2018-06-01

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

Nujod Ali Bakhsh
Brigham Young University

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

Nujod Ali M Bakhsh

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 Male Talker

Nujod Ali M Bakhsh
Department of Communication Disorders, BYU
Master of Science

Speech audiometry is the aspect of audiology that provides critical information on how individuals hear one of the most important sounds of daily life: speech. The speech recognition threshold (SRT) is a measure of speech audiometry that is widely used to provide information on an individual's capacity to hear speech. Over time, researchers and clinicians have worked to improve the SRT by developing and modifying a variety of word lists to be used during testing. Eventually, spondaic words were selected as the best stimuli for the SRT. The spondaic words had to meet four criteria: familiarity, phonetic dissimilarity, normal sampling of English sounds, and homogeneity with respect to audibility. This study examined the aspect of homogeneity with regard to slope of the psychometric function. Specifically, whether slope of the psychometric function had an effect on the number of words used to obtain the SRT, and thus reduce test time, as well as whether slope had an effect on the relationship between the SRT and the pure-tone average (PTA). It was hypothesized that words with a steep slope would significantly reduce test time and yield a close SRT-PTA agreement. Three word lists (steep, medium, and shallow sloping words), all recorded by a male talker, were used to obtain the SRT on 40 participants (ages 18-30 years). Statistical analysis showed significant differences in the number of words to obtain the SRT and the SRT-PTA agreement. However, when the differences were examined from a clinical perspective, the results were negligible. When compared with words with medium and steep slopes, words with shallow slope required an average of four extra words to obtain the SRT, which does not result in a meaningful reduction in test time. For clinical purposes, it appears that the slope of the psychometric function does not need to be taken into consideration for the SRT. Clinicians may use a variety of words as long as they meet the original four criteria for selection of spondees.

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

ACKNOWLEDGMENTS

I am grateful to my family and my husband, Cameron, for supporting and encouraging me throughout my education. I am grateful to Dr. Harris for guiding me and providing insight as I completed this project. Additionally, thank you to the faculty of Brigham Young University's Department of Communication Disorders. I would not have succeeded without their education and guidance. Thank you to all.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS iv

LIST OF APPENDICES v

DESCRIPTION OF THESIS STRUCTURE vi

Introduction 1

Method 7

Participants 7

Materials 8

Procedure 8

Results 9

Table 1 11

Discussion 12

References 16

LIST OF APPENDICES

Appendix Page

A: Annotated Bibliography 19

B: Informed Consent 27

C: List of Words Spoken by Male Speaker by Slope 29

DESCRIPTION OF THESIS STRUCTURE

This thesis, *The Effect of the Slope of the Psychometric Function on the Measurement of Speech Recognition Threshold Using a Male 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 audiology. An annotated bibliography is presented in Appendix A.

Introduction

Audiological testing serves to determine the presence, type, and degree of hearing loss in individuals. An audiologist's evaluation may include the following components: otoscopic examination, pure-tone audiometry, air conduction testing, bone conduction testing, tympanometry, acoustic reflexes, and speech audiometry. Individually, these tests provide unique information about different aspects of an individual's hearing system. Together, they provide a complete analysis of the type and degree of hearing loss.

Pure-tone audiometry, a measure of an individual's hearing sensitivity across different frequencies and intensities, is used in both audiological screenings and evaluations (American Speech-Language-Hearing Association, 1978). While pure-tone audiometry data are valuable, the test does not reflect what humans hear in everyday life, or more specifically, what sounds are important to humans. In contrast, speech audiometry provides valuable information that captures the impact of hearing loss on speech comprehension, and consequently, quality of life (Carhart, 1952; Hirsh et al., 1952; Ramkissoon, 2001). In fact, Carhart stated that "speech audiometry is … the most useful advance in hearing testing since the introduction of pure-tone audiometry" (Carhart, 1951, p. 62).

Speech audiometry includes three types of measures: the speech recognition threshold (SRT), the speech detection threshold (SDT), and the word recognition score. The SRT is defined as the level at which an individual can correctly repeat 50% of the presented speech stimuli (ASHA, 1988; Beattie, Svihovec, & Edgerton, 1975; Brandy, 2002; Hanekom, Soer, & Pottas, 2015). The SDT is the level at which an individual can correctly identify the presence of speech 50% of the time (ASHA, 1988). Word recognition scores are a measure of how well an individual can identify words at a specific suprathreshold level (Brandy, 2002). This investigation will focus on the SRT component of speech audiometry.

The SRT has several purposes. First, it confirms the accuracy of the pure-tone average (PTA), measured at 500, 1000, and 2000 Hz (Brandy, 2002). The SRT should be no more than 6 dB higher or lower than the PTA. If the difference is larger, there may be need for further evaluation (Hanekom et al., 2015; Olsen & Matkin, 1979). Second, the SRT indicates how sensitive hearing is to speech. Third, it provides a baseline for suprathreshold speech recognition tests, such as word recognition scores (Brandy, 2002). The SRT also provides information about whether an individual is benefitting from hearing aids (Beattie & Warren, 1983; Carhart, 1952).

In 1988, the ASHA established guidelines for determining the SRT. These guidelines provide information regarding calibration of audiometers, test environment, test materials, method of presentation, response mode, recording results and masking of the nontest ear (ASHA, 1988). The SRT testing begins with a familiarization phase and the SRT is then measured in a descending method. A full explanation of methodology is provided in the Methods section. ASHA recommends the use of recorded testing materials rather than monitored live-voice materials (ASHA, 1988). As a result, this investigation used recorded material to reduce variation during testing.

When determining the SRT, spondaic words, which are two syllable words with equal stress, are recommended as the speech stimuli (ASHA, 1988; Hanekom et al., 2015; Ramkissoon, 2001). There are four basic criteria for selection of spondaic words (Bilger, Matthies, Meyer, & Griffiths, 1998; Olsen & Matkin, 1979; Ramkissoon, 2001). First, the spondaic words must be familiar to the listener because the SRT is a test of hearing, rather than of intelligence or vocabulary. Second, the words must be phonetically dissimilar, meaning none of the words should sound similar or rhyme. Third, the words must have a normal sampling of English sounds; however, this criterion is often considered of lesser importance, as long as there is a good representation of different sounds. Lastly, the spondaic words must be homogenous with respect to basic audibility. This is a measure of the ease at which words can be understood at the same intensity level (Hanekom et al., 2015; Hudgins, Hawkins, Karlin, & Stevens, 1947; Olsen & Matkin, 1979). Spondees are highly audible due to having equal stress on both syllables, which provides the listener with auditory cues (Hudgins et al., 1947). The slope of the psychometric function, a measure of how well an individual hears the speech stimuli as the presentation level increases, for spondees tends to be steep due to the high audibility (Carhart, 1951; Hudgins et al., 1947; Wilson & Carter, 2001). Homogeneity, both in terms of audibility and slope, is considered important as it allows for precision during SRT testing as well as decreases the length of time needed to determine the SRT (Olsen & Matkin, 1979; Young, Dudley, & Gunter, 1982).

As various researchers and audiologists studied and developed standardized tests for speech audiometry, the recommendation for using spondaic words with steep slopes of the psychometric function emerged. Standardization was necessary for audiologists to compare speech audiometry testing results between clinics (Hirsh et al., 1952). Early forms of speech audiometry testing were available in the 19th century, when individuals were asked to repeat words or sentences that were whispered or spoken at varying distances. In 1929, Fletcher, Steinberg, French, and others at the Bell Telephone Laboratory began the process of standardizing speech audiometry and laid the foundation for subsequent work (Brandy, 2002; Carhart, 1951; Olsen & Matkin, 1979). Fletcher also developed the Western Electric 4-A Audiometer which was used for school hearing screenings involving speech audiometry. The audiometer utilized phonograph recordings of pairs of digits at different intensity levels. The digits were presented to the listener, who was asked to write down the digits (Olsen & Matkin, 1979). In 1947, Hudgins et al. reviewed the development of speech stimuli for determining SRTs. They suggested the use of spondaic words because of their high audibility, which provided the best results. They created a list of 84 spondaic words which adhered to the four criteria for spondees. The words were then recorded on a phonographic disc by a male speaker. This test was known was the Psycho-Acoustic Laboratory (PAL) No. 9. On this test, each set of six spondees was 4 dB lower than the previous set. A similar test, PAL Test No. 14 was recorded with the same words but with a fixed intensity level (Brandy, 2002; Olsen & Matkin, 1979).

In 1952, Hirsh et al. made modifications to PAL No. 9 and No. 14 based on the idea that some of the words were unfamiliar. The list of 84 words was reduced to 36 familiar words (Hirsh et al., 1952; Olsen & Matkin, 1979). When the 36 words were presented to different listeners, the results showed that some words were easier or harder than others. As a result, the easier or harder words were adjusted by ± 2 dB to allow for homogeneity in terms audibility. This new list was recorded at a constant level (minus the adjustments) by a male talker at the Central Institute for the Deaf (CID). The recording became known as the CID-W-1 test (Hirsh et al., 1952; Olsen & Matkin, 1979). An alternate version of this test, the CID-W-2, was recorded with each set of three spondees being 3 dB lower than the previous set (Brandy, 2002; Olsen & Matkin, 1979). Hirsh et al.'s list of 36 spondaic words is currently widely used and recommended by ASHA (1988).

Researchers and audiologists continued to examine the list of 36 spondaic words and found discrepancies related to homogeneity in regard to audibility. Curry and Cox (1966) studied the CID W-1 list of 36 spondaic words and determined that the range of audibility (8 dB) between the spondaic words was too large and negatively impacted homogeneity in terms of audibility. As a result, they recommended dropping nine of the 36 words to form a more homogenous list of 27 words with a range of ± 2 dB. Young et al. (1982) also examined the CID W-1 word list and determined that it was not homogenous, especially when only one section of the list was presented. They claimed the words were not equal in difficulty, and thus not homogenous in terms of audibility. They reduced the 36 words to 15 words and compared the SRTs obtained from each list. They determined that there was no significant difference between the SRTs using the 15-word list compared to the 36-word list. They suggested that fewer than 36 words can be used to accurately determine the SRT as well as reduce memory and learner fatigue effects (Young et al., 1982). Other researchers also attempted to produce sublists with improved audibility with respect to homogeneity (Beattie et al., 1975; Curry & Cox, 1966; Meyer & Bilger, 1997). Bilger et al. (1998) believed homogeneity should be considered both in terms of audibility and slope of the psychometric function. Bilger et al. examined the work of previous researchers (Hirsh et al., 1952; Young et al., 1982) and found that individual words were homogenous in terms of audibility and slope, but the overall word lists were not homogenous in regard to slope. As a result, 30 spondaic words with similar slopes were selected and adjusted to improve homogeneity in terms of slope (Bilger et al., 1998). Wilson and Strouse (1999) also examined the CID W-1 words in terms of the slope of the psychometric function. They attempted to improve homogeneity by equating the words at 0 VU and then equating them to intelligibility and calculating the slopes. They found that with digital adjustments, the words were homogenous with respect to both slope and audibility (Wilson & Strouse, 1999).

While the homogeneity of the CID W-1 spondees was debated, another issue arose regarding the validity of the words. The English language had evolved since the words were first selected, and some were no longer considered familiar. Chipman (2003) reexamined the CID W-1 words using two modern corpora of English language to measure their frequency of usage in the English language. Chipman determined that 14 of the original 36 words were no longer familiar and proposed a list of 98 familiar, homogenous words with steep slopes of the psychometric function. The 33 words with the highest slope were then selected and adjusted to further improve homogeneity.

With all four criteria for spondaic words met, one final factor to consider is slope of the psychometric function of the words. The consensus is that steeply sloping words allow for the shortest testing time, precise scores, and reduce listener fatigue (Chipman, 2003; Harris et al., 2007; Ramkissoon, 2001; Wilson & Carter, 2001). Researchers have shown the CID W-1 words to have an average slope ranging between 8-12%/dB (Beattie et al., 1975; Hirsh et al., 1952; Hudgins et al., 1947; Young et al., 1982). Young et al. determined that the slope of each individual word should be ± 1 SD of the mean to be considered homogenous, which resulted in words that were steeply sloping (12.5%/dB and above) and words which were not steeply sloping. Chipman (2003) further broke down words into steep, medium, and shallow sloping words. To date, no researchers have investigated the effect of words with different slopes of the psychometric function (shallow, medium, steep) on obtaining the SRT using a male talker. This study aims to do so by investigating two main questions. First, does the slope of the psychometric function affect how many words it takes to accurately establish the SRT? Second, does the slope of the psychometric function affect the SRT-PTA relationship? It is believed that words with steep slopes will yield a more reliable and time-effective SRT when compared to words with shallow and medium slopes. Steeply sloping words will also produce higher SRT-PTA agreement.

Method

Participants

A total of 40 (20 male, 20 female) native English speakers with normal hearing took part in this study. Participants were tested bilaterally and data from each ear were regarded as individual data sources; thus presented and analyzed data is based on 80 subjects. The participants ranged from age 18 to 30 years (*M*=24.8). This age group was chosen to reduce the likelihood of hearing impairment or phonemic regression based on age. All participants possessed normal hearing with pure-tone air conduction thresholds of ≤ 15 dB HL at all frequencies ranging from 125-8000 Hz, including mid-octaves. A Grason-Stadler model 1761 clinical audiometer was used. The participants also had normal type A tympanograms and an ipsilateral acoustic reflex at 1000 Hz. A Grason-Stadler GSI TympStar Middle Ear Analyzer was used. All participants reported no history of hearing loss. Participants were recruited through flyers, e-mail, social media, and word of mouth. An informed consent form, approved by the Brigham Young University Institutional Review Board, was read and signed by each subject. A copy of the consent form can be found in Appendix B.

There were 22 additional subjects whose data were not used. Eleven subjects (5 male, 6 female) had pure-tone air conduction thresholds >15 dB HL and as a result, did not complete the remainder of the study. For the other 11 subjects (3 male, 8 female), an SRT could not be established for at least one word list due to the individual continually providing correct responses at the audiometer's lowest presentation level (-10 dB HL). As a result, these participants did not miss the 5/6 words needed to discontinue SRT testing.

Materials

The current study used Chipman's (2003) spondaic wordlist recorded by a male talker to compile three lists of words with shallow (7.4-10.8%/dB), medium (12-14.8 %/dB), and steep (16-20.5 %/dB) slopes. The lists were divided into 23 shallow words, 34 medium words, and 24 steep words. The lists of words can be found in Appendix C. The recordings of the words were made at Brigham Young University by Chipman. Further detail regarding the recordings can be found in Chipman.

Procedure

After normal hearing was established through audiometric testing, tympanometry, and acoustic reflex testing, participants filled out an informed consent form to proceed with the study. Participants were then familiarized with the list of spondaic words (presented by a male speaker and a female speaker) while wearing headphones and listening to the words at 50 dB HL as well as reading the words on a paper. After familiarization, participants were asked if they were familiar with all the words. If a participant was unfamiliar with a word, the researcher explained the word and asked the participant if he or she would be comfortable with having that word used in testing. All subjects stated they would be comfortable with that word.

After familiarization, the participants were oriented to the nature of the task and asked to respond by repeating the word they heard. The instructions also encouraged the subject to guess if they were unsure. Prior to the test phase, the ear to be tested first was determined randomly for each subject. The slope and the talker gender were also previously determined randomly. Custom wavPlayer software developed at Brigham Young University was used for playback and randomization of the words. The SRT was measured six times in each ear. The total number of SRTs measured was 12 for each participant. Each SRT took approximately two minutes to measure. The entire procedure took approximately 30 minutes per participant. Each ear was tested with a single headphone to minimize variability.

ASHA (1988) details two methods for obtaining the SRT: a 5 dB method and a 2 dB method. This investigation used the 2 dB method. The SRT was then determined by presenting the stimuli in a descending method, starting at 30 dB HL. One spondaic word was presented, and if the subject's response was correct, then the presentation level was decreased by 10 dB. This process continued until the subject responded incorrectly. If the subject did not respond correctly to the first spondaic word presented, the spondees were increased in 20 dB increments. Once the subject missed one spondee at a certain level, a second spondee was presented at that same level. When two consecutive spondees were missed at the same level, the presentation level was then increased by 10 dB and was defined as the starting level for testing. Two spondaic words were then presented at the starting level and then in 2 dB decrements. This process was discontinued when subject missed five out of the last six words presented. The SRT was calculated using the Spearman-Kärber method with a correction factor of 1 (ASHA, 1988). The simplified formula is SRT = starting level − #correct + correction factor

Results

The primary purpose of the current study was to examine the effect of slope of the psychometric function on the number of words to obtain the SRT, thus making SRT testing more efficient. The effects of the listener's test ear on the number of words to obtain the SRT was also investigated. The following dependent variables were also examined: the actual measured SRT, the difference between the SRT and the PTA, and the number of words needed to obtain the starting level. The data were analyzed using SAS v 9.4.

Table 1 presents the descriptive statistic from the analysis. The results show medium sloping words required the lowest average number of words to obtain the SRT and yielded the highest average SRT in dB HL. Additionally, medium sloping words had the smallest average SRT-PTA difference. Steeply sloping words required the lowest number of words on average to obtain the starting level for the SRT. Shallow sloping words had the highest average number of words to obtain the SRT and the lowest average SRT. Additionally, shallow words had the largest SRT-PTA difference and the highest number of words to obtain the starting level. The range of the number of words to obtain the SRT is 2.35. The range between the lowest and highest average SRT is 1.36 dB HL. With all three slopes combined, the average difference between the SRT and PTA was 2.24, which is within the acceptable 6 dB difference (Hanekom et al., 2015; Olsen & Matkin, 1979).

A multivariate analysis of variance (MANOVA) was used to examine the variables of slope of the psychometric function and test ear. The MANOVA showed no significant differences or interactions in the test ear variable. As a result, it was eliminated from further statistical analysis.

The MANOVA revealed significant differences in the number of words required to obtain the SRT due to slope of the psychometric function; $F(2, 158) = 13.02$, $p < .001$. A Tukey-Kramer post hoc analysis revealed a significant difference between steep and shallow sloping words; $t(158) = 3.75$, $p < .001$) as well as a significant difference between words of shallow and medium slopes; $t(158) = 4.93$, $p < .001$. There was no significant difference between steep and medium sloping words; $t(158) = 1.32$, $p = .384$.

Table 1

Summary of Descriptive Statistics of SRT Testing Using a Male Speaker, N = 80

Slope	Variable	*M*	*SD*	Min	Max
Steep	Words for SRT	17.56	3.99	10.00	31.00
	SRT	0.05	3.56	-6.00	12.00
	SRT-PTA Diff	-1.24	4.03	-11.33	11.00
	Words to Start	7.90	2.76	4.00	16.00
Medium	Words for SRT	16.70	3.26	10.00	28.00
	SRT	0.80	3.33	-6.00	14.00
	SRT-PTA Diff	-0.49	3.75	-10.67	7.33
	Words to Start	8.60	2.44	4.00	13.00
Shallow	Words for SRT	19.91	4.94	11.00	30.00
	SRT	-0.56	3.76	-8.00	15.00
	SRT-PTA Diff	-1.85	3.84	-10.33	8.33
	Words to Start	9.41	2.98	5.00	18.00
Average[a]	Words for SRT	18.06	4.32	10.00	31.00
	SRT	0.10	3.58	-8.00	15.00
	SRT-PTA Diff	-2.24	3.76	-14.67	9.33
	Words to Start	8.64	2.80	4.00	18.00

Note: Words for SRT = the number of words required to obtain the SRT; SRT (dB HL) = actual measured SRT; SRT-PTA (dB) = difference between the measured SRT and the measured PTA; Words to Start = the number of words required to obtain the starting level

[a]The average data for all three slopes: steep, medium and shallow (*N* = 240)

Additionally, the MANOVA revealed significant differences in the actual measured SRT due to slope of the psychometric function; $F(2, 148) = 13.0$, $p < .001$. The Tukey-Kramer analysis showed a significant difference between steep and medium sloping words; $t(148) = -2.80$, $p = .016$. It also showed a significant difference between medium and shallow sloping words; $t(158) = -.509$, $p < .001$). There was no significant difference between steep and shallow sloping words; $t(158) = 2.29$, $p = .060$.

The MANOVA also indicated significant differences in the SRT-PTA relationship due to slope of the psychometric function; $F(2, 158) = 13.00$, $p < .001$. The Tukey-Kramer revealed a significant difference between the steep and medium sloping words; $t(148) = -2.80$, $p = .016$. A significant difference between the shallow and medium sloping words was also noted; $t(158) = -.509$, $p < .001$. There was not a significant difference between steep and shallow words; $t(158) = 2.29$, $p = .060$.

The MANOVA also revealed significant differences in the number of words needed to obtain the starting level due to slope of the psychometric function; $F(2, 158) = 13.02$, $p < .001$. The post hoc Tukey-Kramer revealed a significant difference between the steep and shallow sloping words; $t(158) = -3.75$, $p < .001$. No statistical difference was found between steep and medium sloping words; $t(158) = -1.73$, $p = .196$. Similarly, no statistical difference was found between shallow and medium sloping words; $t(158) = 2.01$, $p = .113$.

Discussion

The primary purpose of this study was to examine the relationship between the slope of the psychometric function and the SRT, specifically whether slope affects the number of words used to establish the SRT. The general consensus in the field has been that steeply sloping words provide the most reliable SRT in the shortest amount of time (Chipman, 2003; Harris et al., 2007; Ramkissoon, 2001; Wilson & Carter, 2001). As a result, it was hypothesized that steeply sloping words would yield the most precise SRTs using fewer words when compared to shallow and medium sloping words. The secondary purpose of this study was to examine whether the slope of the psychometric function had an effect on SRT-PTA agreement. It was hypothesized that steeply sloping words would lead to higher SRT-PTA agreement. If there were significant differences due to slope, then slope would be a factor to consider when assessing the homogeneity of words, to ensure spondees are homogenous both in terms of audibility and slope.

The results from the MANOVA and subsequent post hoc analyses showed significant differences among the number of words to obtain the SRT and the SRT-PTA difference due to the slope of the psychometric function. For the number of words to obtain the SRT, descriptive statistics reveal that shallow words had the highest average number (M = 19.91) of words needed to establish the SRT, and that this number was statistically different than the averages of medium and steeply sloping words. Based on this, it is statistically justifiable to eliminate shallow sloping words from a word set. However, for clinical purposes, the differences are negligible. While shallow sloping words yielded an average of 19.91 words, the medium sloping words yielded the lowest average (16.70 words), making a difference of just 3.21 words. Since the process of presenting each word is relatively fast (approximately 5 seconds per word), administering four (rounded up from 3.21) additional words would add approximately 20 seconds to the time it takes to obtain the SRT. Clinically, including shallow sloping words in a word set would not negative impact test efficiency or the length of SRT duration.

Regarding the SRT-PTA difference, shallow sloping words yielded the highest average SRT-PTA difference (1.85) while medium sloping words had the smallest difference (0.49). Although statistical analysis did reveal significant differences in the SRT-PTA variable due to slope of the psychometric function, the clinical difference again seems negligible. The range for the highest and lowest average SRT-PTA agreement is 2.24. Clinically, this difference is small and insignificant, especially when considering the PTA has a ±5 dB test-retest margin of error (Schlauch & Nelson, 2009).

The actual measured SRT varied due to slope of the psychometric function by a range of 1.36 dB. Although statistical analysis revealed differences in actual measured SRT due to slope, the differences seem to be clinically negligible because the SRT has a ±5 dB test-retest margin of error as specified by Schlauch & Nelson (2009).

Lastly, in regard to the number of words to obtain the starting level, the range between the highest and lowest average is 1.51, with the shallow sloping words having the highest average number of words (9.41) and steep sloping words having the average lowest number (7.90). Only two additional words would have to be administered if shallow sloping words were used in SRT testing. The total average number of words (number of words to obtain the starting level and the number of words to obtain the SRT) can also be examined: 25.46 for steep slope, 25.3 for medium slope, and 29.32 for shallow slope. With numbers rounded up, shallow sloping words add four extra words on average, which adds approximately 20 seconds.

In conclusion, this study revealed that the slope of the psychometric function does not need to be a consideration for spondees in SRT testing when using a male talker. While statistical analysis revealed that slope of the psychometric function did have a significant effect on the tested variables, the clinical significance is slight. Including shallow sloping wordlists in SRT testing will only add four words on average, increasing testing time by approximately 20 seconds. However, if slope is a necessary consideration for clinical or research purposes, it may be beneficial to use words that are as steep or steeper than medium sloping words. Future research using a female talker would provide a greater analysis of the subject. Since this study used participants with normal hearing between the ages of 18 and 30 years, future studies could design a similar study for pediatric populations, older populations, and hearing impaired populations. Another avenue for future research is to examine the relationship between slope of the psychometric function and the SRT in languages other than English. With this knowledge, future researchers can develop word lists of varying slopes to increase the range of words available for testing as long as the words meet the four criteria: familiarity, phonetic dissimilarity, normal sampling of English sounds, and homogeneity with respect to audibility.

References

American Speech-Language Hearing Association. (1978). Guidelines for manual pure-tone threshold audiometry. *ASHA, 20*(4), 297-301. doi:10.1044/policy.GL2005-00014

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

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.

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.

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.

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

Carhart, R. (1952). Speech audiometry in clinical evaluation. *Acta Oto-Laryngologica, 41*(1-2), 18-48.

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

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

Hanekom, T., Soer, M., & Pottas, L. (2015). Comparison of the South African spondaic and CID W-1 wordlists for measuring speech recognition threshold. *South African Journal of Communication Disorders, 62*(1*)*, 1-10. doi:10.4102/sajcd.v62i1.97

Harris, R. W., Nissen, S. L., Pola, M. G., McPherson, D. L., Tavartkiladze, G. A., & Eggett, D. L. (2007). Psychometrically equivalent Russian speech audiometry materials by male and female talkers. *International Journal of Audiology, 46*(1), 47-66. doi:10.1080/14992020601058117

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.

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.

Olsen, W. O., & Matkin, N. D. (1979). Speech audiometry. In W. F. Rintelmann (Ed.), *Hearing assessment* (1st ed., pp. 39-140). Baltimore, MD: University Park Press.

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

Schlauch, R. S., & Nelson, P. (2009). Puretone evaluation. In J. Katz, L. Medwetsky, R. Burkard, & L. J. Hood (Eds.), *Handbook of clinical audiology* (6th ed., pp. 30-49). Philadelphia, PA: Lippincott Williams & Wilkins.

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

Annotated Bibliography

Abdulhaq, N. (2005). *Speech perception test for Jordanian Arabic speaking children.* (Doctor of Philosophy Dissertation), University of Florida, Gainesville, FL.
Purpose: This study involved two experiments. First, to determine if there was an improvement in word recognition scores of Jordanian Arabic speaking children if the presentation level is increased and to develop four lists of 40-words in Jordanian Arabic to be used for speech audiometry. Second, to examine if word recognition scores differ when Jordanian versus Saudi dialect words are used.
Method: Jordanian Arabic speaking children, aged 6-9, with either normal hearing or moderate to severe hearing loss participated. They listened to randomized lists of monosyllabic words that were recorded by native Jordanian Arabic speakers. The children were asked to repeat and their word recognition scores were collected.
Results: The investigator developed four lists of 50 words with no significant differences in curve shapes and slope. There was a significant difference when the children listened to Jordanian versus Saudi Arabic dialects.
Relevance to Current Study: The lists that were developed were considered homogenous in terms of audibility. Speech audiometry materials need to be homogenous to provide reliable results.

American Speech-Language Hearing Association. (1978). Guidelines for manual pure-tone threshold audiometry. *ASHA, 20*(4), 297-301. doi:10.1044/policy.GL2005-00014
Purpose: ASHA guidelines for a recommended set of procedures for measuring thresholds using pure-tones. Guidelines and standardization of procedures help clinicians and clinics to communicate information more effectively.
Summary: These guidelines discuss manual pure-tone audiometry. The guidelines include the instructions given to the participant, the response mode, and the interpretation by the clinician. The guidelines also discuss how to determine the pure-tone threshold through familiarization and threshold measurements.
Relevance to Current Study: Pure-tone audiometry was completed as part of the "normal hearing" check prior to testing the SRT.

American Speech-Language Hearing Association. (1988). Guidelines for determining threshold level for speech. *ASHA, 30*(3), 85-89.
Purpose: ASHA guidelines for a recommended set of procedures for determining threshold levels of speech. Guidelines help to standardize speech audiometry and provide more accurate results.
Summary: The guidelines discuss history and scope of speech audiometry over time. The guidelines also establish definitions for speech audiometry terms. Considerations such as calibration of audiometers, test material (spondaic words), response mode, and recorded/live voice presentation are also mentioned. The guidelines then establish a basic procedure for determining the speech recognition threshold in four steps: instructions, familiarization, descending threshold determination, and calculation. Both a 5-dB and 2-dB descending method were discussed.

Relevance to Current Study: The current study followed the 2 dB method dictated by the ASHA guidelines for speech audiometry.

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: To provide the history of SRT testing and to measure homogeneity of the 36 CID spondees when presented with monitored live voice. Additionally, to examine the relationship between the SDT and SRT.
Methods: 75 subjects, aged 19 to 34, with normal hearing were tested for their SDT. They were then familiarized with the 36 CID spondaic words. They were then tested for the SRT using an ascending approach through monitored live voice by three different speakers.
Results: The range of homogeneity in terms of audibility between words was 7.9 dB. The researchers believed this number to be excessive and reduced the 36 spondaic words to 18 words to improve homogeneity within the list. Differences between the three speakers were limited and could be generalized to other clinics.
Relevance to Current Study: In order to reduce variation and maintain homogeneity, monitored live voice was not used in the current investigation.

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: To investigate data on slope of monosyllabic words in the sensorineural hearing loss population. The study examines two topics: 1) whether there is a relationship between the magnitude of hearing loss and the slope of the intelligibility function, and 2) whether there is a relationship between the pure-tone audiometric configuration and the slope of the intelligibility function.
Method: 60 adult subjects with mild to moderate hearing loss participated. Pure-tone audiometry and SRT testing was conducted with monosyllabic words from CID Test W-22. Slopes were then calculated between the 20 and 80% intelligibility functions.
Results: The results showed that the magnitude of the hearing loss did have affect the slope of intelligibility. There was a weak relationship between audiometric configuration and slope. It was found that to obtain an accurate slope, the slope of each individual function needed to be calculated separately.
Relevance to Current Study: Slope is an independent factor in the present study. The slope of words should have been calculated in the correct way.

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 authors believed that other researchers only examined the CID W-1 words homogeneity in terms of threshold, but did not examine psychometric function of slope. As a result, homogeneity is compromised. Their goal was to create a sublist of words that were homogenous in terms of audibility and slope.
Method: Recordings of spondaic words were analyzed. The data were fitted to a logistic function (psychometric function) and then 50% point (threshold) and slope were

calculated. 30 spondees were chosen to be homogenous. These spondees were adjusted to vary less in both slope and threshold.
Results: Individual words between lists were found to be homogenous in terms of both intelligibility and slope. However, the word lists as whole were not homogenous in terms of slope. The list of 30 words showed homogeneity in terms of slope and audibility.
Relevance to Current Study: Homogeneity both in terms of audibility and slope are crucial for selecting spondees. The current study examines the variable of slope to determine whether slope has an effect on duration and accuracy of SRT testing.

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: To provide history and background of speech audiometry, including modifications to speech recognition testing over time.
Summary: Prior to the 20th century, speech recognition tests were informal and measured without standardized guidelines. Several researchers including Hudgins et al (1947) and Hirsch et al (1952) have made modifications to the lists of words that are used in SRT testing.
Relevance to Current Study: The current study is about speech recognition testing and examines the effect of different words of varying slopes on SRT testing. The history of the SRT is useful information.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*(1-2), 62-71.
Purpose: To provide history of speech audiometry and to discuss its significance. Also defines speech audiometry and discusses requirements for testing.
Summary: Four events in history have contributed to the growth of speech audiometry. The four events are: Fletcher and associates laying the foundation for speech audiometry at the Bell Telephone Laboratories, analysis of speech and hearing conducted at Harvard University, development of military programs in the US for aural rehabilitations, and continuation of research. Also indicates that prior to testing, audiometer/equipment must be calibrated, the test construction must be valid and consistent with the aspect of hearing that is being tested.
Relevance to Current Study: The history of SRT testing and the principles behind it is useful information because the current study is about improving SRT testing.

Carhart, R. (1952). Speech audiometry in clinical evaluation. *Acta Oto-Laryngologica, 41*(1-2), 18-48.
Purpose: To discuss the different purposes of speech audiometry and why it’s important.
Summary: The SRT supplies more information than pure-tone audiometry and gives specific information about how an individual hears speech. The SRT can also provide information about hearing improvement after treatment and surgery. It can also give information regarding benefit of hearing aids. The SRT can also be compared to an individual’s PTA to see if they are aligned.
Relevance to Current Study: The current study examines the SRT. The purposes of SRT testing is important. Additionally, the relationship between the SRT and PTA is a variable in the current study.

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

Purpose: To develop a list of words that were steep in slope as well as familiar to listeners. These words would improve the list of 36 spondaic words used for SRT testing.

Methods: Chapman used two English corpora to determine the 10,000 most frequently used English words. In addition, the 36 W-1 spondaic words and 12 words from BYU's children picture board in the audiology clinic were added. The list was refined to 98 words. 20 participants listened to pre-recorded versions of the words and repeated them. This was used to calculate the slope of each word.

Results: 33 words, which included 19 of the CID W-1 words, were determined to be steeply sloping. To improve the homogeneity of the 33 words, the intensity of each word was digitally adjusted.

Relevance to Current Study: The current study used the 98 words originally selected by Chipman, the words were divided into shallow, medium, and steep slope in the current study.

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

Purpose: To investigate the 36 words from CID Auditory Test W-1 to determine the range of intelligibility of the spondaic words and if the 36 words could be reduced to fewer words while still making the test reliable.

Methods: 50 subjects participated and their SRT was determined using the CID W-1 words. The means of the relative intelligibility were calculated.

Results: The overall means ranged from 2.8 dB to 10.8 dB. The range between the means (8 dB) exceeded the 6 dB criterion which can be accepted clinically as the range of intelligibility. They reduced the list to 27 words which fell within ± 2 dB.

Relevance to Current Study: Researchers have claimed that the 36 CID W-1 words are not homogenous in terms of intelligibility. This is important to the current investigation as it provides background to the history of SRT testing.

Hanekom, T., Soer, M., & Pottas, L. (2015). Comparison of the South African spondaic and CID W-1 wordlists for measuring speech recognition threshold. *South African Journal of Communication Disorders, 62*(1), 1-10. doi:10.4102/sajcd.v62i1.97

Purpose: To compare the PTA and SRT results of the English Second Language (ESL) population in South Africa. Three wordlists were used: a South African Spondaic (SAS) wordlist, a CID W-1 list with less familiar words, and a CID W-1 list with more familiar words.

Methods: The different wordlists were presented to 101 normal hearing ESL individuals. The SRT was collected based on the different lists.

Results: There was a strong correlation between the PTA and the SRT using the SAS wordlist as well as the CID W-1 more familiar wordlist. There was a weak correlation between the PTA and the CID W-1 less familiar wordlist. The strongest correlation was with words that are familiar, indicating the importance of words that are familiar to the listener.

Relevance to Current Study: The words that were used during SRT testing were required to be "familiar words". The current study also examined the relationship between the SRT and PTA.

Harris, R. W., Nissen, S. L., Pola, M. G., McPherson, D. L., Tavartkiladze, G. A., & Eggett, D. L. (2007). Psychometrically equivalent Russian speech audiometry materials by male and female talkers. *International Journal of Audiology, 46*(1), 47-66. doi:10.1080/14992020601058117
Purpose: To develop speech audiometry materials in Russian to be used for word recognition and SRT testing.
Method: Native speakers of Russian recorded both monosyllabic and bisyllabic words that were considered familiar words. These words were than evaluated by a different group of native Russian speakers.
Results: The list of monosyllabic words were digitally adjusted to make the words more homogenous in terms of audibility and slope. There were 25 bisyllabic words with steep slope. The slopes are similar to values for English and other languages as well.
Relevance to Current Study: Bisyllabic words with steep slopes are encouraged by most researchers. The present investigation examines steep, medium, and shallow slopes.

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: To provide a history of speech audiometry and modifications that have occurred to improve and standardize testing. The CID W-1, W-2, and W-22 Auditory Test lists were compared.
Summary: Articulation testing methods developed for the military during World War II were applicable to clinical evaluation of hearing. Research on speech audiometry then began with the Psycho-Acoustic Laboratory Tests No 9 and 12. These simple tests were modified over time to become more refined in terms of audibility and familiarity. The ability to record spoken material also contributed to the standardization of testing. The CID W-1, W-2, and W-22 tests all had higher intelligibility and better audibility than tests that were previously used.
Relevance to Current Study: History of the SRT and development of different wordlists is important to the current study.

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.
Purpose: To provide a background of speech audiometry and delineate the characteristics of test materials. Additionally, to explore difficulties with construction of speech audiometry test and to create a test that allows precise measurement for all degrees of hearing loss.
Summary: The four characteristics of test material are: familiarity, phonetic dissimilarity, normal sampling of English speech sounds, and homogeneity in terms of audibility. The development of PAL Auditory Test no. 9 and no. 12 is discussed. Regarding the tests' reliability, the scores for both tests tend to be within 2 dB of the true score on the tests.

Relevance to Current Study: The current study focuses on homogeneity in terms of audibility, which is one of the four characteristics for speech audiometry materials.

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: To examine the effect of set size (number of words in the wordlist) on SRTs.
Method: 12 normal hearing subjects participated in the study. SRTS were obtained in sets of 2, 4, 8, and 16 words using two different methods: 1) a modified version of the 1988 ASHA guidelines, and 2) a forced-choice, adaptive-staircase procedure that estimated threshold in terms of d'= 1.00.
Results: Using the modified version of the ASHA guidelines, the SRT was found to be dependent on set size. In the second method, there was no significant relationship between SRT and set size.
Relevance to Current Study: The current study follows the ASHA 1988 guidelines for determining the SRT. Meyer & Bilger state that the set size should not be decreased beyond what ASHA has set forth, unless a different methodology is used.

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
Purpose: To develop and evaluate SRT materials for native speakers of Taiwan Mandarin.
Method: Familiar trisyllabic words were digitally recorded by male and female speakers. 20 native listeners evaluated the words at 14 intensity levels.
Results: 28 words with steep slopes were selected and digitally adjusted to become more homogenous in terms of audibility and slope. The mean slope for female words was 11.7% dB and 11.3% dB for male words.
Relevance to Current Study: Researchers have consistently aimed to develop and use steeply sloping words. In the current study, steeply sloping words are compared to shallow/medium sloping words to see if there is a difference.

Olsen, W. O., & Matkin, N. D. (1979). Speech audiometry. In W. F. Rintelmann (Ed.), *Hearing assessment* (1st ed., pp. 39-140). Baltimore, MD: University Park Press.
Purpose: To provide information and history regarding different audiometric speech materials and methods.
Summary: Results from pure-tone testing do not accurately capture an individual's hearing abilities. Speech audiometry provides more comprehensive information. In 1804, Pgfinsten was the first to report degree of hearing loss based on speech tests. Since then, research on speech audiometry has continued to grow. The current material for speech audiometry is spondees, which require four characteristics: familiarity, phonetic dissimilarity, normal sampling of English sounds, and homogeneity with respect to audibility. The relationship between the SRT and PTA is also explained.
Relevance to Current Study: This chapter provides background information on SRT. The current study focused on homogeneity in regard to audibility, one of the four characteristics of spondaic words. Additionally, the current study investigates The SRT-PTA relationship.

Ramkissoon, I. (2001). Speech recognition thresholds for multilingual populations. *Communication Disorders Quarterly, 22*(3), 158-162.
Purpose: To discuss appropriate stimuli for SRT testing in multilingual population.
Summary: Familiarity is a key tenant of the spondaic words used in SRT testing. If the words are unfamiliar, the validity of the measurement is greatly reduced. For multilingual or limited English individuals who are tested by English-speaking audiologists, pairs of digits (numbers) may be used as stimuli. Also, stimulus materials must have a steep rise in intelligibility.
Relevance to Current Study: The article emphasizes the need for familiarity in SRT testing. Also, the article notes that steep rise is important for homogeneity in terms of audibility.

Schlauch, R. S., & Nelson, P. (2009). Puretone evaluation. In J. Katz, L. Medwetsky, R. Burkard, & L. J. Hood (Eds.), *Handbook of clinical audiology* (6th ed., pp. 30-49). Philadelphia, PA: Lippincott Williams & Wilkins.
Purpose: To provide background on pure-tone audiometry and its uses and benefits. The chapter provides a comprehensive guide on how to use and interpret pure-tone thresholds.
Summary: Pure-tone audiometry is more complex than it is originally perceived. It is a widely used diagnostic tool.
Relevance to Current Study: The current study screen participants using pure-tone audiometry and derived the participant's pure-tone averages as well.

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: To determine how slope influences homogeneity of CID W-22 and Rus Hughes PB-50 materials.
Method: Randomized 25-word lists with eight presentation levels were given to 12 normal hearing participants. The subjects responded verbally and their answers were recorded.
Results: The results showed that the W-22 words had steeper slopes than the PB-50 word lists. The words were also more homogenous in terms of audibility, which is a criterion for SRT materials.
Relevance to Current Study: The general consensus in the field is that steep words should be used in steep audiometry. The current study compares the effect of different slopes on SRT testing.

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: To evaluate the CID W-1 words for homogeneity, establish psychometric functions and adjust them digitally to improve homogeneity.
Method: There were two experiments involving a female speaker and the original Hirsh recording. First, the words were equated to 0 vu to improve audibility. In the second experiment, the words were equated to intelligibility. The slopes were calculated in the second experiment.

Results: The slope values were homogenous and similar to previous studies that used the CID W-1 words.
Relevance to Current Study: Using words that are homogenous provides steeper slopes, which is one of the variables for the current study.

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: To examine the homogeneity of the 36 commonly used spondaic words through two experiments. The first experiment involved determining the sound-pressure level at which each words is intelligible 50% of the time and to determining the slope of each word. The second experiment examined the SRT using 15 words that were considered homogenous in the first experiment and compared it to the SRT using the standard 36 words.
Method: For the first experiment, 20 normal-hearing participants were familiarized with the 36 spondaic words and were presented with 10 randomized lists of the words at different presentation levels. The participants were asked to repeat each words. For the second experiment, 24 normal-hearing subjects were tested for the SRT using two separate lists of spondaic words.
Results: For the first experiment, only 15 out of the 36 words were found to be homogenous in terms of threshold intensities and their slopes. For the second experiment, the two lists yielded the same SRTs.
Relevance to Current Study: The article highlights the importance of homogeneity, which is the aspect studied in the current investigation. Additionally, this study shows that lists shorter than 36 words can yield accurate SRTs.

Appendix B

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 (XXX) XXX-XXXX or Jessica Reese at (XXX) XXX-XXXX, or through email at xxxxxxxxx@gmail.com or xxxxxxxx@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 C

List of Words Spoken by Male Speaker by Slope

Shallow

Word	%/dB
airplane	18.7
although	18.3
armchair	10.3
farewell	8.8
football	10.8
framework	10.8
Friday	10.6
hotdog	10.1
hothouse	8.7
inside	10.6
lifetime	8.4
maybe	10.5
network	7.7
outcome	8.2
outline	9.1
outlook	9.0
padlock	7.9
platform	8.6
popcorn	9.0
snowman	9.4
somewhat	7.4
toothbrush	9.5
wildlife	7.8

Medium

Word	%/dB
aircraft	12.5
baseball	13.1
bathtub	13.2
childhood	12.7
classroom	14.8
cupcake	14.4
daybreak	13.1
doormat	12.4
duckpond	12.4
eardrum	12.2
elsewhere	14.7
firetruck	13.8
grandson	13.4
highchair	12.0
horseshoe	13.8
inkwell	12.1
instead	14.0
maintain	14.2
meanwhile	13.3
midwest	12.2
mousetrap	13.1
northwest	13.9
nowhere	12.3
outside	14.2
program	14.1
railroad	13.4
something	14.0
southeast	14.6
stairway	14.4
Sunday	14.0
sunset	12.2
whitewash	12.9
widespread	12.9
woodwork	12.3

Steep

Word	%/dB
airport	19.6
always	17.9
birthday	20.2
broadway	17.0
cowboy	18.1
daylight	18.2
doorway	16.2
forecast	16.6
greyhound	16.0
hardware	20.2
highway	18.2
iceberg	17.0
mailman	20.5
mankind	17.2
playground	17.8
sailboat	16.0
schoolboy	19.4
somehow	16.0
somewhere	20.5
sunlight	19.4
weekend	18.7
welfare	16.2
withdraw	18.1
workshop	20.2