2018-06-01

The Effect of the Slope of the Psychometric Function on the Measurement of Speech Recognition Threshold Using a Female Talker

Jessica Lee Reese

The Effect of the Slope of the Psychometric Function on the Measurement of Speech Recognition Threshold Using a Female Talker

Jessica Lee Reese

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Richard Harris, Chair
David McPherson
Christopher Dromey

Department of Communication Disorders

Brigham Young University

ABSTRACT

The Effect of the Slope of the Psychometric Function on the Measurement of Speech Recognition Threshold Using a Female Talker

Jessica Lee Reese
Department of Communication Disorders, BYU
Master of Science

Speech audiometry has long been a component of a thorough audiological examination. The speech recognition threshold (SRT) measurement is perhaps the most widely used measurement in speech audiometry. For decades, researchers and clinicians have worked to create and fine-tune word lists that for use in SRT testing; their aim being to improve the accuracy for classifying a client's ability to hear and comprehend speech. Experts in the field have agreed to follow four tenets of speech audiometry when selecting word sets. This study examined whether improvement to stimulus lists for SRT measurement could be made in regards to the tenet of homogeneity with respect to audibility if the slope of the psychometric function were a selection consideration. The study was performed with the hypothesis that steeply sloping words would significantly reduce the number of words needed to obtain the SRT. Three word lists, all recorded by a female talker, comprising of steeply sloping words, medium sloping words, and shallow sloping words, were used in the study. Participants with normal hearing between the ages of 18 and 30 years provided data that was used to calculate SRT measurements for all three lists from each ear. The results showed a significant difference in the number of words needed to obtain the SRT when comparing the steep and shallow word sets and the shallow and medium word sets. Steeply sloping words required fewer words to obtain the SRT, $M = 17.02$. Shallow sloping words required the most words, $M = 18.88$, amounting to a difference of 1.86 words. While statistically different, a reduction by fewer than 2 words during the course of SRT testing will not equate to a substantial saving of time for the clinician. For clinical application, the slope of the psychometric function of the words used in SRT measurement need not be a primary consideration when developing stimulus lists.

Keywords: psychometric function of slope, speech recognition threshold, speech audiometry

ACKNOWLEDGMENTS

Thanks must be given to the faculty and staff of the Brigham Young University's Department of Communication Disorders. I am grateful to have had the privilege of learning from you and feel blessed to have been a part of this program. I thank my children, Kate, Daniel, and Laura, for their love and patience while "Mommy did her school."

TABLE OF CONTENTS

TITLE i

ABSTRACT ii

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES v

DESCRIPTION OF THESIS STRUCTURE vi

Introduction 1

Method 10

Participants 10

Materials 10

Procedure 11

Results 12

Discussion 15

References 18

APPENDIX A: Informed Consent 21

APPENDIX B: Spondaic Words Spoken by a Female Talker 23

APPENDIX C: Annotated Bibliography 24

LIST OF TABLES

Table 1: Summary of Descriptive Statistics from a Female Talker Based on Slope............. 13

DESCRIPTION OF THESIS STRUCTURE

This thesis, *The Effect of the Slope of the Psychometric Function on the Measurement of Speech Recognition Threshold Using a Female Talker*, is part of a larger research project, and portions of this thesis may be published as part of articles listing the thesis author as a co-author. The body of this thesis is written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology. A copy of the informed consent is found in Appendix A, the stimuli used for testing is in Appendix B, and an annotated bibliography is presented in Appendix C.

Introduction

An audiological evaluation is used to determine if a hearing loss exists and to what degree, to characterize the type of hearing loss, and to establish appropriate intervention. A proper audiological evaluation utilizes a battery of tests and assessments that typically include a comprehensive case history, otoscopic examination, comprehensive audiometric examination that includes air conduction thresholds, bone conduction thresholds, speech audiometry, otoacoustic emissions, tympanometry, and acoustic reflex thresholds and reflex decay.

Speech audiometry is one of the basic components of an audiological evaluation. It has been recognized that to properly evaluate the hearing function for verbal communication, an individual's hearing of verbal output should also be evaluated (Carhart, 1946; Egan, 1979; Hirsh, 1947; Ramkissoon, 2001). Speech audiometry is used to verify the results of the pure tone average (American Speech-Language-Hearing Association [ASHA], 1988; Beattie, Svihovec, & Edgerton, 1975; Egan, 1979). Determining a client's abilities to discriminate speech is reliant on the results of speech audiometry (Beattie et al., 1975; Egan, 1979). Also, speech audiometry can simply indicate whether a hearing aid is needed based on the degree of hearing loss (Carhart, 1951; Egan, 1979).

Within speech audiometry, there are three common tests for hearing of speech the speech recognition threshold (SRT), the speech detection threshold (SDT), and the word recognition score (WRS). The SRT is the minimum intensity level at which 50% of the presented speech sample can be understood (ASHA, 1988). The SDT is the minimum intensity level at which 50% of the presented speech stimuli can be detected (ASHA, 1988). The SDT is usually reserved for testing of young children, for those who cannot respond verbally, or for those suspected of malingering. For testing in English, the WRS is typically obtained by presenting a list of monosyllabic words to each ear; the percentage correct is tallied to calculate the WRS (Schlauch, Anderson, & Micheyl, 2014). The SRT component of speech audiometry will be the focus of this study.

Current methods of establishing the SRT follow the guidelines put forth by ASHA in 1988. The term *speech recognition threshold* is interchangeable with *speech reception threshold*, although ASHA gives preference to speech recognition threshold, due to its more accurate description of the task procedures. Considerations, as stated by ASHA guidelines, to account for prior to administration of a speech threshold test are: instrumentation and calibration, test environment, test material, response mode, recording of results, and masking of the nontest ear. Once these variables are properly managed, the client is given instructions regarding the procedure and is familiarized with the word list; the SRT is then determined using a descending technique. A more detailed explanation of the threshold procedure is outlined in the Method section.

Fletcher and Steinberg (1930) at the Bell Laboratories worked to standardize tests of speech recognition. Originally, speech recognition lists developed for the Bell Laboratories were used to test telephone systems (Hirsh, 1947). These early lists were developed randomly, without organization or order for the sounds within the lists (Fletcher & Steinberg, 1930). Continued refinement of the lists allowed them to become more accurate tests of hearing (Hirsh, 1947). As concern grew for the content of the testing lists, two objections arose. The first was the difficulty in making the lists equally hard without being too long (Fletcher & Steinberg, 1930). The second was a concern regarding the learning effect and the desire to reduce or eliminate this effect. It was believed that a large number of lists would be required to achieve that result (Fletcher & Steinberg, 1930).

As the use of speech recognition tests grew, the question arose concerning the effectiveness of speech recognition tests in determining hearing loss as compared to results from pure tone testing. A second concern was normalizing the results from speech recognition tests. Pure tone testing and the resulting audiogram have a universal continuity that makes that test particularly effective in describing hearing loss. There is no ambiguity or confusion when comparing audiograms from various audiologists and clinics. Hughson and Thompson wanted to "establish a normal threshold for the hearing of speech as related to normal audiograms" (1942, p. 527). Simplicity of testing procedures was also a consideration for the development of the speech recognition tests (Hudgins, Hawkins, Karlin, & Stevens, 1947; Hughson & Thompson, 1942). Hughson and Thompson experimented with testing procedures to find one that met the two aforementioned criteria. The procedure Hughson and Thompson detailed in their study is similar to the current established method for obtaining an SRT. The testing was done within a soundproof audiometric booth, the intensity levels of the stimuli were carefully controlled, the subject was required to repeat back what was heard, and an intensity level of normal hearing was established at which testing began, followed by systematic decreases in intensity until a predetermined number of incorrect responses was made (Hughson & Thompson, 1942). Hughson and Thompson found a correlation between the results of speech reception testing and the results of traditional audiometric testing.

Carhart (1946) also investigated whether speech reception tests were a good measure of hearing. To accomplish this aim, Carhart compared the results of speech recognition tests to pure tone results. Carhart concluded that the correlations between the two tests were strong enough to indicate a high commonality between them and that the results did not favor one test over the other. Within the same study, Carhart explored the reliability of testing speech recognition with disyllabic words versus sentences. It had been standard practice to use the sentence lists from the Bell Laboratories when testing speech reception; however, the time advantage of using single words was recognized. Researchers from the Psycho-Acoustic Laboratory (PAL) at Harvard University developed a series of articulation tests for use of the military during and after World War II (Hirsh et al., 1952). A few of these tests were found to be efficient in also testing speech reception, particularly the PAL Auditory Test No. 9, which used spondaic words (Hirsh et al., 1952). Spondaic words are bisyllabic words with equal stress on each of the two syllables. Carhart (1946) concluded that words and sentences were equal in their ability to accurately measure the speech reception threshold when the stimuli followed proper testing parameters.

In development of the PAL lists, consideration was given to the vast nature of possible materials: nonsense syllables, phonetically balanced monosyllabic words, spondaic bisyllabic words, and sentences (Hirsh, 1947). Auditory Test No. 9 was developed using the bisyllabic spondees. Spondaic words were chosen over iambic words and trochaic words because, for the latter two classes of words, recognition may be affected not only by the phonetic content but also by the stress pattern (Hirsh, 1947). Auditory Test No. 9 consisted of 84 spondaic words, chosen because the range of intelligibility thresholds for all of the words was relatively small (Hirsh, 1947). The homogeneity with respect to audibility obtained by Auditory Test No. 9 became one of the tenets for SRT testing.

Four stimulus criteria have been named as essential when testing for SRT: (a) familiarity, (b) phonetic dissimilarity, (c) normal sampling of English speech sounds, and (d) homogeneity with respect to basic audibility (Hudgins et al., 1947; Young, Dudley, & Gunter, 1982). Familiarity refers to the words chosen as stimuli; they must be familiar to the patient. The SRT is not a test of intelligence; hence, the words themselves need to be familiar to the general population. Phonetic dissimilarity prevents the SRT from being a test of auditory discrimination. When the words are dissimilar the client will not have to try to distinguish the stimuli from a possible second stimulus. A normal sampling of English speech sounds ensures that a test of speech recognition is testing all speech sounds an individual may encounter. This criterion, historically, has been the least considered of the four and adherence is sufficient if reasonable effort has been made to comply (Hudgins et al., 1947).

It would be difficult to obtain an accurate threshold for speech recognition if the words presented at similar intensities did not have similar audibility. To comply with homogeneity with respect to audibility, the words chosen must all have the same level of amplitude when spoken at similar intensities, or the level of amplitude must be adjusted through audiometric instrumentation so that homogeneity is reached (Hudgins et al., 1947). Both of these solutions may be used to create a word list that has homogeneity with respect to audibility (Hudgins et al., 1947). Homogeneity with respect to audibility benefits the SRT in the time required to obtain the SRT. Knowing all words are homogeneous with respect to audibility, the lists can be subdivided into smaller lists from which accurate SRTs may still be obtained (Hudgins et al., 1947). The resulting scores will be based on intensity levels and not differences in audibility (Hudgins et al., 1947).

As clinicians and researchers tried to adhere to the four tenets of speech audiometry, deficiencies in the Harvard PAL tests were noted (Hirsh et al., 1952). It was found that the thresholds for the words in Auditory Test No. 9 varied slightly and that size of vocabulary was too large for regular clinical use (Hirsh et al., 1952). Improvements were made to Auditory Test No. 9 and a new list was created and was used to form the Central Institute of the Deaf (CID)

Tests W-1 and W-2 (Hirsh et al., 1952). The 84 words from the PAL test were narrowed down to 36, improving the homogeneity of the words' thresholds (Hirsh et al., 1952). The 36 words were divided into six word orders. Previously, the entire list, or a portion of the list, was presented at a fixed intensity level. This added to the time required to establish the SRT. With the CID Test W-2, the intensity level was attenuated 3 dB after every three words (Hirsh et al., 1952). The change in testing procedures reduced the time needed for SRT establishment yet still provided accurate results (Hirsh et al., 1952).

While the CID Test W-1 did improve homogeneity with respect to audibility, variability was still found to exist (Wilson & Strouse, 1999; Young et al., 1982). Several attempts have been made to create sublists with better homogeneity with respect to audibility (Beattie et al., 1975; Bilger, Matthies, Meyer, & Griffiths, 1998; Curry & Cox, 1966; Wilson & Strouse, 1999; Young et al., 1982). Young et al. investigated whether a list of words with greater homogeneity with respect to audibility would yield different SRTs as well as test/retest reliability. The results of the study led to the conclusion that a more homogeneous list of words did not produce SRTs that were significantly different when compared to the entire list of 36 words (Young et al., 1982). Nor did increased homogeneity with respect to audibility yield different test/retest reliability than the 36-word list (Young et al., 1982). The shorter lists did not improve on the establishment of an SRT but neither did the shorter lists fall short of the results of the entire 36 words. Hence, a reliable SRT could be obtained using a list of just 15 words rather than 36 (Young et al., 1982).

The concern over the homogeneity with respect to audibility of the word lists used to establish SRTs continued. Bilger et al. (1998) argued that homogeneity based on the threshold of intelligibility was not enough. Bilger et al. considered psychometric functions in determining a list of words that would more closely adhere to the tenet of homogeneity with respect to audibility. Slope of the psychometric function refers to the rate at which intelligibility is affected with changes in intensity. A steeper slope will yield greater increases in intelligibility with smaller increases in intensity (MacPherson and Akeroyd, 2014). Slope of the psychometric function is calculated as a relationship between the change in correct recognition of the stimulus and the change in the presentation level (Wilson & Carter, 2001).

Cambron, Wilson, and Shanks (1991) calculated the mean slope of the psychometric function for the 36 spondaic words as spoken by a male and female talker. The average for the male talker was 9.2%/dB and for the female, 9.1%/dB. Cambron et al. then examined the slope of the psychometric function for each individual word when spoken by a female speaker. The slopes were less homogeneous than the averaged slopes implied. Individual slopes ranged from 6.5%/dB to 26.1%/dB (Cambron et al., 1991). Bilger et al. applied a logit function to the data collected by Young et al. (1982) and compared the results to Young and colleagues' linear function results. Bilger et al. created a list of 30 words taken from their findings. They argued that their list of 30 words, which was formed using psychometric function measurements as a selection criteria, was more homogeneous than other published word lists. Through digital manipulation, they further reduced variability in the thresholds and the slopes of the psychometric function of the word list.

Despite the efforts of the previously mentioned researchers to improve homogeneity with respect to audibility, Wilson and Strouse (1999) argued that too much variability still existed in the CID Test W-1 list with a female talker. Wilson and Strouse compared thresholds of intelligibility and psychometric functions of the female talker with those of the original recording done by Hirsh et al. in 1952. They found significant differences in the thresholds between the two recordings. The next step was to create a recording of the female talker which produced words with equal thresholds. Cambron et al. (1991) noted a connection between the threshold and the slope of the psychometric function. An inverse relationship existed between the standard deviation of the threshold and the slope; a more gradual slope was found when the standard deviation was larger. The increased homogeneity with respect to audibility of the list developed by Wilson and Strouse (1999) naturally included words with steep psychometric function slopes, leading to improved reliability and reduced testing time (Nissen, Harris, & Slade, 2007).

The tenet of familiarity has been revisited in research, particular for the 36 spondees of CID Test W-1. Hirsh et al. (1952) established this list in the early 1950s. Chipman (2003) researched whether these words were still, in fact, “familiar” to the general public. By comparing the CID Test W-1 lists to current corpora of English language usage, Chipman discovered that 14 of the 36 CID Test W-1 words were no longer frequently occurring words. Chipman developed a list of 98 words that were frequently used in American English; the threshold and psychometric function slope were measured for each word. The 98 words were recorded digitally following established and recognized professional guidelines (American National Standards Institute [ANSI], 1996). The list was condensed to 33 words that met the other tenets of homogeneity with respect to audibility.

When looking at the slope of the psychometric function of stimulus words and using it as a means to create lists that are homogeneous with respect to audibility, the distinction of what constitutes a steeply sloping word versus a shallow sloping word must be considered. Young et al. (1982) determined the mean slope of the 36 spondees of the CID Test W-1 to be 10.0%/dB. Young et al. reasoned that for a list to be homogeneous, the slope of each individual word should be within ± 1 *SD* of the mean. The resulting list excluded words below 8.0%/dB and words above 12.5%/dB. Hence, words below 8.0%/dB may be considered shallow sloping, words within ± 1 *SD* of 10.0%/dB may be considered medium sloping, and words above 12.5%/dB may be considered steeply sloping. Another way to use the work performed by Young et al. as a guide is to create word lists of shallow slope, medium slope, and steep slope that fall within ± 1 *SD* of a predetermined mean for shallow, medium, and steeply sloping words.

The idea that the slope of the psychometric function improves the time required to establish an SRT provided the basis for the purpose of this study. This study examined whether it makes sense to use spondaic words with a steep slope over spondaic words with a shallow or medium slope. In addition to the effect on the amount of time required to obtain the SRT, the effect the slope of the psychometric function has on the accuracy of the SRT measurement and how it agrees with the pure tone average (PTA) will be analyzed. There is no current, published research documenting the effect of slope on the number of words required to obtain the SRT nor concerning the effect on the results of SRT measurement. Current research on the measurement of the SRT with English spondaic words and a focus on the slope of the psychometric function is lacking. This study will add to the field current research findings based on SRT measurement. Using both a male and female talker, the effect of the slope of the psychometric function on SRT testing will be determined. The SRT will be measured along with the number of words required to obtain the SRT. The hypothesis is that a set of steeply sloping words will enable the accurate measurement of SRT using fewer words when compared to sets of words with shallow slopes and medium slopes.

Method

Participants

The subjects of this study were native English speakers from the United States. A total of 40 subjects (20 male, 20 female) participated in the study of the effect of slope on SRT measurement. The right and left ears were tested and data from each ear were regarded as individual data sources; subsequent presented and analyzed data is then based on 80 subjects. The subjects' ages ranged from 18 to 30 years (*M* = 24.8). The subjects had normal hearing, with pure tone air conduction thresholds of ≤15 dB HL at all octave and mid-octave frequencies from 125 to 8000 Hz in both ears. Each subject also had acoustic admittance between 0.3 and 4.3 mmhos with peak pressure between -10 and +50 daPa (ASHA, 1990). Subjects were recruited through flyers, word of mouth, and social media. An informed consent document, approved through the Brigham Young University Institutional Review Board, was signed by each of the subjects.

In addition to the 40 subjects from whom the collected data were analyzed, there were 22 potential subjects whose data could not be used. Eleven potential subjects failed the hearing screen, 5 male and 6 female. A threshold of speech recognition for at least one of the sublists presented could not be established for 11 of the potential subjects, 3 male and 8 female. The individuals continued to give correct responses to the stimuli at the lowest dB levels at which it was possible to present words given the range of the audiometer (-10 dB HL).

Materials

The 98-word-list established by Chipman (2003) was used to compile lists of words with a shallow slope (6.0 – 11.0%/dB), a medium slope (12.3 – 15.0%/dB), and a steep slope (16.0 – 22.2%/dB). There were 19 words in the set of steeply sloping words, 33 in the medium sloping set, and 25 in the shallow sloping set. Appendix B contains the lists of steep, medium, and shallow sloping words used in the study. The recordings used were those made by Chipman at Brigham Young University. A review of Chipman's study is recommended for further information on the recordings.

Procedure

Once normal hearing was verified, each participant was familiarized with the list of words that were used to calculate the SRT. Both male and female recordings were played to the subjects to familiarize them with the word lists. The SRT was measured in both the right and left ears. Each subject was presented two lists (one male talker and one female talker) of shallow sloping words, medium sloping words, and steeply sloping words. The total number of SRTs measured from each subject was 12. The approximate time needed to measure one SRT was 2 minutes, for a total approximate time of 30 minutes per subject. One headphone was used to eliminate a possible variable of differences in headphone quality and one ear was tested at a time, eliminating the variable of a constant need to change the headphone. It was randomly decided if the right ear was tested first or the left ear. The presentation of talker and type of list was randomly determined.

ASHA (1988) guidelines allow for SRT measurement using a 5 dB descending approach and/or a 2 dB descending approach; this study utilized the 2 dB method. The starting point for each subject was 30 dB HL. One spondaic word was presented and if the response was correct, the presentation level was reduced by 10 dB increments until an incorrect response was given. If the initial response was incorrect, the presentation level was increased by 20 dB until a correct response was given, and then the 10 dB decrement process began. When a spondee was missed, a second was presented at the same level. When two spondees were missed at the same level, the process was stopped. The starting level was 10 dB above the level of the two missed responses. Two spondaic words were presented at the starting level with the presentation level descending 2 dB after every two spondees. The process was continued until five of the last six response were incorrect. Calculation of the SRT used the Spearman-Karber method and the simplified formula: SRT = starting level - # correct + correction factor.

Results

Understanding of the effect of the slope of the psychometric function, talker gender, and subject ear on the number of words needed to obtain the SRT, or the effect on the amount of time needed to for the SRT measurement, was the primary purpose of this within-subjects study. Second, the slope of the psychometric function, talker gender, and subject ear were examined for their effects on the accuracy of SRT measurement, the difference between the measured SRT and the measured PTA, and the number of words needed to obtain the starting level. The data were analyzed using SAS version 9.4.

Table 1 reports the descriptive statistics. The results indicate that steeply sloping words required the fewest words for obtaining both the SRT and the starting level, (17.02 words and 8.31 words, respectively). Steeply sloping words also produced the lowest average in dB HL for the SRT (M = -0.26 dB HL). Medium sloping words produced the highest average in dB HL for the SRT (M = 0.54 dB HL), had the smallest difference between the SRT and PTA (-0.75 dB), but had the greatest range for both the SRT (-7.00 – 16.00 dB HL) and the difference between the SRT and PTA (-9.67 – 8.33 dB). Shallow sloping words required the greatest number of words for obtaining both the SRT and the starting level (18.88 words and 8.58 words, respectively). As well, shallow sloping words had the highest range of words needed to obtain the SRT (11.00 – 31.00 words). Analysis of the combined data revealed that the average measured SRT was lower in dB HL than the measured PTA (-1.13 dB); however, the average for the difference between the SRT and PTA fell within accepted limits (Brandy, 2002; Martin & Clark, 2006).

Table 1

Summary of Descriptive Statistics from a Female Talker Based on Slope, n = 80

Slope	Variable	*M*	*SD*	Minimum	Maximum
Steep	Words for SRT	17.02	3.73	10.00	25.00
	SRT (dB HL)	-0.26	3.87	-8.00	16.00
	SRT-PTA (dB)	-1.55	3.75	-9.67	6.67
	Words for SL	8.31	2.47	4.00	13.00
Medium	Words for SRT	17.20	3.43	10.00	29.00
	SRT (dB HL)	0.54	3.49	-7.00	16.00
	SRT-PTA (dB)	-0.75	3.66	-9.67	8.33
	Words for SL	8.55	2.89	5.00	16.00
Shallow	Words for SRT	18.88	4.69	11.00	31.00
	SRT (dB HL)	0.23	3.65	-8.00	15.00
	SRT-PTA (dB)	-1.07	3.57	-8.33	6.00
	Words for SL	8.58	2.81	4.00	15.00
Average[a]	Words for SRT	17.70	4.06	10.00	31.00
	SRT (dB HL)	0.17	3.67	-8.00	16.00
	SRT-PTA (dB)	-1.13	3.66	-9.67	8.33
	Words for SL	8.48	2.72	4.00	16.00

Note. Words for SRT = the number of words required to obtain the SRT; SRT-PTA = difference between the measured SRT and the measured PTA; Words for SL = the number of words required to obtain the starting level.
[a]The average data for all three slopes: steep, medium, and shallow (*n* = 240).

A multivariate analysis of variance was employed to examine any differences caused by the slope of the psychometric function, talker gender, and/or subject ear. The effects of talker gender and subject ear were determined to have no significant effect on the variables being measured and thus they were not included in further statistical analyses. The data were analyzed again using an ANOVA to examine the effect of the slope of the psychometric function on the measured words needed to obtain the SRT, the SRT measurement, the difference between the measured SRT and measured PTA, and the number of words needed to obtain the starting level.

The ANOVA revealed that the slope of the psychometric function significantly influenced the number of words to measure the SRT. There was a statistically significant difference among the three slopes with regard to the number of words needed to obtain the SRT; $F(2, 158) = 5.48$, $p = .005$. A Tukey-Kramer posthoc analysis showed a significant difference between the steeply sloping set of words with the shallow sloping set of words; $t(158) = -3.00$, $p = .009$. A significant difference existed between the medium sloping set of words and the shallow sloping set of words; $t(158) = 2.71$, $p = .020$. Additionally, the slope of the psychometric function had a statistically significant effect on the measured SRT; $F(2, 158) = 4.52$, $p = .012$. A Tukey-Kramer posthoc analysis revealed this difference was between the steeply sloping words and medium sloping words; $t(158) = -2.98$, $p = .009$. A statistically significant difference was revealed for the effect of the slope of the psychometric function on the difference between the measured SRT and the measured PTA; $F(2, 158) = 4.52$, $p = .012$. This difference existed between the steeply sloping words and the medium sloping words as shown by a Tukey-Kramer posthoc analysis; $t(158) = -2.98$, $p = .009$. A statistical difference was not found to exist for the effect of slope on the number of words to obtain the starting level; $F(2, 158) = 0.28$, $p = .759$.

Discussion

The primary purpose of this investigation was to examine, experimentally, whether the slope of the psychometric function had a significant effect on the number of words needed to obtain an SRT measurement. It was hypothesized that a set of steeply sloping words would produce SRTs with fewer words needed for the testing than word sets of medium and shallow slopes. Secondarily, the effect of the slope of the psychometric function was examined for any effect on the accuracy of the SRT, the difference between the measured SRT and the measured PTA, and the number of words needed to obtain the starting level for SRT measurement. The investigation stemmed from one of the four tenets for SRT word selection regarding homogeneity with respect to audibility. If the slope did, in fact, have a significant effect, then consideration of the slope of the word sets would be necessary to adhere to the required homogeneity with respect to audibility. For purposes of the study, steeply sloping words were considered to have a slope between 16.0%/dB to 22.2%/dB; medium sloping words were considered to have a slope between 12.3%/dB to 15.0%/dB; and shallow sloping words were considered to have a slope between 6.0%/dB to 11.0%/dB.

The results of the ANOVA and subsequent posthoc analyses show a significant effect of the slope of the psychometric function on the number of words needed to obtain the SRT, or the time required to administer of the test. The mean number of words with steep psychometric function slopes was 17.0; this was significantly different from the mean number of words with shallow psychometric function slopes, 18.9. The mean number of words with medium psychometric function slopes, 17.2, was also significantly different from the mean number of words with shallow psychometric function slopes. There was no statistically significant difference between steep slope and medium slope in the number of words to obtain the SRT.

While statistically different, the actual difference was less than two words, for both steep slope versus shallow slope and medium slope versus shallow slope. A savings of less than two words is not likely to equate to noticeable difference in the time required to obtain a SRT measurement in the clinical setting.

The effect of the slope of the psychometric function was significant for the measured SRT. There was a statistical difference between the mean SRT measured using words with steep psychometric function slopes, -.0.3 dB HL, and the mean SRT measured using words with medium psychometric function slopes, 0.5 dB HL. However, the actual difference was 0.8 dB, and this is an acceptable difference for SRT measurement in a clinical setting (Brandy, 2002; Martin & Clark, 2006).

A statistical difference was revealed between the measured SRT and the measured PTA. The mean difference between the measured SRT and the measured PTA using words with steep psychometric function slopes was -1.6 dB, which was significantly different from the mean difference between the measured SRT and the measured PTA using words with medium psychometric function slopes, -0.8 dB. As with the measured SRT, although the difference was statistically significant, the actual difference of 0.8 dB between the measured SRT and the measured PTA is within acceptable limits for clinical audiology (Brandy, 2002; Martin & Clark, 2006).

The effect of the slope of the psychometric function was not significant on the number of words needed to obtain the starting level to begin the SRT assessment. The difference between the mean number of words for the starting level using words with steep, medium, and shallow slopes was less than 0.3 words. In the clinical setting this difference is not likely to effect the time required to administer the SRT test.

Primarily, this study revealed whether the slope of the psychometric function needs to be considered when selecting stimuli for SRT testing. There is not a need to set an ideal slope for words used to measure the SRT. A difference of approximately two words will not save a clinician meaningful time to perform the measurement. Second, this study revealed the greatest differences existed among the words with a shallow slope of the psychometric function. If slope must be considered for research or clinical purposes, then the division may be between words with a shallow slope and the remaining stimuli with a medium and steep slope.

Considering that the slope of the psychometric function had a significant effect on SRT measurement, further study may be warranted. This study was limited to participants with normal hearing between the ages of 18 and 30 years. The effect may be different among a subject group with hearing impairment or a pediatric subject group. Clinical testing of both populations would benefit from shorter administration time. This study was also limited to stimuli consisting of English spondaic words. Speech audiometry in areas where languages other than English are spoken often use stimuli consisting of trisyllabic words (Nissen et al., 2007). The slope of the psychometric function may have a greater effect on SRT measurement utilizing trisyllabic word lists.

In conclusion, while the slope of the psychometric function did have a statistical effect on the number of words needed to obtain the SRT and on the value of the SRT when compared to the PTA value, the effect of the slope of the psychometric function did not make a clinical difference. When creating word lists for the measurement of the SRT, researchers and clinicians can then focus on other factors of selection that adhere to the tenets of familiarity, phonetic dissimilarity, a normal sampling of English speech sounds, and homogeneity with respect to basic audibility.

References

American National Standards Institute. (1996). *American national standard: Specification for audiometers* (S3.6-1996). New York, NY: Acoustical Society of America.

American Speech-Language-Hearing Association. (1988). Guidelines for determining threshold level for speech. *ASHA, 30*(3), 85-89.

American Speech-Language-Hearing Association. (1990). Guidelines for screening for hearing impairments and middle-ear disorders. *ASHA, 32*(Supplement 2), 17-24.

Beattie, R. C., Svihovec, D. V., & Edgerton, B. J. (1975). Relative intelligibility of the CID spondees as presented via monitored live voice. *Journal of Speech and Hearing Disorders, 40*(1), 84-91.

Bilger, R. C., Matthies, M. L., Meyer, T. A., & Griffiths, S. K. (1998). Psychometric equivalence of recorded spondaic words as test items. *Journal of Speech, Language, and Hearing Research, 41*(3), 516-526.

Brandy, W. T. (2002). Speech Audiometry. In J. Katz (Ed.), *Handbook of clinical audiology* (5th ed., pp. 96-110). Philadelphia, PA: Lippincott Williams & Wilkins.

Cambron, N. K., Wilson, R. H., & Shanks, J. E. (1991). Spondaic word detection and recognition functions for female and male speakers. *Ear and Hearing, 12*(1), 64-70.

Carhart, R. (1946). Monitored live-voice as a test of auditory acuity. *Journal of the Acoustical Society of America, 17*(4), 339-349.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*(1-2), 62-71.

Chipman, S. (2003). *Psychometrically equivalent English spondaic words.* (Master's thesis), Brigham Young University, Provo.

Curry, E. T., & Cox, B. P. (1966). The relative intelligibility of spondees. *Journal of Auditory Research,* 6, 419-424.

Egan, J. P. (1979). Basic aspects of speech audiometry. *Ear, Nose, and Throat Journal, 58*(5), 190-193.

Fletcher, H., & Steinberg, J. C. (1930). Articulation testing methods. *Bell Telephone Systems Technical Publications, 1*(2), 17-21.

Hirsh, I. J. (1947). Clinical application of two Harvard auditory tests. *Journal of Speech Disorders, 12*(2), 151-158.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., & Benson, R. W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders, 17*(3), 321-337.

Hudgins, C. V., Hawkins, J. E., Karlin, J. E., & Stevens, S. S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope, 57*, 57-89.

Hughson, W., & Thompson, E. (1942). The correlation of hearing acuity for speech with discrete frequency audiograms. *Archives of Otolaryngology, 36*, 526-540.

MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*, *18*, 2331216514537722. doi:10.1177/2331216514537722

Martin, F. N., & Clark, J. G. (2006). *Introduction to audiology* (9th ed.). Boston: Allyn & Bacon.

Nissen, S. L., Harris, R. W., & Slade, K. B. (2007). Development of speech reception threshold materials for speakers of Taiwan Mandarin. *International Journal of Audiology, 46*(8), 449-458. doi:780762686 [pii]10.1080/14992020701361296

Ramkissoon, I. (2001). Speech recognition thresholds for multilingual populations. *Communication Disorders Quarterly, 22*(3), 158-162.

Schlauch, R. S., Anderson, E. S., & Micheyl, C. (2014). A demonstration of improved precision of word recognition scores. *Journal of Speech Language and Hearing Research, 57*(2), 543-555. doi:10.1044/2014_JSLHR-H-13-0017

Wilson, R. H., & Carter, A. S. (2001). Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *Journal of the American Academy of Audiology, 12*(1), 7-14.

Wilson, R. H., & Strouse, A. (1999). Psychometrically equivalent spondaic words spoken by a female speaker. *Journal of Speech, Language, and Hearing Research, 42*(6), 1336-1346.

Young, L. L., Dudley, B., & Gunter, M. B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech and Hearing Research 25*(4), 586-593.

APPENDIX A: Informed Consent

Consent to be a Research Subject

Introduction

This research study is being conducted by Richard Harris, PhD at Brigham Young University; Nujod Bakhsh, BS Communicative Sciences and Disorders, Communication Disorders graduate student at BYU; and Jessica Reese, BS Communication, Communication Disorders graduate student at BYU to evaluate a word list recorded using improved digital techniques. You were invited to participate because you are a native speaker of English with normal hearing.

Procedures

If you agree to participate in this research study, the following will occur:

- You will receive an audiometric exam, which will act as a screening process, to qualify you for further participation. The exam will consist of a hearing test where you will hear beeps and indicate whether or not you heard them, a tympanogram, and a test of acoustic reflex.
- Those that do not meet qualifications: participation in the study will cease.
- Those meeting the qualifications: you will listen to English words and repeat the words you hear.
- The total time commitment will be approximately 60 minutes.
- This will take place in a laboratory in the Taylor Building also known as the Comprehensive Clinic. The laboratory is located in room 110 of the Taylor Building on the BYU Campus.

Risks/Discomforts

There are minimal risks associated with this study. The researchers will be present at all times to make sure that you are not experiencing any problems during any portion of the study. If you indicate in any way that you do not want to participate, we will stop immediately.

Benefits

The primary benefit to you is finding out whether you have normal hearing or not throughout the course of the study. There may be benefits to society in general in that this study may result in more effective methods for hearing evaluations.

Confidentiality

Your participation will be confidential. The data will be stored in file cabinets within locked laboratories or offices in the Taylor building on the campus of Brigham Young University. Only the researchers will have access to the data. All names will be removed from research materials. Your name will never be used in association with this research.

Information will be kept for three years after the study is completed. The files will remain in a locked laboratory only accessible by the researchers. Internet data will be saved as a Microsoft Excel document with no subject identifiers. Participants will be identified only by number with no names or any other identifying referents.

Compensation

You will receive $20 for your participation; compensation will not be prorated. You will receive a free hearing exam and you will be provided a printed copy of your hearing evaluation. If you have questions about the results of your exam, you may consult with Dr. Richard Harris, PhD, CCC-A, a state licensed audiologist, at (801) 422-6460.

Participation
Participation in this research study is completely voluntary. You are free to decline to participate in this research study. In addition, you may withdraw your participation at any point without loss of compensation.

Questions about the Research
Please direct any further questions about the study to Richard Harris at (801) 422-6460 or richard_harris@byu.edu. You may also contact Nujod Bakhsh at (848) 459-1722 or Jessica Reese at (559) 905-6611, or through email at nujodprice@gmail.com or reesejessica328@gmail.com.

Questions about Your Rights as Research Participants
Should you have any questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent
I have read, understood, and received a printed copy of this entire consent document. I desire of my own free will to participate in this study.

Your Name (Printed):__________________ Signature: _________________________ Date:______

APPENDIX B: Spondaic Words Spoken by a Female Talker

Steep		Medium		Shallow	
Word	%/dB	Word	%/dB	Word	%/dB
broadway	18.7	aircraft	13.1	childhood	10.1
doorway	18.3	airplane	12.8	classroom	10.3
downtown	16.1	although	12.5	cupcake	11.0
eardrum	16.0	bathtub	12.6	duck pond	10.8
firetruck	22.2	birthday	13.6	farewell	8.6
hardware	17.7	broadcast	12.8	football	7.6
highchair	17.0	cowboy	13.7	framework	6.0
instead	16.2	daylight	14.2	grandson	9.9
midwest	17.0	doorman	14.4	headlight	9.7
mousetrap	17.6	drawbridge	13.3	hotdog	11.0
northwest	16.2	elsewhere	13.3	hothouse	8.5
nowhere	19.5	Friday	14.6	inkwell	9.3
outside	19.1	greyhound	14.8	lifetime	10.5
playground	16.5	horseshoe	13.8	outcome	9.0
railroad	17.6	iceberg	12.6	outline	10.2
somewhere	17.6	ice cream	12.5	outlook	9.6
sunlight	16.2	inside	15.0	padlock	10.3
Tuesday	17.1	maintain	14.6	platform	9.8
woodwork	20.3	mankind	14.1	schoolboy	9.6
		oatmeal	13.1	snowman	10.8
		popcorn	12.7	something	9.9
		sailboat	12.8	somewhat	8.7
		somehow	13.5	Sunday	8.0
		southeast	14.7	therefore	10.4
		suitcase	13.7		
		toothbrush	13.4		
		welfare	14.1		
		whitewash	13.2		
		wildlife	13.9		
		withdraw	14.6		
		without	12.3		
		workshop	13.9		

APPENDIX C: Annotated Bibliography

Abdulhaq, N. (2005). *Speech Perception Test for Jordanian Arabic Speaking Children.* (Doctoral Disseration), University of Florida, Gainesville, FL.
Purpose: The purpose of the study was to develop speech audiometric measures in Jordanian Arabic.
Methods: Children ages 6-9 years old, native to Jordan, listened to randomized lists of 250 monosyllabic words spoken by native Jordanian speakers. The children were to respond by verbally repeating the word. Words were presented in an ascending manner.
Results: Four 50-word lists were established as appropriate for speech audiometric measures of native speaking Jordanian children.
Relevance to current study: When comparing speech audiometric material from other languages, it was noted that steeper sloping words were found to be more homogenous with respect to audibility and comprised the final lists established by those researchers. Steeper sloping words are preferred by current developers of speech audiometric materials.

American National Standards Institute. (1996). *American national standard: Specification for audiometers* (S3.6-1996). New York, NY: Acoustical Society of America.
Purpose: Guidelines for professionals in audiology and speech language pathology to follow for use of audiometers in the clinical setting to determine hearing thresholds.
Summary: The specifications are for calibration of audiometers that will be used in testing with pure-tones and speech stimuli.
Relevance to current study: These specifications were followed in the creation of the stimulus recordings by Chipman. Chipman's recordings were used to present stimuli in this study.

American Speech-Language-Hearing Association. (1988). Guidelines for determining threshold level for speech. *ASHA, 30*(3), 85-89.
Purpose: Guidelines for professionals in audiology and speech language pathology to follow when performing speech audiometry. A standardarization of procedures would allow for easier and improved comparison of data between clinics and clinicians.
Summary: The guidelines give four parts to SRT testing: instruction to the client, familirazation of words to be used, descending threshold determination, and calucation of the threshold. Descending threshold determination can be done using a 2 dB-step method or a 5 dB-step method, each has its own formula for calucation of the threshold.
Relevance to current study: The current study followed these guidelines during testing of participants. The 2 dB-step method was chosen and exact adherence to the guidelines was executed.

American Speech-Language-Hearing Association. (1990). Guidelines for screening for hearing impairments and middle-ear disorders. *ASHA, 32*(Supplement 2), 17-24.
Purpose: Guidelines for professionals in audiology and speech language pathology to use as appropriate for hearing screens. The screening procedures are to aid in identifying individuals with hearing loss as well as individuals with disorders of the middle-ear.

Summary: The document gives guidelines for visual inspection of the ear, identification audiometry, and acoustic immitance measurements. There are also guidelines for the type of equipment to use, the type of personnel who may perform the screens, and a recommended protocol.
Relevance to current study: Acoustic immitance measurements were made for each subject in the current study. The ASHA guidelines were followed in making the measurements and in determining if the measurement fell within normal limits of hearing.

Beattie, R. C., Svihovec, D. V., & Edgerton, B. J. (1975). Relative intelligibility of the CID spondees as presented via monitored live voice. *Journal of Speech and Hearing Disorders, 40*(1), 84-91.
Purpose: The purpose of the study was to investigate the intellibility of the spondees used to measure SRT from the CID word list and to determine the words' homogeniety with respect to audibility. In addition, the researchers also wanted to answer two questions: "What is the shape of the articulation function for spondees?" and "What is theh relationship between speech detection threshold (SDT) and the sofest level at which 50% of the 36 words are identified (SRT)?"
Method: 75 subjects were used in the study, 10 male and 65 female, and were all of normal hearing. Each subject's SDT was first determined. The subjects then were measured for SRT via monitored live voice by trained and experienced male examiners.
Results: Homogeneity with respect to audibility was not maintained throughout all 36 words. 18 words were found to be more homogeneous with respect to audiblity and were suggested to be used exclusively for assessments of SRT. The researchers also found that the SRT may be estimated by adding 8 dB to an individual's SDT.
Relevance to current study: The article reaffirmed the use of speech audiometry in verifying results of pure tone audiometry. The current study is focused on improving the homogeniety with respect to audibility in terms of slope. The article demonstrated that long established lists for measuring SRT may not adhere to the tenet of homogeniety with respect ot audibility.

Beattie, R. C., & Warren, V. (1983). Slope characteristics of CID W-22 word functions in elderly hearing-impaired listeners. *Journal of Speech and Hearing Disorders, 48*(2), 119-127.
Purpose: The purpose of the study was to gather data regarding variables that affect the slope of monosyllabic word functions for subjects with sensorineural hearing loss. The study aimed to answer two questions: "Is there a relationship between the magnitude of hearing loss and the slope of the intelligibility function?" "Is there a relationship between the pure-tone audiometric configuration and the slope of the intelligibility function?"
Methods: Participants included 60 elderly individuals with sensorineural hearing loss. SRTs were measured using recordings of the CID W-22 word sets. The slopes of the individual functions were calculated using the 20% and 80% intelligibility points.
Results: The magnitude of hearing loss did not significantly affect the slope of the intelligibility function. Slope was found to be dependent on how it was measured. Steeper slopes were calculated using the 20-50% intelligibility points over the 50-80% segment for participants with flat and falling audiogram configurations. When slope is calculated by averaging the individual functions at fixed intensities, the resultant slope is

actually underestimated. To calculate a more accurate slope, the slopes of each individual function should be averaged.
Relevance to current study: The manner of calculation of slope can affected the outcome. As the current study is concerned with how slope affects the measurement of SRTs, it is necessary to use the best method of calculation

Bilger, R. C., Matthies, M. L., Meyer, T. A., & Griffiths, S. K. (1998). Psychometric equivalence of recorded spondaic words as test items. *Journal of Speech, Language, and Hearing Research, 41*(3), 516-526.
Purpose: The focus of this study was to create a homogeneous list of spondees that could be used to measure SRT, using a n-alternative, forced-choice procedure.
Methods: The researchers applied a logit function to the data reported by Young et al (1982). The CID W-1 words recorded by Tillman et al (1963) were digitized and then analyzed for RMS amplitude of each syllable, RMS amplitude for each spondee, duration of each syllable, and duration for each spondee. To obtain equal intelligibilty 30 of the 36 spondees were digitally multiplied. The psychometric function of the the 30 words was then calculated following testing using normal hearing individuals. The five participants listened to 10 computer-generated lists of 90 spondaic words, containing 3 repetitions of the 30 words being studied. The words were presented 10, 14, 18, 22, and 26 dB SPL.
Results: RMS amplitude and threshold were found to not correlate with word thresholds. Homongeniety with respect to threshold and slope was improved when the threshold of each word was adjusted to equal the mean of the threshold of the set of spondees.
Relevance to current study: Bilger et al reported that homogeniety of the slope of the psychometric function is important in adhering to the tenet of homogeniety with respect to audibility. It is not sufficient to calculate the mean slope of the set of words, but to calculate the slope for each word and then find the mean. The current study hypothesizes that a set of words with a similar steep slope will produce a more accurate SRT in shorter time.

Brandy, W. T. (2002). Speech Audiometry. In J. Katz (Ed.), *Handbook of clinical audiology* (5th ed., pp. 96-110). Philadelphia, PN: Lippincott Williams & Wilkins.
Purpose: The purpose of the chapter was to give an introduction into the practice of speech audiometry. A history of speech audiometry was given as well as a description of the various speech audiometric tests
Relevance to current study: In his discussion of the purpose of the SRT, Brandi described how the SRT is used to check the results of the PTA. A general consensus exists that a difference of ±6 dB indicates a good agreement between the SRT and PTA and a difference of ±7 to ±12 dB indicates a fair agreement. One of the purposes of this study was to examine the effect the psychometric funtion of slope had on the difference between the SRT and PTA.

Cambron, N. K., Wilson, R. H., & Shanks, J. E. (1991). Spondaic word detection and recognition functions for female and male speakers. *Ear and Hearing, 12*(1), 64-70.

Purpose: The purpose of the study was to compare the normative detection and recognition data of a, then, newly published disc of spondees spoken by a female talker to those of the original recording made by Hirsh et al (1952). Further, the slope of the psychometric function of each word in the female recording was measured and analyzed for interstimilus, intertrial, and intersubject variability.

Methods: The recordings made by Hirsh et al (1952) were digitized along with the recording of the female talker. The recordings were presented to 22 participants with normal hearing starting at 40 dB SPL with decreasing increments of 2 dB. The participant was to press a button when the word was detected and repeat the word. Testing stopped when no correct responses at two consecutive levels were given. To calculate the slope of the words spoken by the female talker, two experienced listeners with normal hearing were used. Each listened to randomized lists of the 36 words at 10 different levels of 2 dB increments, verbally repeating each word.

Results: Between the two recordings, no signficant differences were found. Both are equal in proficiency and efficacy for measurement of the SRT. The slopes of the psychometric function of the words spoken by the female talker were found to be less homogeneous than the threshold points. When variability increased in terms of interstimilus, intertrial, and intersubject data, the slope became more gradual.

Relevance to current study: The focus of the current study is on the effect the slope of the psychometric function has on SRT measurement. These researchers reported on the relationship between slope and the threshold of words; a set of words more shallow slope will have a threshold with a greater standard deviation, thus, less homogeneous.

Carhart, R. (1946). Monitored live-voice as a test of auditory acuity. *Journal of the Acoustical Society of America, 17*(4), 339-349.

Purpose: The purpose of the study was to determine and define the efficacy of the speech audiometry as part of an audiological evaluation. Further the study looked evaluated the PTA in terms of speech reception.

Methods: The data used to compare PTA and SRT measurement came from clinical reports of patients served. The population consisted of normal hearing individuals and individuals with hearing loss.

Results: No difference was found in SRT measurments made with sentences compared to those made with words. SRT measurements do provide an accurate description of hearing loss, though not the pattern of hearing loss.

Relevance to current study: Carhart demonstrated the appropriate placement of speech audiometry in an audiolical evaluation. The current study is concerned with improving the measurement of SRT as part of speech audiometry.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*(1-2), 62-71.

Purpose: Carhart described the need for speech audiometry, calling it the most useful audiological evaluation since the advent of pure tone testing. A brief history of speech audiometry was given followed by, what was at the time, a current definition for speech audiometry. Carhart explained the ariticulation function graph and how a displaced graph showed the level of hearing loss. Two criterion for words used in testing were laid

forth: phonemic balance and an equal distribution of phonemes when compared to the langauge used in testing.
Relevance to current study: Carhart's article provided valuable backround and historical information to speech audiometry. A proper understanding of what was accomplished before was necessary to design the current study.

Chipman, S. (2003). *Psychometrically equivalent English spondaic words.* (Master's thesis), Brigham Young University, Provo.
Purpose: The purpose of the study was to develop a recorded list of steeply sloping words to be used in speech audiometry.
Methods: Corpora of American English words were used to select spondaic words that are most frequently among the American population. Added to this selection were the 36 CID Test W-1 words and the 11 words from the children's picture board for a total of 98 words. 20 participants with normal hearing were presented the recorded words at 13 predetermined levels. The participants verbally repeated the words. The psychometric functions of slope was calculated for each word.
Results: 33 spondees were determined to have a steep slope; at 50% the mean slope for words recorded by a male talker was 16.2%/dB and for words recorded by a female talker 15.2%/dB. The words were recorded and digitally adjusted so the mean threshold was equal to the mean pure-tone average for the subjects.
Relevance to current study: The 98 words orinally selected by Chipman, with the results of the calculations for the pyschometric funciton of slope, were used the in current study to determine the effect slope has the on the measurement of SRT.

Curry, E. T., & Cox, B. P. (1966). The relative intelligibility of spondees. *Journal of Auditory Research, 6*(419), 424.
Purpose: The purpose of this study was to determine the relative intelligibility of the CID W-1 spondees used in speech audiometry. The researchers wanted to determine the range of intelligibility as well as determine if there was a subgroup of words within the list that had a more homogeneous range of intelligibility and when used exclusively yield a more accurate SRT.
Methods: Recordings of the CID W-1 word lists were used. 4 randomizations of the 36 words were used. 50 participants with normal hearing were presented the recordings with the instruction to repeat the word. Presentation started at -10 dB HL and increased incrementally by 2 dB until 100% correct identification.
Results: The calculated range of intelligibility was 8 dB. The researchers felt this range was too large. 27 of the 36 words fell within a range ±2 dB. The researchers proposed the use of those 27 words would produce greater homogeniety among test items.
Relevance to current study: The current study is concerned with improving homogeniety with respect to audibilty of speech audiometry materials. This article demonstrated that the list of words commonly used in speech audiometry does not contain ideal homogeniety with respect to audibility.

Egan, J. P. (1979). Basic aspects of speech audiometry. *Ear, Nose, and Throat Journal, 58*(5), 190-193.

Purpose: The purpose of the article was to discuss the need of speech audiometry in addition to pure-tone testing to determine the true nature of a individual's hearing loss. Egan also discusses the history and the then current practices of measuring the speech recognition threshold and the word-discrimination score.
Relevance to current study: The current study's focus is the speech recognition threshold. This article gave insight into the history of speech audiometry and its appropriate place in an audiological evaluation.

Fletcher, H., & Steinberg, J. C. (1930). Articulation testing methods. *Bell Telephone Systems Technical Publications, 1*(2), 17-21.
Purpose: This article reported on the work by Fletcher and Steinberg to standardize tests for speech audiometry. Two primary factors in developing the word lists were that the words be representative of the language being tested and that the words must be suitable for making tests. The authors discovered that randomly selecting words resulted in test with words that were difficult to pronounce. Fletcher and Steinberg also discuss the various methods of testing and subsequent data analysis equations.
Relevance to current study: The work of Fletcher and Steinberg was among the first in producing modern materials for speech audiometry. For the current study to proceed, an accurate knowledge of the history of the field was necessary.

Hirsh, I. J. (1947). Clinical application of two Harvard auditory tests. *Journal of Speech Disorders, 12*(2), 151-158.
Purpose: The purpose of the article to was to describe the history of speech audiometry and the types of word lists developed as speech audiometry materials. Hirsh was concerned with how these practices could be transferred to the clinical setting, conceding the disadvantages the then current clinics and schools had compared to military facilities where many advances in speech audiometry were made. Hirsh proposed the use of the PAL Audiometric Tests No 9 and 12 in clinical audiometry.
Relevance to current study: Hirsh was a principal figure in the history of speech audiometry. His work directed the path of that field. A proper understanding of his work was necessary to proceed with the current study of speech recognition threshold.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., & Benson, R. W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders, 17*(3), 321-337.
Purpose: The purpose of the article was to describe modifications to recorded speech audiometry materials. The article focused on new tests that improved measurements for hearing loss for speech and word discrimination, specifically the CID Auditory Test W-1, the CID Auditory Test W-2, and the CID Audiotry Test W-22. All three test were performed in laboratorys studies and the results analized.
Methods: 6 experienced listeners participated in the analysis of the W-1 and W-2 tests, measurements of speech recognition. Both tests were administer per perscribed means. The same 36 words were used. The W-1 test presented all 36 words at 6 different levels, the order of which was determined randomly. The W-2 test presented the words 3 at a time at 3 dB descending increments. The W-22 test, a measurement of word discrimination, consisted of 200 phonetically balanced words divided into 4 sublists. 3

groups of five listeners participated in the study. Words were presented in 10 dB increments from 20 dB to 70 dB.
Results: All three test provided were found to have higher intelligibility and steeper psychometric functions than older tests. It was found that the use of the W-1 test resulted in a more accurate threshold than the W-2 test method.
Relevance to current study: The article provided important historical information needed to understand the development of word lists for speech audiometry. The current study's focus is also concerned with improvement of speech audiometry materials.

Hudgins, C. V., Hawkins, J. E., Karlin, J. E., & Stevens, S. S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *The Laryngoscope, 57*, 57-89.
Purpose: The purpose of the article was to describe the development of the PAL Audiometry Test No. 9 and Audiometry Test No. 12. The development of the test aimed to follow the four characteristics deemed essential to appropriate speech audiometry materials: familiarity, phonetic dissimilarity, normal sample of English speech sounds, and homogeneity with respect to audibility. The authors described the two tests and the test reliability for both. Each test would two-thirds of the time produce scores with 2.8 dB of true threshold.
Relevance to current study: The four essential characteristics for speech audiometry material has guided the field since their publication. The current study is focusing on the tenet of homogeneity with respect to audibility and how that may be improved through the use of slope as a guide for material selection.

Hughson, W., & Thompson, E. (1942). The correlation of hearing acuity for speech with discrete frequency audiograms. *Archives of Otolaryngology, 36*, 526-540.
Purpose: The purpose of the study was to establish a normal threshold for hearing speech and to relate that threshold to a normal audiogram. In addition, the study aimed to correlate the audiogram of hearing loss to the results of speech audiometry in a simple manner so that the one may easily and accurately predict the other.
Methods: 56 healthy, young adults participated in the study. Each participated in pure-tone testing and speech audiometry in the manner prescribed by Fletcher. The results were compared to those obtain from 43 individuals with varying degrees and types of hearing loss.
Results: A correlation between the results of audiometric test and speech audiometry was found. When the results are reduced to the audiometric percentage of loss and the percentage of loss of speech, the two more easily predict each other.
Relevance to current study: Hughson and Thompson produced evidence of the correlation between an audiogram and the results of speech audiometry. This evidence has been important in the continuation of research in and development of speech audiometry. The current study is focused on the speech recognition threshold and a proper understading of the work by Hughson and Thompson was necessary to proceed.

Kumar, S. B. R., & Mohanty, P. (2012). Speech recognition performance of adults: A proposal for a battery for Telugu. *Theory and Practice in Language Studies, 2*(2), 193-204. doi:10.4304/tpls.2.2.193-204

Purpose: The purpose of the study was to develop speech recognition performance assessment materials in the Telugu language.
Methods: Four lists of bisyllabic words were developed for the study. The lists were used to assess speech recognition performance of 3 groups of participants, each group representing and different region of the county where Telugu is spoken. All participants were of normal hearing and between the ages 18-25 years.
Results: All four lists developed were found to be reliable and valid in assessing speech recognition performance.
Relevance to the current study: The researchers detail their efforts in developing appropriate speech materials for speech audiometry. Kumar and Mohanty's focus was on using words that were familiar to the intended population and that also conveyed meaning. In developing the word sets, the researchers kept in line with several other findings regarding the mean slope of the function of the words.

MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*, *18*, 2331216514537722. doi:10.1177/2331216514537722
Purpose: This meta-analysis of 139 studies examined reported psychometric function data to quantify range of slope and to identify how slope is effected with changes in listening conditions.
Methods: 885 psychometric functions were examined with a logistic equation to calculate the slope of each signal. A variety of signals were included: valid sentences, invalid sentences, words, digits, continuous speech, and short tokens.
Results: Slope of the psychometric function varied from 1%/dB to 44%/dB. Listening conditions did have an effect on the calculated slope. The authors suggested care for researchers in speech audiometry to consider the stimulus and presentation conditions, as threshold and intelligibility will be affected.
Relevance to current study: The authors clearly define slope of the psychometric function and its role in intelligibility of stimulus in speech research. This study's focus is on the effect of slope of the psychometric function on a specific test of speech recognition.

Martin, F. N., & Clark, J. G. (2006). *Introduction to audiology* (9th ed.). Boston: Allyn & Bacon.
Purpose: The purpose of this book was to give an introduction into the field of audiology. Two chapters focused on speech audiometry and gave history of speech audiometry as well as a description of the various speech audiometric tests
Relevance to current study: The authors discussed how the use of the SRT to check the results of the PTA. They state the SRT and PTA should have an agreement of between ±5 to ±10 dB for a generally flat audiogram. One of the purposes of this study was to examine the effect the psychometric funtion of slope had on the difference between the SRT and PTA.

Meyer, T. A., & Bilger, R. C. (1997). Effect of set size and method on speech reception thresholds in noise. *Ear and Hearing, 18*(3), 202-209.
Purpose: The study looked at the effect set size has the results of SRT testing.

Methods: SRTs were measured using two different methods. One method followed the guidelines for obtaining SRTs put forth by ASHA in 1988. The second method defined the threshold in terms of d' = 1.00. All testing was performed with white noise at 80 dB HL and with four sets of words containing 2, 4, 8, and 16 words respectively. 12 female subjects, with self-reported normal hearing, participated in the study.
Results: Using the ASHA 1988 method, the speech to noise ratio threshold was found to be dependent on set size. The speech to noise ratio threshold was found to be independent of set size when the threshold was defined in terms of d' = 1.00.
Relevance to the current study: The authors caution against the use of using smaller sets in the clinical setting while employing the ASHA 1988 method. The resultant SRTs may differ up to 2 to 3 dB from SRTs obtained using a full set of words.

Nissen, S. L., Harris, R. W., & Slade, K. B. (2007). Development of speech reception threshold materials for talkers of Taiwan Mandarin. *International Journal of Audiology, 46*(8), 449-458. doi:780762686 [pii]10.1080/14992020701361296
Purpose: The purpose of this study was to develop appropriate materials for the measurment of SRTs in the Taiwan Mandarin dialect.
Methods: 20 native speakers of Taiwan Mandarin with normal hearing participated in the study. 130 familiar trisyllabic words in the Taiwan Mandarin dialect were chosen for the study. 6 native speakers of Taiwan Mandarin recorded the words. The words and recording were evaluate and 89 words were selected to use in the testing. The words were presented randomly, starting a -10 dB and increasing 2dB increments until 16 dB. The participants were to repeat back the word they heard.
Results: A final list was developed of 28 words that adhered to the four tenets of speech audiometry.
Relevance to current study: The researchers in this article documented a steep slope as desirable to achieve homogeneity with respect to audibility. The current study is focused on the effect slope has on SRT measurement, theorizing steeper slopes are statistically superior to shallow and medium sloping words.

Punch, J. L., & Howard, M. T. (1985). Spondee recognition threshold as a function of set size. *Journal of Speech and Hearing Disorders, 50*(2), 120-125.
Purpose: The study focused on the effect reducing the number of stimuli in speech audiometric testing has on the magnitude and accuracy of SRT results.
Methods: The 36-word CID-W-1 list was used to create six separate lists of stimuli items. One list was comprised of the entire 36 words, the other five had reduced numbers, 27, 18, 9, 6, and 3. Words for each of these lists were selected based on homogeneity in level and slope characteristics. Six groups of 12 participants provided SRT results for all six stimuli sets, using the 2-dB step method.
Results: Recognition of the presented spondees improved systematically with the decrease in set size.
Relevance to current study: The sublists were carefully developed to maintain homogeneity with respect to audibility. This was accomplished by selecting words for each sublist based on a median slope for that set. To adhere to the tenet of homogeneity with respect to audibility, stimuli lists must developed around a similar slope.

Ramkissoon, I. (2001). Speech recognition thresholds for multilingual populations. *Communication Disorders Quarterly, 22*(3), 158-162.

Purpose: The purpose of the article was to describe the development of speech audiometry and to emphasize the need of proper materials for non-native English speaking individuals. The article argues that the tenet of familiarity is the most important of the four tenets of speech audiometry and that it is not currently being adhered to when testing multilingual clients.

Relevance to the current study: The article points out the necessity of a steep rise in intelligibility to adhere to the tenet of homogeneity with respect to audibility.

Ramkissoon, I., Proctor, A., Lansing, C. R., & Bilger, R. C. (2002). Digit speech recognition thresholds (SRT) for non-native speakers of English. *American Journal of Audiology, 11*(1), 23-28. doi:10.1044/1059-0889(2002/005)

Purpose: To establish a test of speech recognition in English, for non-native English speaking individuals using digit stimuli.

Methods: Participants were 12 non-native English speaking individuals and 12 native English speaking individuals. The stimuli consisted of non-duplicated digit pairs using the digits 1-9, excluding & as it is a disyllable. Each subject participated in SRT testing using the digit pairs and stimuli from the CID W-1 list.

Results: Digit pairs were as effective at obtain SRT results as the CID W-1 lists for native English speaking participants. For the non-native English speaking participants, SRT results using digit pair stimuli were more accurate than results from the CID W-1 lists.

Relevance to current study: This study highlights the need of familiarity of SRT test materials to obtain accurate results. Familiarity is one of the four tenets of proper SRT testing.

Schlauch, R. S., Anderson, E. S., & Micheyl, C. (2014). A demonstration of improved precision of word recognition scores. *Journal of Speech Language and Hearing Research, 57*(2), 543-555. doi:10.1044/2014_JSLHR-H-13-0017

Purpose: The purpose of the study was to improve the accuracy of WRS assessment by increasing the length of the word lists.

Methods: Participants include 24 individuals with normal hearing. PTAs were obtained for all subjects. The WRS was measured using the full list and a half list of the Northwestern University Test No. 6.

Results: Results from use of the full list were more sensitive to identifying the change in hearing than results from use of half the list.

Relevance to current study: Schlauch, Anderson, and Micheyl define clearly the word recognition score assessment, how it should be obtained, and how it is measured in the clinical setting.

Wilson, R. H., & Carter, A. S. (2001). Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *Journal of the American Academy of Audiology, 12*(1), 7-14.

Purpose: The purpose of this study was to establish the psychometric function of the 36 CID words commonly used to measure SRT.
Methods: The study was divided into two parts. For both parst 20 participants with normal hearing were used. In the first part of the study, each participant gave SRT measurements using recordings by Hirsh and recordings of a female talker. The psychometric function of the 36 words were established using a recorded male talker and a recorded female talker. The recordings from the female talker were then digitally adjusted to be equilavent in intelligibility. For the second part, subjects' SRTs were measured using the adjusted female recording.
Results: The results of the study indicate that when words are equated to intelligibility rather than 0 vu, the threshold of intelligibility is reduced and the words become more homogeneous with respect to audibility.
Relevance to current study: This article demonstrates the room for improvement in standard lists to measure SRT, particularly in regard to homogeneity with respect to audibility. The article reports steeper slopes calculated at the 50% correct points compared to slopes calculated from the 20 to 80% correct points. As the current study's focus is on the slope of the psychometric function, consideration of point of calculation should be taken and noted.

Wilson, R. H., & Strouse, A. (1999). Psychometrically equivalent spondaic words spoken by a female speaker. *Journal of Speech, Language, and Hearing Research, 42*(6), 1336-1346.
Purpose: This study investigated whether the homogeneity of individual word functions was directly related to the mean psychometric functions of the word sets used in the study.
Methods: The psychometric function of each word in the PB-50 lists 8 and 9 recorded by Rush Hughes and the CID W-22 lists 3 and 4 recorded by Hirsh were established by participant listening and response. The words were presented in randomized lists at predetermined presentation levels. The verbal responses were recorded in a spreadsheet used for data analysis.
Results: The functions for the individual words from the W-22 lists were more homogeneous than those from the PB-50 lists. The mean function of the W-22 lists had a steeper slope than the slope from the mean function of the PB-50 lists.
Relevance to current study: Word lists with a steeper slope are more homogeneous than word lists with shallower slopes. To comply with the tenet of homogeneity with respect to audibility, steeper sloping word lists should be employed.

Young, L. L., Dudley, B., & Gunter, M. B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech and Hearing Research 25*(4), 586-593.
Purpose: This study investigated the 36 words from the CID list commonly used to measure SRTs to determine the extent of homogeneity amongst the words; homogeneity based on the sound-pressure level needed to yield a 50% correct response and homogeneity based on the slope of the psychometric funtion of each word.
Method: The study was accomplished through two experiments. The first experiment used 20 adult subjects with normal hearing. The 36 CID words were randomized into 10 different lists. The words were presented to the subjects at 10 established decibel levels. The subjects were to repeat back the word presented. The second experiment was to

determine if a significant difference existed in SRTs measured using the full 36 CID list compared to SRTs measured using the 15 words determined to be homogenous. 24 subjects with normal hearing participated in this experiment. Each subject was assessed for SRT using both word lists, randomly given.

Results: In the first experiment, it was found that based on sound-pressure level measurements and the slope of each word, the 36 CID words were not homogeous. Words were considered homogeneous if they fell within ±1 SD of the mean value of the slope of the pschometric function. This resulted in a list of 15 words that was considered to be homogeneous. The results of the second experiment found that both the full 36 CID list and the list of 15 homogeneous words yielded the same SRT.