TOWSON UNIVERSITY
COLLEGE OF GRADUATE STUDIES AND RESEARCH

RELIABILITY OF THE AMERICAN ENGLISH MATRIX TEST

by

Holly A. Rose

A thesis

presented to the faculty of

Towson University

in partial fulfillment

of the requirements for the degree

Doctor of Audiology

Department of Audiology, Speech-Language
Pathology and Deaf Studies

Towson University
Towson, Maryland 21252

May 2013

THESIS APPROVAL PAGE

This is to certify that the thesis prepared by Holly Rose entitled "Reliability of the American English Matrix Test" has been approved by the thesis committee as satisfactorily completing the thesis requirements for the degree Doctor of Audiology.

Brian Kreisman, Ph.D. *Chair, Thesis Committee*	May 3, 2012 *Date*
Andrew John, Ph.D. *Committee Member*	May 3, 2012 *Date*
Candace Robinson, Au.D. *Committee Member*	May 3, 2012 *Date*
Janet DeLany, D.Ed. *Dean of Graduate Studies*	*Date*

Acknowledgements

I would like to acknowledge the chair of my thesis committee, Dr. Brian Kreisman, for all of his guidance and support throughout the entire thesis process. I would also like to thank my committee members, Dr. Andrew John and Dr. Candace Robinson for their input to improve my thesis. To all of the individuals who took time out of their busy schedules to participate in my thesis, thank you. Last but not least, I would like to thank my husband and family members for all of their support throughout this process.

Abstract

Reliability of the American English Matrix Test

Holly A. Rose

The current study assessed the list equivalency of six lists from the American English Matrix test as well as training effects when using this test with an adaptive speech stimulus and a fixed speech stimulus. Twenty (*N*=20) participants (5 males, 15 females) with normal hearing were included in this study. Repeated-measures ANOVA results showed the six test lists were equivalent; in other words, the speech reception threshold (SRT) scores from the participants were comparable. Repeated-measures ANOVA results between the three fixed signal-to-noise ratios (SNRs) used in the fixed speech stimulus condition, -6 dB SNR, -8.5 dB SNR and -11 dB SNR, showed significant differences in scores. More specifically, when the fixed SNRs increased in difficulty, the participants' scores decreased. However, the participants' scores were higher than the predicted intelligibility scores of 80%, 50%, and 20%, respectively. Repeated-measures ANOVA results within each session and between sessions showed an increase in SRT scores indicating strong training effects. Further research is needed to determine if gender effects as well as ear effects impact the scores for each list. It is recommended that future studies use lists that do not contain duplicate sentences, in order to truly look at list

equivalency and training effects. Following the collection of the data with normal hearing individuals, further research should be conducted to determine how individuals with varying degrees of hearing loss, as well as varying ages, perform on the American English Matrix test.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

Understanding speech in noisy listening environments is often difficult compared to understanding speech when in quiet listening conditions. Numerous tests have been created to assess the speech-perception abilities of individuals in both quiet conditions and conditions with competing noise. Several studies have tested diverse groups of individuals with various hearing abilities, as well as with various age differences, speech stimuli, and masking noise stimuli. Other factors that have been considered when creating these tests include, but are not limited to, accuracy of the test, difficulty in scoring the test, and practicality of the test in the clinical world.

The current study analyzes an American English version of the Oldenburg sentence test, German version (Wagener & Brand, 2005), and the Dantale II sentence test, the Danish version (Wagener, Josvassen, & Ardenkjar, 2003). All of these tests contain a base list with five different categories: name, verb, number, adjective and object. Each of the categories is made up of 10 words. For every sentence, a word is chosen from each of the categories from the base list to make up a nonsense sentence (e.g. Peter wants four old toys). However, every sentence will have a similar structure in that each contains a name, verb, number, adjective and object. The noise competition for these tests is steady-state speech spectrum shaped noise. The purpose of the present study is to assess the list equivalency of the American English Matrix test as well as determine if there are any training effects from this test for individuals with normal hearing.

Chapter 2

Review of Literature

Because of the specific focus of this thesis project, the scope of this literature review is limited to the development and validity of different speech perception tests. Specifically, this chapter will review variability within speech-in-noise test material, between hearing abilities and performance in speech-in-noise testing, between age and performance in speech-in-noise testing, between different speech materials, between different scoring methods, and between the types of competing noise in speech-in-noise testing.

Variability within Speech-in-Noise Test Material

It is important that the test itself has little variability, such as comparable results between different lists, within groups of various age ranges, as well as groups of various hearing abilities. One goal of developing a reliable speech-in-noise test is little variability between different test lists found for individual subjects as well as within groups of individuals with similar hearing abilities and age (Cox, Alexander, & Gillmore, 1987; Dirks, Morgan, & Dubno, 1982; McArdle & Wilson, 2006; Plomp & Mimpen, 1979; Versfeld & Dreschler, 2002; Wagener, Josvassen & Ardenkjar, 2003; Wilson & McArdle, 2007).

In a study by McArdle and Wilson (2006), the consistency of performance from 18 lists of the Quick Speech-In-Noise (QuickSIN) test was analyzed with monaural presentation via headphones resulting in deletion of several lists in order to achieve minimal variability between lists. This was accomplished by testing the performances of two different groups: (1) a normal-hearing group of 24 listeners with a mean age of 22.0 years and, (2) a hearing-impaired group of 72 listeners with a mean age of 66.4 years. All of the individuals in group 2 had confirmed high-frequency sensorineural hearing loss (SNHL) with no greater than a mild hearing loss in the low frequencies. Furthermore, in order for members of group 2 to participate in the study, it was required for them to have at least a 40% or higher score on speech testing in quiet.

On average, group 2 demonstrated a mean SNR of 8.7 dB higher (i.e., poorer performance) on the 18 different lists than group 1, with greater variability in SNRs in group 2. Of the 72 individuals with hearing loss, five performed equally as well as the participants in group 1. The remaining subjects in the hearing-impaired group performed poorer than individuals with normal hearing abilities.

Given the results with list variability, it was decided to maintain 12 of the original 18 lists for the QuickSIN test. Determining the deviation from the mean of the thresholds of all lists and eliminating the three lists with the highest mean thresholds along with the three lists with the lowest mean thresholds accomplished this. This eliminated lists 4, 5, and 13 (lowest thresholds) and 7, 14, and 16 (highest thresholds). Furthermore, lists 3, 9, and 18 were also discarded from the final QuickSIN battery, as their results were irregular. Of the original 18 lists, only nine were kept with variability of less than 1 dB overall between one another.

Similar procedures were used by Plomp and Mimpen (1979) when reducing the variability within different test lists for assessing speech recognition thresholds (SRTs) with sentences in the presence of background noise. These lists contained a total of 220 sentences that were initially analyzed by several audiologists and speech pathologists, ultimately resulting in the removal of 50 sentences. To test the variability between the sentences, the speech stimuli were presented binaurally under the headphone condition to 20 normal-hearing individuals at roughly 50 dBA. The noise utilized was equivalent to the average spectrum of speech. After obtaining data on the different sentences, 40 of the easiest and most difficult sentences were eliminated to reduce variability. The final number of sentences for the study was 130. The sentences were then broken up into 10 different lists. To assess the consistency between the lists, SRTs were obtained in 10 listeners for each list. Variability between the lists was found to be less than 0.9 dB, which was comparable to the variability found in another study (McArdle & Wilson, 2006).

In a study by Wilson and McArdle (2007), variability was assessed for the words-in-noise (WIN) test under the headphone condition. This study was conducted with two experiments. In experiment 1 the WIN was presented with multitalker babble used as the masking noise to 315 individuals with SNHL. Speech scores in quiet were also assessed. Furthermore, the participants were tested on two separate occasions in order to examine the test-retest reliability of the WIN. Little variability was seen in the mean scores between the two sessions; however, the 0.3 dB difference found was determined to be significant.

Moreover, roughly half of the listeners showed improvement in the second session.

In the second experiment by Wilson and McArdle (2007), two groups of subjects were used: (1) 48 individuals with mild-to-severe SNHL, and (2) 48 individuals with moderate-to-severe SNHL. For all participants, there were no significant differences within each session or between the two sessions. These results suggested the WIN test produced consistent thresholds for individuals with SNHL as well as being reproducible. Moreover, individuals with mild-to-severe hearing loss obtained better scores when compared to individuals with greater degrees of hearing impairment. The results of this study demonstrated that the WIN test could be relied upon for testing all levels of hearing loss as little variability was found between lists with test-retest scores remaining stable.

In a study by Wagener et al. (2003), list variability and participant variability was assessed under the headphone condition for 60 normal hearing individuals with an age range of 19-40 years and a mean age of 27.5 years. Two groups were formed out of the 60 participants. In the study described, the Dantale II test was utilized, which is a Danish version of the Oldenburg sentence test (Wagener et al., 2003). This test consisted of five words per sentence with similar sentence structure and 10 sentences per test list with 16 lists. The first group received half of the test material at one SNR, -10 dB, and the remaining material at another SNR, -6 dB. Likewise, the second group also received half of the test lists at one SNR, -6 dB, and the remaining lists at another SNR, -10 dB. When the SRTs were compared between test lists, no significant differences in performance were found. However, individuals tended to perform better on the test when they were re-tested at a later time, showing a slight training effect. In addition to acceptable list variability, there were no significant differences among the scores of the participants. This suggested the Dantale II test to be reliable for normal hearing individuals.

Variability Between Hearing Abilities and Performance in Speech-in-Noise Testing

In addition to variability in the test material, greater variability between normal hearing and hearing-impaired listeners is expected when comparing overall results from a speech-in-noise test. In a study by McArdle et al. (2005), word understanding abilities were assessed with monaural presentation via headphones in quiet and also in multitalker babble when using the NU-6 word list. When the participants were asked to rate their perceived difficulty of hearing speech in noise and in quiet on a scale of 1 to 10 (10 being most problematic), listeners with hearing loss perceived greater difficulty in both listening conditions while showing more variability in scores when compared to normal-hearing individuals. Moreover, both groups perceived greater difficulty understanding speech when in noise compared to when in quiet.

Given this information, results from normal hearing listeners serve as a reference point to which performances for listeners with hearing impairment are compared. When analyzing results in speech-in-noise tests, normal-hearing listeners tend to perform better than listeners with hearing loss regardless of noise and speech stimuli used (Beattie, 1989; Eisenberg, Dirks, & Bell, 1995;

Festen & Plomp, 1990; Hygge, Ronnberg, & Arlinger, 1992; McArdle & Wilson, 2006; McArdle et al., 2005; Nelson, Jin, Carney, & Nelson, 2003; Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999; Wilson, Bell, & Koslowski, 2003; Wilson & Cates, 2008; Wilson, Carnell, & Cleghorn, 2007). Beattie (1989) compared the results from individuals with mild-to-moderate SNHL to those with normal hearing when speech was presented in noise using the CID W-22 word test under the headphone condition. SNRs tested included 0, 6, 12, 18, and 24 dB. Differences in thresholds were found between the two groups with the normal-hearing group having better scores than the hearing-impaired group. It was determined that hearing-impaired listeners needed higher SNRs than normal-hearing listeners in order to obtain word understanding in noise scores of 50% or better.

Results from Hygge et al. (1992) had similar findings. This study consisted of a female speaker reading a passage presented in the soundfield condition with masking noise present. The participants' task was to increase the level of the female speaker's voice to a level where they could just barely hear and understand what was being said. For all noise conditions, the normal-hearing group performed better than the hearing-impaired group.

Eisenberg et al. (1995) conducted a study in order to determine whether differences in the amount of release from masking (gaining benefit from the valleys in the noise stimuli) occurred for individuals with hearing impairment compared to normal-hearing listeners when using steady and fluctuating noise maskers. Participants were divided into three groups: (1) 12 normal-hearing individuals, (2) four individuals with mild-to-moderate SNHL, and (3) 12 normal-hearing subjects (same participants as in group 1) presented with broadband noise used to simulate a hearing loss equivalent to group 2.

The speech stimulus utilized was nonsense syllables spoken by both a male and female speaker and presented in the headphone condition. It was found that group 1 performed better than group 2 when using the steady noise masker. In addition, group 3 had a similar release from masking to that of group 2 in fluctuating noise. Since the same participants were used for the normal-hearing and simulated hearing loss group, it was determined that audibility levels played a factor in speech understanding when in the presence of noise.

Variability Between Age and Performance in Speech-in-Noise Testing

In addition to differences in performance based on hearing abilities, age may also play a role in variability of responses (Dubno, Horwitz, & Ahlstrom, 2002; Dubno, Horwitz, & Ahlstrom, 2003; Versfeld & Dreschler, 2002). Dubno et al. (2002) studied whether differences in release from masking were affected by differences in age when presented monaurally via headphones. The study included two groups: (1) eight normal-hearing individuals with an age range of 20 to 27 years and a mean age of 24.1 years, and (2) eight normal-hearing individuals with an age range of 60 to 74 years and a mean age of 67.3 years.

This study included the use of low-level noise to reduce the variability in audibility differences among the different subjects. In the steady-state noise, when presentation level of the masker rose, speech understanding performance decreased. This observation was made regardless of age. However, older subjects performed 11.7% worse than younger subjects when the noise levels increased. In fluctuating noise, both groups showed a decrease in performance when the noise level was raised, with group 1 scoring 15.2% better than group 2; however, older listeners' performance decreased more substantially than younger listeners. These findings indicated that, regardless of similar hearing levels, older listeners did not benefit as much from the release of masking compared to the amount of release seen in the younger subjects.

Likewise, similar observations were reported by Dubno et al. (2003) when the effect of masking on older versus younger adults with normal hearing was compared. This study was conducted using two groups: (1) 10 normal-hearing individuals with an age range of 21 to 30 years and a mean age of 25.1 years, and (2) 10 normal-hearing individuals with an age range of 61 to 72 years and a mean age of 65.7 years. Subjects were tested with consonant-vowel/vowel-consonant syllables using various noise conditions, steady-state, forward and simultaneous maskers under the headphone condition. To minimize audibility variability within each of the two groups as well as between groups, a threshold matching noise was used throughout the testing.

Dubno and colleagues found that younger subjects performed better when threshold matching noise and fluctuating noise were used during the speech intelligibility portion than the performance with older listeners under similar conditions. This showed older subjects did not benefit as much from release from masking as younger subjects. Additionally, the amount of release from masking decreased as forward masked thresholds worsened. Differences were found between both groups when the speech presentation level was increased. Specifically, the younger listeners performed better than the older listeners.

Versfeld and Dreschler (2002) studied the effects of steady-state and fluctuating noise on the speech-understanding abilities of younger and older listeners with normal hearing along with those with hearing difficulties when using the sentences developed by Plomp and Mimpen (1979). The participants were separated into four groups: (1) 18 normal-hearing individuals with an age range of 20 to 29 years and a mean range of 23 years, (2) 11 normal-hearing individuals with an age range of 58 to 70 years and a mean age of 64 years, (3) eight individuals with SNHL with an age range of 15 to 35 years and a mean age of 18 years, and (4) 12 individuals with SNHL with an age range of 55 to 73 years and a mean age of 63 years. All individuals were tested monaurally via headphones.

For group 1, the SRTs when using the fluctuating noise was better than when using the steady-state noise, -11.1 dB and -4.8 dB respectively. As SRTs with steady-state noise improved, time-compressed thresholds (TCT) improved. Similar trends were established in group 2; however, their thresholds, -7.8 dB for fluctuating noise and -2.6 dB for steady-state noise, were not as high as group 1. Groups 3 and 4 did not show any significant differences between the two different noise conditions when measuring SRTs; however, younger hearing-impaired listeners performed better than the older hearing-impaired listeners. Versfeld and Dreschler (2002) speculated the difference in groups 3 and 4 occurred as a result of unequal language abilities, although this was not assessed during this study.

Type of Competing Noise in Speech-in-Noise Testing

Another factor to consider when assessing speech understanding performance in noise is the type of noise competition being utilized. Studies have shown a trend toward better performance in speech-in-noise testing when fluctuating noise is used as opposed to steady noise (Danhauer, Doyle, & Lucks, 1985; Danhauer & Leppler, 1979; Dubno et al., 2002; Dubno et al., 2003; Eisenberg et al., 1995; Festen & Plomp, 1990; Hygge et al., 1992; Nelson et al., 2003; Versfeld & Dreschler, 2002; Wagener & Brand, 2005; Wilson et al., 2007). Moreover, normal-hearing listeners show a greater release from masking than listeners with hearing loss, indicating better performance in fluctuating noise conditions for normal-hearing individuals (Eisenberg et al., 1995; Festen & Plomp, 1990; Hygge et al., 1992; Nelson et al., 2003; Versfeld & Dreschler, 2002; Wagener & Brand, 2005; Wilson et al., 2007).

Wagener and Brand (2005) assessed performance on the Oldenburg sentence test of 10 normal-hearing individuals ranging in age from 22-40 years and 10 individuals with varying degrees of SNHL who ranged in age from 59-79 years. Testing was conducted monaurally via headphones. The test consisted of 10 lists with 10 sentences per list and five words per sentence. Each sentence contained the same structure in that they all contained a name, verb, number, adjective and an object. It was found that the normal-hearing participants performed better than the hearing-impaired listeners when different fluctuating noise maskers were utilized. In other words, individuals with normal hearing had a greater release from masking than those with hearing loss.

Wagener and Brand (2005) also assessed the type of presentation of noise. When the noise was played continuously compared to synchronous with the speech material, performance for normal-hearing individuals was better. When referring to synchronous presentation, this indicated the speech and noise stimuli were presented at the same time as one another only. Similar results were found with the hearing-impaired group; however, the difference in performance was not as large when compared to the difference in performance in the normal-hearing group. In contrast, no difference in performance was found for either group when the speech was fixed and the noise was adaptive compared to when the speech was adaptive and the noise was fixed.

Nelson et al. (2003) assessed whether cochlear implant wearers would have the same release from masking as individuals with normal hearing. This study was conducted with three groups: (1) eight normal-hearing individuals, (2) eight normal-hearing individuals who listened to a narrower speech-spectrum range, and (3) nine individuals with cochlear implants, all of whom had been wearing their implants for over two years. All participants were tested in the soundfield condition. Both steady-state and fluctuating noises were used in this experiment in order to evaluate whether cochlear implant users benefit from the variations in modulation depths as normal-hearing listeners. Modulation depths are the changes in loudness of the noise stimulus, which may permit an improved listening condition. Ultimately, this brief decrease in loudness may allow the listener to interpret the speech stimulus better. Five different SNRs were tested, ranging from -16 to +16 dB.

Normal-hearing listeners scored nearly 100% when speech was presented in the quiet condition. In addition, when an SNR of 0 dB was used, higher scores were found in the fluctuating noise condition compared to the steady-state noise condition: 100% and 80% respectively. On the other hand, listeners in group 2 scored approximately 55% in the quiet condition. When an SNR of 16 dB was employed, scores of approximately 40% were observed in the fluctuating noise condition, which was poorer than the score in quiet. However, this score was still better than the score in steady-state noise, which was approximately 30% with an SNR of 16 dB. Furthermore, as the SNR decreased, minimal, if any, masking was observed. For listeners in group 3, scores of 80% were established when speech was presented in the quiet condition. When 16 dB SNR was used, scores ranged from 55% to 65% with fluctuating noise and 60% with steady-state noise, resulting in an unremarkable difference between the masking noises. These results showed that normal-hearing subjects obtained more release from masking than the other two groups with no significant differences found between masked normal listeners and cochlear implant wearers. With regards to the implant wearers, either minimal or no release at all from masking was observed.

Similar findings were obtained by Wilson et al. (2007) who analyzed differences in multitalker babble (MTB) noise and speech-spectrum noise (SSN) when using the WIN test under the headphone condition. The study consisted of two groups: (1) 24 normal-hearing listeners, and (2) 48 listeners with high-frequency SNHL. Variability between the two groups for the two different noise conditions were found, with a greater difference observed when using the MTB noise. The SSN showed to be a more efficient noise stimulus than the MTB due to the fluctuations in the MTB, which allowed for a temporarily improved SNR. More specifically, there were peaks and valleys in the noise masker, which allowed for the listeners to possibly obtain more information during the valley periods. This was seen more so for the normal-hearing subjects, as for individuals with hearing loss, the benefit of the temporarily-improved SNR in the MTB noise was not as pronounced. Therefore, the normal-hearing group showed greater benefit from the MTB noise than the hearing-impaired group. These results showed a greater release from masking when using the MTB noise. In addition, subjects with normal hearing had a larger amount of masking release than individuals with hearing loss.

These findings were supported by the results from Festen and Plomp (1990), who studied the effects of modulated masking noise compared to steady noise under the headphone condition. Two different groups were analyzed: (1) 20 normal hearing participants, and (2) 20 hearing-impaired participants. Speech material consisted of sentences ranging from eight to nine syllables in length. Four noise conditions were used: steady-state noise, single-band modulated noise, two-band modulated noise, and time-reversed speech. For the normal-hearing subjects, the trend was fluctuating noise resulted in better speech thresholds when compared to steady-state noise. Similarly, this trend was found in the individuals with hearing loss, although the benefit of the fluctuating noise was not as drastic. These trends are supported by the findings of other studies (Danhauer et al., 1985; Danhauer & Leppler, 1979; Dubno et al., 2002; Dubno et al., 2003; Eisenberg et al., 1995; Festen & Plomp, 1990; Hygge et al., 1992; Nelson et al., 2003; Versfeld & Dreschler, 2002; Wagener & Brand, 2005; Wilson et al., 2007).

Variability between Speech Materials

In addition to various maskers being analyzed, different speech materials have been compared in numerous studies (Danhauer et al., 1985; Duncan & Aarts, 2006; McArdle et al., 2005; Wilson, et al., 2005). McArdle et al. (2005) scrutinized variability in speech understanding tests with background noise present when the speech stimuli ranged from digits to words to sentences under the headphone condition. The study consisted of two groups: (1) 36 normal-hearing participants, and (2) 72 participants with a high-frequency SNHL.

Speech understanding abilities were assessed at the word level in quiet (NU-6, and also in multitalker babble, WIN) while sentences, (QuickSIN, and digit speech stimuli) were assessed only in multitalker babble. Significant differences were found in the results between the normal-hearing and hearing-impaired listeners for all conditions tested in noise. An 8 dB difference was noted between the normal-hearing and the hearing-impaired groups with the individuals with normal hearing performing better. Both groups performed significantly better on the digits tasks than on words or sentences while in noise. When the speech stimuli consisted of sentences, the listening task was more difficult than if listening to words or digits. However, using sentences as the speech stimuli of choice was more representative of what people encounter on a daily basis. In other words, sentences were more “real-world.”

In addition to having speech stimuli that would represent the most accurate abilities an individual has when in noise, the test itself needs to be practical for clinical purposes. Wilson and Burks (2005) conducted a study in order to create two words lists from the WIN that allowed clinicians to determine the speech understanding performance of individuals when in noise. Moreover, the goal was to obtain little variability between the lists, showing the use of just one of the lists to be an accurate representation of the individual’s hearing in noise abilities. This was done with two experiments, both under the headphone condition.

The first experiment consisted of 72 subjects with SNHL. Two 35-word lists from the WIN were presented. Each word list included seven SNRs ranging from 24 to 0 dB, which were presented in sequential order. Within each individual tested, variability of 0.5 dB was found in their performances on the two different lists with the better performance occurring on list 2. As mentioned earlier, consistency between lists were found by other studies with variability being less than 0.9 dB (McArdle & Wilson, 2006; Plomp & Mimpen, 1979; Wilson & McArdle, 2007). Therefore, results from this study are consistent to the findings of others (McArdle & Wilson, 2006; Plomp & Mimpen, 1979; Wilson & McArdle, 2007).

In the second experiment, 48 mild-to-moderate SNHL individuals were tested using the same methods as the first experiment, except that the SNRs were randomized throughout the experiment. Likewise, this group showed little variability between performances between both lists. It was found that the order of the list presentation did not matter, whether the SNRs were randomized or presented in sequential order of difficulty. Given that both lists were reliable, it was found that the WIN test list could be reduced to one 35 word list in order to save time in the clinical environment. Therefore, this study established a more practical speech in noise test with regard to use in the clinical world.

With regard to the presentation of the speech material, Wilson, Burks, and Weakley (2005) assessed whether differences in word understanding with multitalker noise occurred when using a random versus descending speech signal level approach. This study consisted of 40 high-frequency SNHL subjects with 70 NU-6 words used as the speech stimuli. Seven SNRs, 0 to 24 dB, were evaluated (10 words per SNR) using both approaches under the headphone condition. Despite better results in 32 words when the SNRs were presented randomly, performance in the two approaches did not differ significantly. Therefore, the study determined that either approach would be acceptable, similar to the findings by Wilson and Burks (2005). Despite no differences between the two protocols, the participants found the random protocol to be a harder task. All participants expressed greater perceived difficulty understanding speech in noise than in quiet with similar ratings as McArdle et al. (2005). Given this information, the descending approach may be a better protocol in order to reduce patient stress.

Variability in Performance between the Hearing-in-Noise-Test (HINT) and the QuickSIN

Duncan and Aarts (2006) studied the practical use of various speech-in-noise tests. The purpose of this study was to compare scores from the hearing-in-noise test (HINT) and the QuickSIN to normative data obtained from a previous study. Fifteen individuals with normal hearing participated in the comparison of these two tests. Both tests use sentences as their speech stimuli. Testing was conducted in the soundfield condition and test orders were randomized.

Larger variability was found in the HINT scores compared to the QuickSIN, with the HINT having a significant difference in values when compared to the normative data. In contrast, the QuickSIN results were consistent with normative data. The discrepancy between the HINT scores found in this study compared to the normative data may have been due to the acoustics of the test environment. Therefore, if testing the HINT under the soundfield condition, it would be necessary to obtain normative data specific to that test environment, whereas such data were not needed for the use of the QuickSIN, as normative data was comparable to the SNRs found in the study. Given this information, the use of the QuickSIN as opposed to the HINT would require less set-up in the clinical environment, making it a more ideal test.

Variability in Performance between the Nonsense Syllable Test (NST) and the NU-6 Test

Similar to Duncan and Aarts, Danhauer et al. (1985) compared the results of the nonsense syllable test (NST) when several noise stimuli were used to results from the NU-6 test under similar conditions for 20 normal hearing individuals; however, testing was conducted under the headphone condition. This study was done to see if one of the speech stimuli was preferred for clinical purposes over the other. Regardless of the noise, the subjects performed better on the NU-6 test material than the NST material. These results indicate the NST to be a more difficult task than the NU-6 speech task.

Variability between Scoring Methods

Danhauer et al. (1985) also scrutinized differences in preferred scoring method of speech understanding performance. When comparing the scores obtained at the phonemic level to the scores obtained at the word level, higher scores were found at the phonemic scoring level, regardless of noise presented. For example, when using the multitalker noise, a score of 85% was obtained at the phonemic level while 54% was obtained at the word level. This indicated phonemic scoring to be more precise than word scoring; however, this was a more tedious and time-consuming method than analyzing the results at a word level. Therefore, for clinical purposes, the use of word scoring would be more clinically applicable than phonemic scoring.

Dirks et al. (1982) tested the performance of normal-hearing and hearing-impaired subjects in speech in noise tests when NU-6 and spondaic words were used. All participants were tested under the headphone condition. The authors found a slight increase in SNR as the speech level of the speech stimulus increased for the spondaic words task in normal-hearing listeners. In contrast, participants with hearing loss required larger SNRs in order to obtain a performance score of 50% for the spondaic task. Similar findings occurred when monosyllables (NU-6 words) in noise were used. Results were comparable with both test stimuli; however, more variability was found when using the monosyllabic test material compared to spondaic words.

Statement of Purpose

As seen in the literature, various factors play a role in the performance of speech understanding in noise. Different outcomes are expected when comparing results from individuals with assorted hearing levels. Likewise, types of background noise can also influence how well or poorly a person can understand speech in noisy conditions. These factors, and many more, have been assessed in order to determine the best possible tests to use in clinical settings for reliable speech in noise results.

With these considerations in mind, the purpose of the present investigation is to assess the list equivalency on the American English Matrix test for normal hearing individuals. Due to the similarities in the stimuli between lists, it is predicted that there will be no differences between lists. Furthermore, it is expected for the participants to perform poorer as the SNRs become less favorable.

Chapter 3

Methods

Participants

Inclusion criteria for this study consisted of individuals between the ages of eighteen and thirty-five with normal hearing (better than or equal to 15 dB HL) as well as normal outer and middle ears at the time of testing as determined by clear canals via otoscopy and normal tympanograms. Pure-tone air-conduction thresholds were obtained for frequencies 125-8000 Hz. Participants were recruited via flyers, as well as personal contact throughout the Audiology program at Towson University. Participants were required to read and sign an informed consent form approved by the Towson University IRB (see Appendix B).

All participants met the inclusion criteria of age, hearing, outer and middle ear status, and completed all test conditions as well as both sessions. The study sample consisted of twenty individuals (5 males, 15 females) ranging from 18-27 years of age (*M* = 23.75, *SD* = 1.94).

Speech Stimuli

The American Matrix sentence test was utilized as the speech material during this study. This test is similar in sentence structure/speech material to the Oldenburg sentence test (Wagener & Brand, 2005) and the Dantale II sentence test (Wagener et al., 2003). It consists of twenty sentences containing five words each. The sentence structure of all the sentences is identical in that they all contain a name, verb, number, adjective and an object (e.g., Doris likes fourteen red desks).

Noise Stimuli

Noise from the American English Matrix test (USmatrixnoise) was used as the noise competition. This noise consists of steady speech spectrum shaped noise, similar to the noise used in the Oldenburg sentence test (Wagener & Brand, 2005) and the Dantale II sentence test (Wagener et al., 2003); however, it is matched to the speech spectrum of the English sentences. This noise was presented synchronous to the speech stimulus, with the noise beginning 0.5 seconds prior to the speech stimuli sentence and ending 0.5 seconds after the conclusion of each speech stimuli sentence. The noise was held constant at 65 dB SPL for all test subjects.

Study Equipment

Test material from the American Matrix software was sent from a Dell Latitude D410 computer to a Gigaport HD sound card, and then routed to Sennheiser HDA 200 circumaural headphones worn by the participant. Test material was presented to one of the participants' ears, with ten participants being tested in their right ear and ten participants being tested in their left ear. The selection of the ear to be tested for the participants was chosen in a pseudo random order. All testing was conducted in an Industrial Acoustics Company, Inc. sound-treated booth. The Matrix computer software also served as means to storing the test data for each participant.

Procedures

Each participant underwent an otoscopic examination, pure-tone air conduction screening with insert earphones via a Grason-Stadler 61 clinical audiometer, and tympanograms via a Madsen Otoflex 100 prior to the actual test. All individuals were required to have normal Jerger type A tympanograms (Margolis & Heller, 1987).

Testing included two test sessions for each participant. During the first session, the participant was presented with six test lists in noise with a starting signal-to-noise ratio (SNR) of 0 dB. The level of sentence presentation was varied adaptively. Following the adaptive testing in noise, the participants were presented with nine test lists at fixed SNRs. The subjects' role during testing was to repeat back the speech material presented to them to the best of their ability, regardless of whether they could repeat back the entire sentence or just specific words of the sentence. The clinician, who was located on the outside of the sound booth, marked all of the words repeated correctly by the participant in the response window of the American Matrix software. Target words for each sentence being presented were displayed in the software. All responses were analyzed at the word level, i.e. word scoring automatically by the Matrix software. Therefore, if a subject missed one of the five target words in a sentence, they scored 4 out of 5 for that sentence versus getting the entire sentence wrong if sentence scoring was used. This allowed for a more precise measurement of the subjects' ability to understand speech in the presence of competing noise. Testing consisted of the following conditions:

1. **Adaptive speech in 65 dB SPL competing USmatrixnoise**. All six lists were presented to each participant with the competing noise. List order was randomized within and between participants. The tester selected on the Matrix software display the words the participant got correct for each sentence. The software then automatically adjusted the speech stimuli levels based on the response from the participant.
2. **Fixed signal-to-noise ratios with 65 dB SPL competing USmatrixnoise**. Three fixed SNRs were used, specifically -6 dB SNR, -8.5 dB SNR, and -11dB SNR in order to reach intelligibilities of about 80%, 50%, and 20% respectively. The level of the speech was held constant for the fixed SNR measures. The SNR conditions were randomized between subjects with each SNR being presented three times during each session, while the test lists were assigned by the researcher to ensure lists were presented in each SNR approximately the same number of times across participants.

Statistics

The mean SRTs and standard deviations were determined for each test list in the adaptive test condition. Similarly, the mean scores, in percent correct, and standard deviations were determined for each fixed SNR condition. Results from each list in the adaptive condition were compared using a repeated-measures analysis of variance (ANOVA) to determine list equivalency. Results from the first session were compared to the results from the second session, for both test conditions, using a repeated-measures ANOVA to determine if there were training effects. The order of presentation was also analyzed using a repeated-measures ANOVA to determine possible training effects based on the order of presentation within a session. The scores, in percent correct, for all of the fixed SNRs were compared using a repeated measures ANOVA to determine if the scores varied across the SNRs. An alpha level of $\alpha = .01$ was utilized for all statistical analyses.

Chapter 4

Results

Adaptive Speech Stimuli Test Condition

Descriptive statistics. Mean speech reception threshold (SRT) scores (in dB SNR) and standard deviations (SD) were calculated separately for each list within each session. Mean scores for session one were as follows: List 1 (*M* = -8.46, *SD* = 1.04); List 2 (*M* = -8.74, *SD* = 0.85); List 3 (*M* = -8.40, *SD* = 1.04); List 4 (*M* = -8.37, *SD* = 0.94); List 5 (*M* = -8.19, *SD* = 1.14); List 6 (*M* = -8.28, *SD* = 1.02). The overall mean SRT for session one was -8.40 dB with a standard deviation of 1.00. Mean scores for session two were as follows: List 1 (*M* = -9.71, *SD* = 0.68); List 2 (*M* = -9.77, *SD* = 1.02); List 3 (*M* = -9.83, *SD* = 1.03); List 4 (*M* = -9.41, *SD* = 0.77); List 5 (*M* = -9.75, *SD* = 1.00); List 6 (*M* = -9.76, *SD* = 0.88). The overall mean SRT for session two was -9.70 dB with a standard deviation of 0.90. The difference in overall SRTs for session one and session two was 1.29 dB. Table 1 displays the mean SRT values and standard deviations, along with the minimum and maximum scores for each test list across participants for sessions one and two. See Appendix C Table 1 for the SRT scores for each list in both sessions for every participant.

List equivalency. The SRT scores for each test list, across sessions, were compared using a repeated-measures ANOVA to determine if there were any statistically-significant differences among the scores from the six test lists.

No statistically-significant differences in test scores were found among any of the six test lists ($F(5, 195) = 1.09$, $p = .365$). To further ensure list equivalency, post-hoc analyses via paired-samples t-tests compared each list to the others. Results confirmed that there were no significant differences in the scores between any of the test lists. See Table 2 for specific *t*-values and *p*-values for each list pair.

Training effects within session one. The SRT scores for each list within session one were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. A statistically-significant difference was found in SRT scores within session one ($F(5,95) = 54.04$, $p < .001$) suggesting a significant training effect. A paired samples t-test was conducted to confirm the training effect. More specifically, as the presentation order progressed, first list presented versus last list presented in session one, the SRTs improved significantly ($t(19) = 10.90$, $p < .001$). See Figure 1.

Training effects within session two. The SRT scores for each list within session two were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. A statistically-significant difference was found in SRT scores within session two ($F(5,95) = 24.19$, $p < .001$) suggesting a significant training effect. A paired samples t-test was conducted to confirm the training effect. More specifically, as the presentation order progressed, first list presented versus last list presented in session two, the SRTs improved significantly ($t(19) = 8.68$, $p < .001$). See Figure 2 and Table 3.

Training effects between sessions. A paired samples t-test was calculated to compare the mean SRT scores for all of the lists from session one (M = -8.41, SD = 1.01) to the mean SRT scores for all of the lists from session two (M = -9.70, SD = 0.90). A statistically-significant increase was found in SRT scores from session one to SRT scores from session two ($t(119) = 20.07$, $p < .001$) suggesting a significant training effect. See Figure 3.

Fixed Speech Stimuli Test Condition

Descriptive statistics. The mean speech intelligibility percentage when using a fixed SNR of -6 dB for session one was 85% with a standard deviation of 0.07 and the mean percentage for session two was 90% with a standard deviation of 0.06 with a difference of 5%. Mean speech intelligibility percentages ranged from 68% to 94% during session one and 73% to 98% during session two. The overall mean speech intelligibility percentage across both sessions was 90% with a standard deviation of 7.24.

The mean speech intelligibility percentage when using a fixed SNR of -8.5 dB for session one was 65% with a standard deviation of 0.14 and the mean percentage for session two was 77% with a standard deviation of 0.11 with a difference of 12%. Mean speech intelligibility percentages ranged from 23% to 81% during session one and 48% to 93% during session two. The overall mean speech intelligibility percentage across both sessions was 71% with a standard deviation of 14.39.

The mean speech intelligibility percentage when using a fixed SNR of -11 dB for session one was 30% with a standard deviation of 0.13 and the mean percentage for session two was 41% with a standard deviation of 0.17 with a difference of 11%. Mean speech intelligibility percentages ranged from 6% to 49% during session one and 10% to 78% during session two. The overall mean speech intelligibility percentage across both sessions was 35% with a standard deviation of 16.43. Table 4 displays the mean speech intelligibility percentage values, the minimum scores, the maximum scores, and the standard deviations for each fixed SNR across participants for sessions one and two.

Training effects within session one. The scores, in percent correct, for the fixed SNR of -6 dB within session one were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. A statistically-significant difference was found between the scores within session one ($F(2,38) = 7.43$, $p = .002$) suggesting a significant training effect. A paired samples t-test confirmed that as the presentation order progressed, first list presented versus last list presented in session one, the scores improved significantly ($t(19) = -3.65$, $p = .002$).

The scores, in percent correct, for the fixed SNR of -8.5 dB within session one were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. No statistically-significant difference was found in the scores within session one ($F(2,38) = 2.23$, $p = .121$), suggesting a lack of a training effect. Despite the lack of statistical difference, a trend was noticed. More specifically, as the presentation order progressed, first list presented versus last list presented in session one, the scores improved.

The scores, in percent correct, for the fixed SNR of -11 dB within session one were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. No statistically-significant difference was found in the scores within session one ($F(2,38) = 2.10$, $p = .136$), suggesting a lack of a training effect. Despite the lack of statistical difference, a trend was noticed. More specifically, as the presentation order progressed, first list presented versus last list presented in session one, the scores improved.

Training effects within session two. The scores, in percent correct, for the fixed SNR of -6 dB within session two were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. No statistically-significant difference was found in the scores within session two ($F(2,38) = 0.19$, $p = .822$), suggesting a lack of a training effect. Additionally, there was not a trend noticed with the scores as the presentation order progressed.

The scores, in percent correct, for the fixed SNR of -8.5 dB within session two were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. No statistically-significant difference was found in the scores within session two ($F(2,38) = 1.68$, $p = .200$), suggesting a lack of a training effect. Despite the lack of statistical difference, a trend was noticed. More specifically, as the presentation order progressed, first list presented versus last list presented in session two, the scores improved.

The scores, in percent correct, for the fixed SNR of -11 dB within session two were compared to one another using a repeated-measures ANOVA to determine if there were any training effects. No statistically-significant difference was found in the scores within session two ($F(2,38) = 0.17$, $p = .845$), suggesting a lack of a training effect. Additionally, there was not a trend noticed with the scores as the presentation order progressed.

Training effects between sessions. A paired samples t-test was calculated to compare the scores, in percent correct, for the fixed SNR of -6 dB from session one ($M = 85.40$, $SD = 7.36$) to the scores from session two ($M = 90.40$, $SD = 6.24$). A statistically-significant increase was found in scores from session one to scores from session two ($t(59) = -7.11$, $p < .001$) suggesting a significant training effect. See Figure 4.

A paired samples t-test was calculated to compare the scores, in percent correct, for the fixed SNR of -8.5 dB from session one ($M = 65.07$, $SD = 14.41$) to the scores from session two ($M = 76.82$, $SD = 11.83$). A statistically-significant increase was found in scores from session one to scores from session two ($t(59) = -7.89$, $p < .001$) suggesting a significant training effect. See Figure 4.

A paired samples t-test was calculated to compare the scores, in percent correct, for the fixed SNR of -11 dB from session one ($M = 30.10$, $SD = 13.16$) to the scores from session two ($M = 40.88$, $SD = 17.67$). A statistically-significant increase was found in scores from session one to scores from session two ($t(59) = -7.26$, $p < .001$) suggesting a significant training effect. See Figure 4.

Speech intelligibility scores. The scores for each of the fixed SNR conditions (-6 dB, -8.5 dB, -11 dB) collapsed across both test sessions were compared using a repeated measures ANOVA to determine if there were any statistically-significant differences between the scores for each fixed SNR. A statistically-significant difference was found when comparing the scores from the three fixed SNR conditions ($F(2, 238) = 1324.23$, $p < .001$). Paired samples t-tests were conducted to determine which SNR scores were significantly different from one another. More specifically, statistical differences were found when comparing the scores in -6 dB SNR to the scores in -8.5 dB SNR ($t(119) = 19.14$, $p < .001$), the scores in -6 dB SNR to the scores in -11 dB SNR ($t(119) = 44.69$, $p < .001$), and the scores in -8.5 dB SNR to the scores in -11 dB SNR ($t(119) = 34.09$, $p < .001$).

Table 1: *Mean Scores, Standard Deviations, Minimum Scores Maximum Scores for the Six Lists for Session One and Session Two*

Session & List Number	Mean Score (dB SNR)	Standard Deviation (SD)	Minimum Score (dB SNR)	Maximum Score (dB SNR)
Session One				
List 1	-8.46	1.04	-6.30	-10.10
List 2	-8.74	0.85	-6.90	-10.00
List 3	-8.40	1.04	-5.80	-10.40
List 4	-8.37	0.94	-6.70	-10.00
List 5	-8.19	1.14	-5.80	-10.00
List 6	-8.28	1.02	-6.30	-9.90
Session Two				
List 1	-9.71	0.68	-8.10	-10.70
List 2	-9.77	1.02	-7.30	-11.40
List 3	-9.83	1.03	-7.70	-12.10
List 4	-9.41	0.77	-7.90	-11.30
List 5	-9.75	1.00	-8.10	-11.20
List 6	-9.76	0.88	-8.00	-11.20

Note. **A lower speech reception threshold indicates better performance.**

Table 2: *Comparison of SRT Scores Between List Pairs*

List Pair	***t***	***df***	***p***
List 1-List 2	1.53	39	.133
List 1-List 3	0.17	39	.863
List 1-List 4	-1.09	39	.281
List 1-List5	-0.70	39	.485
List 1-List 6	-0.38	39	.701
List 2-List 3	-0.90	39	.372
List 2-List 4	-2.27	39	.290
List 2-List 5	-1.83	39	.074
List 2-List 6	-1.54	39	.130
List 3-List 4	-1.44	39	.157
List 3-List 5	-0.75	39	.456
List 3-List 6	-0.52	39	.603
List 4-List 5	0.41	39	.680
List 4-List 6	0.69	39	.489
List 5-List 6	0.24	39	.811

Table 3: *Mean Scores, Standard Deviations, Minimum Scores and Maximum Scores Minimum Scores, Maximum Scores for the First and Last Lists Presented for Session One and Session Two*

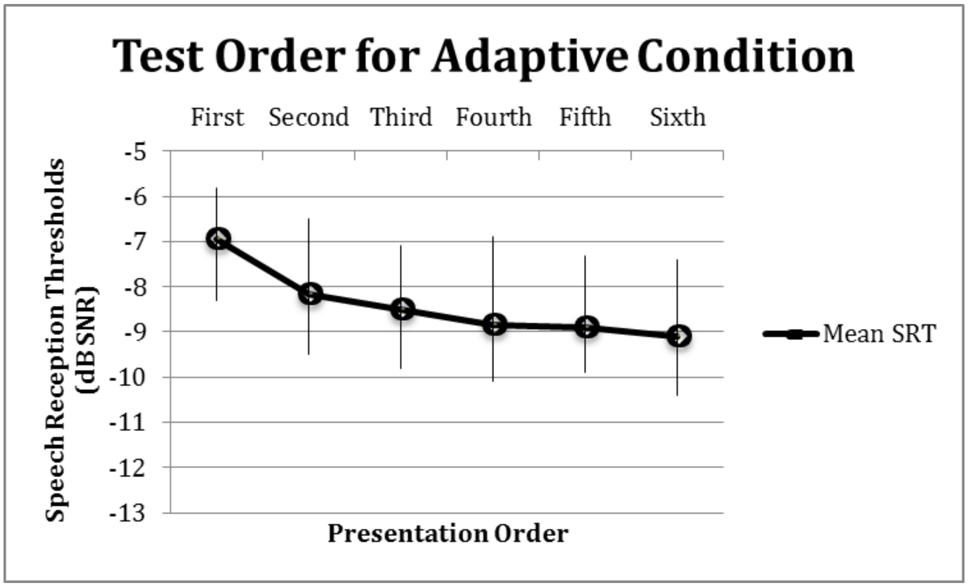
Session & Test Order	Mean Score (dB SNR)	Standard Deviation (SD)	Minimum Score (dB SNR)	Maximum Score (dB SNR)
Session One				
First List	-6.95	0.72	-5.80	-8.30
Last List	-9.18	0.69	-7.40	-10.40
Session Two				
First List	-8.91	0.75	-7.30	-10.40
Last List	-10.12	0.85	-8.20	-12.1

Note. **A lower speech reception threshold indicates better performance.**

Table 4: *Mean Scores, Standard Deviations, Minimum Scores and Maximum Scores for the Three Fixed SNRs for Session One and Session Two*

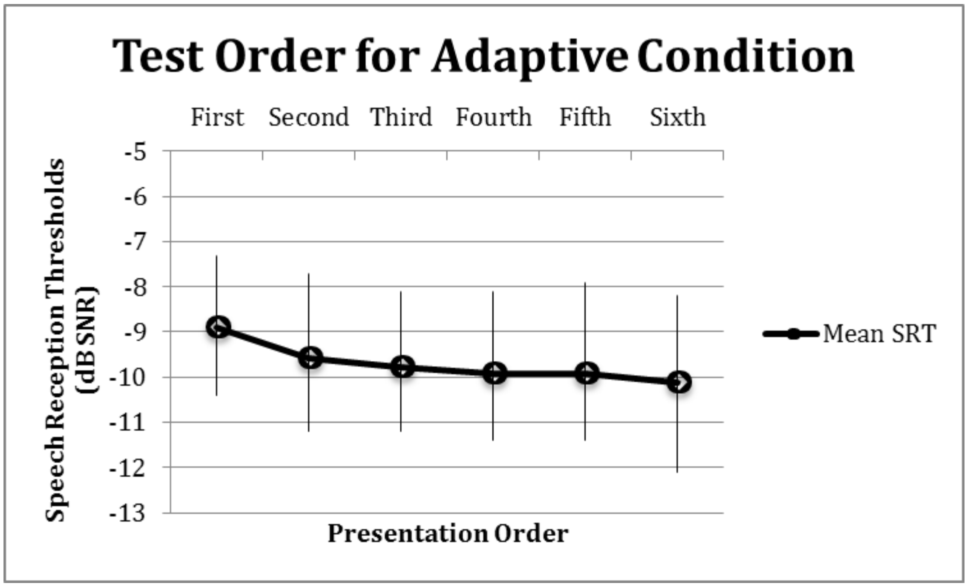
Session & Fixed SNR	Mean Score (%)	Standard Deviation (SD)	Minimum Score (%)	Maximum Score (%)
Session One				
-6 dB SNR	85	0.07	65	98
-8.5 dB SNR	65	0.14	20	90
-11 dB SNR	30	0.13	3	57
Session Two				
-6 dB SNR	90	0.06	70	99
-8.5 dB SNR	77	0.11	44	96
-11 dB SNR	41	0.17	8	81

Figure 1: *Mean SRTs and SRT Range for Test Order in the Adaptive Condition in Session One*



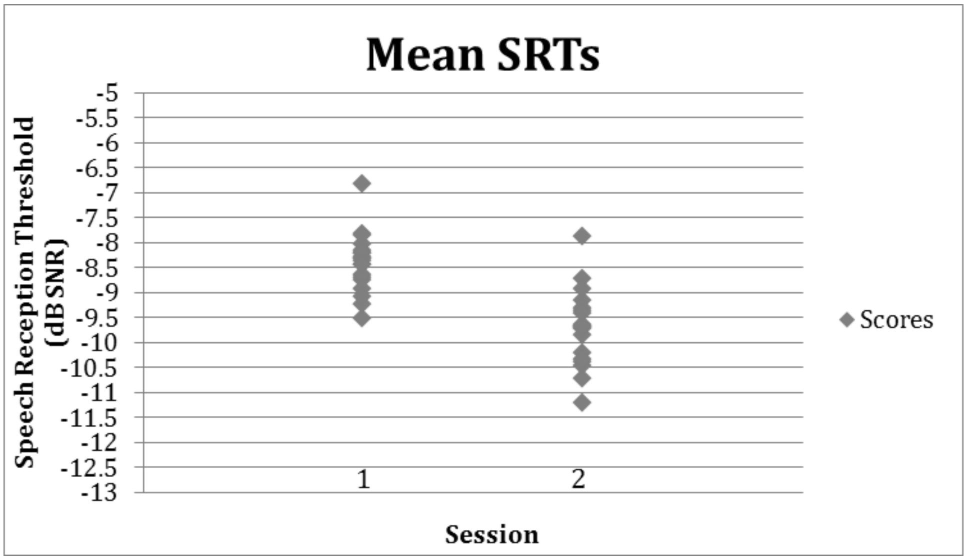
Note. **A lower speech reception threshold indicates better performance.**

Figure 2: *Mean SRTs and SRT Range for Test Order in the Adaptive Condition in Session Two*



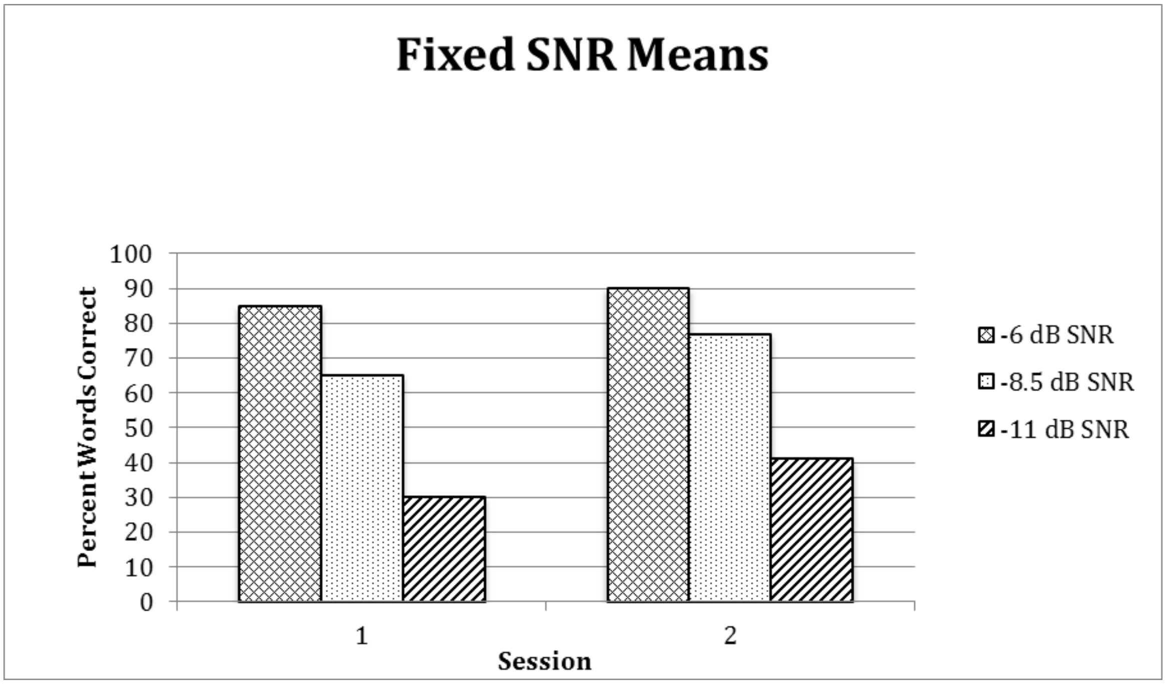
Note. **A lower speech reception threshold indicates better performance.**

Figure 3: *Mean SRTs for Each Participant for Session One and for Session Two*



Note. **A lower speech reception threshold indicates better performance.**

Figure 4: *Mean Scores of the Three Fixed SNR Conditions for Session One and Session Two*



Chapter 5

Discussion

The present study was looking at how individuals with normal hearing performed on the American English Matrix test when an adaptive speech stimulus and a fixed speech stimulus were used. Moreover, the researcher assessed the list equivalency as well as possible training effects when using this test. More specifically, the study assessed whether training effects had occurred based on the presentation order of the lists within each of the two sessions as well as between both sessions for an adaptive speech stimuli and a fixed speech stimuli condition.

Findings

Adaptive speech stimuli test condition. As hypothesized, individuals' performances in the six test lists on the American English Matrix test were similar. Because there were no statistically-significant differences in speech reception threshold (SRT) scores between the six lists, it could be concluded that the lists were equivalent.

Within each session, there were statistically-significant increases in SRT scores, indicating training effects had occurred. Similarly, training effects were seen within the Dantale II and the Oldenburg tests (Wagener & Brand, 2005; Wagener, Josvassen & Ardenkjar, 2003). More specifically, the Dantale II test found a training effect of 2.2 dB when determining the differences in mean SRTs from the first list presented compared to the eighth list (last list) presented (Wagener, Josvassen & Ardenkjar, 2003). The present study found a training effect of 2.23 dB when determining the differences in mean SRTs from the first list presented compared to the sixth list (last list) presented in session one. Given this information, it would be recommended that at least one practice list should be presented to a patient in order to train them to the task. Once patients have been trained appropriately, their true SRT scores could be measured accurately.

When the SRT scores from session one were compared to the SRT scores in session two, statistically-significant differences were found, indicating training effects had occurred. These findings are in agreement with the findings from the Oldenburg study (Wagener & Brand, 2005). More specifically, the SRT scores in session two were significantly better than the scores in session one with a training effect of 5 dB occurring. Wagener and Brand (2005) found a training effect of 0.67 dB when comparing the SRTs from one session to the SRTs from the second session, although statistical significance was not noted. Given this information, different test lists should be utilized with a patient if they are to be presented with the American English Matrix hearing in noise test repeatedly in order to reduce skewed SRT scores, despite the low predictability of each sentence (Wagener & Brand, 2005; Wagener, Josvassen & Ardenkjar, 2003).

Fixed speech stimuli test condition. In the present study, as the fixed SNRs were reduced, poorer performance was seen across individuals, as supported by the findings from the Dantale II study (Wagener, Josvassen & Ardenkjar, 2003). In that study, when tested at fixed SNRs of -10 dB and -6 dB, participants' scores were approximately 30% and 75% respectively (Wagener, Josvassen & Ardenkjar, 2003). In the present study, statistically-significant differences when comparing the scores from each fixed SNR to one another supports the hypothesis of poorer participant performance as the SNR became less favorable.

Within session one, a statistically-significant difference was found in scores for the fixed SNR of -6 dB, indicating training effects. In contrast, there were no statistically-significant differences found in scores for the fixed SNRs of -8.5 dB and -11 dB. Despite the lack of significance, a trend was noticed. More specifically, the results suggested that the participants' scores improved as they were presented with additional lists. Given this information, a recommendation should be made for at least one practice list to be given prior to actually measuring a participant's score at fixed SNRs in order to reduce potential training effects.

Within session two, there were not any statistically-significant differences found in scores in any of the fixed SNRs. Despite the lack of significance, a trend was found when for the fixed SNR of -8.5 dB. More specifically, the participants' scores improved as they were presented with additional lists. In contrast, there were not any trends noticed when the SNRs of -6 dB and -11 dB were used. This may be due to ceiling effects. More specifically, with the SNR of -6 dB, the scores were relatively high to begin with. It is possible that there was not much room for improvement, or at least not significant improvement with the fixed SNR of -6 dB. With regards to the SNR of -11 dB, it is not clear as to why there was not a trend or statistically significant difference in test scores. Given the trend noticed for the SNR of -8.5 dB, a recommendation should be made to present at least one practice list prior to actually measuring a participant's score in order to reduce potential training effects.

When the scores from session one were compared to the scores in session two, statistically-significant differences were found, indicating training effects had occurred. More specifically, the scores in session two were significantly better than the scores in session one. Given this information, different test lists should be utilized with a patient if they are to be presented with the American English Matrix hearing in noise test repeatedly in order to reduce skewed scores.

Limitations

Sentence variability. One major limitation of this study was that within each of the six test lists used, ten out of the twenty sentences that made up each list were found in all of the lists. Since participants were tested with the lists a total of fifteen times each session, the duplicated ten sentences may have impacted the finding of the lists being equivalent, despite the low predictability of each sentence (Wagener & Brand, 2005; Wagener, Josvassen & Ardenkjar, 2003).

Test duration. Each test session lasted approximately 1 hour to 1 hour and 30 minutes, depending on the amount of breaks participants used as well as their quickness of responses during testing. Throughout testing, several participants expressed fatigue, which may in fact have impacted their scores.

Tester error. Another limitation of this study was the possibility of tester error. The tester needed to mark the participants' correct responses for each sentence. Since this was completed subjectively, it is possible the tester may have misheard or marked the incorrect responses.

Future Research

Due to the limitations of this study, further research is needed to determine if gender effects as well as ear effects impact the scores for each list. It is recommended that future studies use lists that do not contain duplicate sentences, in order to truly look at list equivalency and training effects. Following the collection of the data with normal hearing individuals, further research should be conducted to determine how individuals with varying degrees of hearing loss, as well as varying ages, perform on the American English Matrix test.

Conclusion

The present study verified the six American English Matrix test lists used were equivalent. Additionally, when the fixed SNRs increased in difficulty, participants' scores decreased. However, with SNRs of -6 dB, -8.5 dB, and -11 dB participants' scores were higher than the predicted intelligibility scores of 80%, 50%, and 20%, respectively. Given this information, less favorable SNRs should be used to reach the desired intelligibility scores. With regard to the training effects within sessions, participants should be presented with at least one practice list in order to accurately reflect their true listening abilities when tested with the American English Matrix test. Moreover, since participants performed better in the second session compared to the first session, different test lists should be used when re-testing someone with the American English Matrix test in order to reduce training effects, despite the low predictability of each sentence. Since the lists were found to be equivalent, any of them could serve as the practice list.

Appendices

Appendix A

IRB Approval

TOWSON UNIVERSITY

APPROVAL NUMBER: 12-A029

To: Holly Rose
8000 York Road
Towson MD 21252

From: Institutional Review Board for the Proctection of Human Subjects, Justin Buckingham, Member

Date: Friday, January 20, 2012

RE: Application for Approval of Research Involving the Use of Human Participants

APPROVED
JAN 20 2012
JB/ wnw
TOWSON IRB

Office of University
Research Services

Towson University
8000 York Road
Towson, MD 21252-0001

t. 410 704-2236
f. 410 704-4494

Thank you for submitting an Application for Approval of Research Involving the Use of Human Participants to the Institutional Review Board for the Protection of Human Participants (IRB) at Towson University. The IRB hereby approves your proposal titled:

Reliability of the American English Matrix Test

If you should encounter any new risks, reactions, or injuries while conducting your research, please notify the IRB. Should your research extend beyond one year in duration, or should there be substantive changes in your research protocol, you will need to submit another application for approval at that time.
We wish you every success in your research project. If you have any questions, please call me at (410) 704-2236.

CC: B. Kreisman
File

Appendix A (continued)

TOWSON
UNIVERSITY

Date: Friday, January 20, 2012

NOTICE OF APPROVAL

TO: Holly Rose **DEPT:** ASLD

PROJECT TITLE: *Reliability of the American English Matrix Test*

SPONSORING AGENCY:

APPROVAL NUMBER: 12-A029

The Institutional Review Board for the Protection of Human Participants has approved the project described above. Approval was based on the descriptive material and procedures you submitted for review. Should any changes be made in your procedures, or if you should encounter any new risks, reactions, injuries, or deaths of persons as participants, you must notify the Board.

A consent form: [✓] is [ ] is not required of each participant

Assent: [ ] is [ ] is not required of each participant

This protocol was first approved on: 20-Jan-2012
This research will be reviewed every year from the date of first approval.

Justin Buckingham

Justin Buckingham, Member
Towson University Institutional Review Board
WCP

Appendix B

INFORMED CONSENT FORM

The Center for Amplification, Rehabilitation, and Listening (CARL) is carrying out research on the reliability of a speech understanding in noise test. Your role in this project will consist of attending two, one-and-a-half hour experimental session.

At the first session, you will have a hearing and middle ear screening completed. You will wear headphones and indicate when you have heard a tone. After the hearing screening, you will a probe tip will be inserted into your ear and you will feel a brief change in pressure in your ear. You do not have to do anything for this screening. You will be told if you did not pass either of the screening tests and recommendations for follow-up testing will be provided if needed.

Following the screenings, you will complete several listening tasks, including:

- Listening to sentences in noise and repeating back what you heard.
- Listening to sentences in noise with the level of the sentences getting quieter/louder and repeating back what you heard.

At the second session, you will be tested under the same conditions as your first session but in the other ear.

After completion of their second session, participants will be entered into a drawing for a chance to win one of four gift cards (3-$25 gift cards and 1-$50 gift card) provided by the researcher. The drawing will occur after completion of all research.

There are no known physical risks or discomforts associated with this testing; however, an individual who has an unknown hearing loss may experience surprise and/or emotional distress upon finding out about the hearing loss.

Participation in this study is voluntary. All information will remain strictly confidential. Although the descriptions and findings may be published, at no time will your name or any other identifying information be used. You are at liberty to withdraw your consent to the experiment and discontinue participation at any time without prejudice. If you have any questions after today, please feel free to call 410-704-3620 and ask for Dr. Brian Kreisman, or contact Dr. Debi Gartland, Chairperson of the Institutional Review Board for the Protection of Human Participants at Towson University at (410) 704-2236.

I, ___________________________________, affirm that I have read and understood the above statement and have had all of my questions answered.

Date: _____________________

Signature: ____

Witness: ____

Appendix C

Adaptive Test Condition Scores

Table 1: *SRTs (in dB SNR) for Each List in Session One for Each Participant*

Session & Participant	List 1	List 2	List 3	List 4	List 5	List 6
Session One						
1	-8.7	-8.9	-9.0	-7.6	-8.1	-7.4
2	-7.5	-9.7	-9.2	-9.4	-8.0	-8.8
3	-10.1	-9.7	-9.1	-8.3	-10.0	-9.9
4	-9.3	-10.0	-9.1	-6.7	-8.4	-8.7
5	-8.6	-9.5	-8.4	-8.4	-8.9	-8.1
6	-8.8	-8.4	-8.1	-7.6	-6.2	-7.9
7	-9.4	-9.5	-7.0	-8.7	-9.2	-9.8
8	-6.3	-7.6	-9.8	-9.8	-7.4	-9.0
9	-6.7	-7.5	-9.1	-8.4	-9.3	-8.9
10	-8.3	-8.3	-7.2	-8.5	-8.4	-8.5
11	-9.1	-8.3	-8.2	-8.3	-8.0	-6.3
12	-8.0	-9.2	-8.1	-8.6	-7.1	-9.7
13	-9.2	-8.2	-5.8	-8.3	-9.2	-8.3
14	-9.3	-9.5	-10.4	-10.0	-8.8	-7.4
15	-8.6	-8.5	-8.8	-9.4	-8.3	-6.6
16	-8.8	-8.6	-8.1	-8.9	-6.6	-8.2
17	-9.4	-9.7	-9.1	-7.9	-9.1	-9.3
18	-6.5	-6.9	-7.3	-7.1	-5.8	-7.4
19	-8.9	-8.8	-8.1	-9.1	-7.5	-8.2
20	-7.7	-8.0	-8.2	-6.5	-9.6	-7.2

Note. **A lower speech reception threshold indicates better performance.**

Appendix C (continued)

Adaptive Test Condition Scores

Table 1: *SRTs (in dB SNR) for Each List in Session Two for Each Participant*

Session & Participant	List 1	List 2	List 3	List 4	List 5	List 6
Session Two						
1	-9.7	-9.6	-8.6	-9.2	-10.1	-9.0
2	-10.0	-10.3	-10.3	-10.1	-9.5	-9.0
3	-10.4	-11.4	-10.6	-10.6	-11.0	-10.4
4	-9.6	-9.6	-9.4	-9.7	-10.4	-9.2
5	-8.5	-9.8	-9.2	-8.9	-9.4	-9.1
6	-9.8	-8.8	-9.9	-9.6	-8.4	-9.3
7	-10.7	-11.3	-10.4	-9.2	-10.8	-10.4
8	-10.3	-10.9	-11.4	-9.2	-9.6	-10.9
9	-9.9	-9.7	-10.6	-9.6	-10.6	-8.0
10	-9.7	-9.8	-8.9	-9.2	-10.3	-10.1
11	-8.8	-8.1	-8.4	-8.7	-8.4	-9.9
12	-10.2	-10.0	-10.5	-9.3	-11.0	-10.3
13	-9.8	-9.8	-9.6	-9.9	-8.9	-10.3
14	-10.3	-11.2	-12.1	-11.3	-11.2	-11.2
15	-10.2	-10.0	-9.5	-8.2	-9.9	-10.5
16	-10.2	-9.4	-10.3	-9.1	-9.0	-10.1
17	-10.0	-10.3	-10.3	-10.1	-10.6	-10.7
18	-8.1	-7.3	-7.7	-7.9	-8.1	-8.2
19	-9.3	-9.2	-9.6	-8.9	-9.8	-9.7
20	-8.8	-8.9	-9.4	-9.5	-8.1	-8.9

Note. **A lower speech reception threshold indicates better performance.**

Appendix D

Transcription of American English Matrix Test Lists

List One

William sold two green sofas
Peter gives four large tables
Steven wants two dark spoons
Kathy prefers nine pretty flowers
Peter wants four old toys
Alan ordered sixty red rings
Alan ordered nineteen white houses
Thomas bought twelve small chairs
Thomas prefers fifteen heavy spoons
Doris sees three old windows
Lucy sold eight cheap toys
Doris sees nineteen white desks
Rachel got seven heavy rings
Steven got nine pretty windows
Kathy kept seven large tables
Nina gives eight cheap flowers
Nina kept fifteen red desks
William has sixty green sofas
Rachel has three small chairs
Lucy bought twelve dark houses

List Two

Rachel has three small chairs
Steven wants seven white rings
Kathy kept seven large tables
William sold two red desks
Lucy has three green windows
Rachel got four large toys
Alan gives sixty heavy flowers
William sold two green sofas
Peter prefers twelve dark sofas
Thomas prefers fifteen heavy spoons
Nina bought nineteen old tables
Peter wants four old toys
Lucy bought twelve dark houses
Alan ordered sixty red rings
Doris sees eight small houses
Steven got nine pretty windows
Nina gives eight cheap flowers
Kathy ordered nine pretty spoons
Doris sees nineteen white desks
Thomas kept fifteen cheap chairs

List Three

Nina gives eight cheap flowers
Doris sold seven white houses
Alan ordered sixty red rings
Thomas gives four old spoons
Peter got nine pretty rings
Doris sees nineteen white desks
Peter wants four old toys
Lucy has sixty dark flowers
Kathy kept eight small tables
Lucy bought twelve dark houses
Steven prefers three large windows
William sees fifteen heavy desks
Rachel has three small chairs
Alan ordered two cheap toys
Kathy kept seven large tables
Rachel wants nineteen red chairs
Steven got nine pretty windows
Nina bought twelve green sofas
William sold two green sofas
Thomas prefers fifteen heavy spoons

List Four

Lucy kept twelve large toys
Lucy bought twelve dark houses
Doris sees nineteen white desks
Alan ordered seven red tables
Rachel wants four pretty chairs
Thomas bought two white sofas
Steven got nine pretty windows
Nina has nineteen cheap spoons
William prefers nine small houses
Doris sold eight heavy desks
Rachel has three small chairs
Steven sees three green flowers
Peter wants four old toys
Kathy gives fifteen old rings
Nina gives eight cheap flowers
Peter got sixty dark windows
Kathy kept seven large tables
William sold two green sofas
Alan ordered sixty red rings
Thomas prefers fifteen heavy spoons

List Five	List Six
Thomas kept four large sofas	Kathy kept seven large tables
Steven has three heavy flowers	Doris sees nineteen white desks
Steven got nine pretty windows	Kathy prefers eight cheap houses
Nina gives eight cheap flowers	Nina gives eight cheap flowers
William sold two green sofas	Thomas has three small flowers
Alan ordered sixty red rings	Steven bought twelve green toys
Lucy bought twelve dark houses	William sees fifteen dark chairs
Kathy kept seven large tables	Alan ordered sixty red rings
Alan sold twelve white toys	Doris kept nineteen white windows
Doris sees nineteen white desks	Nina wants seven heavy tables
Peter wants four old toys	Steven got nine pretty windows
Peter wants eight green chairs	Lucy gives sixty red spoons
Rachel has three small chairs	Lucy bought twelve dark houses
Nina gives nineteen old desks	Rachel has three small chairs
Doris ordered sixty small windows	Rachel sold two old desks
Kathy sees nine cheap spoons	Thomas prefers fifteen heavy spoons
William got two pretty rings	Peter got four large rings
Thomas prefers fifteen heavy spoons	William sold two green sofas
Rachel prefers fifteen dark houses	Alan ordered nine pretty sofas
Lucy bought seven red tables	Peter wants four old toys

REFERENCES

Beattie, R. (1989). Word recognition functions for the CID W-22 test in multitalker noise for normally hearing and hearing-impaired subjects. *Journal of Speech and Hearing Disorders, 54,* 20-32.

Cox, R., Alexander, G., & Gillmore, C. (1987). Development of the connected speech test (CST). *Ear and Hearing, 8*(Suppl. 5), 119S-126S.

Danhauer, J., Doyle, P., & Lucks, L. (1985). Effects of noise on NST and NU 6 stimuli, *Ear and Hearing, 6*(5), 266-269.

Danhauer, J. & Leppler, J. (1979). Effects of four noise competitors on the California consonant test. *Journal of Speech and Hearing Disorders, 34,* 354-362.

Dirks, D., Morgan, D., & Dubno, J. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47*, 114-123.

Dubno, J., Horwitz, A. & Ahlstrom, J. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *Journal of the Acoustical Society of America, 111*(6), 2897-2907. doi: 10.1121/1.1480421

Dubno, J., Horwitz, A., & Ahlstrom, J. (2003). Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. *Journal of the Acoustical Society of America, 113*(4), 2084-2094. doi:10.1121/1.1555611.

Duncan, K., & Aarts, N. (2006). A comparison of the HINT and QuickSIN tests. *Journal of Speech-Language Pathology and Audiology, 30*(2), 86-94.

Eisenberg, L., Dirks, D., & Bell, T. (1995). Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing. *Journal of Speech and Hearing Research, 38*, 222-233.

Festen, J., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America, 88*(4), 1725-1736.

Hygge, S., Ronnberg, J., & Arlinger, B. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *Journal of Speech and Hearing Research, 35,* 208-215.

Margolis, R. & Heller, J. (1987). Screening tympanometry: Criteria for medical referral. *Audiology, 26,* 197-208.

McArdle, R. & Wilson, R. (2006). Homogeneity of the 18 QuickSIN™ lists. *Journal of the American Academy of Audiology, 17*(3)*,* 157-167.

McArdle, R., Wilson, R., & Burks, C. (2005). Speech recognition in multitalker babble using digits, words, and sentences. *Journal of the American Academy of Audiology, 16*(6)*,* 726-739.

Nelson, P., Jin, S., Carney, A. & Nelson, D. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 113*(2), 961-968. doi: 10.1121/1.1531983

Plomp, R., & Mimpen, A. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18,* 43-52.

Studebaker, G., Sherbecoe, R., McDaniel, D. & Gwaltney, C. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *Journal of the Acoustical Society of America, 105*(4), 2431-2444.

Versfeld, N., & Dreschler, W. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *Journal of the Acoustical Society of America, 111*(1), 401-408. doi: 10.1121-1.1426376

Wagener, K. & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology, 44*(3), 144-156. doi: 10.1080/14992020500057517

Wagener, K., Josvassen, J. & Ardenkjar, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *International Journal of Audiology, 42*(1), 10-17.

Wilson, R., Bell, T., & Koslowski, J. (2003). Learning effects associated with repeated word-recognition measures using sentence materials. *Journal of Rehabilitation Research and Development, 40*(4), 329-336.

Wilson, R., & Burks, C. (2005). Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol. *Journal of Rehabilitation & Development, 42*(6), 839-852. doi: 10.1682/JRRD.2005.01.0009.

Wilson, R., Burks, C., & Weakley, D. (2005). Word recognition in multitalker babble measured with two psychophysical methods. *Journal of the American Academy of Audiology, 16*(8), 622-630.

Wilson, R., Carnell, C., & Cleghorn, A. (2007). The words-in noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *Journal of the American Academy of Audiology, 18*(6), 522-529.

Wilson, R., & Cates, W. (2008). A comparison of two word-recognition tasks in multitalker babble: speech recognition in noise test (SPRINT) and words in noise test (WIN). *Journal of the American Academy of Audiology, 19*(7), 548-556.
doi: 10.3766/jaaa.19.7.4

Wilson, R., & McArdle, R. (2007). Intra- and inter-session test, retest reliability of the words-in-noise (WIN) test. *Journal of the American Academy of Audiology, 18*(10). 813-825.

Curriculum Vita

NAME: Holly A. Rose

PERMANENT ADDRESS: 2956 Raking Leaf Drive, Abingdon, Maryland 21009

PROGRAM OF STUDY: Applied Doctor in Audiology

DEGREE AND DATE TO BE CONFERRED: Doctorate in Audiology, May 2013

Secondary education: The Catholic High School of Baltimore, Baltimore, Maryland, May 2005

Collegiate institutions attended

Doctor of Audiology Candidate
Towson University, Doctor of Audiology Program
Towson, Maryland
Dates attended: August 2009-Present
Expected graduation: May 2013

Bachelor of Science in Speech-Language Pathology & Audiology
Towson University, Department of Audiology and Speech-Language Pathology
Towson, Maryland
Dates attended: August 2005-May 2009
Graduation: May 2009