adults.

Linguistic intelligence refers to an individuals' ability to understand spoken and written language, as well as their ability to speak and write (Weintraub et al., 2013). A segment of linguistic ability refers to individual differences in vocabulary knowledge. Language was measured using the Toolbox Picture Vocabulary Test. The mean for the Toolbox Picture Vocabulary Test for individuals between 18 and 29 years is 110.5 (SD = 20.07). The scores for participants in this study fell within one standard deviation of the mean (M = 108.9, SD = 6.5), and thus are considered typical performance for young adults.

Working memory refers to a temporary storage center that can manipulate a small amount of information over a brief period of time (Baddeley, 1992). Working memory was measured using the List Sorting Working Memory Test. The mean for the List Sorting Working Memory Test for individuals between 18 and 29 years is 111.6 (SD = 21.5). The scores for participants in this study fell within one standard deviation of the mean (M = 106.7, SD = 5.8), and thus are considered typical performance for young adults.

Processing speed refers to the time it takes for an individual to do a mental task (Dryden et al., 2017). Processing speed was measured using the Pattern Comparison Processing Speed Test. The mean for the Pattern Comparison Processing Speed Test for individuals between 18 and 29 years is 115.5 (SD = 23.7). The scores for participants in this study fell within one standard deviation of the mean (M = 113.2, SD = 20.3), and thus are considered typical performance for young adults.

4.4 Correlational Analysis

Correlational analyses were performed between the nine covariates alone and between the covariates and independent variables to identify the extent to which two variable types were related to each other.

For the covariates alone, the correlational analysis revealed that few covariates were related to one another. Table 8 shows the correlations among the covariates.

Table 8. Correlations Among the Covariates

	PTA	AFS_500	AFS_4000	TW_500	TW_4000	ATN	LNG	WM	PS
PTA	-	0.20	-.09	-.04	-.16	.02	-.15	**-.39***	**-.40***
AFS_500		-	-.26	.01	**.54****	.10	.18	.16	-.12
AFS_4000			-	.03	.32	-.14	.27	.02	-.04
TWS_500				-	.13	-.25	-.07	-.19	.04
TWS_4000					-	-.2	.09	-.15	-.13
ATN						-	.35	.06	-.07
LNG							-	.12	**-.37***
WM								-	0.30
PS									-

* denotes $p < .05$
** denotes $p < .01$

Note. Table 8 provides a summary of the correlations among the covariates. PTA = pure tone average, AFS = auditory filter slope, TWS = temporal window slope, ATN = attention, WM = working memory, LNG = language, PS = processing speed. Significant correlations in bold with stars.

Of the 36 possible correlations among covariates, only four were significant. A significant correlation was found between pure tone average and working memory, $r(30) = -.39$, $p < .05$, pure tone average and processing speed, $r(30) = -.40$, $p < .05$, auditory filter slope at 500 Hz and temporal window slope at 4000 Hz, $r(30) = .54$, $p < .05$, and language and processing speed, $r(30) = -.37$, $p < .05$. Because the covariates were not highly correlated, a Principle Component Analysis (PCA) was not undertaken.

For the covariates and independent variables, the correlational analysis revealed that very few (3 out of 72) covariates and independent variables were related to one another. Table 9 summarizes the correlations between the covariates and independent variables. Significant correlations were observed between attention and the sentence speech noise high noise level listening condition, $r(30) = .41$, $p < .05$, attention and the sentence babble noise high noise level listening condition, $r(30) = .39$, $p <.05$, and processing speed and the sentence babble noise low noise level listening condition, $r(30) = .36$, $p < .05$.

Table 9. Correlations Among the Covariates and Independent Variables

Listening Environment	PTA	AFS_500	AFS_4000	TWS_500	TWS_4000	ATN	LNG	WM	PS
wd_sn_lo	.01	.04	-.16	.02	.13	-.01	.19	-.14	-.31
wd_sn_hi	.28	.04	.11	.10	.04	-.27	-.006	-.07	-.29
wd_bn_lo	.09	.21	-.04	.10	.05	.10	.17	-.11	-.30
wd_bn_hi	-.10	.06	.21	.03	.18	.14	.21	-.04	.00
st_sn_lo	.24	.15	.11	-.01	.35	-.26	-.03	.03	-.08
st_sn_hi	.32	.26	.09	.22	.25	**.41***	-.05	.06	-.01
st_bn_lo	-.15	.14	-.01	-.03	.21	-.10	-.24	.30	**.36***
st_bn_hi	-.11	.04	-.06	.11	.06	**.34***	-.13	.17	.08

* denotes $p < .05$

Note. Table 9 shows the correlations among the covariates and the independent variables. wd = words, st = sentences, sn = speech noise, bn = babble noise, lo = favorable noise level, hi = unfavorable noise level, Pure tone average (PTA), auditory filter slope (AFS), temporal window slope (TWS), attention (ATN), working memory (WM), language (LNG), processing speed (PS). Significant correlations in bold.

4.5 Effect of the Covariates on Speech Perception in Noise

4.5.1 Regression: Single Words

A linear regression analysis was performed to examine the effect of the covariates variables on the perception of words as measured by percent correct recognition in the four listening environments. Each model included speech perception scores for one of the listening conditions as the dependent variable (i.e., wd_sn_lo, wd_bn_lo, wd_sn_hi, and wd_bn_hi) along with all sensory: PTA, AFS_500 and AFS_4000, and TWS_500 and TWS_4000 and cognitive: ATN, WM, LNG, PS variables as predictor variables.

Table 10 provides the standardized beta coefficients and R squared values for single words. The coefficients describe the strength of the association of each sensory and cognitive variable (covariate) with the dependent variable (percent correct score in a specified listening condition).

Table 10. Regression for Single Words

	Standardized Beta Coefficients			
Words	SN_LO	BN_LO	SN_HI	BN_HI
PTA	-0.105	-0.035	0.232	-0.131
AFS_500	0.001	0.261	0.042	-0.012
AFS_4000	-0.292	-0.122	0.089	0.186
TWS_500	-0.022	0.118	0.084	-0.035
TWS_4000	0.162	-0.106	-0.105	0.112
ATN	-0.093	0.005	-0.298	-0.059
WM	0.206	0.136	0.011	-0.090
LNG	-0.151	-0.079	0.141	-0.092
PS	-0.233	-0.228	-0.257	-0.039
R squared	0.200	0.168	0.224	0.086

* denotes p < .05
** denotes p < .01

Note. SN = speech noise, BN = babble noise, HI = high noise level, LO = low noise level. PTA = pure tone average, AFS = auditory filter slope, TWS = temporal window slope, ATN = attention, WM = working memory, LNG = language, PS = processing speed.

Table 10 reveals that of the four listening environments including single words, none of the sensory or cognitive variables emerged as significant predictors of speech in noise performance. The R-squared values at the bottom of the columns in the table indicate the amount of variance accounted for by sensory and cognitive predictors for each listening environment. Of the variance in speech perception in noise as measured with single words, the sensory and cognitive factors only accounted for 9-22 percent.

4.5.2 Regression: Sentences

A linear regression analysis was performed to study the effect of the covariates on the perception of sentences as measured by percent correct recognition in the four listening environments. Each model included speech perception scores for one of the listening conditions as the dependent variable (i.e., wd_sn_lo, wd_bn_lo, wd_sn_hi, and wd_bn_hi) along with all sensory: PTA, AFS_500 and AFS_4000, and TWS_500 and TWS_4000 and cognitive: ATN, WM, LNG, PS variables as independent variables.

Table 11 provides the standardized beta coefficients and R-squared values for sentences. The coefficients describe the strength of the association of each sensory and cognitive variable (covariate) with the dependent variable (percent correct score in a specified listening condition).

Table 11. Regression Analysis for Sentences

	Standardized Beta Coefficients			
Sentences	SN_LO	BN_LO	SN_HI	BN_HI
PTA	0.413	0.098	**0.463***	-0.119
AFS_500	-0.079	0.064	0.223	0.145
AFS_4000	-0.022	-0.097	0.004	-0.073
TWS_500	-0.081	-0.060	0.172	0.072
TWS_4000	0.420	0.235	0.089	-0.120
ATN	-0.225	-0.084	-0.360	-0.321
WM	0.094	0.025	-0.074	-0.198
LNG	0.098	0.161	0.229	0.214
PS	0.119	0.395	0.089	-0.124
R squared	0.278	0.242	0.445	0.204

* denotes $p < .05$

Note. SN = speech noise, BN = babble noise, HI = high, LO = lo. PTA = pure tone average, AFS = auditory filter slope, TWS = temporal window slope, ATN = attention, WM = working memory, LNG = language, PS = processing speed. Significant effect in bold with a star.

Table 11 shows that of the four listening environments including sentences, only pure-tone average emerged as a significant predictor in the listening condition containing speech noise at a high noise level. The R-squared values at the bottom of the columns in the table indicate the amount of variance accounted for by the sensory and cognitive predictors for each listening environment. Of the variance in speech perception in noise as measured with sentences, the sensory and cognitive factors accounted for 20 to 28 percent. The exception was the condition of high-level speech noise with 45 percent of the variance accounted for.

4.5.3 Mixed-Model Analysis: Single Words

A linear mixed model analysis takes advantage of all the information in the statistical analysis included the fixed effects, random effects and all observations. Thus, a linear mixed-model analysis of variance was performed to examine the effect of the covariates on the perception of words as measured by percent correct recognition.

Table 12. Fixed Effects: Single Words

Source	Numerator df	Denominator df	F	Significance
Intercept	1	22	2.666	0.117
PTA	1	22	0.255	0.618
AFS_500	1	22	0.211	0.651
AFS_4000	1	22	1.552	0.226
TWS_500	1	22	0.221	0.643
TWS_4000	1	22	1.027	0.322
ATN	1	22	0.05	0.826
LNG	1	22	1.445	0.242
WM	1	22	0.17	0.205
PS	1	22	0.051	0.823
sn_bn	1	93	2.957	0.089
lo_hi	1	93	632.124	**0.000**
sn_bn*lo_hi	1	93	12.428	**0.001**

Note. Table 12 provides the results for a linear mixed-model ANOVA for words. PTA = pure tone average, AFS = auditory filter slope, TWS = temporal window slope, ATN = attention, WM = working memory, LNG = language, PS = processing speed

The results of the ANOVA are summarized in Table 12. The analysis revealed no significant main effect for sensory or cognitive variables when the target speech context was single words. The main effects for noise level and type were significant and marginally significant, respectively. The noise type-by-level interaction was also significant. These results were mentioned previously in section 4.2.1.

4.5.2 Mixed-Model Analysis: Sentences

A linear mixed-model analysis of variance was performed to study the effect of the covariates on the perception of sentences as measured by percent correct recognition. The results of the ANOVA are summarized in Table 13.

Table 13. Fixed Effects: Sentences

Source	Numerator df	Denominator df	F	Significance
Intercept	1	22	0.53	0.476
PTA	1	22	3.4	0.079
AFS_500	1	22	4.6	**0.043**
AFS_4000	1	22	1.9	0.173
TWS_500	1	22	0.36	0.554
TWS_4000	1	22	1.3	0.272
ATN	1	22	4.4	**0.048**
LNG	1	22	0.05	0.828
WM	1	22	0.07	0.792
PS	1	22	0.96	0.336
sn_bn	1	22	479	**0.022**
lo_hi	1	93	7.4	**0.000**
sn_bn*lo_hi	1	22	4.6	**0.001**

Note. Table 13 provides the results for a linear mixed-model ANOVA for sentences. The analysis revealed a significant main effect AFS_500 and ATN. PTA = pure tone average, AFS = auditory filter slope, TWS = temporal window slope, ATN = attention, WM = working memory, LNG = language, PS = processing speed. Significant effect in bold with a star.

The main effect of auditory filter slope at 500Hz (AFS_500) was statistically significant $F(1,22) = 4.6$, $p < .05$. Additionally, the main effect of attention (ATN) was statistically significant $F(1,22) = 4.4$, $p < .05$. The main effects for noise level and type were significant. The noise type-by-level interaction was also significant. These results were mentioned previously in section 4.2.1.

CHAPTER 5. Discussion

Speech perception in noise was investigated by asking young adult participants to perform listening tasks in two target speech contexts. Each speech context was presented in four listening environments. External factors within each listening environment were manipulated, and internal characteristics of the listener were measured. The external factors within the environment were selected because they have been shown to affect speech perception in noise (Mattys et al., 2012). These factors included target speech context, background noise type and level. Target speech varied in complexity from single words to sentences, background type varied from speech noise to babble, and background noise level varied from low to high. The internal characteristics of a listener that were measured included sensory abilities such as hearing sensitivity, frequency resolution, and temporal acuity, along with cognitive abilities such as attention, working memory, language, and processing speed. The internal factors were chosen because investigations of speech perception have found these abilities to be associated with speech perception in noise (Dryden et al., 2017)

The primary aim of this study was to investigate a listener's sensory and cognitive abilities that may contribute to speech perception in noise. Of particular interest were the sensory and cognitive factors that best predict speech perception in noise and whether the factors vary with changes in the listening environment. Heinrich et al. (2015) reported that cognition is a predictor of speech in noise performance only when the speech context is sentences. Therefore, I predicted a greater number of listener abilities would contribute to word recognition in sentences compared to single words.

5.1 Effects of the Independent Variables on Speech Perception in Noise

The speaker factor selected for study was target speech context because speech stimuli such as words and sentences convey the meaning of language and are typically used as target speech stimuli in clinical measures of speech perception in noise. Additionally, variations in target speech complexity may engage cognitive processes differently (Dryden et al., 2017; Heinrich, Henshaw, Ferguson, & 2015). The environmental factors were related to the presence of simultaneous background sounds. More specifically, the type and level of the background were varied.

Background levels were based on pilot work designed to achieve approximately 25- and 75-percent correct performance with high and low levels, respectively. Increasing the level of the background should increase speech perception difficulty. Additionally, speech noise and babble were adjusted so that they contained equivalent energy. Because the energetic masking for both types of backgrounds was comparable at each background level, the perceptual distraction imposed by informational masking was only encountered in conditions containing babble. Thus, a given level of babble noise should cause greater perceptual difficulty than a similar level speech noise. Furthermore, a listener's score on single word recognition should be greater than sentences recognition along the continuum of difficulty due to the increased complexity of sentence stimuli as target speech (Heinrich et al., 2015).

As reported in the results, a significant main effect was observed for background noise level. Listener scores for single words with both speech and babble noise decreased with an increase in level. This result was expected based on the pilot study. The presence of a significant noise type-by-level interaction complicated the results for background type. Scores for speech and babble were not different with the low-level noise; however, with the high-level noise, the scores for babble noise were significantly higher than for speech noise. A similar outcome was observed for listener scores for sentences.

Babble that is made up of two to three talkers is often more difficult for listeners than when the masker is steady noise because competing talker backgrounds can introduce a significant amount of higher-level lexical interference (Carhart et al., 1975; Mattys et al., 2012). Therefore, I expected babble noise should cause greater perceptual difficulty than comparable level speech noise. For single words and sentences, the results indicated that at low noise levels, there were no significant differences in scores for speech noise or babble, and at high noise levels, a listener performed better when the background type was babble over speech noise, and this difference was statistically significant.

Because the energy of speech noise and babble was controlled for, and a difference in scores between speech noise and babble was not observed at a low background level, this result could suggest there was no information masking produced by the two-talker babble in this condition. At high noise levels, while a difference in performance was observed for the two background types, the result is opposite of our prediction. Listeners were not vulnerable to the adverse effects of perceptual interference caused by the two-talker babble, and instead, it appears they were able to take advantage of the silence gaps within the fluctuating background to more accurately perceive the target speech for words *and* sentences. Thus, it is possible that while speech noise and babble had the same physical energy, biologically as they were encoded by the auditory system, they did not have equal energy and listener performed better on babble as they were taking advantage of the interruptions of the fluctuating background at a physiological level.

As mentioned earlier, differences in scores were not observed between noise types at the low noise levels. In addition to the lack of informational masking, another possible reason for this outcome could be that the temporal structure of the noise does not matter as much at low noise levels because there is sufficient acoustic information about the target speech still available to the listener. However, at high noise levels, much less acoustic information regarding the target speech is available to the listener and therefore, they will exploit any fluctuations in the temporal envelope of the babble to accurately perceive the target speech.

Additionally, a difference in scores for single words and sentences was not observed. Noise levels were selected to achieve approximately 25% and 75% correct performance and were different for single words and sentences. Therefore, any difference between target speech context was eliminated due to difference in noise level for the low versus high background sounds.

5.2 Effect of the Covariate Predictors on Speech Perception in Noise

The sensory and cognitive factors selected for study were related to internal characteristics of the listener. Of a listener's sensory and cognitive abilities, the sensory abilities selected for study were hearing sensitivity, frequency selectivity, and temporal acuity, and the cognitive abilities selected were attention, working memory, language, and processing speed. The sensory and cognitive abilities were chosen because previous investigations of speech in noise performance have found the abilities to be associated with speech perception in noise (Dryden et al., 2017).

To evaluate the relationship among the covariate variables, correlation analyses were conducted to determine the strength of the relationship between the covariate predictors and independent variable for each listening condition. Additionally, regression analyses were completed to identify how well the covariates predicted speech perception scores in the different listening environments. Finally, linear mixed-model ANOVAs were conducted to identify the combined contribution of the external factors (listening conditions) and internal factors (sensory and cognitive) to speech perception of single words and sentences. As a reminder, the analyses were carried out separately for single words and sentences because the stimuli for words and sentences came from different corpuses and percent correct was calculated differently for each target speech context.

5.2.1 Correlational Analysis

Correlational analyses identified the extent to which covariates were related to each other. When evaluating these relationships in listening conditions containing single words, no significant correlations were observed for any of the sensory or cognitive variables and the independent variables. However, when evaluating the relationship between a listener's sensory and cognitive factors and listening environments containing sentences, two cognitive abilities: attention (ATN) and processing speed (PS) emerged with significant positive correlations in three of the four listening environments. ATN was significantly correlated with sentence recognition at high noise levels for speech noise and babble. This result suggests that an increase in percent correct score in these two listening environments corresponds with an increase in a listeners score for ATN (indicating greater ability). Additionally, PS was correlated with sentence recognition at low noise levels for babble. Similarly, this result suggests that an increase in percent correct score in this listening environment corresponds with an increase in a listeners score for PS (indicating greater ability). These results are consistent with our prediction that a greater number of listener abilities would emerge with sentence recognition relative to word recognition. Moreover, the correlations coefficients observed for ATN and PS are consistent with those found in the literature. As mentioned, Dryden et al. (2017) found correlation coefficients between ATN and speech in noise performance of $r = .34$ and PS and speech in noise performance of $r = .39$.

5.2.2 Regression Analysis

Linear regression analyses were conducted using the covariates as predictors of speech perception in each of the four listening environments for single words and sentences. Across the eight conditions (2 levels x 2 types x 2 contexts), only PTA emerged as significant in the listening environment containing sentences in speech noise at a high level. While PTA was the only significant predictor in the model, the covariates together in the sentence, speech noise, high noise level model accounted for almost 45% of the variance in scores for this listening environment. Moreover, in comparing the R squared values for listening conditions containing single words relative to conditions containing sentences, the amount of variance that was accounted for in each model by the sensory and cognitive factors was greater for sentences conditions versus single words.

5.2.3 Mixed-Model ANOVA

A linear mixed-model analysis of variance was performed to examine the combined effect of the external factors (listening conditions) and internal factors (sensory and cognitive covariates) on the perception of single words and sentences as measured by percent correct recognition. From the mixed-model ANOVA for sentences, very few covariates reached significance. Significant main effects were observed for auditory filter slope 500 Hz (AFS_500) and attention (ATN). The significant main effects of AFS_500 and ATN for sentences are discussed in the next section.

A significant main effect was observed for auditory filter shape at 500 Hz (AFS_500) for sentences. Our measure of AFS was a slope that was determined from several thresholds obtained from the listener in steady and notched-noise conditions (refer to appendix A for additional information on AFS measurement). Sentence stimuli were selected from the Arizona Bioindustry Association (AZBios) sentence lists. Sentences were presented at two noise levels (low and high) and in two noise types (speech noise and babble). The results of this dissertation suggest that, for young normal hearing listeners, a sensory ability like frequency selectivity plays a critical role in sentence recognition in acoustically adverse listening environments; and the shape of a listener's auditory filter impacts this ability. The steeper the slope of a listener's auditory filter shape measurement, the narrower the width of their auditory filter. As mentioned earlier, narrower auditory filter widths are associated with better speech in noise performance (Moore, 2008; Baer & Moore, 1994). Therefore, it is possible that listeners narrower auditory filter widths were better able to filter out background noise along the length of the cochlea, which increased their ability to identify formants and formant transitions that are fundamental for speech identification.

The importance of frequency selectivity in speech perception has been demonstrated in physiological and psychophysical studies in normal hearing and hearing impaired listeners. For example, listeners with cochlear hearing impairment have traditionally shown significant difficulty understanding speech in noise. The shape of their auditory filter has been suggested to contribute to these difficulties. For listeners with cochlear hearing loss, auditory filter shapes are up to three to four times greater than listeners with normal hearing (Baer & Moore, 1994). Baer and Moore (1994) used spectral smearing to simulate the effects of impaired frequency selectivity to evaluate its impact on speech intelligibility in normal hearing listeners. Spectral smearing simulated a broadened auditory filter shape by a factor of 3 or 6. Results showed that spectral streaming had little effect on the intelligibility of speech in quite, however, had a large effect on the intelligibility of speech in noise. These results, along with our findings, are consistent with previous studies suggesting that a listener's auditory filter shape contributes significantly to speech perception in noise.

A significant effect was also observed for attention (ATN) for sentences. As mentioned, sentence stimuli were selected from AZBio sentence lists and were presented at two noise levels (low and high) and in two noise types (speech noise and babble). Our measure of ATN was the Flanker Inhibitory Control and Attention Test. This measure assesses a subset of attention called inhibitory control by evaluating a listener's ability to inhibit visual attention to irrelevant flanking stimuli. On each trial, a central directional target was flanked by similar stimuli on the left and right. The listener's task was to indicate the direction of the central stimulus.

Inhibitory control has been suggested to play a role in speech perception in noise (Dryden et al., 2017; Janse, 2012). As mentioned, Janse (2012) investigated hearing loss and a measure of inhibitory control to predict listening performance for speech in quiet and in a competing talker background. Janse (2012) used phoneme monitoring to assess speech perception in quiet and two-talker babble, and the Stroop color-naming test to measure inhibitory control. Janse (2012) found that even after hearing loss and the Stroop effect measure were orthoganalized, the individual Stroop measure was significantly associated with speech perception performance in the two-talker babble condition. Thus, Janse (2012) concluded that poor inhibition increases a listener's susceptibility to competing background sounds, particularly in informational masking conditions imposed by babble. Additionally, poor inhibition has been found to be associated with a reduced ability for listener's to select an appropriate target during the retrieval process from the mental lexicon (Sommers & Danielson, 1999), and inhibition may be associated with degraded signal restoration (Janse & Jesse, 2014; Mattys et al., 2012). Thus, our finding is consistent with studies suggesting that attention contributes significantly to speech perception in noise.

5.3 Lack of Significance

Our selection of sensory and cognitive factors that influence speech perception in noise was based on a review of the literature. Remarkably, the results of this study did not find significant effects for seven of the nine variables selected to study. The following discussion will consider each of these variables and speculate why significant correlations were not found.

Working Memory. In recent years, there has been an increase in interest in the role of cognition in speech in noise performance. A critical component in this area of research is working memory (Dryden et al, 2017; Fullgrabe & Rosen, 2015; Fullgrabe & Rosen, 2016). Working memory has been shown to be involved in a range of complex behaviors such as reasoning, comprehension, and attention (Weintraub, et al., 2013, Dryden et al., 2017; Fullgrabe & Rosen, 2015, Fullgrabe & Rosen 2016), and is considered the cognitive construct behind processing information across tasks and modalities, and holding, storing, and manipulation that information in a short-term storage (Weintraub, et al., 2013; Fullgrabe & Rosen, 2015; Fullgrabe & Rosen 2016).

The link between working memory and higher-order complex behavior has led researchers to use complex span tasks as a tool to explain individual differences on speech in noise performance. For example, using the Reading Span test, Lunner (2003) and Rudner, Ronnberg, and Lunner (2011) predicted unaided and aided speech perception in noise ability for hearing impaired listeners with moderate to strong correlations. However, when referring to the association between working memory and speech in noise performance, it is generally not recognized that the participants in these studies are hearing impaired listeners with an average age of 55 years or older (Fullgrabe & Rose, 2015; Fullgrabe & Rosen, 2016).

A listener's working memory is measured through complex span tasks that require them to temporarily store and simultaneously manipulate and recall stored information (Fullgrabe & Rosen, 2016). In the present study, working memory was measured using the List Sorting Working Memory Test. Participants were presented with a series of stimuli and instructed to remember each stimulus, mentally reorder them from smallest to largest, and repeat back the names of the stimuli in this order.

This study did not find an association between working memory and speech in noise performance.

Recently, a meta-analysis by Fullgrabe and Rosen (2016) was conducted to assess the notion that individual differences in working memory account for variability in speech in noise performance, even when hearing loss is not present. They studied 132 English-speaking participants ranging from 18 – 91 years of age. Hearing thresholds for their participants were $\leq$ 20 dB HL for the octave frequencies between 0.125 and 4k-Hz. Working memory was assessed by the Reading-Span test and speech in noise was measured using the Matrix sentence test. To investigate effect of age on the association between working memory and speech in noise performance, participants were divided into four groups based on age: young (18 – 39 years), middle age (40 – 50 years), old (60 – 69 years), and old-old (70 – 91 years), and a separate correlational analysis was conducted for each age group. The researchers found that correlations between working memory and speech in noise performance were weak and not significant for the young participants, but moderately strong and significant for the three older groups. Thus, consistent with the findings of this study, the meta-analysis did not reveal that in adverse listening environments, working memory is a reliable predictor of speech in noise performance in young normal hearing listeners (Fullgrabe & Rosen, 2016).

Language. Listeners are constantly matching the acoustic information they hear to internal representations of what they have heard before (Kaandorp, et al., 2016). Thus, I hypothesized the number of words in a listener's mental lexicon could be important for this process and selected vocabulary size as evaluated through the

Toolbox Picture Vocabulary Test (TPVT) as our measure of language. In the TPVT, participants were presented with an audio recording of a word and four photographs on an iPad screen. They were instructed to select the photograph that matches the meaning of the word they heard. The results of this study did not find vocabulary size to be a predictor of speech in noise performance in young normal hearing listeners.

A study by Kaandorp and colleagues (2016) investigated the effect of two linguistic abilities, lexical-access and vocabulary size, on speech in noise performance in normal hearing listeners with different levels of language proficiency (native and non-native language backgrounds). Lexical access refers to the retrieval of information from a pool of mentally stored information (ie: mental lexicon), whereas vocabulary size refers to the receptive vocabulary of a listener (Kaandorp et al., 2016). The authors found that lexical-access explained about 60% of the variance in their speech in noise measures in listeners with normal hearing. They concluded that lexical access over vocabulary size is an important predictor of speech recognition in noise. Lexical access is measured with lexical-decision and word-naming tasks that measure how quickly participants can classify stimuli as words and non-words (Kaandorp et al., 2016); and these measures are clinically feasible. Our measure of linguistic ability was the TPVT, which measures a listener's vocabulary size, not lexical-access ability. Vocabulary knowledge was selected by the NIH Toolbox as a measure of language because of its high correlation with general measures of intelligence (Weintraub et al., 2013). However, given the findings of Kaandorp and colleagues (2016), it is possible that a measure of lexical-access may have produced a significant main effect for language in our speech in noise listening outcomes over a measure of vocabulary size. This is a consideration for future projects.

Processing Speed. While processing speed was correlated with a speech in noise measure (sentences presented in babble and a low noise level), it did not produce a significant main effect for the linear mixed-model analyses for sentences.

Processing speed refers to the rate at which information is processed in order to perform a mental task (Dryden et al., 2017). Processing speed is associated with speech perception due to the sequential nature of speech and it is suggested that quick processing is even more important when speech is linguistically complex or degraded in some way (Dryden et al., 2017). Our measure of processing speed was the Pattern Comparison Processing Speed Test. This test required participants to identify whether two visual patterns were the same or different. The results of this study did not find processing speed to be a predictor of speech in noise performance in young normal hearing listeners.

Dryden and colleagues (2017) explored the relationship between cognition and speech in noise perception and found a moderate correlation ($r = .39$) between processing speed and speech perception in noise. The review included four studies of processing speed, however, much like the literature on working memory, the review did not differentiate between normal hearing and hearing impaired listeners in their analysis. Additionally, participants in the review studies were middle age to older listeners ($\geq$ 55 years of age). Supra-threshold changes in sensory abilities have been document in older listeners, even in the absence of hearing loss; and these changes in sensory abilities in listeners could place a greater reliance on cognitive abilities like processing speed to understand speech in noise.

Moreover, the trajectory for processing speed changes with age, reaching its peak in young adulthood (Zekveld, Kramer, Festen, & 2014). For example, Zekveld and colleagues (2014) examined cognitive load during speech perception in noise and examined the effects of hearing loss, age, and cognition. Participants included middle age normal hearing and hearing impaired adults. Processing speed was measured with the letter digit substitution test (LDST) and speech perception in noise was measured using sentence stimuli in stationary noise. In the LDST, letters are presented to participants and they are asked to write the corresponding digit, according to a key, in a blank space below the letter (Zekveld et al., 2014). Zekveld et al. (2014) compared processing speed results from the current study using middle age participants to an earlier study (Zekveld, Kramer, & Festen, 2010) using young normal hearing participants. The researchers found that young normal hearing participants from the previous study (Zekveld et al., 2010) had higher (better) processing speed than both groups of middle-aged participants from Zekveld et al. (2014). This finding is consistent with the literature on the trajectory for processing speed, which peaks at young adulthood and steadily declines as people age (Zekveld et al., 2014; Weintraub, et al., 2013). The participants in the present study were young normal hearing adults (< 29 years of age). Thus, while individual differences in processing speed likely exist, it association with speech in noise outcomes might not be measurable until a listener is older.

Hearing Sensitivity. Participants were required to have normal hearing sensitivity for the octave frequencies between 250 Hz and 8000 Hz, including normal middle ear function to participate in this study. As a result, hearing sensitivity varied within a narrow range, with all listeners having better than 20 dB HL hearing. Thus, hearing sensitivity was not a significant factor as a source of individual variability in speech in noise performance in young normal hearing listeners.

Temporal Acuity. The current study did not find an association between temporal acuity and speech perception in noise. This observation might be due to the fact that the participants in this study all performed reasonably well on temporal acuity measures. For example, the range of scores for temporal window slope is fairly small when compared to the range for auditory filter shape. Additionally, participants in this study performed significantly better on speech in noise tasks at high noise levels when the background noise was babble vs. speech noise, as revealed in the significant noise type-by-noise level interaction for both single words and sentences. As mentioned, competing talker backgrounds that are composed of two to three talkers are often more difficult for listeners than when the masker is stationary, due to the effects of perceptual interference caused by information masking (Carhart et al., 1975). However, listeners in this study were able to overcome the higher-level lexical interference cause by two-talker babble and take advantage of the temporal gaps in the fluctuating background to increase their speech perception in noise performance at high noise levels. Thus, it is possible that all listeners in this study had good temporal acuity and therefore a significant main effect for temporal acuity was not observed.

5.4 Clinical Implications

The following section discusses the clinical implications related to the study. The discussion begins with a review of clinical measures of auditory properties followed by considerations of the clinical relevance of this study.

5.4.1 Clinical Measures of Auditory Properties

Classical measures of auditory filter and temporal window shape are conducted using psychoacoustic procedures. For example, a measure of auditory filter shape is a psychophysical tuning curve. Psychophysical tuning curves are measured with a tone that is fixed in frequency and level, presented simultaneously with a masker at varying frequencies and levels (Charaziak, Souza, & Siegel, 2012). The masker level that causes the signal to be just audible is determined and plotted as a function of its frequency, forming a "V" shaped psychophysical tuning cure (Charaziak, et al., 2012). The tuning curve shape describes the ability of the auditory system to filter out one stimulus (the signal) from the others (the masker) as a function of frequency (Charaziak, et al., 2012).

Similarly, the shape of a listener's temporal window is traditionally measured using a psychophysical method. A listener's temporal window can be estimated by measuring the threshold for a brief sinusoidal signal presented in a temporal gap between two bursts of noise as a function of the duration of the gap and the position of the signal within the gap (Moore, 1998). The temporal window shape is associated with a listener's temporal acuity (Moore, 1998). Standard psychophysical procedures that are used to measure auditory filter and temporal window shape are traditionally used in laboratory settings, and because of the time requirements necessary to administer these measures, they are unsuitable for clinical use.

A tertiary goal of this project was to develop clinically feasible measures of auditory filter and temporal window shape. Our measures were established using a threshold measurement obtained in steady noise and several notched noise and amplitude modulated (AM) noise conditions for auditory filter width and temporal window shape, respectively. Noise was delivered monaurally to the same ear and the slope of the function for each measure was an estimate of a listener's auditory filter or temporal window shape. These measures utilize equipment already found in audiology clinics, measurement techniques familiar to an audiologist, and can be completed within a reasonable amount of time making them clinically practical.

Auditory Filter Slope. To confirm the validity of our auditory filter shape scores, we compared our data to data found in the literature using the same measurement technique and notch-width (Unoki, Ito, Ishimoto, & Tan, 2006). Data from Unoki and colleagues (2006) found auditory filter shape slopes at 500 Hz and 4000 Hz to be -6.8 dB and -8 dB, respectively. Our slope measurements for auditory filter shape at 500 Hz and 4000 Hz were -5.6 dB and -9.7 dB respectively. Thus, data from this dissertation follows the trend found in the literature: a listener's thresholds decreases with increasing notch-width and steeper auditory filter slopes are obtained at 4000 Hz when compared to 500 Hz. Additionally, our data is within one standard deviation of the data by Unoki and colleagues (2006). Taken together, I believe our auditory filter slope scores are valid measures of auditory filter shape.

Temporal Window Slope. Similarly, to confirm the validity of our temporal window slope scores, we compared our data to data in the literature. Shen and Richards (2013) used a similar technique to estimate the shape of a listener's temporal window. Temporal window slope measures were averaged for center frequencies between 30 and 200 Hz at modulation rates of 8, 16, 32, 64, 128, 256, and 512 Hz. Our temporal window slope measurements were calculated at center frequencies of 500 and 4000 Hz with modulation rates of: 0, 80, 160, 320 Hz. To compare our data to that of Shen and Richards (2013), slope measurements from Shen and Richards (2013) were calculated for modulation rates of: 8, 64, 128, and 256 Hz and compared to our data. Average temporal window slope scores from Shen and Richards (2013) were -2.2 dB. Our temporal window slope at 500 and 4000 Hz were -2.3 and -2.7, respectively. Our data follows the trend noted in the literature that increasing modulation rate decreases the detection threshold for the listener. Additionally, our data is within one standard deviation of the averaged data by Shen & Richards (2013). Thus, I believe our temporal window slope scores are valid measures of a listener's temporal window.

5.6.2 Clinical Significance

The percentage of the population with difficulties perceiving speech approximately doubles with every decade in age (Fullgrabe & Rosen, 2016; Plomp, 1978). While hearing loss and changes in sensory abilities in this population are major contributors, they are generally not sufficient to fully account for the individual differences in speech in noise performance that are commonly seen in clinical environments. Thus, from a clinical perspective, more thorough assessments of speech perception in noise including sensory and cognitive assessments in a range of listening environments may still be valuable in explaining speech in noise performance for hearing impaired patients. Additionally, as suggested by Lunner (2003), cognitive assessments may also be helpful in improving the prediction of aided speech perception in noise abilities for hearing aid patients. The findings of this study do not reduce the practical importance of cognitive assessments in the use of future counseling and prediction of speech in noise performance for older hearing-impaired listeners.

Additionally, traditional measures of speech in noise performance do not fully capture a listener's ability to perceive speech in noise. These assessments typically involve one type of speech context (words vs. sentences), noise type (speech noise vs. babble), and noise level and measure recognition abilities only (Van Engen, et al., 2014). These evaluations emphasize the need to evaluate ecologically valid listening environments that reflect a listener's everyday listening experiences. The present study contributes to the notion that through more thorough assessments of speech in noise performance, which include a greater variety of listening conditions, in addition to measures of sensory *and* cognitive function, it might become possible to identify sources of individual differences in speech in noise outcomes.

A limitation of this project was the homogeneity of the population. A future direction of this project can explore administering a similar battery of measurements in older listeners with normal hearing to evaluate the relationship between cognition and speech perception in noise in listeners who might exhibit greater individual variability based on age. Additionally, while speech perception and sensory assessments were administered in the presence of background noise, cognitive measurements were administered in quite. Therefore, future studies might include evaluating listener's cognitive performance in the presence of background sounds.

5.5 Limitations

A limitation of this study could be that the cognitive tools selected from the NIH Cognitive Toolbox Battery were inappropriate measures of these abilities. The rational for the inclusion of specific cognitive constructs was based on an exhaustive review of the literature (Weintraub, et al., 2013). Additionally, test-retest reliability of the specific instruments that were used from the NIH Cognitive Battery for this study was strong. For example, the test re-test reliability for: the Flanker Inhibitory Control and Attention Test = .94, List Sorting Working Memory Test = .77, Toolbox

Picture Vocabulary Test = .81, and Pattern Comparison Processing Speed Test = .81 (Weintraub, et al., 2013). Additionally, correlations for convergent validity (expressing how well each measure in the battery evaluated the intended construct) ranged from r = -0.48 to 0.78 for all of the cognitive subtests chosen for this study, suggesting that the NIH Cognitive Toolbox measures were tapping into the desired cognitive constructs (Weintraub et al., 2013). Moreover, given the significant main effect for attention for listening environments containing sentences, the convergent validity data of attention suggest that it is a global process, and not modality specific. Taken together, the NIH Cognitive Toolbox Battery is an accessible set of instruments that are psychometrically sound and efficient, and can be used on a wide range of populations. For these reasons, we feel the cognitive toolbox and its subdomains for attention, working memory, language, and processing speed were suitable measures to evaluate cognitive capacity in this study.

Another limitation may pertain to the selection of measures for sensory abilities, specifically frequency resolution and temporal acuity, as either could be measured in different ways (i.e., psychophysical tuning curves, gap detection measures). The approach used to measure frequency selectivity and temporal acuity in this study was appealing as they were analogous to one another, and used standard audiometric equipment. Although our measurement technique and outcomes for frequency resolution and temporal acuity align with data found in the literature (Unoki et al., 2006; Shen & Richards, 2013), our clinical measures of filter shape and temporal window should be validated against traditional psychoacoustical findings.

Moreover, it could be argued that the design of the speech in noise measures used in this study to assess speech perception were not challenging enough, and differences in sensory or cognitive capacity are only identified under more linguistically complex tasks. The purpose of our speech stimuli and listening contexts were to create a diverse range of acoustically adverse listening environments. In the more challenging listening conditions, listener's obtained percent correct scores < 10%; and anecdotally, the participants found the task taxing and exhausting. For these reasons, we believe that that our range of listening environments provided a sufficient amount of linguistic complexity for the participants.

Additionally, a limitation of the study could be that the population was too homogenous and good too (i.e., listeners in this study did not have significant difficulties understanding speech in noise) and the results of the study can only be generalized to a specific subset of the population. While the participants included in the present study were mostly female, of a certain age, and having normal hearing, the speech perception and sensory and cognitive data for this study showed a wide range of performance by the listeners.

Finally, in addition to completing the statistical analysis in the results section, principle component analysis (PCA) was explored. The Kaiser-Meyer-Olkin Test and Bartlett's Test of Sphericity, which are measures of how well the data is suited for PCA (factor analysis), were not significant. Thus, a PCA was not undertaken.

5.6 Conclusions

The primary aim of this investigation was to investigate the sensory and cognitive factors that may contribute to speech perception in noise. Data from the correlational analysis examining the covariates and independent variables indicate that a greater number of listener abilities emerge for speech in noise performance when the speech context is sentences.

While this study did not reveal a significant effect for seven of the nine sensory and cognitive variables, this does not mean that speech perception in noise does not involve a broad range of sensory and cognitive abilities. It's possible that even when individual difference among listeners exist, the sensory and cognitive abilities of most young normal hearing individuals are suitable for the purpose of speech perception in noise. Further systematic efforts will help specify under which acoustic and linguistically complex listening environments the different sensory and cognitive abilities come into play in this population. The results of the discussion emphasize the need for clear labeling of participant characteristics such as age and hearing status when reporting and generalizing results of these studies.

Additionally, with the information from the mixed linear model, the present study supports the notion for a comprehensive approach to speech in noise testing, in which listeners are tested in multiple types of maskers and noise levels. Such an approach would provide clinicians with valuable information about listeners' strengths and weaknesses in understanding speech in noise. Moreover, while this study used young normal hearing listeners, understanding basic knowledge of the factors that influence speech in noise performance will help professionals continue to improve clinical assessments. Thus, as professionals are considering what measures to include in their speech in noise test battery, this investigation suggests sentence recognition tasks matter and varying the type and level of noise also matters. Thus, as clinicians, we may need to increase the number of conditions that sentences in noise are assessed in.

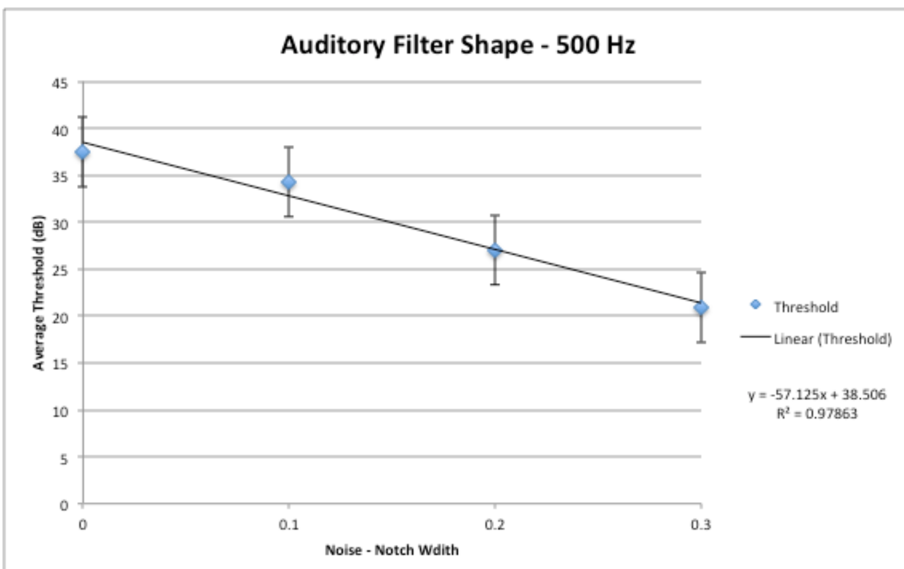
Appendices

Appendix A

Notch-Width Threshold as a Function of Noise-Notch Width

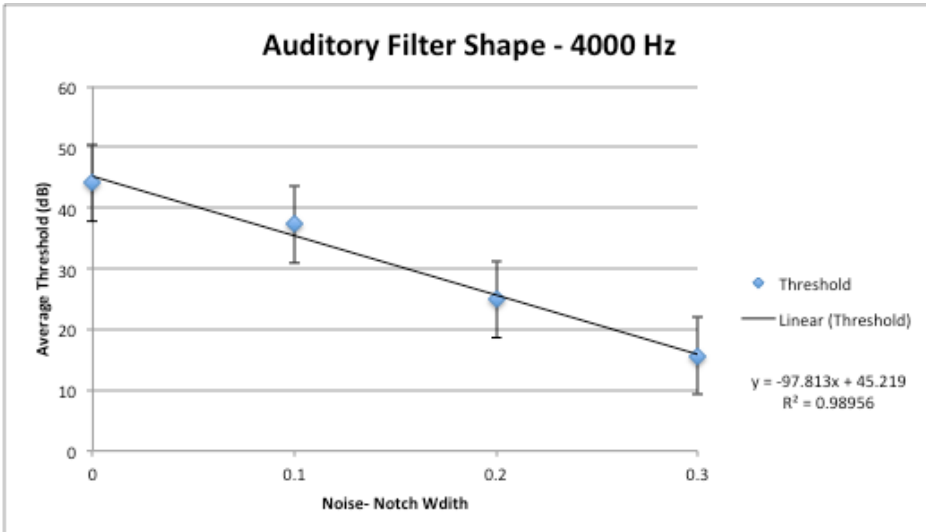
Frequency selectivity refers to the capacity to discriminate successive sounds by frequency can be evaluated through measures of auditory filter shape. Psychophysical procedures that measure auditory filter shape are time consuming and use equipment not regularly found in audiology clinics. Therefore, these approaches are unsuitable for clinical use. Our measure of frequency selectivity can be done clinically and efficiently, using standard audiometric equipment (audiometer and earphones), measures (threshold search paradigm), and stimuli (pure-tones) that are defined by ANSI standards. For this project, auditory filter shape was calculated for each listener as the slope of four thresholds as a function of noise-notch width. A listener's threshold for a pure tone was measured monaurally in a no noise-notch condition as well as three notched noise conditions at 500Hz and 4000Hz. The no-noise notch condition resulted in the greatest amount of masking of the pure tone. For the 500-Hz signal, the three notch widths were 0, 100, 200 and 300-Hz and for the 4000-Hz signal, the notch widths were 0, 800, 1600, and 3200-Hz. The notch within the noise band was centered around the center frequency of the pure tone (either 500-Hz or 4000-Hz), and the width of the notch was calculated as a proportion of the center frequency (Unoki et al., 2006). A listener's threshold in each noise condition was plotted against notch width and their auditory filter shape is represented by the slope of their auditory threshold as a function of notch width. As the notch-width within the noise increases, so does the listener's ability to detect the pure tone, resulting in a decrease in the detection threshold of the pure tone for the listener. The slope of the function estimates one edge of the auditory filter shape. We are considering the filter to be symmetrical, and therefore having a similar slope for the opposite edge. The steeper a listener's slope, the narrower their auditory filter shape, and the better their frequency selectivity ability. Figure 3 and 4 below illustrates our measure of auditory filter shape at 500 Hz and 4000Hz.

Figure 1A. Auditory Filter Shape Estimate at 500 Hz



Note. Figure 1A illustrates our estimate of auditory filter shape at 500 Hz that was calculated for each listener as the slope of four thresholds as a function of noise-notch width. The notch within the noise band was centered around the center frequency of the pure tone (500 Hz), and the width of the notch was calculated as a proportion of the center frequency (.1, .2, or .3).

Figure 2A. Auditory Shape Estimate at 4000 Hz



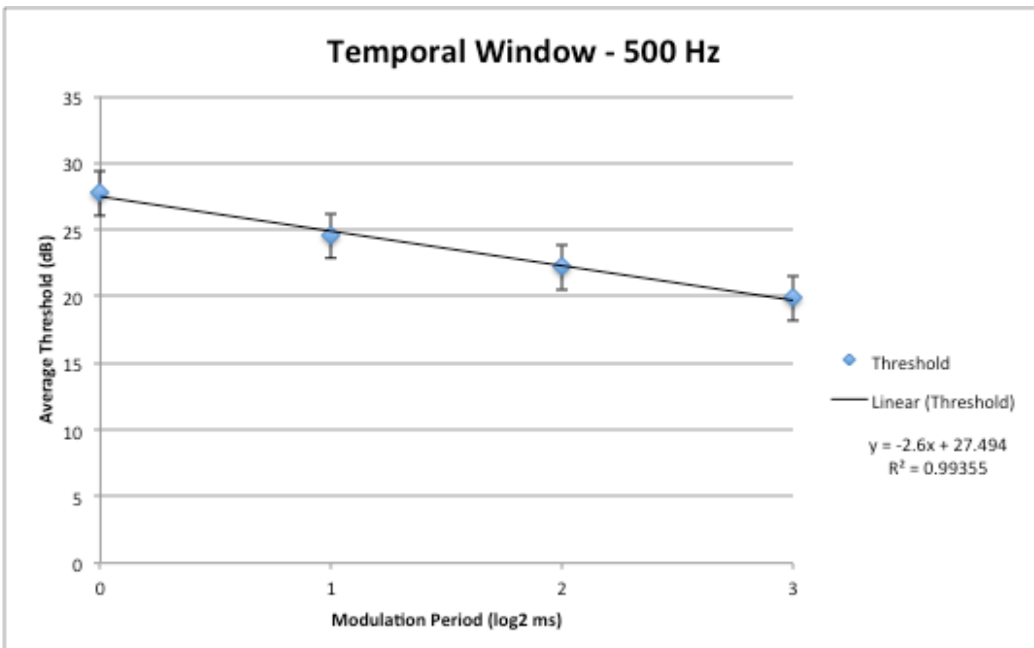
Note. Figure 2A illustrates our estimate of auditory filter shape at 4000 Hz that was calculated for each listener as the slope of four thresholds as a function of noise-notch width. The notch within the noise band was centered around the center frequency of the pure tone (4000 Hz), and the width of the notch was calculated as a proportion of the center frequency (.1, .2, or .3).

Appendix B

Modulation Threshold as a Function of Modulation Frequency

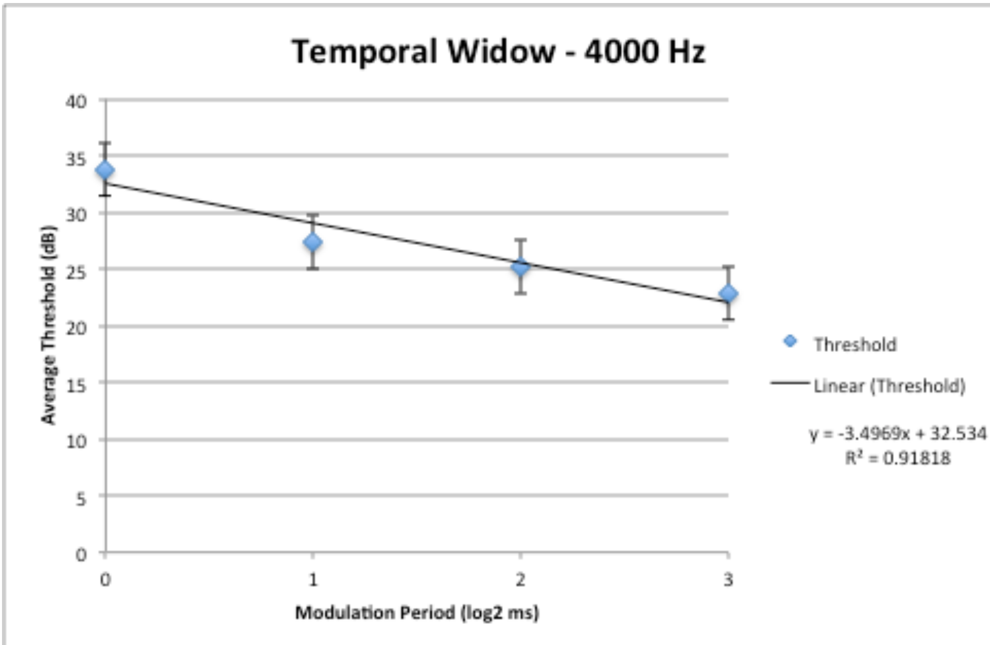
Similarly, psychophysical measures of temporal resolution are time consuming and use equipment not regularly found in audiology clinics. Therefore, these measures are not suitable for clinical use. Temporal resolution refers to a listener's ability to detect changes in stimuli over time, for example, detecting a brief gap between two stimuli or identifying that a sound is modulated in some way (Moore, 2008). A listener's temporal resolution can be evaluated through measures of temporal window shape. Like auditory filter shape, our measure of temporal window can be completed clinically and efficiently, using standard audiometric equipment, measures, and stimuli that are defined by ANSI standards. For this project, temporal resolution was evaluated for each listener as the slope of four thresholds as a function of modulation period. A listener's threshold for a pure-tone signal was measured monaurally in un-modulated noise and three modulated noise conditions (80Hz, 160Hz, 320Hz) at 500Hz and 4000Hz. Un-modulated noise resulted in the greatest amount of masking of the pure-tone. As the modulation period increased, so did the amount of gaps within the noise, increasing the audibility of the pure-tone. A listener's threshold in each modulated noise condition was plotted against modulation period and their temporal resolution ability is represented by the slope of their auditory threshold as a function of modulation period (for scaling purposes, the modulation period is represented as a $\log_2$). The slope of the function estimates one edge of the temporal window. Similar to auditory filter shape, we are considering the temporal window to be symmetrical, and therefore having a similar slope for the opposite edge. The steeper a listener's slope, the narrower their temporal window, and the better their temporal resolution ability. Figure 5 and 6 below illustrates our measure of temporal resolution for 500-Hz and 4000-Hz.

Figure 1B. Temporal Window Shape Estimate at 500 Hz



Note. Figure 1B illustrates our estimate of temporal window shape at 500 Hz. A listener's threshold for 500 Hz was measured in un-modulated noise and three modulated noise conditions (80Hz, 160Hz, 320Hz). Listener's thresholds in each modulated noise condition was plotted against modulation period and their temporal resolution ability is represented by the slope of their auditory threshold as a function of modulation period.

Figure 2B. Temporal Window Shape Estimate at 4000 Hz



Note. Figure 2B illustrates our estimate of temporal window shape at 4000 Hz. A listener's threshold for 4000 Hz was measured in un-modulated noise and three modulated noise conditions (80Hz, 160Hz, 320Hz). Listener's thresholds in each modulated noise condition was plotted against modulation period and their temporal resolution ability is represented by the slope of their auditory threshold as a function of modulation period.

References

Akeroyd, M. (2006). The psychoacoustics of binaural hearing. *International Journal of Audiology*, 45(1), 25-33.

Anderson, S, Parbery-Clark, A., Yi, H.G., Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, 32(6), 750-757.

Baddeley, A. (1992). Working memory. Science, 255, 556-559.

Baddeley, A., Hitch, G. (1974). Psychology of learning and motivation. 8, 47-89.

Baer, T., Moore, B.C.J. (1993). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *Journal of the Acoustical Society of America*, 95(4), 2277 – 2280.

Baer, T., Moore, B.C. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *Journal of the Acoustical Society of America*, 95(4), 2277 -2280.

Broadbent, D. (1958). Perception and Communication. London: Pergamon Press.

Carhart, R., Johnson, C., Goodman, J. (1975). Perceptual masking of spondees by combination of talkers. *Journal of the Acoustical Society of America*, 58, 35.

Charaziak, K., Souza, P., Siegel, J. (2012). The-efficient measures of auditory frequency selectivity. *International Journal of Audiology*, 51(4), 317 – 324.

Dorman, M., Loizou, P., Fitzke, J., Tu, Z. (2004). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. *Journal of the Acoustical Society of America*. 1998;104:3583–3585

Dryden, A., Harriet, A., Henshaw, H., Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic literature review and meta-analysis. *Trends in Hearing*, 21, 1-21.

Edwards B.(2016) A model of auditory-cognitive processing and relevance to clinical applicability, *Ear and Hearing*, 37, 85S – 91S.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton, The Hague.

Firszt, J., Reeder, R., Skinner, M. (2008). Restoring hearing symmetry with two cochlear implants or one cochlear implant and a contralateral hearing aid. *Journal of Rehabilitation Research & Development*, 45(5), 749 – 767.

Fullgrabe, C., Rose, S. (2016). Investigating the role of working memory in speech in noise identification for listeners with normal hearing. *Advances in Experimental Medicine and Biology*, 894, 29 – 36.

Fullgrabe, C., Rosen, S. (2015). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology*. 7, 1268.

Gardner, H., & Hatch, T. (1989). Educational Implications of the Theory of Multiple Intelligences. *Educational Researcher*, *18*(8), 4–10.

Gershon, R., Cook, K., Mungas, D., Manly, J., Slotkin, J., Beaumont, J., Wintraub., S. (2014). Language measures of the NIH toolbox cognitive battery. *Journal of the International Neuropsychological Society*. 20(6): 642-651.

Hoover, E., Pasquesi, L., Souza, P. (2016). Comparision of clinical and traditional gap detection tests. *Journal of the American Academy of Audiology*, 26(6): 540-546.

Hopkins, K., Moore, B.C. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *Journal of the Acoustical Society of America*, 123, 1140-1153.

Hornsby, B., Ricketts, T., Johnson, E. (2006). The effects of speech and speechlike maskers on unaided and aided speech recognition in persons with hearing loss. *Journal of the Academy of Audiology*, 17(6), 432 – 447.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 19, 741 – 758.

Kaandorp, M., De Groot, A., Festen, J., Smits, C., Goverts, S. (2016). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. International Journal of Audiology. 55(3): 157 – 167.

Lehiste I., Peterson G. E. (1959) Linguistic considerations in the study of speech intelligibility. *Journal of the Acoustical Society of America,* 31(3): 280–286.

Lidestam, B., Jolgersson, J., Moradi, S. (2014). Comparison of information vs. energetic masking effects on speechreading performance. *Frontiers in Psychology*, 4, 639.

Liu, C., Kewley-Port, D. (2007). Factors affecting vowel formant discrimination by hearing-impaired listeners. *Journal of the Acoustical Society of America*, 122(5), 2855 – 2864.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. International Journal of Audiology. 42, 49- 58.

MacPherson, A., Akeroyd, M. (2014). Variations in the slope of psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*. 18, 1-26.

Mattys, S., Davis, M., Bradlow, A., Scott, S. (2012). Speech recognition in adverse conditions: a review. *Language and Cognitive Processes*. 27(7/8), 953-978.

McLeod, S. A. (2018, Oct 24). Selective attention. Retrieved from https://www.simplypsychology.org/attention-models.html.

McShefferty, d., Whitmer, W., Akeroyd, M. (2016). The just-meaningful difference in speech-to-nice ratio. *Trends in Hearing*, 20, 1-11.

Miller, G., Licklider, J. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America.* 22(2), 167 – 173.

Moore, B.C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal for the Association of Research in Otolaryngology*, 9(4), 399-406.

Oxenham, A., J. (2012). Pitch perception. *Journal of Neuroscience*, 26:32(39), 13335-13338.

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*. 42, 26– 32.

Rudner, M., Ronnberg, J., Lunner T. (2011). Working memory supports listening in noise for persons with hearing impairment. *Journal of the American Academy of Audiology*. 22(3), 156 – 167.

Sek, A., Moore, B. (1998). Frequency discrimination as a function of frequency , measured several ways. *The Journal of the Acoustical Society of America*, 97(4).

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. *The Handbook of Aging and Cognition*, 155-219.

Shen, Y., Richards, V.M. (2013). Temporal modulation transfer function for efficient assessment of auditory temporal resolution. *Journal of the Acoustical Society of America*. 133(2), 1031 – 1042.

Sommers, M., S., & Danielson, S.M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, 14, 458 – 472.

Sommers, M., Humes, L. (1993). Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners. *Journal of the Acoustical Society of America*, 93(5), 2903 – 2914.

Strait, D., Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontier in Psychology*, 2, 1- 10.

Tamati, T., Gilbert, J., Pisoni, D. (2013). Some factors underlying individual differences in speech recognition on PRESTO: a first report. *Journal of the American Academy of Audiology*, 24(7), 616-634.

Treisman, A., 1964. Selective attention in man. British Medical Bulletin, 20, 12-16.

Tulsky, D., Carlozzi, N., Chevalier, N., Espy, K., Beaumont, J., Mungas, D. (2013). NIH toolbox cognitive function battery (NIHTB_CFB): measuring working memory. *Monographs of the Society for Research in Child Development*. 78(4): 70-87.

Unoki, J., Kazuhito, I, Yuichi, I., Chin-Tuan, T. (2006). Estimate of auditory filter shape using notched-noise masking for various signal frequencies. *Acoustical Science and Technology*, 27(1), 1 – 11.

Van Engen, K., Phelps, E.B., Smiljanic, R., Chandrasekaran, B. (2014). Enhancing speech intelligibility: interactions among context, modality, speech style, and masker. *Journal of Speech Language and Hearing Research*, 57, 1908-1918.

van Wijngaarden, S., Steeneken, H., Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native talkers. *Journal of the Acoustical Society of America,* 112(6), 3004 – 3013.

Weintraub, S., Dikmen, S., Heaton, RK, Tulsky DS, Zelazo PD, Bauer PJ, Gershon RC. (2013). Cognition assessment using the NIH Toolbox. *Neurology*. 80(11), 54 – 64.

Wingfield A. (2016) The evolution of models of working memory and cognitive resources. *Ear Hearing*, 37, 35S–43S.

Winn, M., Edwards, J., Litovsky, R. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and Hearing*, 36(4), 153 – 165.

Zekveld, A., Kramer, S., Festen, J. (2010) Pupil response as an indication of effortful listening: The influence of stenence intelligibility. *Ear and Hearing*, 32(4), 480 - 490.

Zekveld, A., Kramer, S., Festen, J. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response, *Ear and Hearing*, 32(4), 498-510.

Zelazo, P.D., Anderson, J., Richler, J., Wallner-Allen, K., Beaumont, J., Conway, R.G, Weintraub, S. (2014). NIH cognition battery (CB): Validation of executive function measures in Adults. *Journal of the International Neuropsychological Society*, 20(6): 620-629.

VITA

Sangeeta C. Kamdar was born in Scottsdale, Arizona, and lived in Cupertino, California, before her family eventually settled in Plano, Texas. She graduated from Plano West Senior High School in 2002, and attended the University of Texas at Austin. At UT, she earned her Bachelor of Science in Communication Sciences and Disorders in May of 2007 and her Doctor of Audiology in 2011. While enrolled as an Au.D student, she became interested in research and started working Dr. Su-Hyun Jin and Chang Liu's lab. She entered the Ph.D program in 2009. While completing her Ph.D, Sangeeta worked at Austin ENT Clinic as a clinical audiologist from 2010-2016. She is currently a lecturer at the University of Texas at Austin.

Email: kamdarsc@gmail.com

This dissertation was typed by the author.