PSYCHOMETRICALLY EQUIVALENT BISYLLABIC WORD LISTS FOR WORD RECOGNITION TESTING IN TAIWAN MANDARIN

by

Alycia J. Dukes

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Communication Disorders

Brigham Young University

August 2006

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Alycia J. Dukes

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Shawn L. Nissen, Ph.D., Chair
Date	Richard W. Harris, Ph.D.
Date	Ron W. Channell, Ph.D.

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Alycia J. Dukes in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Shawn L. Nissen, Ph.D.
Chair, Graduate Committee

Accepted for the Department

Ron W. Channell, Ph.D.
Graduate Coordinator

Accepted for the College

K. Richard Young, Ph.D.
Dean, David O. McKay School of Education

ABSTRACT

PSYCHOMETRICALLY EQUIVALENT BISYLLABIC WORD LISTS FOR WORD RECOGNITION TESTING IN TAIWAN MANDARIN

Alycia J. Dukes

Department of Communication Disorders

Master of Science

The aim of this study was to develop, digitally record, evaluate, and psychometrically equate a set of Taiwan Mandarin bisyllabic word lists to be used for word recognition testing. Frequently used bisyllabic words were selected and digitally recorded by male and female talkers of Taiwan Mandarin. Twenty normally hearing subjects were presented each word to find the percentage of words which they could correctly recognize. Each word was measured at 10 intensity levels (-5 to 40 dB HL) in increments of 5 dB. Logistic regression was used to include 200 words with the steepest logistic regression slopes in four psychometrically equivalent word lists of 50 words each with eight half-lists of 25 words each. Digital recordings of the psychometrically equivalent bisyllabic word recognition lists are available on compact disc.

ACKNOWLEDGMENTS

I realize that this project is a reflection of all of my academic accomplishments in life and for this reason, I need to thank all those people who have helped to guide and direct me to the person I am today. I would like to publicly thank my parents, family, and dear friends for all their support and love and encouragement through the years. Their belief in me has helped me to believe in myself through many difficult times when I thought I could not read one more book are write one more paper. This project was in no way completed solely by myself, it was in large part completed thanks to the efforts of Dr. Nissen, Dr. Harris, Sabrina Wu, and Katie Slade. I want to specifically thank Dr. Nissen for his constant patience and support throughout the project that has kept me motivated and committed to this research. I also want to thank all my fellow peers in the graduate program who have kept me laughing and encouraged me along this journey. I will truly miss all of the times we have shared throughout our educational pursuit. Lastly I want to thank my Father in Heaven for blessing me with the drive and knowledge to complete this project and all the educational pursuits and endeavors of my life.

TABLE OF CONTENTS

Page

List of Tables ... vii

List of Figures ... ix

Psychometrically Equivalent Bisyllabic Word Lists for Word Recognition

Testing in Taiwan Mandarin ... 1

Literature Review ... 2

History of Speech Audiometry ... 3

Factors Affecting Word Recognition Performance ... 5

Speech Audiometry Materials for other Non-English Languages ... 7

Method ... 10

Participants ... 10

Materials ... 11

Procedure ... 14

Results ... 15

Discussion ... 32

Conclusion ... 37

References ... 38

Appendix ... 44

LIST OF TABLES

Table Page

1. Age (in years) and Pure Tone Threshold (dB HL) Descriptive Statistics for the 20 Taiwan Mandarin Participants....12

2A. Taiwan Mandarin Male Bisyllabic Lists in Rank Order from Easiest to Most Difficult....17

2B. Taiwan Mandarin Male Bisyllabic Lists in Rank Order from Easiest to Most Difficult (Romanized Form)....18

3A. Taiwan Mandarin Female Bisyllabic Lists in Rank Order from Easiest to Most Difficult....19

3B. Taiwan Mandarin Male Bisyllabic Lists in Rank Order from Easiest to Most Difficult (Romanized Form)....20

4A. Taiwan Mandarin Male Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult....21

4B. Taiwan Mandarin Male Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult (Romanized Form)....22

5A. Taiwan Mandarin Female Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult....23

5B. Taiwan Mandarin Male Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult (Romanized Form)....24

6. Mean Performance of Taiwan Mandarin Male Bisyllabic Lists and Half-lists……………………………………………………………………………25

7. Mean Performance of Taiwan Mandarin Female Bisyllabic Lists and Half-lists…...26

LIST OF FIGURES

Figure Page

1. Psychometric functions for the four Taiwan Mandarin bisyllabic lists and eight half-lists for male talker and female talker recordings before intensity adjustments…………………………………………………………………...29

2. Psychometric functions for the four Taiwan Mandarin bisyllabic lists and eight half-lists for male talker and female talker recordings after intensity adjustments to produce 50% performance at 4.5 dB HL…………………………..30

3. Mean psychometric functions for male and female Taiwan Mandarin talker bisyllabic lists before and after intensity adjustment…………………………….....31

Psychometrically Equivalent Bisyllabic Word Lists for Word Recognition Testing in Taiwan Mandarin

Speech audiometry materials have been developed to provide tests which measure an individual's ability to hear speech that is encountered in everyday life. There are two fundamental diagnostic tools used to measure this ability, Speech Reception Threshold (SRT) testing, and word recognition testing. The purpose of SRT testing is to find the lowest level of hearing for speech, where 50% of the speech material can be correctly recognized. The purpose of word recognition testing is to determine the approximate suprathreshold level at which an individual can correctly understand and repeat a list of words (ASHA, 1988).

Speech audiometry is an invaluable part of any comprehensive audiological examination as it provides more information regarding a persons hearing impairment than pure-tone audiometry alone. This is due to the fact that pure-tone audiometry tests an individual's residual hearing threshold, whereas speech audiometry also analyzes sound distortion, loudness, speech comprehension, and localization. Therefore, by using speech audiometry in addition to pure-tone audiometry, diagnostic information can be obtained regarding whether a conductive, sensory, and/or neural hearing impairment exists (Egan, 1979; Hagerman, 1984). Some of the other uses of speech audiometry are to estimate the severity of a hearing impairment (Epstein, 1978; Hagerman, 1984) and to aid in establishing whether or not hearing aids are necessary (Hood & Poole, 1977). Recently researchers have begun to focus on creating high-quality digital recordings of speech stimuli that can be used for speech audiometry testing in languages other than English. Such materials have been developed in the following languages: Italian, Arabic, Brazilian

Portuguese, Polish, Russian, Spanish, Korean, Japanese, and Standard Mandarin Chinese (Aleksandrovsky, McCullough, & Wilson, 1998; Ashoor & Prochazka, 1985; Christensen, 1995; Greer, 1997; Harris, Crawford, & Mastny, 2004; Harris, Goffi, Pedalini, Merrill, & Gygi, 2001; Harris & Kim, 2004; Harris, Nissen, & Jennings, 2004; Harris & Pola, 2002; Mangum, 2005; Ramkissoon, 2001; Ramkissoon, Proctor, Lansing, & Bilger, 2002). Although speech audiometry materials are available in Standard Mandarin Chinese, there is still a need for materials for native speakers of Mandarin who originate from Taiwan.

Thus, this project will aim to develop speech audiometry materials for Taiwan Mandarin Chinese. Specifically the study will focus on digitally recording, evaluating, and psychometrically equating words in Taiwan accent for bisyllabic word lists to be used for word recognition testing.

Literature Review

The purpose of word recognition testing is to measure how well an individual can understand speech once it is loud enough to overcome any possible loss of hearing sensitivity (Epstein, 1978). Word recognition testing provides an estimate of an individual's hearing loss, as well as providing information regarding their speech comprehension abilities. Word recognition testing yields a score which is the percentage of a list of words that are correctly identified at suprathreshold levels. In some sense, word recognition testing is a measure of what the patient does hear, which is in contrast with SRT, which is a measure of what the patient does not hear. However, it is important to note that the SRT and word recognition scores are interrelated. A word recognition score below 50% correlates with a difficult to establish SRT score. Likewise, if a SRT score is easy to obtain, the word recognition score should be 90% or better (Epstein, 1978).

History of Speech Audiometry

Speech audiometry is a widely used method of audiological assessment in the United States. According to a recent study, 98% of the practicing audiologists administer SRT tests, and 99% administer word recognition tests to evaluate an individual's hearing (Martin, Armstrong, & Champlin, 1994). However this has not always been the case. Initially, speech tests began with spoken or whispered messages which were presented at measured distances between the speaker and listener. These tests were not easily quantified and only gave gross estimates of a person's ability to hear speech (ASHA, 1988). Thus clinical speech audiometry was developed in order to have more precise measurements. The first clinical speech test was developed in 1904 by Bryant and was recorded on a phonograph. The test was never commonly administered primarily due to the primitive recording equipment (Hudgins, Hawkins, Karlin, & Stevens, 1947). Thus, the first recorded auditory test which was widely used was The Western Electric 4A (later 4C) test, which was a phonographic recording of spoken digits.

Considerable effort was expended during World War II to develop tests which would aid in evaluating various types of communication equipment for the military. These tests were developed at Harvard University's PAL, and some of them turned out to be applicable for use in clinical hearing evaluations (Hirsh et al., 1952). These tests were the PAL Auditory Test No. 9 and No. 12. The PAL No. 9 measured the threshold of hearing for words, which was done using two lists of 42 spondaic words each. The PAL No. 12 measured the threshold of hearing for sentences using eight lists of sentences which were answered using one word utterances (Hudgins et al., 1947). Initially, these materials were clinically available on phonograph for military rehabilitation centers and then subsequently for general clinical application. With time, new lists were developed which had greater clinical use, as the PAL tests were found to be deficient in word familiarity and phonetic balance (Hirsh et al., 1952). These issues were addressed in modifications to the PAL lists by the Central Institute for the Deaf (CID), thereby resulting in the W-1 and W-2 Auditory Tests, both of which are spondaic word lists (ASHA, 1988).

Word recognition testing was measured by way of phonetically balanced (PB) monosyllabic word lists which were developed using the PAL by Egan (1948). The lists were called the PAL PB-50 lists. These were 20 lists consisting of 50 words each which were phonetically balanced or equivalent (Hirsh et al., 1952). The lists were found to be clinically useful for measuring discrimination loss; however, they did have some shortcomings. The lists contained some words which were unfamiliar to patients, thereby negatively affecting patient performance and the overall reliability of the test (Dennis & Neely, 1991; Hirsh et al., 1952).

A second major set of monosyllabic word lists called the CID auditory test W-22 was created by Hirsh et al. (1952). These materials were developed to create lists with increased listener familiarity, greater phonetic balance, and better clarity through the use of magnetic tape (Epstein, 1978). The CID W-22 test was successful in that the word familiarity was markedly greater in these lists than the PAL PB-50 lists; however neither list was found to be practical in giving prognostic information regarding an individual’s ability to follow content in running speech. Therefore Campbell (1965) took the Hirsh recordings of the CID W-22 lists and redistributed the 200 words into eight 25-item lists, making the lists more homogenous in regard to word difficulty. However, the effect of this redistribution was a loss of phonetic balance. Interestingly, Campbell's follow-up experiments indicated that phonetic balance had only minimal clinical effects on test validity.

Factors Affecting Word Recognition Performance

The reliability and validity of word recognition testing has been found to be influenced by a number of factors, such as word selection and homogeneity, dialect of the talker, method and level of presentation, and type of recording (Nissen et al., 2005). Speech materials that take into account these issues have the potential to provide valid and reliable test results.

Two word selection criteria that may affect auditory word recognition are word familiarity and uniqueness. The familiarity of a given word increases the intelligibility of that word. Research has shown that highly familiar word lists, compared to less familiar word lists, provide a more accurate measure of an individual's comprehension of continuous discourse (Epstein, 1978). Since the point at which the word is recognized (or the optimal discrimination point) takes place early on in the word, uniqueness is also said to increase the ability for an individual to choose or to recall a given word quickly (Luce, 1986). Thus, these two factors should be taken into account when selecting words for word recognition lists.

Homogeneity in regard to speech stimuli relates to equating the audibility amplitude levels for each item tested (Epstein, 1978). As the speech audiometry materials increase in homogeneity, the results of the given speech audiometry test become more precise (Wilson & Carter, 2001). A test list that is considered homogeneous can be divided into shorter lists which have equal audibility. This decreases test time when performing speech audiometry since scores can be established using as few test items as possible (Young, Dudley, & Gunter, 1982). Two important parts of homogeneity are audibility threshold and slope. The slope relates to the percentage of correct recognition and is plotted as a function of intensity. Studies have shown that the more homogeneous or less variable the psychometric characteristics of the words, the steeper the slope of the mean psychometric function. Also the less variable the performance is on each test item in terms of both location and slope, the steeper the slope of the mean psychometric function (Wilson & Carter, 2001).

The dialect of the talker also needs to be taken into account when developing speech audiometry materials. The presentation of dialectally unfamiliar words may negatively affect the validity of subsequent test results. A study by Wilson and Moodley (2000) found that using recordings of a non-native dialect may negatively alter performance levels especially at presentation levels below 50 dB SPL. Weisleder and Hodgson (1989) also had similar findings. In Weisleder and Hodgson’s (1989) study a Spanish word recognition test was evaluated. The talker for this study was of Mexican origin. Research indicated that all of the subjects who were also of Mexican origin obtained better scores particularly on the lower presentation levels than the other subjects of different nationalities. Thus the performance of the listeners who had different dialectical backgrounds from the talker was negatively affected. Therefore when measuring word recognition abilities for individuals from Taiwan it is necessary to use recordings from talkers of Taiwan Mandarin to obtain precise measures.

It is also helpful if speech materials can be presented to listeners through computer software designed for the presentation of speech audiometry materials. This allows for different speaker tracks and the word order within each list to be quickly and effectively randomized which yields more reliable and valid results (Nissen et al., 2005). In addition, recorded presentation of the test materials is recommended over monitored live voice. With live voice it is difficult to monitor the test words to a consistent hearing level and standardization is impossible from client to client and clinic to clinic (ASHA, 1988). These materials are best when recorded using digital technology over tape recordings. Some of the advantages of digital recordings are improved signal-to-noise ratio, increased dynamic range, increased channel separation, reduced harmonic distortion, elimination of wow and flutter from tape recordings, and less deterioration with usage over time which allows for a longer shelf life of the materials (Nissen et al., 2005).

Speech Audiometry Materials for other Non-English Languages

Speech audiometry materials in English are widely available (Martin, Armstrong, & Champlin, 1994). Although English is commonly used in many areas around the world, it is not recommended that English materials be used for speech audiometry testing in those who have acquired English as a second language. Researchers have found that those who have learned English as a second language after puberty perform significantly worse on speech audiometry tests than those who are native speakers of English (Von, Champlin, & Shetty, 2004). Therefore it is necessary to have speech audiometry materials available in each individual's native language for reliable results.

However, to date speech audiometry materials are not as commonly available in languages other than English.

Although speech audiometry materials are presently available in Standard Mandarin Chinese, there is still a deficit in materials for Taiwan Mandarin. Mandarin Chinese has many dialects, with Standard Mandarin or Peking Mandarin, which is used on Mainland China, being the most widely spoken dialect. However, the over 20 million inhabitants of Taiwan speak Mandarin Chinese with a distinct Taiwan accent, also called Taiwan Mandarin (Katzner, 1986).

Standard Mandarin has been found to differ phonologically, lexically, and syntactically from the Mandarin which is spoken in Taiwan. This is due to the influence of Taiwanese, and other indigenous varieties of Chinese. Taiwanese, which is also called Southern Min, is a native dialect spoken by many elderly individuals living in Taiwan and is learned in daily social conversation rather than through compulsory education (Cheng, 1985). Until 1949 Taiwanese was the major language used for communication in Taiwan, however with the influx of mainlanders from China, Mandarin and Cantonese were brought to the island. Thus, from 1949 to 1960 the Taiwan provincial government campaigned to promote the use of Mandarin Chinese as the sole means of communication for instruction in the school systems in Taiwan. The campaign was successful and was able to convert the universal language in Taiwan to Mandarin. This campaign’s effects where so far reaching that Mandarin Chinese has not only been adopted by school children, but is also understood by a majority of the general population (Li, 1985). While Mandarin is now the universal language in Taiwan, the elderly have learned the language secondary to Taiwanese though compulsory education and by way of mass media (Cheng, 1985).

The Mandarin spoken in Taiwan has marked linguistic differences from Standard Mandarin of mainland China. In pronunciation, Taiwan Mandarin exhibits loss of the nonsyllabic final /r/ and substitution of /r/ for the Standard Mandarin /l/. These differences do not negatively affect understanding yet they yield the distinct Taiwanese accent (Kubler, 1985). Another phonological difference between the two dialects is the absence of retroflexion in some consonants. This is linked to the influence of Taiwanese, which has no retroflexion. In addition, the vocalic system of Taiwanese differs significantly from Standard Mandarin, which often affects the learning of Mandarin as a second language by native Taiwanese speakers. The tones of Standard Mandarin also differ from that of Taiwan Mandarin in two ways. In Taiwan Mandarin the third tone generally has no rise even in the final position and the neutral tone occurs less frequently than in Standard Mandarin (Kubler, 1985).

Taiwan Mandarin also exhibits what Standard Mandarin speakers would consider substandard syntax. In Taiwan Mandarin the following words are commonly used and acceptable: yong (*use*), you (*have*), and yao (*want*). In Standard Mandarin those words are replaced in speech, respectively by the following: shi (*be*), le/guo (*aspect markers*), and ma (*question marker*). Thus these syntactic features work to readily differentiate the Mandarin spoken in Taiwan to that of the Standard Mandarin from mainland China.

Another notable difference in the languages is the newly developing slang words or unsystematic lexicons in Taiwan, which are being adopted into the language by various age groups. These idiomatic expressions and lexicons used in Taiwan are so drastically different from those used in mainland China that mutual intelligibility is very difficult (Li, 1985). Due to these differences, Taiwan Mandarin has evolved into a mutually intelligible, yet distinct variety of Mandarin Chinese.

Since the Mandarin which is spoken in Taiwan has a distinct accent, and differs from Standard Mandarin syntactically and lexically, this needs to be taken into account when developing audiometric materials to be used in Taiwan. It is important to have speech materials which accurately resemble the language of the individual being tested in order to obtain precise measures. Thus the specific aims of this study are to develop high quality digitally recorded materials to be used for word recognition testing for native Mandarin Chinese talkers from Taiwan. The objectives of this study are: (a) to identify one female and one male from Taiwan who speak Mandarin Chinese with a standard Taiwan accent to serve as talkers for the recordings; (b) to compose a list of familiar bisyllabic Mandarin words to be used in word recognition testing; (c) to create high-quality digital recordings of the bisyllabic words; (d) to collect normative data on the bisyllabic words; (e) to construct psychometrically equivalent lists (50 words each) and half-lists (25 words each) of bisyllabic Mandarin words from both the female and male talkers.

Method

Participants

A total of 20 native speakers of Mandarin Chinese from Taiwan participated in evaluating the bisyllabic words for this study (3 male and 17 female). The subjects' ages ranged from 18 to 39 years (M = 25.8 years). All of the participants in this study exhibited pure-tone air-conduction thresholds ≤15 dB HL at octave and mid-octave frequencies from 125 to 8000 Hz and had static acoustic admittance between 0.3 and 1.4 mmhos with peak pressure between -100 and +50 daPa (ASHA, 1990; Roup, Wiley, Safady, & Stoppenbach, 1998). Summary statistics of the subject thresholds are listed in Table 1.

Materials

Word lists. Bisyllabic Mandarin words were selected as the stimuli for word recognition testing. Words were selected from an electronic word corpus of Taiwan Mandarin (Academia Sinica Balanced Corpus of Modern Chinese, 1997). Traditionally, the types of words used for word recognition testing in English are monosyllabic. However, due to the nature of Chinese characters and Chinese vocabulary, it is more appropriate to use bisyllabic words. Specifically, each character in Chinese is represented phonologically by a single syllable with a simple and constrained segmental structure (Nissen et al., 2005). Thus, each syllable is a single morpheme which can technically stand free and alone as a monomorphemic word (Zhou & Marslen-Wilson, 1995). However, most monosyllabic words are ambiguous with regard to lexical meaning. Thus, most lexical items in Chinese are polysyllabic, formed by compounding multiple monosyllabic characters. Interestingly, research of spoken Mandarin has shown that individuals mentally store bisyllabic compound words as whole-word representations as opposed to their individual monomorphemic units (Zhou & Marlsen-Wilson, 1995). For these aforementioned reasons, bisyllabic words were chosen in this study to be used for word recognition testing.

Initially, 300 words were selected for recording, avoiding words with the same pronunciation but different meanings (represented by different characters). From the 300

Table 1

Age (in years) and Pure Tone Threshold (dB HL) Descriptive Statistics for the 20 Taiwan Mandarin Participants

	M	*Minimum*	*Maximum*	*SD*
Age	25.8	18	39	6.5
125 Hz	5.8	-5	15	5.9
250 Hz	5.5	-5	15	5.6
500 Hz	5.8	-5	10	4.4
750 Hz	5.8	0	15	5.2
1000 Hz	5.0	0	10	4.0
1500 Hz	6.8	-5	15	5.2
2000 Hz	4.3	-5	15	5.2
3000 Hz	2.0	-5	15	5.5
4000 Hz	2.5	-10	15	6.0
6000 Hz	3.0	-5	15	5.9
8000 Hz	2.3	-5	15	5.0
PTA[a]	5.0	-1.7	10.0	3.5

[a]PTA = arithmetic average of thresholds at 500, 1000 and 2000 Hz.

bisyllabic words that were recorded, 240 words were selected for evaluation in this study. Any words that were judged to be culturally insensitive, considered to be unfamiliar by native judges, or thought to possibly represent inappropriate content were eliminated from the study prior to listener evaluation. This was done using three native speakers who rated all the words on a scale of 1 to 5 based on the how familiar a word would be to a Mandarin speaker from Taiwan (1 = extremely familiar, 2 = very familiar, 3 = somewhat familiar, 4 = infrequency used, and 5 = rarely used). Only those words which received a ranking of 1 or 2 were used in the study.

Talkers. Initial test recordings were made using six native talkers of Taiwan Mandarin (three females and three males). All talkers originated from the country of Taiwan, who self-reported speaking a standard accent of Taiwan Mandarin. After the initial recordings were made, a panel of eight native speakers evaluated the performance of each of the six talkers. The judges were asked to rank order the speakers from best to worst based on vocal quality, standard dialect, and pronunciation. The highest ranked male and female speakers were selected as the talkers for all subsequent recordings.

Recordings. All recordings were made in a double-walled sound booth located on the Brigham Young University campus in Provo, Utah, USA. A Larson-Davis model 2541 microphone, positioned approximately 15 cm from the talker at a 0° azimuth and covered by a 7.62 cm windscreen, was utilized for all recordings. The microphone signal was amplified by a Larson-Davis model 900B microphone preamp, which was coupled to a Larson-Davis model 2200C preamp power supply. The signal was digitized using an Apogee AD-8000 analog-to-digital converter and subsequently stored on a hard drive for later editing. A 44.1 kHz sampling rate with 24-bit quantization was used for all recordings, and every effort was made to utilize the full range of the 24-bit analog-to-digital converter.

During the recording sessions, the talker was asked to pronounce each bisyllabic word four times. The first and last repetition of each word was excluded to avoid any possible list effects. A native judge then rated the medial repetitions of each word for perceived quality of production, and the best production of each word was then selected for inclusion in the evaluation portion of the study. Any words that were judged to be poorly recorded based on peak clipping, extraneous noise etc., were re-recorded or eliminated from the study prior to listener evaluation. After the rating process, the intensity of each bisyllabic word to be included in the listener evaluation was edited as a single utterance using Sadie Disk Editor software (Studio Audio & Video Ltd., 2004) to yield the same average Root Mean Square (RMS) power as that of a 1 kHz calibration tone.

Procedure

Custom software was used to control randomization and timing of the presentation of the words. In addition, the custom software was also used to record and store all of the scoring. The signal was routed from a computer hard drive to the external input of a Grason Stadler model 1761 audiometer. The stimuli were then routed via TDH-50P headphones from the audiometer to the subject, who was seated in a double-walled sound suite meeting ANSI S3.1 standards (American National Standards Institute, 1999) for maximum permissible ambient noise levels for the ears not covered condition using one-third octave-band measurements. Prior to testing each subject, the inputs to the audiometer were calibrated to 0 VU using the 1 kHz calibration tone through customized computer software. In addition, the audiometer was calibrated weekly during and at the conclusion of data collection. Calibration was performed in accordance with ANSI S3.6 standards (American National Standards Institute, 2004). No changes in calibration were necessary throughout the course of data collection.

Evaluation of bisyllabic words. The subjects were not familiarized with the bisyllabic words prior to testing. The 240 bisyllabic words were randomly grouped into ten lists of 24 words each. These ten lists were used for presentation to the first ten subjects. The 240 words were then randomly combined in a second group of ten different lists for presentation to the next group of ten subjects. Ten presentation levels were selected for the lists: -5 to 40 dB HL in 5 dB steps. One list was presented at each level. The order of the presentation of the lists and the order of the words within the list were randomized for each subject. Each word was presented an equal number of times at each intensity level across the entire subject population. Prior to administration of the word recognition test, the following instructions were given to the participants in Mandarin:

> You will hear bisyllabic words at several different loudness levels. At the very soft levels it may be difficult for you to hear the words. Please listen carefully and repeat the words you hear. If you are unsure of the word, you are encouraged to guess. If you have no guess, please remain quiet until the next word is presented. Do you have any questions?

Results

Once the raw data were complied logistic regression was used to obtain regression slopes for all of the 240 bisyllabic words. Rankings were then given to the 240 words from steep to shallow slopes. The 200 words with the highest logistic regression slopes were then divided into four balanced lists of 50 words each. Random block assignment was utilized to counterbalance each of the four lists. Specifically this was performed by randomly assigning the first four words from the rank ordered list of 200 words individually to one of the four lists. This method was continued until four lists were constructed of 50 words each. Table 2 (male) and Table 3 (female) contain the four equivalent Taiwan Mandarin bisyllabic word lists. Tables 2A and 3A present the lists in character form and Tables 2B and 3B present the equivalent lists in Hanyu Pinyin Romanization.

After the four balanced 50-word lists were complied, eight half-lists of 25 words each were constructed. The half lists were constructed by taking each 50 word list and randomly designating either an A or a B to the first word in each list and then subsequently tagging the second word with the remaining letter. The remaining words were assigned as either an A or a B using counterbalancing. Following the letter assignment, each 50-word list could be divided into two 25-word half-lists. These half-lists are presented for both male and female Taiwan Mandarin in character form in Tables 4A and 5A. The equivalent half-lists are presented utilizing Hanyu Pinyin Romanization in Tables 4B and 5B.

Following the construction of the bisyllabic lists and half-lists, regression slopes and regression intercepts were calculated via logistic regression for each of the four lists and eight half-lists for both the male and female talker recordings. These values are presented in Table 6 (male talker) and Table 7 (female talker). The regression slopes and intercept values were analyzed using a modified logistic regression equation (Equation 1). The equation was utilized to calculate percentage of correct performance at any

Table 2A

Taiwan Mandarin Male Bisyllabic Lists in Rank Order from Easiest to Most Difficult

List 1		List 2		List 3		List 4	
後來	我們	開發	方法	內容	清楚	溝通	嚴重
台灣	能夠	加上	昨天	開始	成功	東西	新聞
科技	成立	然後	原來	電話	最近	工程	不會
現在	還是	設備	已經	或者	不要	參加	情況
負責	過去	雖然	國際	應該	美國	感覺	大學
相關	老師	專業	好像	生活	價值	成為	能力
開放	管理	了解	課程	記者	自我	達到	母親
這樣	需要	獲得	動物	支持	注意	過程	自由
就是	事件	事實	如果	社會	怎麼	發生	經過
他們	產生	重要	起來	今天	學習	未來	所以
當然	處理	電腦	一定	其實	非常	目標	面對
電視	這裡	配合	方式	人生	作業	增加	文化
快樂	語言	發現	空間	投資	不同	完全	安全
變成	希望	完成	不能	故事	什麼	媽媽	解決
時間	那麼	可是	同學	受到	不斷	學生	下午
學校	行為	或是	同時	以後	香港	最後	選擇
接受	時候	真正	成長	身體	一樣	幾乎	繼續
甚至	也許	事業	值得	爸爸	原因	交通	可以
改變	但是	工作	只要	問題	有關	觀念	時代
活動	基本	標準	得到	傳統	造成	準備	沒有
包括	機會	程度	都市	階段	這麼	結果	行動
公園	特別	存在	必須	建築	父母	計畫	經濟
朋友	訓練	發展	現象	教授	政治	國家	教育
先生	一直	考慮	喜歡	大家	實在	可能	覺得
家庭	另外	報導	產品	世界	比較	本身	出現

Table 2B

Taiwan Mandarin Male Bisyllabic Lists in Rank Order from Easiest to Most Difficult (Romanized Form)

List 1		List 2		List 3		List 4	
hòulái	wǒmen	kāifā	fāngfǎ	nèiróng	qīngchu	gōutōng	yánzhòng
táiwān	nénggòu	jiāshàng	zuótiān	kāishǐ	chénggōng	dōngxī	xīnwén
kējì	chénglì	ránhòu	yuánlái	diànhuà	zuìjìn	gōngchéng	búhuì
xiànzài	háishì	shèbèi	yǐjīng	huòzhě	búyào	cānjiā	qíngkuàng
fùzé	guòqù	suīrán	guójì	yīnggāi	měiguó	gǎnjué	dàxué
xiāngguān	lǎoshī	zhuānyè	hǎoxiàng	shēnghuó	jiàzhí	chéngwéi	nénglì
kāifàng	guǎnlǐ	liǎojiě	kèchéng	jìzhě	zìwǒ	dádào	mǔqīn
zhèyàng	xūyào	huòdé	dòngwù	zhīchí	zhùyì	guòchéng	zìyóu
jiùshì	shìjiàn	shìshí	rúguǒ	shèhuì	zěnme	fāshēng	jīngguò
tāmen	chǎnshēng	zhòngyào	qǐlai	jīntiān	xuéxí	wèilái	suǒyǐ
dāngrán	chǔlǐ	diànnǎo	yīdìng	qíshí	fēicháng	mùbiāo	miànduì
diànshì	zhèlǐ	pèihé	fāngshì	rénshēng	zuòyè	zēngjiā	wénhuà
kuàilè	yǔyán	fāxiàn	kōngjiān	tóuzī	bùtóng	wánquán	ānquán
biànchéng	xīwàng	wánchéng	bùnéng	gùshì	shénme	māma	jiějué
shíjiān	nàme	kěshì	tóngxué	shòudào	bùduàn	xuésheng	xiàwǔ
xuéxiào	xíngwéi	huòshì	tóngshí	yǐhòu	xiānggǎng	zuìhòu	xuǎnzé
jiēshòu	shíhou	zhēnzhèng	chéngzhǎng	shēntǐ	yīyàng	jīhū	jìxù
shènzhì	yěxǔ	shìyè	zhíde	bàba	yuányīn	jiāotōng	kěyǐ
gǎibiàn	dànshì	gōngzuò	zhǐyào	wèntí	yǒuguān	guānniàn	shídài
huódòng	jīběn	biāozhǔn	dédào	chuántǒng	zàochéng	zhǔnbèi	méiyǒu
bāokuò	jīhuì	chéngdù	dūshì	jiēduàn	zhème	jiéguǒ	xíngdòng
gōngyuán	tèbié	cúnzài	bìxū	jiànzhù	fùmǔ	jìhuà	jīngjì
péngyou	xùnliàn	fāzhǎn	xiànxiàng	jiàoshòu	zhèngzhì	guójiā	jiàoyù
xiānsheng	yìzhí	kǎolǜ	xǐhuan	dàjiā	shízài	kěnéng	juéde
jiātíng	lìngwài	bàodào	chǎnpǐn	shìjiè	bǐjiào	běnshēn	chūxiàn

Table 3A

Taiwan Mandarin Female Bisyllabic Lists in Rank Order from Easiest to Most Difficult

List 1		List 2		List 3		List 4	
開始	不能	電話	不同	安全	不斷	傳統	必須
工程	成功	好像	成為	感覺	但是	過程	處理
問題	得到	媽媽	東西	然後	都市	快樂	非常
報導	改變	專業	溝通	參加	觀念	爸爸	國際
當然	還是	大學	基本	產生	教授	大家	過去
發展	今年	電視	幾乎	發生	機會	管理	階段
建築	考慮	故事	空間	交通	就是	後來	目標
教育	時間	結果	實際	開發	容易	記者	生命
老師	世界	了解	條件	開放	現在	生活	完全
同學	訓練	什麼	新聞	同時	壓力	事實	學習
完成	應該	相關	嚴重	文化	意義	選擇	一直
原來	政治	也許	這樣	增加	知道	自我	原因
程度	資料	成立	成長	作業	直接	包括	投資
動物	朋友	達到	行動	出現	需要	除了	造成
負責	一般	方法	昨天	公園	努力	國家	電腦
課程	美國	活動	存在	或者	自由	加上	時代
可是	或是	可能	不要	科技	語言	可以	如何
面對	現象	沒有	價值	例如	規定	歷史	覺得
清楚	有關	目的	運動	起來	今天	人生	準備
設備	本身	社會	配合	身體	值得	甚至	母親
時候	獲得	受到	經驗	雖然	家庭	所以	醫院
未來	怎麼	特別	父母	我們	希望	他們	其實
行為	這裡	以後	情況	學生	經過	下午	影響
只是	公司	重要	最近	知識	台灣	支持	接受
標準	工作	表現	內容	變成	團體	最後	地方

Table 3B

Taiwan Mandarin Male Bisyllabic Lists in Rank Order from Easiest to Most Difficult (Romanized Form)

List 1		List 2		List 3		List 4	
kāishǐ	bùnéng	diànhuà	bùtóng	ānquán	bùduàn	chuántǒng	bìxū
gōngchéng	chénggōng	hǎoxiàng	chéngwéi	gǎnjué	dànshì	guòchéng	chǔlǐ
wèntí	dédào	māma	dōngxī	ránhòu	dūshì	kuàilè	fēicháng
bàodào	gǎibiàn	zhuānyè	gōutōng	cānjiā	guānniàn	bàba	guójì
dāngrán	háishì	dàxué	jīběn	chǎnshēng	jiàoshòu	dàjiā	guòqù
fāzhǎn	jīnnián	diànshì	jīhū	fāshēng	jīhuì	guǎnlǐ	jiēduàn
jiànzhù	kǎolǜ	gùshì	kōngjiān	jiāotōng	jiùshì	hòulái	mùbiāo
jiàoyù	shíjiān	jiéguǒ	shíjì	kāifā	róngyì	jìzhě	shēngmìng
lǎoshī	shìjiè	liǎojiě	tiáojiàn	kāifàng	xiànzài	shēnghuó	wánquán
tóngxué	xùnliàn	shénme	xīnwén	tóngshí	yālì	shìshí	xuéxí
wánchéng	yīnggāi	xiāngguān	yánzhòng	wénhuà	yìyì	xuǎnzé	yìzhí
yuánlái	zhèngzhì	yěxǔ	zhèyàng	zēngjiā	zhīdao	zìwǒ	yuányīn
chéngdù	zīliào	chénglì	chéngzhǎng	zuòyè	zhíjiē	bāokuò	tóuzī
dòngwù	péngyou	dádào	xíngdòng	chūxiàn	xūyào	chúle	zàochéng
fùzé	yībān	fāngfǎ	zuótiān	gōngyuán	nǔlì	guójiā	diànnǎo
kèchéng	měiguó	huódòng	cúnzài	huòzhě	zìyóu	jiāshàng	shídài
kěshì	huòshì	kěnéng	búyào	kējì	yǔyán	kěyǐ	rúhé
miànduì	xiànxiàng	méiyǒu	jiàzhí	lìrú	guīdìng	lìshǐ	juéde
qīngchu	yǒuguān	mùdì	yùndòng	qǐlai	jīntiān	rénshēng	zhǔnbèi
shèbèi	běnshēn	shèhuì	pèihé	shēntǐ	zhíde	shènzhì	mǔqīn
shíhou	huòdé	shòudào	jīngyàn	suīrán	jiātíng	suǒyǐ	yīyuàn
wèilái	zěnme	tèbié	fùmǔ	wǒmen	xīwàng	tāmen	qíshí
xíngwéi	zhèlǐ	yǐhòu	qíngkuàng	xuésheng	jīngguò	xiàwǔ	yǐngxiǎng
zhǐshì	gōngsī	zhòngyào	zuìjìn	zhīshi	táiwān	zhīchí	jiēshòu
biāozhǔn	gōngzuò	biǎoxiàn	nèiróng	biànchéng	tuántǐ	zuìhòu	dìfāng

Table 4A

Taiwan Mandarin Male Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
後來	台灣	加上	開發	內容	開始	東西	溝通
現在	科技	然後	設備	或者	電話	工程	參加
負責	相關	專業	雖然	應該	生活	成為	感覺
這樣	開放	了解	獲得	支持	記者	達到	過程
就是	他們	重要	事實	社會	今天	未來	發生
電視	當然	電腦	配合	人生	其實	目標	增加
快樂	變成	完成	發現	投資	故事	媽媽	完全
學校	時間	可是	或是	以後	受到	學生	最後
接受	甚至	事業	真正	身體	爸爸	交通	幾乎
活動	改變	工作	標準	傳統	問題	觀念	準備
包括	公園	存在	程度	階段	建築	計畫	結果
先生	朋友	發展	考慮	大家	教授	國家	可能
家庭	我們	方法	報導	世界	清楚	嚴重	本身
成立	能夠	昨天	原來	最近	成功	新聞	不會
還是	過去	國際	已經	不要	美國	大學	情況
管理	老師	好像	課程	自我	價值	能力	母親
需要	事件	如果	動物	注意	怎麼	經過	自由
處理	產生	起來	一定	非常	學習	所以	面對
這裡	語言	空間	方式	作業	不同	安全	文化
那麼	希望	不能	同學	不斷	什麼	解決	下午
行為	時候	成長	同時	香港	一樣	繼續	選擇
但是	也許	值得	只要	有關	原因	可以	時代
基本	機會	都市	得到	造成	這麼	行動	沒有
訓練	特別	必須	現象	政治	父母	經濟	教育
一直	另外	產品	喜歡	實在	比較	出現	覺得

Table 4B

Taiwan Mandarin Male Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult (Romanized Form)

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
hòulái	táiwān	jiāshàng	kāifā	nèiróng	kāishǐ	dōngxī	gōutōng
xiànzài	kējì	ránhòu	shèbèi	huòzhě	diànhuà	gōngchéng	cānjiā
fùzé	xiāngguān	zhuānyè	suīrán	yīnggāi	shēnghuó	chéngwéi	gǎnjué
zhèyàng	kāifàng	liǎojiě	huòdé	zhīchí	jìzhě	dádào	guòchéng
jiùshì	tāmen	zhòngyào	shìshí	shèhuì	jīntiān	wèilái	fāshēng
diànshì	dāngrán	diànnǎo	pèihé	rénshēng	qíshí	mùbiāo	zēngjiā
kuàilè	biànchéng	wánchéng	fāxiàn	tóuzī	gùshì	māma	wánquán
xuéxiào	shíjiān	kěshì	huòshì	yǐhòu	shòudào	xuésheng	zuìhòu
jiēshòu	shènzhì	shìyè	zhēnzhèng	shēntǐ	bàba	jiāotōng	jīhū
huódòng	gǎibiàn	gōngzuò	biāozhǔn	chuántǒng	wèntí	guānniàn	zhǔnbèi
bāokuò	gōngyuán	cúnzài	chéngdù	jiēduàn	jiànzhù	jìhuà	jiéguǒ
xiānsheng	péngyou	fāzhǎn	kǎolǜ	dàjiā	jiàoshòu	guójiā	kěnéng
jiātíng	wǒmen	fāngfǎ	bàodào	shìjiè	qīngchu	yánzhòng	běnshēn
chénglì	nénggòu	zuótiān	yuánlái	zuìjìn	chénggōng	xīnwén	búhuì
háishì	guòqù	guójì	yǐjīng	búyào	měiguó	dàxué	qíngkuàng
guǎnlǐ	lǎoshī	hǎoxiàng	kèchéng	zìwǒ	jiàzhí	nénglì	mǔqīn
xūyào	shìjiàn	rúguǒ	dòngwù	zhùyì	zěnme	jīngguò	zìyóu
chǔlǐ	chǎnshēng	qǐlai	yīdìng	fēicháng	xuéxí	suǒyǐ	miànduì
zhèlǐ	yǔyán	kōngjiān	fāngshì	zuòyè	bùtóng	ānquán	wénhuà
nàme	xīwàng	bùnéng	tóngxué	bùduàn	shénme	jiějué	xiàwǔ
xíngwéi	shíhou	chéngzhǎng	tóngshí	xiānggǎng	yīyàng	jìxù	xuǎnzé
dànshì	yěxǔ	zhíde	zhǐyào	yǒuguān	yuányīn	kěyǐ	shídài
jīběn	jīhuì	dūshì	dédào	zàochéng	zhème	xíngdòng	méiyǒu
xùnliàn	tèbié	bìxū	xiànxiàng	zhèngzhì	fùmǔ	jīngjì	jiàoyù
yìzhí	lìngwài	chǎnpǐn	xǐhuan	shízài	bǐjiào	chūxiàn	juéde

Table 5A

Taiwan Mandarin Female Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
工程	開始	電話	好像	感覺	安全	傳統	過程
問題	報導	專業	媽媽	然後	參加	爸爸	快樂
發展	當然	大學	電視	發生	產生	大家	管理
建築	教育	結果	故事	交通	開發	記者	後來
同學	老師	了解	什麼	同時	開放	生活	事實
完成	原來	也許	相關	文化	增加	自我	選擇
動物	程度	成立	達到	出現	作業	包括	除了
負責	課程	活動	方法	公園	或者	加上	國家
面對	可是	可能	沒有	例如	科技	可以	歷史
清楚	設備	社會	目的	起來	身體	甚至	人生
未來	時候	受到	特別	我們	雖然	所以	他們
行為	只是	重要	以後	學生	知識	支持	下午
不能	標準	表現	不同	不斷	變成	最後	必須
成功	得到	東西	成為	但是	都市	非常	處理
還是	改變	溝通	基本	教授	觀念	國際	過去
今年	考慮	空間	幾乎	機會	就是	目標	階段
世界	時間	實際	條件	現在	容易	生命	完全
訓練	應該	嚴重	新聞	壓力	意義	一直	學習
資料	政治	這樣	成長	直接	知道	原因	投資
朋友	一般	昨天	行動	需要	努力	電腦	造成
或是	美國	存在	不要	語言	自由	時代	如何
現象	有關	運動	價值	規定	今天	準備	覺得
獲得	本身	配合	經驗	家庭	值得	母親	醫院
怎麼	這裡	情況	父母	希望	經過	影響	其實
工作	公司	最近	內容	團體	台灣	接受	地方

Table 5B

Taiwan Mandarin Male Bisyllabic Half-lists in Rank Order from Easiest to Most Difficult (Romanized Form)

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
gōngchéng	kāishǐ	diànhuà	hǎoxiàng	gǎnjué	ānquán	chuántǒng	guòchéng
wèntí	bàodào	zhuānyè	māma	ránhòu	cānjiā	bàba	kuàilè
fāzhǎn	dāngrán	dàxué	diànshì	fāshēng	chǎnshēng	dàjiā	guǎnlǐ
jiànzhù	jiàoyù	jiéguǒ	gùshì	jiāotōng	kāifā	jìzhě	hòulái
tóngxué	lǎoshī	liǎojiě	shénme	tóngshí	kāifàng	shēnghuó	shìshí
wánchéng	yuánlái	yěxǔ	xiāngguān	wénhuà	zēngjiā	zìwǒ	xuǎnzé
dòngwù	chéngdù	chénglì	dádào	chūxiàn	zuòyè	bāokuò	chúle
fùzé	kèchéng	huódòng	fāngfǎ	gōngyuán	huòzhě	jiāshàng	guójiā
miànduì	kěshì	kěnéng	méiyǒu	lìrú	kējì	kěyǐ	lìshǐ
qīngchu	shèbèi	shèhuì	mùdì	qǐlai	shēntǐ	shènzhì	rénshēng
wèilái	shíhou	shòudào	tèbié	wǒmen	suīrán	suǒyǐ	tāmen
xíngwéi	zhǐshì	zhòngyào	yǐhòu	xuésheng	zhīshi	zhīchí	xiàwǔ
bùnéng	biāozhǔn	biǎoxiàn	bùtóng	bùduàn	biànchéng	zuìhòu	bìxū
chénggōng	dédào	dōngxī	chéngwéi	dànshì	dūshì	fēicháng	chǔlǐ
háishì	gǎibiàn	gōutōng	jīběn	jiàoshòu	guānniàn	guójì	guòqù
jīnnián	kǎolǜ	kōngjiān	jīhū	jīhuì	jiùshì	mùbiāo	jiēduàn
shìjiè	shíjiān	shíjì	tiáojiàn	xiànzài	róngyì	shēngmìng	wánquán
xùnliàn	yīnggāi	yánzhòng	xīnwén	yālì	yìyì	yìzhí	xuéxí
zīliào	zhèngzhì	zhèyàng	chéngzhǎng	zhíjiē	zhīdao	yuányīn	tóuzī
péngyou	yībān	zuótiān	xíngdòng	xūyào	nǔlì	diànnǎo	zàochéng
huòshì	měiguó	cúnzài	búyào	yǔyán	zìyóu	shídài	rúhé
xiànxiàng	yǒuguān	yùndòng	jiàzhí	guīdìng	jīntiān	zhǔnbèi	juéde
huòdé	běnshēn	pèihé	jīngyàn	jiātíng	zhíde	mǔqīn	yīyuàn
zěnme	zhèlǐ	qíngkuàng	fùmǔ	xīwàng	jīngguò	yǐngxiǎng	qíshí
gōngzuò	gōngsī	zuìjìn	nèiróng	tuántǐ	táiwān	jiēshòu	dìfāng

Table 6

Mean Performance of Taiwan Mandarin Male Bisyllabic Lists and Half-lists

List	Intercept[a]	Slope[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	50% Threshold[e]	Change to Midpoint[f]
1	2.2178	-0.4078	10.2	8.8	5.4	1.0
2	2.2178	-0.4078	10.2	8.8	5.4	1.0
3	1.8779	-0.3496	8.7	7.6	5.4	0.9
4	1.9625	-0.3649	9.1	7.9	5.4	0.9
M	2.0690	-0.3825	9.6	8.3	5.4	0.9
Minimum	1.8779	-0.4078	8.7	7.6	5.4	0.9
Maximum	2.2178	-0.3496	10.2	8.8	5.4	1.0
Range	0.3399	0.0582	1.5	1.3	0.1	0.1
SD	0.1753	0.0299	0.7	0.6	0.0	0.0
1A	2.0085	-0.3733	9.3	8.1	5.4	0.9
1B	2.4885	-0.4531	11.3	9.8	5.5	1.0
2A	2.0573	-0.3822	9.6	8.3	5.4	0.9
2B	2.4104	-0.4389	11.0	9.5	5.5	1.0
3A	1.8019	-0.3359	8.4	7.3	5.4	0.9
3B	1.9625	-0.3649	9.1	7.9	5.4	0.9
4A	1.8779	-0.3496	8.7	7.6	5.4	0.9
4B	2.0573	-0.3822	9.6	8.3	5.4	0.9
M	2.0830	-0.3850	9.6	8.3	5.4	0.9
Minimum	1.8019	-0.4531	8.4	7.3	5.4	0.9
Maximum	2.4885	-0.3359	11.3	9.8	5.5	1.0
Range	0.6866	0.1172	2.9	2.5	0.1	0.1
SD	0.2432	0.0410	1.0	0.9	0.1	0.1

[a]regression intercept. [b]regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the 50% threshold of a list to the mean 50% threshold for male and female lists (4.5 dB HL).

Table 7

Mean Performance of Taiwan Mandarin Female Bisyllabic Lists and Half-lists

List	Intercept[a]	Slope[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	50% Threshold[e]	Change to Midpoint[f]
1	1.0684	-0.3054	7.6	6.6	3.5	-1.0
2	1.1768	-0.3288	8.2	7.1	3.6	-0.9
3	1.0794	-0.3082	7.7	6.7	3.5	-1.0
4	1.1157	-0.3133	7.8	6.8	3.6	-0.9
M	1.1101	-0.3139	7.8	6.8	3.5	-0.9
Minimum	1.0684	-0.3288	7.6	6.6	3.5	-1.0
Maximum	1.1768	-0.3054	8.2	7.1	3.6	-0.9
Range	0.1083	0.0233	0.6	0.5	0.1	0.1
SD	0.0488	0.0104	0.3	0.2	0.0	0.0
1A	1.1137	-0.3169	7.9	6.9	3.5	-1.0
1B	1.0265	-0.2950	7.4	6.4	3.5	-1.0
2A	1.2164	-0.3434	8.6	7.4	3.5	-0.9
2B	1.1412	-0.3156	7.9	6.8	3.6	-0.9
3A	1.1627	-0.3295	8.2	7.1	3.5	-0.9
3B	1.0067	-0.2900	7.3	6.3	3.5	-1.0
4A	1.2164	-0.3434	8.6	7.4	3.5	-0.9
4B	1.0332	-0.2892	7.2	6.3	3.6	-0.9
M	1.1146	-0.3154	7.9	6.8	3.5	-0.9
Minimum	1.0067	-0.3434	7.2	6.3	3.5	-1.0
Maximum	1.2164	-0.2892	8.6	7.4	3.6	-0.9
Range	0.2097	0.0542	1.4	1.2	0.1	0.1
SD	0.0843	0.0224	0.6	0.5	0.0	0.0

[a]regression intercept. [b]regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the 50% threshold of a list to the mean 50% threshold for male and female lists (4.5 dB HL).

specified intensity level. The percent of correct values which were yielded from Equation 1 were subsequently used to construct psychometric functions.

$$P = (1 - \frac{\exp(a + b \times i)}{1 + \exp(a + b \times i)}) \times 100 \qquad (1)$$

The following is a description of Equation 1: *P* is percentage of correct recognition, *a* is the regression intercept, *b* is the regression slope, *i* is the intensity level of presentation in dB HL. The percentage of correct word recognition at any specified intensity level is predictable when inserting the regression slope, regression intercept, and intensity level into Equation 1. Thus the percentage of correct word recognition was predicted via Equation 1 for each of the bisyllabic lists and half-lists. The range of presentation intensity levels was -5 to 40 dB HL in 5 dB increments. After the percentages were predicted using Equation 1, psychometric functions were constructed. Equation 2 was then used to find the threshold (presentation intensity required for 50% word recognition performance), the slope at threshold, and the slope from 20 to 80% for the bisyllabic lists and half-lists. The calculation was performed by inserting specific proportions into Equation 2.

$$i = \frac{\log \frac{p}{1 - p} - a}{b} \qquad (2)$$

In Equation 2, *i* is the presentation level in dB HL, *p* is the proportion of correct recognition, *a* is the regression intercept, and *b* is the regression slope. Presented in Table 6 (male) and Table 7 (female) are the results for threshold, slope at threshold, and slope from 20 to 80% for each list and half-list.

After the lists and half-lists were compiled, a two-way Chi-Square (χ^2) analysis (intensity and list as independent variables with response as the dependent variable) was completed in order to discern any statistically significant differences among the bisyllabic 50-word lists or 25-word half-lists. This Chi-Square analysis was conducted as part of the logistic regression analysis of the lists and half-lists. The results of the analysis indicated that there were no significant differences among the 50-word lists for the male and female talkers, $\chi^2(3, N = 20) = 1.97$, $p = 0.578$, and $\chi^2(3, N = 20) = 0.26$, $p = 0.965$. Results also indicated that there was no significant differences found among the 25-word half-lists for the male and female talkers, $\chi^2(7, N = 20) = 3.75$, $p = 0.808$ and $\chi^2(7, N = 20) = 0.98$, $p = 0.995$. There were no significant intensity x list interactions, which indicated that there were no differences among the psychometric function slopes for the lists or half-lists. Although there were not any statistically significant differences among the 50-word lists or 25-word half-lists, intensity level adjustments were digitally completed by way of Sadie Disk Editor software (Studio Audio & Video Ltd., 2004). This was performed in order to increase the psychometric equivalency of the lists and half-lists. The intensity of each word from the male and female bisyllabic lists and half-lists was adjusted digitally so that the 50% threshold of each list was equal to the midpoint (4.5 dB HL) between the mean threshold of the eight male half-lists and the mean threshold of the eight female half-lists. Presented in Table 6 (male) and Table 7 (female) are the intensity adjustments which were made to each word in the four lists and eight half-lists. Figure 1 exhibits the psychometric functions for the male talker and female talker bisyllabic lists and half-lists prior to the intensity adjustments. Figure 2 represents the psychometric functions for the female talker and male talker bisyllabic lists and half-lists after the intensity adjustments were performed to produce 50% performance at 4.5 dB HL. Figure 3 shows the mean psychometric functions for the combined male

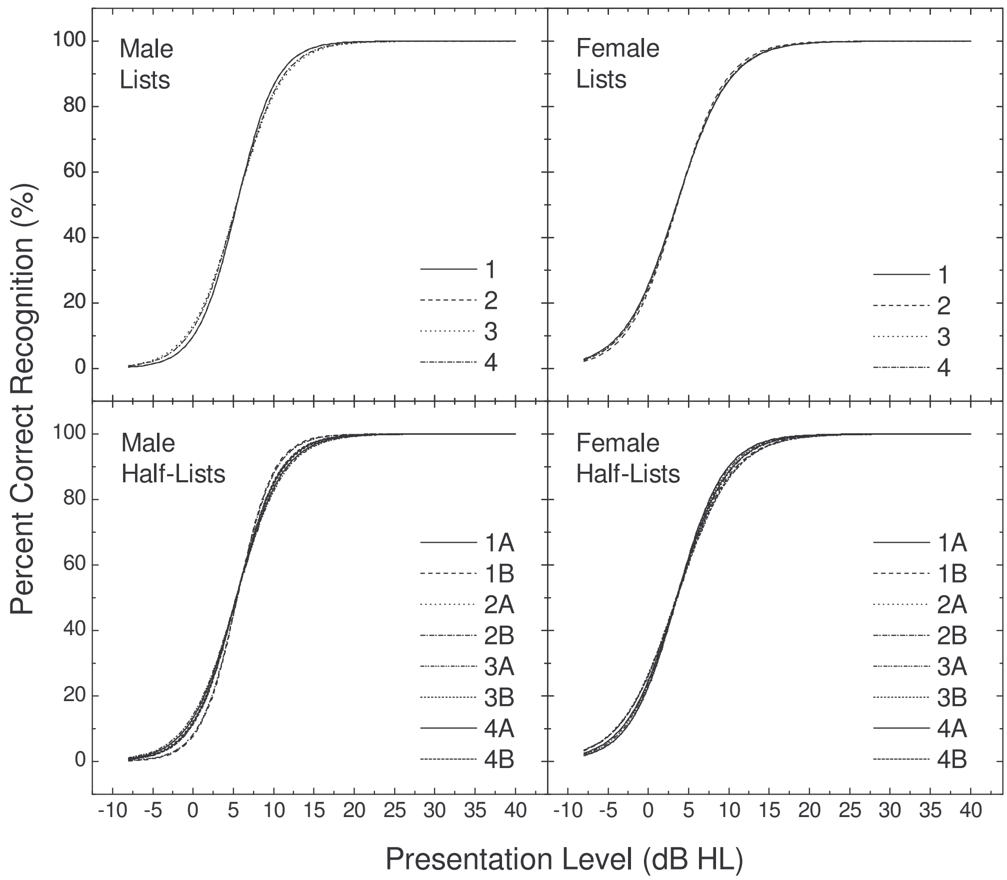

Figure 1. Psychometric functions for the four Taiwan Mandarin bisyllabic lists and eight half-lists for male talker and female talker recordings before intensity adjustments.

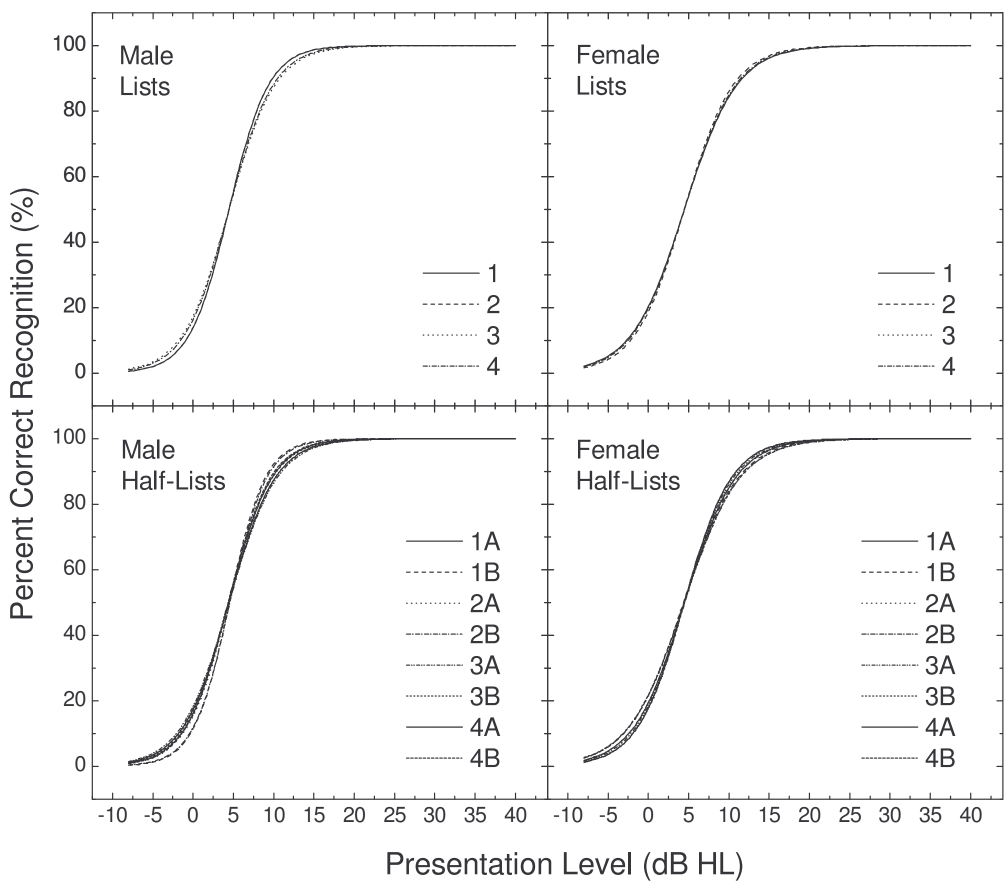

Figure 2. Psychometric functions for the four Taiwan Mandarin bisyllabic lists and eight half-lists for male talker and female talker recordings after intensity adjustments to produce 50% performance at 4.5 dB HL.

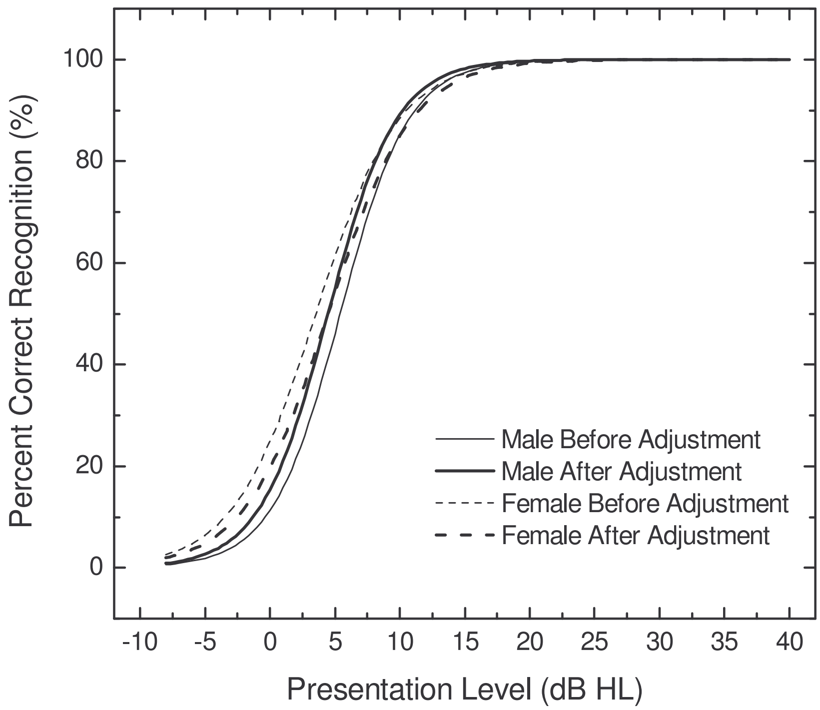

Figure 3. Mean psychometric functions for male and female Taiwan Mandarin talker monosyllabic lists before and after intensity adjustment.

and female talker bisyllabic lists both before and after the intensity adjustments. The predicted psychometric functions and those created after the intensity adjustments were very similar for both the male and female talker lists. When comparing Figure 1 and Figure 2, it is evident that only minor adjustments (≤1 dB) to the words in the lists and half-lists were necessary to equate the 50-word lists and 25-word half-lists for the male and female talkers.

Discussion

The purpose of the current study was to construct a set of bisyllabic Taiwan Mandarin word recognition lists and half-lists for use in word recognition testing. Figures 1-3 indicate that these 50-word lists and 25-word half-lists have been developed which are relatively homogenous in performance with regard to audibility and psychometric function slope. Results from a two-way Chi-square (χ^2) analysis indicated that there were no statistically significant differences among any of the bisyllabic 50-word lists or 25-word half-lists for subjects with normal hearing.

In review of Table 6 (male) and Table 7 (female), it is evident that at the 50% location the psychometric function slopes for the bisyllabic lists and half-lists ranged from (M = 9.6 %/dB) for the male recordings, which were somewhat steeper than those for the female recordings (M = 7.9 %/dB). From 20 to 80%, the psychometric function slopes followed a similar pattern, with the male recordings (M = 8.3 %/dB) having steeper slopes than the female recordings (M = 6.8 %/dB). These results indicate that the psychometric function slopes which were calculated at the linear (50%) point of the function were approximately 1% higher than those measured from the 20 to 80% points of the psychometric functions.

Currently the means of the Taiwan Mandarin word recognition materials are slightly higher than those of English word recognition materials when measured from the 20 to 80% points on each function. Specifically the English word lists have a reported mean slope of 4.2 %/dB for the NU-6 word lists and a mean slope of 4.6 %/dB for the CID W-22 word lists (Beattie, Edgerton, & Svihovec, 1977). Wilson and Oyler in 1997 found when using recordings from the Auditec of St. Louis CD, that the NU-6 word lists had a mean of 4.4 %/dB and CID W-22 word lists had a mean of 4.8 %/dB.

In this study the four lists and eight half-lists were not phonemically balanced. In part this is due to the fact that it is extremely difficult to represent all the phonemes in a given language in a short list of 25 or 50 words. Also it is almost impossible to design a short list of words in citation form that can account for the coarticulation that naturally occurs in connected speech. Therefore the lists were not phonemically balanced, however they were constructed to adhere to the criteria of familiarity and auditory homogeneity.

Although this study has succeeded in developing an initial set of digitally recorded word recognition materials for speakers of Taiwan Mandarin, there remains a need for additional research. For example, the word recognition lists were made equivalent from materials spoken by both male and female talkers using subjects with normal hearing. Thus the lists may not accurately assess word recognition abilities in subjects with hearing impairments. Recently a study was conducted where 24 young adults with normal hearing and 72 older adults with sensorineural hearing loss were administered speech audiometry tests. The lists which were evaluated were made equivalent on a population of young adult subjects with normal hearing. The results of the study found that there was high performance variability across the lists for those subjects with hearing loss, but for the subjects with normal hearing the variability was relatively low. The results indicated that lists which are equivalent for normal hearing individuals may not be equivalent for those with hearing loss (McArdle & Wilson, 2006). Jerger (2006) reiterated the importance of this type of "population specific" research when he concluded that it is imperative that speech testing materials be developed on a sample of the intended test population and not on a group of normal hearing young adults. For our study the lists were made equivalent using listeners with normal hearing, and not on those with hearing loss due largely to the limited availability of subjects from Taiwan in the Utah area. Thus, an extension of this study would be to create equivalent lists of Taiwan Mandarin words utilizing listeners with a hearing impairment.

Wilson and McArdle (2005) found that word recognition abilities in background noise are difficult to predict using word recognition performance which was obtained in quiet. Based on this study it would also be of benefit for patients with hearing losses to be administered word recognition tests in the presence of background noise to determine proper intervention, thus indicating a need for word recognition lists and half-lists to be created with background noise to more accurately determine how the hearing loss is affecting intelligibility of words in everyday life. In order to establish equivalent lists in regard to familiarity and auditory homogeneity with background noise, the Taiwan Mandarin words could be equated in the presence of multi-talker babble.

Historically, word recognition testing utilized full lists (50 words) during testing. With time however, audiologists have began presenting only half of the full lists in an effort to avoid patient fatigue and save time during testing (Penrod, 1980). Although in this study both full lists and half-lists were provided, further research should study which length is most effective. Gelfand (1998) found that based on curves representing the variability of word recognition test scores based on the binomial model, 450 scoreable items are needed to achieve optimal reliability for a word recognition test. In order to save time during testing, Gelfand utilized phonemic scoring to decrease the number of items on the test yet still achieve the target 450 scorable items. Thus some audiologists have turned to phonemic scoring as a means for increasing the number of items scored while decreasing the number of words presented, thereby striking a balance between validity and testing time.

Another area for further research is the validity of the lists and half-lists for use in testing children. This study has aimed to produce speech audiometry materials for adult listeners however such materials may be inappropriate for younger individuals. Currently it is unknown how a child would perceive the words on the lists. If children are found to not understand the words in the lists and half-lists, then the words become nonsense stimuli, which may change the validity of the list equivalency. Thus it is vital that the stimuli be familiar to the subject. Currently the lists contain words such as 經濟 (*economy*) which is generally not a familiar word to the average child. However some words which are a part of the test may indeed be familiar to a child such as 學校 (*school*). Ashoor and Prochazka (1985) developed Saudi Arabic speech audiometry materials which were created specifically for children. These materials were selected from children's books and evaluated by younger listeners. It is evident that the materials developed in this study need further examination to determine if they are appropriate for children.

This study utilized logistic regression to calculate the psychometric slope of each word. However, in the literature three different types of data analysis are commonly used: logistic regression, 3rd order polynomial, and simply plotting the raw data. It is possible that utilizing a different calculation technique may change the psychometric functions of the words. For instance, utilizing logistic regression as used in this study compared to 3rd order polynomial on the same data will likely result in different psychometric functions (S. L. Nissen, personal communication, May 15, 2006). This is due to the fact that with 3rd order polynomial the data is forced into an s-curve so it creates tails at the beginning and end of the data. It would be of interest to know which of the three methods for figuring the slopes is most accurate and what the advantages or possible disadvantages are of using each of the three different methods for figuring the slopes.

In the current study each subject only listened to the stimuli on one occasion. Therefore there is some question as to the test-retest reliability of the data gathered in this study. Generally there is a natural amount of slight variability when subjects are re-tested. However, further study is needed to assess to what degree performance of the words would change if each subject were to be re-tested and in turn how this variability would affect the validity of the list equivalency.

Finally, further research is needed to detect whether regional dialect has a real and measurable affect on the results of word recognition testing. To date there is limited research on how a regional dialect might affect word recognition testing. In the study by Weisleder and Hodgson (1989), findings indicated that a regional dialect can affect the validity of word recognition testing for the various Spanish dialects; however it is unknown whether this is also true for Mandarin. Therefore, a study should be conducted in which an individual from Mainland China and an individual from Taiwan are administered both a word recognition test in Standard Mandarin and in Taiwan Mandarin. The results of the tests can then be utilized to determine the true effects of regional dialect of Mandarin on word recognition testing performance.

Creating the lists and half lists of Taiwan Mandarin bisyllabic words to be used in word recognition testing required a considerable amount of time to identify, evaluate and digitally record. However, the use of digital speech audiometry materials on CD allows the audiologist many benefits, such as the ability to select which words or test stimuli to give from a longer list of items, varying inter-stimulus intervals, and presentation order randomization with assistance from a computer program.

Conclusion

In review, the specific aims of this project were to create digitally recorded bisyllabic Taiwan Mandarin word recognition lists and half-lists of familiar words. The recorded 50-word lists and 25-word half-lists for both the male and female talkers are relatively homogenous in regard to audibility and psychometric function slope. The results from this study will be used to create a compact disk of digitally recorded materials to be used for word recognition testing for individuals whose native language is Taiwan Mandarin.

References

Academia Sinica Computing Center. (1997). *Academia Sinica Balanced Corpus of Modern Chinese*. Retrived March 2, 2005, from http://www.sinica.edu.tw/ftms-bin/kiwi1/mkiwi.sh?language=1

Aleksandrovsky, I. V., McCullough, J. K., & Wilson, R. H. (1998). Development of suprathreshold word recognition test for Russian-speaking patients. *Journal of the American Academy of Audiology, 9,* 417-425.

American National Standards Institute. (1999). *Maximum permissible ambient noise levels for audiometric test rooms*. ANSI S3.1-1999. New York: ANSI.

American National Standards Institute. (2004). *Specification for audiometers*. ANSI S3.6-2004. New York: ANSI.

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

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

Ashoor, A. A., & Prochazka, T. (1985). Saudi Arabic speech audiometry for children. *British Journal of Audiology, 19,* 229-238.

Beattie, R. C., Edgerton, B. J., & Svihovec, D. V. (1977). A comparison of the Auditec of St. Louis cassette recordings of NU-6 and CID W-22 on a normal-hearing population. *Journal of Speech and Hearing Disorders, 42,* 60-64.

Campbell, R. A. (1965). Discrimination test word difficulty. *Journal of Speech and Hearing Research, 8,* 13-22.

Cheng, R. L. (1985). A comparison of Taiwanese, Taiwan Mandarin and Peking Mandarin. *Language, 61,* 352-377.

Christensen, L. K. (1995). *Performance intensity functions for digitally recorded Spanish speech audiometry.* Unpublished master's thesis. Brigham Young University, Provo, UT.

Dennis, J. M., & Neely, J. G. (1991). Basic hearing tests. *Otolaryngologic Clinics of North America, 24*, 253-276.

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

Epstein, A. (1978). Speech audiometry. *Otolaryngologic Clinics of North America, 11,* 667-676.

Gelfand, S. A. (1998). Optimizing the reliability of speech recognition scores. *Journal of Speech, Language, and Hearing Research, 41,* 1088-1102.

Greer, L. F. (1997). Performance intensity functions for digitally recorded Italian speech audiometry materials. Unpublished master's thesis. Brigham Young University, Provo, UT.

Hagerman, B. (1984). Some aspects of methodology in speech audiometry. *Scandinavian Audiology Supplementum, 21,* 1-25.

Harris, R. W., Crawford, T., & Mastny, P. (2004). Japanese speech audiometry materials. [Compact Disc]. Provo, UT: Brigham Young University.

Harris, R. W., Goffi, M. V. S., Pedalini, M. E. B., Merrill, A., & Gygi, M. A. (2001). Psychometrically equivalent Brazilian Portuguêse bisyllabic word recognition materials spoken by male and female talkers. *Pró-Fono, 13,* 249-262.

Harris, R. W., & Kim, E. (2004). Korean speech audiometry materials. [Compact Disc]. Provo, UT: Brigham Young University.

Harris, R. W., Nissen, S. L., & Jennings, L. (2004). Mandarin Chinese speech audiometry materials: Beijing dialect. [Compact Disc]. Provo, UT: Brigham Young University.

Harris, R. W., & Pola, M. G. (2002). Russian speech audiometry materials. [Compact Disc]. Provo, UT: Brigham Young University.

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,* 321-337.

Hood, J. D., & Poole, J. P. (1977). Improving the reliability of speech audiometry. *British Journal of Audiology, 11,* 93-101.

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.

Jerger, J. (2006). Are some more equal than others? *Journal of the American Academy of Audiology, 17,* 156.

Katzner, K. (1986). *The Languages of the World*. London & New York: Routledge.

Kubler, C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics, 27,* 156-176.

Li, D. C. (1985). Problems and trends of standardization of Mandarin Chinese in Taiwan. *Anthropological Linguistics, 27,* 122-140.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics, 39,* 155-158.

Mangum, T. C. (2005). Performance intensity functions for digitally recorded Japanese speech audiometry materials. Unpublished master's thesis. Brigham Young University, Provo, UT.

Martin, F. N., Armstrong, T. W., & Champlin, C. A. (1994). A survey of audiological practices in the United States. *American Journal of Audiology, 3,* 20-26.

McArdle, R. A., & Wilson, R. H. (2006). Homogeneity of the 18 QuickSIN Lists. *Journal of the American Academy of Audiology, 17,* 157-167.

Nissen, S. L., Harris, R. W., Jennings, L., Eggett, D. L., & Buck, H. (2005). Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *International Journal of Audiology, 44,* 379-390.

Penrod, J. P. (1980). A comparison of half-vs full-list speech discrimination scores in a hearing-impaired geriatric population. *The Journal of Auditory Research, 20,* 181-186.

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

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,* 23-28.

Roup, C. M., Wiley, T. L., Safady, S. H., & Stoppenbach, D. T. (1998). Tympanometric screening norms for adults. *American Journal of Audiology, 7,* 55-60.

Studio Audio & Video Ltd. (2004). *SADiE disk editor software*. [5.2.2]. *[Computer Program]*. Ely, Cambridgeshire, UK: Studio Audio & Video Ltd.

von, H. D., Champlin, C. A., & Shetty, S. R. (2004). Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. *Journal of the American Academy of Audiology, 15,* 88-98.

Weisleder, P., & Hodgson, W.R. (1989). Evaluation of four Spanish word-recognition-ability lists. *Ear and Hearing, 10,* 387-392.

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,* 7-14.

Wilson, R. H., & Oyler, A. L. (1997). Psychometric functions for the CID W-22 and NU Auditory Test No. 6. Materials spoken by the same speaker. *Ear and Hearing, 18,* 430-433.

Wilson, R. H., & McArdle, R. (2005). Speech signals used to evaluate functional status of the auditory system. *Journal of Rehabilitation Research and Development, 42*, 79-94.

Wilson, W., & Moodley, S. (2000). Use of the CID W22 as a South African English speech discrimination test. *South African Journal of Communication Disorders, 47,* 57-62.

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,* 586-593.

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes, 10,* 545-600.

Appendix A

INFORMED CONSENT
RESEARCH PARTICIPATION FORM

Participant:___________________________________Age:___________

You are asked to participate in a research study sponsored by the Department of Audiology and Speech Language Pathology at Brigham Young University, Provo, Utah. The faculty director of this research is Richard W. Harris, Ph.D. Students in the Audiology and Speech-Language Pathology program may assist in data collection.

This research project is designed to evaluate a word list recorded using improved digital techniques. You will be presented with this list of words at varying levels of intensity. Many will be very soft, but none will be uncomfortably loud to you. You may also be presented with this list of words in the presence of a background noise. The level of this noise will be audible but never uncomfortably loud to you. This testing will require you to listen carefully and repeat what is heard through earphones or loudspeakers. Before listening to the word lists, you will be administered a routine hearing test to determine that your hearing is normal and that you are qualified for this study.

It will take approximately two hours to complete the test. Testing will be broken up into 2 or 3 one hour blocks. Each subject will be required to be present for the entire time, unless prior arrangements are made with the tester. You are free to make inquiries at any time during testing and expect those inquiries to be answered.

As the testing will be carried our in standard clinical conditions, there are no known risks involved. Standard clinical test protocol will be followed to ensure that you will not be exposed to any unduly loud signals.

Names of all subjects will be kept confidential to the investigators involved in the study. Participation in the study is a voluntary service and no payment of monetary reward of any kind is possible or implied.

You are free to withdraw from the study at any time without any penalty, including penalty to future care you may desire to receive from this clinic.

If you have any questions regarding this research project you may contact Dr. Richard W. Harris, 131 TLRB, Brigham Young University, Provo, Utah 84602; phone (801) 422-6460. If you have any questions regarding your rights as a participant in a research project you may contact Dr. Renea Beckstrand, Chair of the Institutional Review Board, 422 SWKT, Brigham Young University, Provo, Utah 84602; phone (801) 422-3873, email: renea_beckstrand@byu.edu.

YES: I agree to participate in the Brigham Young University research study mentioned above. I confirm that I have read the preceding information and disclosure. I hereby give my informed consent for participation as described.

____________________________________ __________________
Signature of Participant Date

____________________________________ __________________
Signature of Witness Date