AUDIOFONOLOGIA 2004, 26

Problemy teorii i praktyki
Problems of theory and practice

Psychometrycznie zrównoważone dwusylabowe polskie listy słowne

Psychometrically equivalent Polish bisyllabic words

Richard W. Harris[1], Wayne S. Nielson[1], David L. McPherson[1], Henryk Skarżyński[2]

[1] Brigham Young University, Provo, Utah, USA
[2] Institute of Physiology and Pathology of Hearing, Warsaw, Poland

Streszczenie

Celem pracy było opracowanie, zrównoważonych psychometryczne list słów dwusylabowych w języku polskim przeznaczonych do audiometrii słownej. Jako materiał słowny wybrano 70 słów dwusylabowych akcentowanych na pierwszą sylabę, potocznie używanych w języku polskim. Materiał słowny nagrano z zastosowaniem techniki zapisu cyfrowego. Słowa były wypowiadane zarówno przez mężczyznę jak i kobietę. Lektorzy pochodzili z obszaru centralnej Polski. W grupie 26 osób dobrze słyszących dla każdego z 70 słów wykreślono krzywą psychometryczną i dla 15 poziomów natężenia (od -10 do 18 dB HL ze skokiem 2 dB). Ostatecznie wybrano 25 dwusylabowych słów o najbardziej stromym nachyleniu krzywej psychometrycznej. W celu poprawy jednorodności tak utworzonej listy dokonano cyfrowej regulacji natężenia każdego ze słów, tak aby próg zrozumiałości każdego z nich wynosił 2,37 dB HL. Wartość średnia nachylenia krzywej zrozumiałości dla wybranych słów dwusylabowych wynosiła 10,1 %/ dB dla list wypowiadanych głosem męskim oraz 9,8 %/dB wypowiadanych głosem żeńskim. W efekcie opracowano listę słów dwusylabowych, które były jednorodne pod kątem wartości progów słyszenia, a także pod kątem nachylenia krzywej psychometrycznej. Opracowane w pracy, zrównoważone psychometrycznie polskie listy słów dwusylabowych służące do badań audiometrii słownej, wypowiadane przez lektora mężczyznę i kobietę dostępne są na płycie CD pt. Polskie Listy Słowne dla Audiometrii Mowy, Brigham Yopung University (Dysk 1.0)

Słowa kluczowe: próg rozumienia mowy, język polski, dwusylabowe, rozróżnianie mowy, słuchowe rozpoznawanie słów.

Summary

This investigation was undertaken to develop, digitally record, evaluate, and psychometrically equate Polish bisyllabic words for use in measurement of the speech reception threshold (SRT). Seventy familiar bisyllabic words, with stress on the first syllable, were selected for the SRT test materials. These words were recorded by both male and female talkers native to Poland who spoke the standard central Polish dialect. Psychometric functions were calculated for each of the 70 bisyllabic words using 26 normally hearing subjects who listened to these words at 15 intensity levels from -10 to 18 dB HL in 2 dB increments. The 25 best bisyllabic words were selected and then digitally adjusted, with respect to intensity, so that the threshold of each word was equal to the mean PTA (2.37 dB HL) of the normally hearing subjects. The mean slope (%/dB) for the bisyllabic words was 10.1 %/db for the male talker and 9.8 %/dB for the female talker. The result was the development of a list of Polish bisyllabic words (male and female talker) which were homogeneous with respect to threshold audibility and also with respect to psychometric function slope. These psychometrically equivalent Polish bisyllabic words spoken by both male and female talkers are included on the Brigham Young University Polish Speech Audiometry Materials (Disc 1.0) compact disc.

Key words: speech reception threshold, Polish, bisyllabic, speech discrimination, auditory word recognition.

The speech reception threshold (SRT) and auditory word recognition score are among the most fundamental measures obtained during speech audiometry evaluations. The SRT is defined as the lowest level, in dB HL, at which an individual can correctly identify 50% of the spondaic words that are presented [American Speech-Language-Hearing Association (ASHA) Committee on Audiologic Evaluation 1988].

Over the years, many researchers have contributed to the development and standardization of speech audiometry materials in English. Following much research and evaluation, spondaic words have been chosen as the preferred stimuli for SRT testing, and monosyllabic words have become the preferred stimuli for auditory word recognition testing in English [ASHA Committee on Audiologic Evaluation 1988]. A spondaic word is defined as a two syllable word with equal stress on both syllables.

Several attributes of recorded word lists have been, or still are considered influential in the standardization of such lists. Included in these variables are word familiarity, word length, dialect of the talker, and mode of presentation. During the course of this review, each of these variables will be discussed.

Word lists may be presented by means of a recording or by monitored live voice (MLV). The ASHA Committee on Audiologic Evaluation [1988] has endorsed recorded materials as the preferred manner for presenting speech stimuli. Of the several recording methods, the digital method is considered the most favorable due to the quality, versatility and durability of the recordings it produces [Kamm, Carterette, Morgan, & Dirks 1980; Ridgway 1986].

Speech audiometry materials in other languages have not been as well developed or standardized as those in English. Recently, there have been efforts to improve the quality of Brazilian Portuguese, Spanish, Italian, and Russian speech audiometry materials [Aleksandrovsky, McCullough, & Wilson 1998; Christensen 1995; Greer 1997; Harris, Goffi, Pedalini, Gygi and Merrill 2001; Weisleder & Hodgson 1989]. In the midst of these efforts, important discoveries have been made as to how the structure of a language affects the efficacy of a word list.

The development of standardized digital recordings of speech audiometry materials in languages other than English is of interest to audiologists within the United States as well as those in other countries. According to a survey of American audiologists, 37% perform speech audiometry in languages other than English [Martin & Sides 1985].

The current project was undertaken in an effort to further contribute to the development of digital speech audiometry materials in languages other than English. The aim of this project was to produce standardized lists in Polish for use in measurement of the SRT and auditory word recognition. These lists may be used within the United States when testing individuals whose native language is Polish and in Poland during standard audiometric evaluations.

Through research efforts at Brigham Young University, digitally recorded word lists have been produced in English, Spanish, Italian and Brazilian Portuguese [Harris and Hilton 1991; Christensen 1995; Greer 1997; Harris et al. 2001]. During the development of these lists, particular attention was paid to the selection of words that were frequently used in each language. After these lists were created, extensive efforts were made to standardize the lists by calculating psychometric functions and eliminating words that produced extreme results. Adjustments in intensity were made to make words or lists more homogeneous with respect to audibility. Similar methods were used in the creation of the Polish speech audiometry materials.

Review of Literature

Stimuli Used in SRT Testing

A variety of stimuli, including digits, sentences, and spondaic words, have been used in the evaluation of the SRT. The Western Electric 4A test, developed in 1926, was the first recorded test that was widely used for the evaluation of the SRT. This test was developed by means of a joint effort between the Bell Telephone Laboratories and the American Federation of Organizations of the Hard of Hearing. Digits were used as stimuli because they are highly familiar and because they are effective in the evaluation of very young children [ASHA 1988; Hudgins, Hawkins, Karlin, & Stevens 1947].

The efficacy of spondaic words was evaluated by Hudgins et al. [1947] at the Psycho-Acoustic Laboratory (PAL) at Harvard. A list of 42 spondaic words was compiled and entitled PAL Auditory Test No. 9. This list was then compared to both bisyllabic and monosyllabic word lists. After performance intensity data were obtained on all three types of materials, Hudgins et al. concluded that spondaic words have the steepest slope on a psychometric function and the highest degree of homogeneity. Hudgins et al. continued working with spondaic words and, as a result, later developed PAL Auditory Test No. 14 [Hirsh et al. 1952].

The primary deficiency of the PAL lists was that some of the words in the lists were not highly familiar to certain populations. In compiling the Central Institute for the Deaf (CID) Auditory Tests W-1 and W-2, Hirsh et al. [1952] ensured that the CID lists included only those words from the PAL lists that were highly familiar. The CID W-1 and CID W-2 word lists were also more rigidly phonetically balanced than the PAL lists.

The ASHA Committee on Audiologic Evaluation [1988] recommended that spondaic words be used for SRT testing. Spondaic words have come to be preferred because of their steep slope on the psychometric function and because of their homogeneity. In regard to the use of other stimuli for evaluation of the SRT, the Committee stated, "It should be recognized... that the use of speech stimuli with less homogeneity than spondaic words may compromise the reliability of this measure" (p. 86).

Stimuli Used in Auditory Word Recognition Testing

Telephone companies sponsored a majority of the initial research on auditory word recognition in an effort to determine which frequencies had to be transmitted over telephone lines to maintain speech intelligibility. After years of research at the research laboratories of the American Telephone and Telegraph Company and the Western Electric Company, Fletcher [1922] developed several monosyllabic word lists. These lists later became known as the Standard Articulation Word Lists.

During World War II, there was a renewal of interest in auditory word recognition testing because it could be used in the evaluation of military communications equipment. As a result of this renewed interest, Harvard Laboratories produced 20 phonetically balanced (PB) lists of 50 monosyllabic words each [Carhart 1965; Hudgins et al. 1947]. One of the problems with the PAL PB-50 word lists was that many of the words were unfamiliar to typical listeners. Hirsh et al. [1952] improved upon the PAL PB-50 word lists by eliminating the PB-50 words that did not appear in a tabulation of familiar words. The new lists created by Hirsh et al. were entitled the CID W-22 word lists.

In contrast to evaluation of the SRT, where redundancy and predictability clues are desirable in order to produce the lowest threshold possible, auditory word recognition testing requires novelty and lack of redundancy in order to make a patient's auditory word recognition difficulties evident [Carhart 1965]. For these reasons, monosyllabic words have become the preferred stimuli for auditory word recognition testing.

History of Polish Speech Audiometry

As reported by Zakrzewski, Pruszewicz, and Rydzewski [1973], the first Polish word lists for use in auditory word recognition testing were developed by Zakrzewski, Suwalski, Antkowski, and Suwalski in 1953. Zakrzewski et al. [1953] created 10 phonetically balanced lists of 50 monosyllabic words each. The methods used for the creation of the phonetically balanced lists were patterned after those used in the creation of the PAL PB 50 word lists and the Standard Articulation Word Lists.

Bystrzanowska [1978] and Zakrzewski et al. [1973] reported that no further developments in the field of Polish speech audiometry were made until 1961 when Taniewski, Kugler, and Wysocki created monosyllabic and bisyllabic word lists that were both phonetically and dynamically balanced. In 1963, Szmeja, Pruszewicz, and Dukiewicz [as cited by Bystrzanowska 1978] created word lists for the evaluation of auditory word recognition in school children. Other notable events in the 1960's included the development of a digits test by Iwankiewicz, the establishment of guidelines for the evaluation of hearing in children by the Institute for Mothers and Children, and the development of word lists for adults and children by Iwankiewicz and Siciński.

In 1971, Zakrzewski, Pruszewicz, and Kubzdela developed a series of Polish word lists. Each list contained both monosyllabic and bisyllabic words and was phonetically and structurally balanced. The series of lists consisted of 10 lists of 29 words each.

Bystrzanowska [1978] and Zakrzewski et al. [1973] reported that in order to make valid comparisons of results from speech audiometry performed at different locations across Poland, an organization was formed to establish guidelines for which word lists should be used by those who perform audiologic evaluations. The name of this organization was the Committee of the Audiology Section of the Polish Otolaryngological Society (PTOL). In 1974, this committee recommended the series of lists created by Zakrzewski et al. in 1971, the digits test by Iwankiewicz, the children's word list developed at the Institute for Mothers and Children by Borkowska-Gaertig, and the word list for children developed by Szmeja et al. in 1963.

After the recommendation of the PTOL in 1974, no developments of great significance occurred in the field of Polish Speech audiometry until the 1990s. In 1994, Pruszewicz, Demenko, Richter, and Wika developed 10 lists of 24 monosyllabic words each. The words for the lists were selected from among the most frequently used monosyllabic Polish nouns. The lists were not only phonemically, acoustically, and structurally balanced, but were semantically balanced as well.

Factors that Affect the Performance of Speech Audiometry Materials

Many factors have been examined in an effort to determine whether or not they affect the performance and test-retest reliability of speech audiometry materials. Such factors include: (a) word selection [Beattie, Svihovec, & Edgerton 1975; Campbell 1965; Hood & Poole 1980]; (b) talker variables [Brandy 1966; Cambron, Wilson, & Shanks 1991; Hood & Poole 1980; Kreul, Bell, & Nixon 1969; Penrod 1979]; (c) method of presentation [Beattie et al. 1975; Brandy 1966; Creston, Gillespie, & Krohn 1966]; and (d) type of recording [Kamm et al. 1980; Ridgway 1986].

Word selection. The words included in a word list influence the difficulty of a speech test. Some words are inherently easier to recognize than others. Words too difficult to recognize can contribute to a poor psychometric function and can negatively skew results [Hood & Poole 1980].

Beattie et al. [1975] evaluated the influence of word difficulty in a study that involved 36 CID spondaic words presented to 75 subjects via MLV. Three male examiners with experience in administering speech audiometry materials via MLV were selected to present the spondaic words. An evaluation of the mean sensation levels at which the words were correctly identified revealed a 7.9 dB range in mean sensation level. Beattie et al. considered large variations in relative intelligibility, such as those obtained with the list of 36 CID spondaic words, detrimental to the measurement of the SRT. Spondaic words that were difficult to identify were characterized by more variability and constituted a key-contributing factor to this large variation in performance. After identifying spondaic words with the least amount of variability, Beattie et al. developed a list using 18 of the CID spondaic words that varied in mean sensation level by only 1.5 dB.

Campbell [1965] also evaluated the influence of word selection in a study involving 140 veterans that took place at the Veterans Administration Outpatient Clinic in Atlanta, Georgia. The veterans were selected for participation in the study because they had mixed or sensorineural hearing loss accompanied by auditory word recognition scores from 10% to 70%. During audiological evaluations, the subjects were presented with words from the CID W-22 word lists. A comparison of the percentage of words that were incorrectly identified on each list revealed a range of 7.5% among lists. From these data, Campbell designed 8 half-lists, each consisting of 25 words. The words that were easiest to identify in each half-list differed from one another in performance on auditory word recognition tests by only 2%, whereas there was a 19% range in performance on auditory word recognition tests among the most difficult to identify words in each half-list. Campbell concluded that the distribution of word difficulty and the efficiency of the CID W-22 word lists could be improved if words that were more homogeneous in performance on auditory word recognition tests were substituted for words with extremes in auditory word recognition performance.

Hood and Poole [1980] selected words from the PAL PB-50 lists in order to create 20 original lists of 25 words each. The new lists were then recorded by a professional announcer and presented to 45 subjects. The test items were rank ordered according to difficulty after the data from 36 presentations were analyzed. Two new word lists were created based on the results of the graded word difficulty. One list contained easy to identify words and the other list contained difficult to identify words. After both lists were recorded by the same announcer and administered to 5 normal hearing subjects, Hood and Poole discovered that the ranking of word difficulty did not change.

Talker variables. Several researchers have investigated whether or not talker variables influence the difficulty of a word list. Talker variables include such factors as variations in intonation and pronunciation. Some researchers have found that significant differences are obtained in the auditory word recognition score when two different talkers read the same list to the same subject [Hood & Poole 1980; Kreul et al. 1969]. Others claim that such differences do not exist [Cambron et al. 1991]. Brandy [1966] is of the opinion that in addition to the significant differences that occur with presentations of two different talkers, multiple presentations of the same list by the same talker also produce significant differences. Penrod [1979] took a unique approach to this issue by considering not only the talker variables but also the effect of the talker-listener interaction. The issue of the influence of variables between talkers or variables within an individual talker is of concern because it affects the extent to which research results can be generalized to other settings. One possible explanation for the conflicting results of such studies is the differences in the procedures used [Kreul et al. 1969].

Hood and Poole [1980] not only evaluated the ranked difficulty of PB-50 words but also examined the effect of various talkers on word intelligibility. In addition to the professional talker who made the first recording of the easy and difficult lists, two non-professional talkers were selected to make recordings of the same lists. Each newly recorded list was presented to 5 different subjects. Psychometric functions collected on the lists recorded by the three talkers depicted a separation between the psychometric functions of the easy and difficult lists for the professional talker, whereas there was no clear separation between the psychometric functions of the two lists for the third talker. Variance ratios indicated no significant differences among groups on scores for the easy list versus the difficult list. A lack of significant differences was also found when the subjects' scores for the easy list were compared with one another. From these results, it can be concluded that the differences among talkers were not due to differences among the groups that listened to the lists or among the lists that were presented. Rather, it appears as though the differences among talkers were due to talker variables. In summarizing the influence of talker variables, Hood and Poole concluded, "In the preparation of recorded lists of words for clinical use, whatever importance is attached to familiarity or otherwise of words, is far outweighed by whatever influence the speaker brings to bear upon them" (p. 449).

Kreul et al. [1969] evaluated subject performance on recordings made by two different talkers as well as recordings made by the same talker on separate occasions. Two of the individuals who were conducting the experiment were selected as the talkers for the study. Twenty-three adult subjects were assigned to one of three groups. The first and second groups listened to words presented by the first talker whereas the third group listened to words presented by the second talker. Comparisons of recordings made by the same talker over two separate recording sessions revealed no significant differences. Significant differences were found, however, among the recordings of the two talkers. This finding suggests that variables that occur between two talkers produce changes in word list difficulty.

Cambron et al. [1991] studied the influence of talker gender on the speech detection threshold (SDT) and on the SRT. Digital recordings of CID W-1 spondaic words were made of both male and female voices. Twenty-two normal hearing individuals were selected to participate in the study. Before listening to the recordings, the subjects studied a list of the CID spondaic words. Prior to the administration of the test, the subjects were instructed to push a button upon the first detection of sound and to repeat the word once they were able to recognize it. Performance intensity functions for the SDT and SRT were similar for both talkers. No significant differences were found between talkers on either the SDT or the SRT.

Brandy [1966] examined the effects of multiple presentations of a word list by the same talker. The word list used for this experiment consisted of 25 words randomly selected from list 3 of the CID W-22 word lists. Three separate randomizations of the 25 words were then produced in order to have uniquely ordered lists for each of three recording sessions. An adult male talker made a series of recordings, over several days, under identical circumstances each day. A few additional lists were created by recording the list from the third day three times, splicing the recordings to produce lists with unique word orders, and then acoustically modifying the recordings so that each word was presented with equal intensity. Those participating in the study were divided into two groups of 12 subjects. One group listened to the original recordings; the other group listened to the modified recordings. An analysis of the variances between the two recording modes (original vs. modified) revealed significant differences. Brandy found that the original recordings had significantly larger variations among presentations than the modified recordings. The results of this study show that presentations of the same list by the same talker on different days cannot be considered equivalent, due to variations within the talker. Examples of such variations include modifications in intensity, word duration, and inflection.

Penrod [1979] examined the performance of 30 adults with sensorineural hearing impairment on auditory word recognition performance using CID W-22 word lists as stimuli. All participants had auditory word recognition scores that were no greater than 90% and an average hearing impairment of 25 dB HL or more at 500, 1000, and 2000 Hz. One female and three male audiologists, with a minimum of two years professional experience, were selected to read the lists. All four audiologists read one of four randomizations of list 2 of the CID W-22 word lists to each participant. The responses of the subjects were recorded on audio tape and were independently scored by each of three speech-language pathology graduate students. Penrod found that the talker-listener interaction accounted for 20.5% of the variance among randomizations of list 2 whereas only 5.7% of the variance was accounted for by talker variables.

While some researchers claim that talker variables are not significant and others claim that talker variables are very significant, the mere existence of talker variables should be noted and guarded against. Variables which exist between talkers should be guarded against because they pose a threat to the ability to generalize test results from one setting to the next. Perhaps the greatest safeguard against talker variables is the use of recorded materials.

Method of presentation. Like the issue of talker variables, the issue of the existence of significant differences in performance between MLV and recorded materials is clouded with controversy and conflicting findings. Beattie et al. [1975] found no significant differences among performance for the MLV presentations of different talkers, whereas Brandy [1966] discovered significant differences among multiple administrations of word lists presented by MLV. Others have found that both MLV and recorded presentations of word lists are equally reliable [Creston et al. 1966].

The study by Beattie et al. [1975] involved the presentation of CID spondaic words by three male talkers using MLV. In relation to the different talkers, Beattie et al. discovered that the scores of the three groups were quite similar despite the fact that a different talker presented the words to each of the three groups. Beattie et al. concluded that results from studies using MLV could be generalized to clinics that use MLV for testing purposes.

Brandy [1966] also compared MLV to prerecorded speech materials. The original recordings used in this study were considered to be similar to MLV testing because they were recorded on three different occasions, were not manipulated in any way, and were all produced by the same talker. In comparison, the modified recordings were considered taped material because they were acoustically manipulated and all three recordings were produced from one original recording. Significant differences were discovered among the three MLV presentations. There were no significant differences, however, among the presentations of the recorded lists.

Creston et al. [1966] compared the efficacy of recorded speech materials to the efficacy of MLV in a study involving 72 adults with sensorineural hearing impairment. All participants in the study were required to have at least a 20 dB HL hearing impairment. The CID W-2 word list of 36 spondaic words was used as stimuli for SRT testing and the CID W-22 word lists were used as stimuli for auditory word recognition testing. Subjects were divided into three groups. The first group was evaluated twice with MLV, the second was administered the tests via taped recordings on two separate occasions, and the third group was evaluated once with MLV and once with the tape recorded lists. No significant differences were found between the first and second presentations for either the first or second group. The only significant difference found for the third group was that the live voice presentation produced SRTs and auditory word recognition scores that were slightly better than those of the taped presentations. Despite the difference found in the third group, test-retest reliability was similar for both MLV and taped presentations. Creston et al. concluded that because both testing procedures were similar in reliability, the choice of whether to use MLV or taped material should be determined by the testing situation. According to Creston et al., tape recorded materials are best in situations where inexperienced clinicians are performing the testing whereas MLV is more effective when young children or those who are difficult to test are being evaluated.

In reference to the controversy over the best presentation method, the ASHA Committee on Audiologic Evaluation [1988] voiced the following opinion: *Recorded presentation of the test material is the preferred procedure. The use of recorded material standardizes the composition and presentation of the test list. It allows for better control of the intensity of the test items and ensures that the speech pattern of the recorded talker will be consistent to each client. (p. 86)*

Differences in intensity, intonation, and pronunciation may occur over several presentations of a word list by MLV. These differences may occur when presentations of speech audiometry materials are made by different speakers or when presentations of such materials are made by the same speaker. In most circumstances, recorded presentations of word lists are preferable to MLV presentations because recorded presentations help to minimize variability and maximize generalizability of test results.

Type of recording. A comparison of the specifications of tape players and compact disc (CD) players demonstrates the advantages of digital recording technology. While a tape player can produce a signal to noise ratio that is better than 64 dB with a channel separation of over 36 dB at 1 kHz, a CD player can produce a signal to noise ratio that is greater than 110 dB with a channel separation in excess of 105 dB [Nakamichi, n.d.; Sony 1991]. In addition, a CD player is capable of a dynamic range of more than 100 dB with less than 0.0025% harmonic distortion [Sony 1991].

Kamm et al. [1980] listed the advantages of digital recordings over other recording methods as follows: (a) arbitrarily high signal-to-noise ratio and dynamic range; (b) zero wow and flutter; (c) no harmonic distortion near upper signal intensity range and no modulation-noise sidebands near the signal; (d) no crosstalk between channels; (e) full bandwidth; (f) no amplitude variations caused by local changes in magnetization; (g) no interaction (print through) between adjacent tape layers (pp. 709-710).

Since its introduction to the United States in 1983, the CD has gradually become the preferred medium for presentation of recorded materials. The tracks of a CD, unlike recordings on tape, can be presented in any order, and certain tracks can be repeated if desired. A plastic lacquer that is applied to the surface of the CD serves as a protective coating, making it fairly resistant to minor scratches, fingerprints, and dust. Unlike analog audiocassettes, CDs have a potentially unlimited life span that makes it possible to have high quality recordings that do not vary in sound quality over periods of extensive use [Ridgway 1986]. Perhaps the greatest advantage of a digital recording is that with the use of computers, the digital signal can be modified in a highly efficient and uniform manner [Kamm et al. 1980; Ridgway 1986].

Factors that Affect Speech Audiometry in Other Languages

Several factors have been found to be important in the development of speech audiometry materials in other languages, including (a) word familiarity [Comstock & Martin 1984; Weisleder & Hodgson 1989; Zakrzewski, Jassem, Pruszewicz, & Obrębowski 1975]; (b) variations in dialect and pronunciation [Weisleder & Hodgson 1989]; and (c) the linguistic structure of a word list [Aleksandrovsky et al. 1998; Cancel 1968; Weisleder & Hodgson 1989].

Word familiarity. The familiarity of the words selected for a particular list is thought to play a major role in determining the difficulty of that list. Studies of word lists in both Spanish and Polish show that subjects perform more consistently with words that are familiar than with words that are unfamiliar [Comstock & Martin 1984; Weisleder & Hodgson 1989; Zakrzewski et al. 1975].

Comstock and Martin [1984] developed and evaluated a auditory word recognition test for Spanish-speaking children. Twenty children, ranging in age from three to eight years, participated in the study. The dominant language of every participant was Spanish. All participants had lived in the central region of Texas for their entire lives. Comstock and Martin developed four word lists, each of which consisted of 25 bisyllabic words that were within the vocabularies of Spanish-speaking preschool children in central Texas. A native Texan who spoke Spanish fluently was selected as the talker for the lists. The lists were presented in a picture identification format that required the child to point to the picture that represented the word that was spoken. Auditory word recognition scores demonstrated a tendency to improve with age. This tendency was attributed to a vocabulary limitation in the younger children. Comstock and Martin highlighted the importance of assessing whether test items are within a child's vocabulary before administering a test.

Weisleder and Hodgson [1989] evaluated the effect of word familiarity on 16 native Spanish-speaking adults with normal hearing. Four lists, each consisting of 50 bisyllabic Spanish words, were used as stimuli. A native Spanish-speaking male was selected as the talker for the recordings. Each list was presented at four different presentation levels (8, 16, 24, & 32 dB HL) to each participant. After consulting a Spanish frequency usage dictionary, the authors discovered that 90% of the least missed words could be found in this dictionary, whereas only 45% of the most frequently missed words were in the dictionary.

In a study involving 297 Polish school children ranging in age from 9 to 14 years, Zakrzewski et al. [1975] evaluated the difference between auditory word recognition scores obtained with monosyllabic words and those obtained with nonsense words. A total of 20 lists, consisting of 10 words each, were used during the study. Ten of the lists contained meaningful monosyllabic Polish words; the other ten contained nonsense words. Scores obtained on the nonsense syllable word lists were significantly lower than those obtained on the monosyllabic word lists.

Research has shown that unfamiliar words, nonsense words, and words that are not within an individual's vocabulary are missed more frequently than familiar words in SRT and auditory word recognition tests. SRT and auditory word recognition tests should be designed so that only words that are familiar to the test population are included in the word lists. If a word list is limited to frequently-used words, the test eliminates the influence of vocabulary and linguistic skills and becomes a strict test of auditory reception.

Variations in dialect and pronunciation. The talker who participated in the study by Weisleder and Hodgson [1989] was a native of Mexico. Nine of the 16 subjects involved in the study were also natives of Mexico. The remaining subjects were natives of countries in either Central or South America. An analysis that compared the scores obtained by native Mexicans to those obtained by the other subjects revealed that the native Mexicans performed significantly better at low presentation levels.

Talkers who have a dialect that is different from that of the test population may negatively affect auditory word recognition scores. In order to avoid this negative affect, talkers selected to make recordings of word lists should be native speakers of the standard dialect of the test population.

The linguistic structure of a word list. The number of syllables per word and the arrangement of phonemes within the words influence the difficulty of speech audiometry materials. Research has shown that words that have a large number of syllables are easier to identify than words with fewer syllables [Aleksandrovsky et al. 1998]. Cancel [1968] and Weisleder and Hodgson [1989] found that the presence of certain phonemes in Spanish words contributed to their difficulty.

Aleksandrovsky et al. [1998] evaluated the performance of 21 Russian-speaking subjects with normal hearing on the Russian Picture Identification Task (RPIT). Each time the subject heard a word during the RPIT, a computer monitor depicted a set of four pictures. One of the pictures represented the word presented, whereas the other three were pictures of words that rhymed with the test word. The subject was asked to identify the picture of the word presented. Aleksandrovsky et al. rated the words represented by each set of pictures as either minimally phonetically varied or maximally phonetically varied. A significant difference was noted in subject performance in the two different conditions. Subjects experienced more difficulty in selecting a test word from a foil of items that were minimally phonetically varied. Differences in performance due to the number of syllables in test items were also discovered. During the analysis of test results, Aleksandrovsky et al. found that bisyllabic words were identified at lower presentation levels than monosyllabic words.

Cancel [1968] found that the presence of certain phonemes in Spanish words made them less intelligible. A particular difficulty was noted when the phoneme /s/ was in the final position. Subjects frequently omitted /s/ when it occurred at the end of a word and always omitted /s/ when its presence in the final position indicated plurality. One possible explanation for this phenomenon is that aspiration of /s/ occurs in several regions of Central and South America.

The results of the study by Weisleder and Hodgson [1989] verify the findings of Cancel [1968]. The presence of the phoneme /s/ was noted in 18 of the 20 most frequently missed Spanish words. Weisleder and Hodgson also attributed this finding to regional variations in pronunciation.

The number of syllables per word and the presence of certain phonemes in a word can affect the difficulty of a word list. Care should be taken to avoid words that are less intelligible due to the presence of certain phonemes in certain positions. The number of syllables per word should be chosen in accordance with the purpose and desired difficulty of a word list.

In summary, it is important to consider a number of factors when developing speech audiometry materials. Careful attention must be given to the selection of words that are both familiar and moderately difficult to identify [Campbell 1965]. Words that are unfamiliar or are too difficult to distinguish may hinder the development of a standardized word list [Comstock & Martin 1984; Weisleder & Hodgson 1989; Zakrzewski et al. 1975]. A list of familiar words can be compiled by consulting a frequency usage dictionary of the chosen language [Harris et al. 2001; Weisleder & Hodgson 1989]. Talkers should be selected who use the standard or most common dialect of that language because talkers with unusual dialects may negatively affect test results [Weisleder & Hodgson 1989]. If the word lists are to be recorded, a high quality digital recording method should be used. Digital recording methods offer the greatest longevity, quality, and versatility of any recording method available.

In accordance with the aforementioned requirements for creation of speech audiometry materials in other languages, this Polish speech audiometry project was designed to meet the following objectives: (1) to create word lists for SRT and auditory word recognition testing that are composed of frequently used words; (2) to identify a native male and a native female Pole who use a standard dialect of Polish to serve as talkers for the recordings; (3) to create high-quality digital recordings of the SRT and auditory word recognition lists; (4) to obtain normative data on both the SRT and auditory word recognition lists; (5) to select 25 words with the steepest slope for use in evaluation of the SRT; (6) to develop four equivalent, phonemically balanced auditory word recognition lists and eight equivalent half-lists. This manuscript will present the development of and results for the Polish bisyllabic SRT materials.

Method

Subjects

All subjects participating in this study were natives of Poland. A total of 26 subjects (7 male & 19 female), ranging in age from 20 to 29 years (M = 23.7 years), participated in the evaluation of the bisyllabic words. Summary statistics of the subject thresholds are listed in Table 1. Each participant had pure tone air-conduction thresholds 15 dB HL at octave and midoctave 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, and Stoppenbach 1998].

Tab. 1. Age (in years] and Pure Tone Thresholds (dB HL] Descriptive Statistics for the 26 Subjects that Participated in the Bisyllabic Study

	M	Minimum	Maximum	SD
Age	23.7	20.0	29.0	3.3
125 Hz	7.5	-5.0	15.0	6.7
250 Hz	6.2	-5.0	15.0	6.7
500 Hz	4.4	-5.0	15.0	5.5
750 Hz	3.8	-5.0	15.0	5.3
1000 Hz	3.7	-5.0	15.0	5.6
1500 Hz	-1.1	-5.0	15.0	12.0
2000 Hz	-1.0	-10.0	5.0	4.2
3000 Hz	0.4	-5.0	10.0	5.1
4000 Hz	2.5	-10.0	15.0	7.4
6000 Hz	8.1	-10.0	15.0	6.0
8000 Hz	9.2	0.0	15.0	4.6

Materials

Word lists. Bisyllabic words with stress on the first syllable were selected for the SRT materials on the basis of what is currently used for the evaluation of the SRT in Poland [E. Michałowska, personal communication, October 12 1999]. A total of 260 bisyllabic words were initially selected from the frequency usage dictionary by Zgołkowa [1983] and from a bisyllabic word list by Taniewski, Kugler, and Wysocki [cited by Bystrzanowska 1969]. Of the 260 bisyllabic words considered for evaluation, 70 words were selected for recording and evaluation in this study. The 190 words which were initially selected but not evaluated were eliminated for one, or more, of a number of reasons which included: a) undesireable vocabulary, b) considered to be unfamiliar by some of the judges, c) questionable stress, or d) confustion with other words.

Talkers. Initial recordings were made using five native Polish-speaking individuals, three males and two females. All talkers were from central Poland and spoke a standard Polish dialect. After the recordings were made, a panel of nine native Polish judges evaluated the performance of each talker. The judges were asked to indicate whether the vocal quality and accent of the talker was acceptable or unacceptable and then were asked to rank order the talkers from best to worst. The highest ranked talkers (one male & one female) were selected as the talkers for the recordings. Neither of the talkers who were selected received any unacceptable ratings, whereas two of the remaining three talkers not selected were considered to be unacceptable by one or more of the judges.

Recording. All recordings were made in a large anechoic chamber located on the Brigham Young University campus in the Eyring Science Center in Provo, Utah, USA. The ambient background noise levels in the anechoic chamber were approximately 60-65 dB down from the speech levels measured during recording. The microphone was positioned approximately 15 cm from the talker. A Larson-Davis model 2541 microphone was positioned at a 0° azimuth and was covered by a 3" windscreen. The microphone was connected to a Larson-Davis model 900B preamp, and the preamp was coupled to a Larson-Davis model 2200C preamp power supply. The signal from the preamp power supply was routed through an Apogee AD-8000 24-bit analog-to-digital converter; the digitized signal was 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 recording. Once recorded, the words were edited using Sadie Disk Editor software [Studio Audio and Video Limited 1996].

During the recording sessions, the talker was asked to pronounce each word several times. A native Polish judge rated each word for perceived goodness of production, and the best production of each word was then selected for inclusion on the CD. If there were no satisfactory recordings of a word, that word was recorded a second time. After the rating process, the intensity of each word to be included on the CD was edited to yield the same intensity as that of the 1000 Hz calibration tone contained on the CD (ANSI 3.6-1996). The evaluation CD was produced on a Yamaha CDE 100II recordable CD-ROM drive using a 44.1 kHz sampling rate and 16-bit quantization. The NS high dither option in the Sadie Disk Editor software was used to convert the recordings from 24 to 16-bit quantization.

Procedures

Custom software was used to control randomization and timing of the presentation of the words. The signal was routed from a computer-controlled CD-ROM drive to the external inputs of a Grason Stadler model 1761 (GSI-61) audiometer. The stimuli were routed from the audiometer to the subject via TDH-50P headphones. Prior to testing each subject, the inputs to the audiometer were calibrated to 0 VU using the 1000 Hz calibration tone on track 1 of the Polish evaluation CD. All testing was carried out in a sound suite that met ANSI [1991] standards for maximum permissible ambient noise sound pressure levels in one-third octave bands for the ears not covered condition.

Evaluation of bisyllabic words. Each subject listened to the entire list of bisyllabic words at 50 dB HL in order to become familiar with the words before testing commenced. After familiarizing the subject to the words, the subject was given the following instructions:

(Polish) *Będziesz słyszał słowa dwusylabowe w zestawach różniących się między sobą głośnością, od bardzo cichych do dobrze słyszalnych. Te najcichsze słowa mogą być trudne do usłyszenia. Zielony wskaźnik świetlny będzie się pojawiał podczas wymawiania danego słowa. Proszę słuchać jak najuważniej i powtarzać każde słowo. Jeżeli nie jesteś pewien jakie słowo usłyszałeś, zachęcamy abyś zgadywał. Jeżeli nie domyśliłeś się słowa, nie mów nic i czekaj na następne słowo. Czy masz jakieś pytania?*

(English) *You will hear bisyllabic words (2 syllables) at a number of different loudness levels. These loudness levels will vary from very soft to a more comfortable loudness level. At the very soft loudness levels it may be difficult for you to hear the words. The green indication light will come on, informing you that a word has been presented. Please listen carefully and repeat the word that you hear. If you are unsure of the word, you are encouraged to guess. If you have no guess, please be quiet and listen for the next word. Do you have any questions?*

The entire bisyllabic word list (70 words) was presented at each of 15 different intensity levels, ranging from -10 to 18 dB HL in 2 dB steps. Word order within the list was randomized prior to each presentation. Each subject listened to both the male and female recordings of the bisyllabic list. The order of presentation of the male and female recordings was randomly determined for each subject.

Calibration

The audiometer was calibrated prior to, weekly during and at the conclusion of data collection. Calibration was performed in accordance with the specifications of the American National Standards Institute (ANSI S3.6-1996). No calibration adjustments were required throughout the duration of the study.

Results

Bisyllabic Words

After the raw data were collected, logistic regression was used to obtain the regression slope and regression intercept for each of the 70 bisyllabic words. The regression slope and regression intercept values were then inserted into a modified logistic regression equation that was designed to calculate the percent correct at each intensity level. The original logistic regression equation is as follows:

$$\log\frac{p}{1-p} = a + b * dB \qquad (1)$$

In Equation 1, p is the proportion correct at any given intensity level, a is the regression slope, b is the regression intercept, and dB is the intensity level in dB HL. When Equation 1 is solved for p and multiplied by 100, we obtain Equation 2.

$$\% = (1 - \frac{\exp(a + b * dB)}{1 + \exp(a + b * dB)}) * 100 \qquad (2)$$

By inserting the regression slope, regression intercept, and intensity level into Equation 2, it is possible to predict the percent correct at any specified intensity level. The percent correct was predicted for each of the bisyllabic words for a range of -10 to 18 dB HL in 1 dB increments. Smoothed psychometric functions were then produced using the predicted percentages. The smoothed psychometric functions for each of the 70 bisyllabic words can be found in Figures 1 (male) and 2 (female).

$$dB = \frac{\log\frac{p}{1-p} - a}{b} \qquad (3)$$

$$dB = \frac{-a}{b} \qquad (4)$$

In order to calculate the intensity level required for a given proportion, Equation 1 was solved for dB (see Equation 3). By inserting the desired proportions into Equation 3, it is possible to calculate the threshold (the intensity required for 50% intelligibility), the slope (%/dB) at threshold, and the slope (%/dB) from 20 to 80% for each psychometric function. Table 2 (male) and Table 3 (female) contain the calculations for the threshold, slope at threshold, and slope from 20 to 80% for each psychometric function. When solving for the threshold (p = 0.5), Equation 3 can be simplified to Equation 4.

Thresholds for the 70 bisyllabic words ranged from 4.3 to 15.6 dB HL (M = 2.1 dB HL) for the male words, and from 3.7 to 14.0 dB HL (M = 5.3 dB HL) for the female words. The slope from 20 to 80% for each psychometric function encompassed a range of 2.7 to 12.9 %/dB (M = 8.5 %/dB) for the male words and a range of 4.7 to 11.9 %/dB (M = 8.6 %/dB)

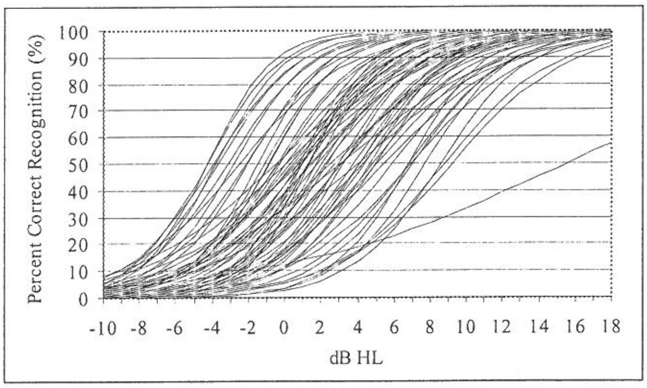

Fig. 1. Psychometric functions for 70 male Polish talker bisyllabic words

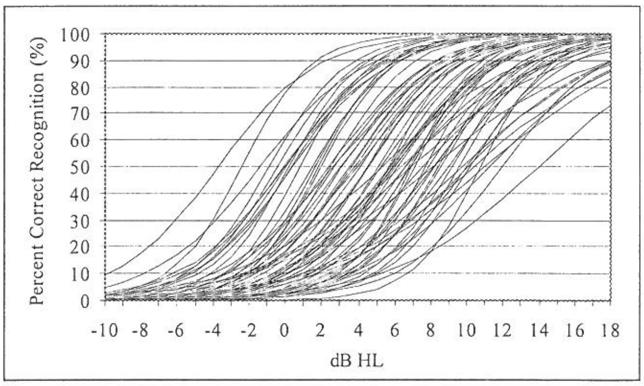

Fig. 2. Psychometric functions for 70 female Polish talker bisyllabic words

Tab. 2. Mean Performance for Polish Male Bisyllabic SRT Words

Word	a[a]	b[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	Thres-hold[e]	Change in dB[f]
chłopiec	-0.10	-0.35	8.7	7.5	-0.3	-2.7
chwila	3.27	-0.48	12.1	10.5	6.8	4.4
córka	1.55	-0.41	10.3	8.9	3.8	1.4
człowiek	-0.11	-0.42	10.5	9.1	-0.3	-2.7
dobrze	-0.18	-0.43	10.8	9.3	-0.4	-2.8
drzewo	-0.88	-0.49	12.2	10.6	-0.9	-3.3
dziecko	-0.77	-0.52	13.1	11.3	-1.5	-3.9
forma	1.34	-0.32	7.9	6.8	4.2	1.8
grupa	1.58	-0.32	7.9	6.8	5.0	2.6
kino	3.26	-0.42	10.6	9.2	7.7	5.3
kłopot	0.21	-0.27	6.6	5.7	0.8	-1.6
koniec	2.02	-0.41	10.2	8.9	4.9	2.5
książka	-1.32	-0.47	11.8	10.2	-2.8	-5.2
księżyc	-1.65	-0.45	11.1	9.6	-3.7	-6.1
lampa	0.11	-0.34	8.5	7.4	0.3	-2.1
lekarz	-1.33	-0.42	10.6	9.1	-3.2	-5.6
ludzie	1.65	-0.40	10.1	8.7	4.1	1.7
mama	-0.81	-0.33	8.2	7.1	-2.5	-4.9
miejsce	1.37	-0.46	11.4	9.9	3.0	0.6
miłość	0.43	-0.43	10.8	9.4	1.0	-1.4
nazwa	0.44	-0.30	7.5	6.5	1.5	-0.9
numer	2.08	-0.45	11.3	9.8	4.6	2.2
okno	-0.52	-0.47	11.8	10.2	-1.1	-3.5
okres	0.48	-0.31	7.7	6.6	1.6	-0.8
palec	0.88	-0.38	9.6	8.3	2.3	-0.1
pani	2.83	-0.32	8.0	7.0	8.8	6.4
pismo	3.50	-0.41	10.2	8.8	8.6	6.2
pokój	1.03	-0.33	8.2	7.1	3.1	0.7
praca	2.10	-0.30	7.6	6.6	6.9	4.5
prawda	-0.12	-0.25	6.3	5.5	-0.5	-2.9
próba	2.84	-0.31	7.7	6.7	9.2	6.8
problem	1.64	-0.34	8.6	7.4	4.8	2.4
przerwa	0.54	-0.42	10.5	9.1	1.3	-1.1
przykład	-1.68	-0.43	10.6	9.2	-4.0	-6.4
rama	1.95	-0.13	3.1	2.7	15.6	13.2
rodzaj	1.15	-0.36	9.1	7.9	3.1	0.7
rola	2.04	-0.35	8.7	7.5	5.9	3.5
ściana	-2.39	-0.56	14.0	12.1	-4.3	-6.7
słońce	0.20	-0.29	7.3	6.4	0.7	-1.7
słowo	0.01	-0.26	6.5	5.7	0.1	-2.3
sposób	2.45	-0.36	9.0	7.8	6.8	4.4
środek	-0.42	-0.49	12.3	10.7	-0.9	-3.3
stopień	1.50	-0.46	11.6	10.0	3.2	0.8
światło	-2.12	-0.49	12.3	10.7	-4.3	-6.7
sygnał	0.32	-0.42	10.6	9.1	0.8	-1.6
system	1.09	-0.35	8.6	7.5	3.1	0.7
szkoła	-0.89	-0.42	10.4	9.0	-2.1	-4.5
szpital	0.63	-0.46	11.5	9.9	1.4	-1.0
sztuka	0.83	-0.60	14.9	12.9	1.4	-1.0
temat	0.04	-0.36	9.0	7.8	0.1	-2.3
teren	0.60	-0.33	8.3	7.2	1.8	-0.6
termin	1.75	-0.41	10.1	8.8	4.3	1.9
tydzień	1.84	-0.45	11.3	9.7	4.1	1.7
uczeń	1.07	-0.45	11.2	9.7	2.4	0.0
wczoraj	0.88	-0.42	10.6	9.2	2.1	-0.3
wieczór	0.51	-0.52	12.9	11.1	1.0	-1.4
władza	0.71	-0.23	5.7	5.0	3.1	0.7
woda	0.33	-0.35	8.7	7.5	1.0	-1.4
wojna	0.47	-0.55	13.8	11.9	0.8	-1.6
worek	0.96	-0.24	6.0	5.2	4.0	1.6
wtorek	0.16	-0.34	8.4	7.3	0.5	-1.9
wynik	2.69	-0.38	9.6	8.3	7.0	4.6
zakład	-1.95	-0.51	12.8	11.1	-3.8	-6.2
zboże	1.51	-0.28	7.0	6.0	5.4	3.0
zdanie	0.82	-0.29	7.3	6.3	2.8	0.4
zdrowie	0.90	-0.51	12.7	11.0	1.8	-0.6
zespół	0.34	-0.38	9.4	8.1	0.9	-1.5
zmiana	0.67	-0.26	6.6	5.7	2.6	0.2
żona	0.68	-0.52	12.9	11.2	1.3	-1.1
życie	-0.93	-0.46	11.5	9.9	-2.0	-4.4
M	0.66	-0.39	9.8	8.5	2.1	-0.3
Minimum	-2.39	-0.60	3.1	2.7	-4.3	-6.7
Maximum	3.50	-0.13	14.9	12.9	15.6	13.2
Range	5.89	0.47	11.8	10.2	19.9	19.9
SD	1.31	0.09	2.3	2.0	3.6	3.6

(table continues)

a[a] = regression slope. b[b] = regression intercept. [c]Performance intensity function slope (%/dB) at 50% was calculated from 49.99 to 50.01%. [d]Performance intensity function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of a word to the mean PTA of the subjects in the bisyllabic study.

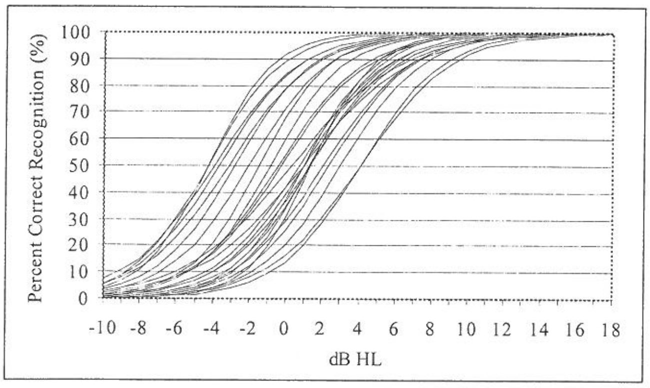

Fig. 3. Psychometric Functions for the 25 selected male Polish talker bisyllabic words

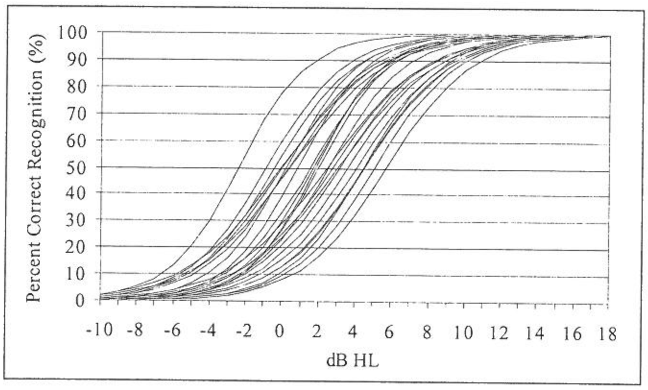

Fig. 4. Psychometric Functions for the 25 selected female Polish talker bisyllabic words

for the female words. In comparison to the slopes from 20 to 80% for each psychometric function, the slopes at the 50% threshold were slightly steeper. The ranges for the slopes at threshold were 3.1 to 14.9 %/dB (M = 9.8 %/dB) for the male words and 5.5 to 13.7 %/dB (M = 9.9 %/dB) for the female words.

Due to the wide range in slopes among the 70 bisyllabic words, the 25 words that had the steepest psychometric function slopes for both the male and female recordings were selected for inclusion in the final Polish speech audiometry CD. The threshold, the slope at threshold, and the slope from 20 to 80% for the 25 selected bisyllabic words are listed in

Tab. 3. Mean Performance for Polish Female Bisyllabic SRT Words

Word	a[a]	b[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	Thres-hold[e]	Change in dB[f]
chłopiec	2.07	-0.35	8.7	7.5	5.9	3.5
chwila	2.95	-0.52	13.1	11.3	5.6	3.2
córka	3.26	-0.41	10.4	9.0	7.9	5.5
człowiek	1.03	-0.51	12.9	11.1	2.0	-0.4
dobrze	1.20	-0.47	11.8	10.2	2.5	0.1
drzewo	0.94	-0.34	8.5	7.4	-1.4	-3.8
dziecko	1.53	-0.42	10.6	9.2	3.6	1.2
forma	2.79	-0.34	8.6	7.4	8.1	5.7
grupa	1.49	-0.40	9.9	8.6	3.8	1.4
kino	3.72	-0.53	13.1	11.4	7.1	4.7
kłopot	2.27	-0.39	9.7	8.4	5.9	3.5
koniec	2.01	-0.35	8.7	7.5	5.8	3.4
książka	-1.26	-0.52	13.1	11.3	-2.4	-4.8
księżyc	0.11	-0.51	12.7	11.0	0.2	-2.2
lampa	1.34	-0.43	10.7	9.2	3.2	0.8
lekarz	0.86	-0.51	12.7	11.0	1.7	-0.7
ludzie	2.26	-0.44	11.1	9.6	5.1	2.7
mama	2.08	-0.40	9.9	8.6	5.2	2.8
miejsce	2.47	-0.43	10.8	9.4	5.7	3.3
miłość	1.70	-0.42	10.5	9.1	4.1	1.7
nazwa	2.21	-0.29	7.3	6.3	7.6	5.2
numer	4.30	-0.42	10.5	9.1	10.2	7.8
okno	0.87	-0.32	8.1	7.0	2.7	0.3
okres	1.60	-0.25	6.3	5.5	6.3	3.9
palec	1.77	-0.22	5.5	4.7	8.1	5.7
pani	3.17	-0.28	7.0	6.1	11.3	8.9
pismo	3.79	-0.55	13.7	11.9	6.9	4.5
pokój	1.50	-0.42	10.6	9.2	3.5	1.1
praca	3.14	-0.38	9.6	8.3	8.2	5.8
prawda	2.45	-0.24	5.9	5.1	10.4	8.0
próba	1.99	-0.35	8.7	7.6	5.7	3.3
problem	3.04	-0.45	11.2	9.7	6.8	4.4
przerwa	1.89	-0.42	10.6	9.2	4.5	2.1
przykład	-0.01	-0.38	9.5	8.2	0.0	-2.4
rama	3.50	-0.25	6.2	5.4	14.0	11.6
rodzaj	3.77	-0.44	11.0	9.5	8.6	6.2
rola	2.79	-0.41	10.2	8.8	6.9	4.5

(table continues)

Word	a[a]	b[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	Thres-hold[e]	Change in dB[f]
ściana	-0.40	-0.47	11.6	10.1	-0.9	-3.3
słońce	1.39	-0.35	8.8	7.7	3.9	1.5
słowo	2.00	-0.30	7.6	6.6	6.6	4.2
sposób	3.91	-0.51	12.7	11.0	7.7	5.3
środek	1.00	-0.54	13.6	11.8	1.8	-0.6
stopień	4.13	-0.45	11.2	9.7	9.2	6.8
światło	-0.18	-0.46	11.4	9.9	-0.4	-2.8
sygnał	0.91	-0.37	9.3	8.1	2.4	0.0
system	2.55	-0.39	9.7	8.4	6.6	4.2
szkoła	-1.35	-0.36	9.1	7.9	-3.7	-6.1
szpital	0.16	-0.38	9.5	8.2	0.4	-2.0
sztuka	0.02	-0.41	10.3	9.0	0.1	-2.3
temat	3.66	-0.37	9.1	7.9	10.0	7.6
teren	2.40	-0.25	6.1	5.3	9.8	7.4
termin	5.74	-0.53	13.3	11.5	10.8	8.4
tydzień	2.11	-0.47	11.8	10.2	4.5	2.1
uczeń	0.71	-0.46	11.5	10.0	1.5	-0.9
wczoraj	0.15	-0.42	10.5	9.1	0.4	-2.0
wieczór	0.44	-0.52	13.1	11.3	0.8	-1.6
władza	1.65	-0.24	5.9	5.1	6.9	4.5
woda	1.16	-0.38	9.5	8.2	3.1	0.7
wojna	4.40	-0.46	11.5	9.9	9.6	7.2
worek	4.30	-0.36	9.0	7.8	12.0	9.6
wtorek	2.29	-0.38	9.5	8.3	6.0	3.6
wynik	3.88	-0.46	11.6	10.0	8.4	6.0
zakład	0.01	-0.35	8.8	7.6	0.0	-2.4
zboże	2.49	-0.23	5.6	4.9	11.0	8.6
zdanie	1.36	-0.26	6.4	5.6	5.3	2.9
zdrowie	3.38	-0.38	9.4	8.1	9.0	6.6
zespół	2.45	-0.40	9.9	8.6	6.2	3.8
zmiana	2.72	-0.27	6.8	5.8	10.1	7.7
żona	1.93	-0.43	10.6	9.2	4.5	2.1
życie	1.12	-0.40	10.1	8.7	2.8	0.4
M	1.99	-0.40	9.9	8.6	5.3	2.9
Minimum	-1.35	-0.55	5.5	4.7	-3.7	-6.1
Maximum	5.74	-0.22	13.7	11.9	14.0	11.6
Range	7.10	0.33	8.2	7.1	17.7	17.7
SD	1.41	0.09	2.1	1.8	3.8	3.8

a[a] = regression slope. b[b] = regression intercept. [c]Performance intensity function slope (%/dB) at 50% was calculated from 49.99 to 50.01%. [d]Performance intensity function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of a word to the mean PTA of the subjects in the bisyllabic study.

Table 4 (male) and Table 5 (female). Figures 3 and 4 contain the psychometric functions for the final 25 words for the male and female recordings, respectively.

After the final 25 bisyllabic words were selected, there was still a wide range of variability in threshold of audibility among the words. In order to improve homogeneity among the 25 bisyllabic words, the intensity of each bisyllabic word was digitally adjusted so that the threshold of each word was equal to the mean PTA of the subjects in the bisyllabic study (2.37 dB HL). The adjustments necessary for each bisyllabic word for the male and female recordings are presented in Table 4 and Table 5, respectively. Figures 5 (male) and 6 (female) contain the predicted psychometric functions for the 25 selected bisyllabic words after intensity adjustment. After editing of the 25 male and 25 female bisyllabic words mean psychometric functions were calculated and are presented in Figure 7.

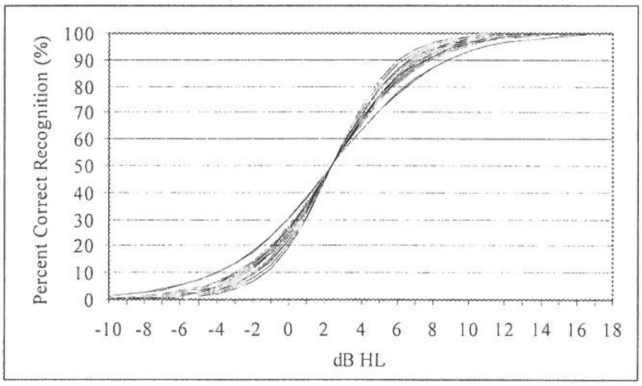

Fig. 5. Psychometric functions for the 25 selected male Polish talker bisyllabic words after intensity adjustment

Tab. 4. Mean Performance for Selected Polish Male Bisyllabic SRT Words

Word	a[a]	b[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	Thres-hold[e]	Change in dB[f]
człowiek	1.03	-0.51	12.9	11.1	2.0	-0.4
dobrze	1.20	-0.47	11.8	10.2	2.5	0.1
dziecko	1.53	-0.42	10.6	9.2	3.6	1.2
lampa	1.34	-0.43	10.7	9.2	3.2	0.8
książka	-1.26	-0.52	13.1	11.3	-2.4	-4.8
księżyc	0.11	-0.51	12.7	11.0	0.2	-2.2
lekarz	0.86	-0.51	12.7	11.0	1.7	-0.7
ludzie	2.26	-0.44	11.1	9.6	5.1	2.7
miejsce	2.47	-0.43	10.8	9.4	5.7	3.3
miłość	1.70	-0.42	10.5	9.1	4.1	1.7
przerwa	1.89	-0.42	10.6	9.2	4.5	2.1
przykład	-0.01	-0.38	9.5	8.2	0.0	-2.4
ściana	-0.40	-0.47	11.6	10.1	-0.9	-3.3
środek	1.00	-0.54	13.6	11.8	1.8	-0.6
światło	-0.18	-0.46	11.4	9.9	-0.4	-2.8
sygnał	0.91	-0.37	9.3	8.1	2.4	0.0
szpital	0.16	-0.38	9.5	8.2	0.4	-2.0
sztuka	0.02	-0.41	10.3	9.0	0.1	-2.3
tydzień	2.11	-0.47	11.8	10.2	4.5	2.1
uczeń	0.71	-0.46	11.5	10.0	1.5	-0.9
wczoraj	0.15	-0.42	10.5	9.1	0.4	-2.0
wieczór	0.44	-0.52	13.1	11.3	0.8	-1.6
woda	1.16	-0.38	9.5	8.2	3.1	0.7
żona	1.93	-0.43	10.6	9.2	4.5	2.1
życie	1.12	-0.40	10.1	8.7	2.8	0.4
M	-0.07	-0.46	11.4	9.8	-0.1	-2.5
Minimum	-2.39	-0.60	8.5	7.4	-4.3	-6.7
Maximum	1.84	-0.35	14.9	12.9	4.1	1.7
Range	4.23	0.25	6.4	5.5	8.4	8.4
SD	1.18	0.05	1.5	1.3	2.5	2.5

a[a] = regression slope. b[b] = regression intercept. [c]Performance intensity function slope (%/dB) at 50% was calculated from 49.99 to 50.01%. [d]Performance intensity function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of a word to the mean PTA of the subjects in the bisyllabic study.

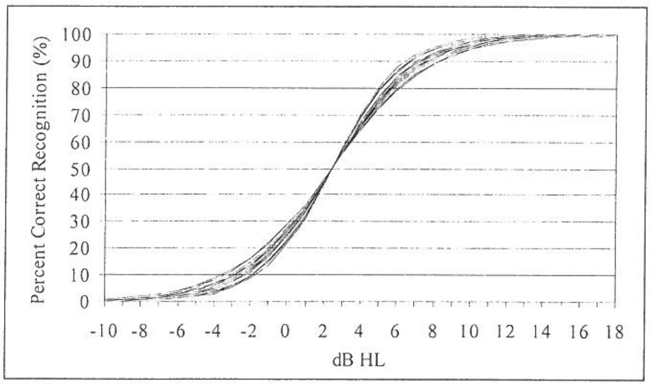

Fig. 6. Psychometric functions for the 25 selected female Polish talker bisyllabic words after intensity adjustment

Tab. 5. Mean Performance for Selected Polish Female Bisyllabic SRT Words

Word	a[a]	b[b]	Slope at 50%[c]	Slope from 20 to 80%[d]	Thres-hold[e]	Change in dB[f]
człowiek	1.03	-0.51	12.9	11.1	2.0	-0.4
dobrze	1.20	-0.47	11.8	10.2	2.5	0.1
dziecko	1.53	-0.42	10.6	9.2	3.6	1.2
książka	-1.26	-0.52	13.1	11.3	-2.4	-4.8
księżyc	0.11	-0.51	12.7	11.0	0.2	-2.2
lekarz	0.86	-0.51	12.7	11.0	1.7	-0.7
lampa	1.34	-0.43	10.7	9.2	3.2	0.8
ludzie	2.26	-0.44	11.1	9.6	5.1	2.7
miejsce	2.47	-0.43	10.8	9.4	5.7	3.3
miłość	1.70	-0.42	10.5	9.1	4.1	1.7
przerwa	1.89	-0.42	10.6	9.2	4.5	2.1
przykład	-0.01	-0.38	9.5	8.2	0.0	-2.4
ściana	-0.40	-0.47	11.6	10.1	-0.9	-3.3
środek	1.00	-0.54	13.6	11.8	1.8	-0.6
światło	-0.18	-0.46	11.4	9.9	-0.4	-2.8
sygnał	0.91	-0.37	9.3	8.1	2.4	0.0
szpital	0.16	-0.38	9.5	8.2	0.4	-2.0
sztuka	0.02	-0.41	10.3	9.0	0.1	-2.3
tydzień	2.11	-0.47	11.8	10.2	4.5	2.1
uczeń	0.71	-0.46	11.5	10.0	1.5	-0.9
wczoraj	0.15	-0.42	10.5	9.1	0.4	-2.0
wieczór	0.44	-0.52	13.1	11.3	0.8	-1.6
woda	1.16	-0.38	9.5	8.2	3.1	0.7
żona	1.93	-0.43	10.6	9.2	4.5	2.1
życie	1.12	-0.40	10.1	8.7	2.8	0.4
M	0.89	-0.45	11.2	9.7	2.0	-0.4
Minimum	-1.26	-0.54	9.3	8.1	-2.4	-4.8
Maximum	2.47	-0.37	13.6	11.8	5.7	3.3
Range	3.73	0.17	4.3	3.7	8.1	8.1
SD	0.92	0.05	1.3	1.1	2.1	2.1

a[a] = regression slope. b[b] = regression intercept. [c]Performance intensity function slope (%/dB) at 50% was calculated from 49.99 to 50.01%. [d]Performance intensity function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of a word to the mean PTA of the subjects in the bisyllabic study.

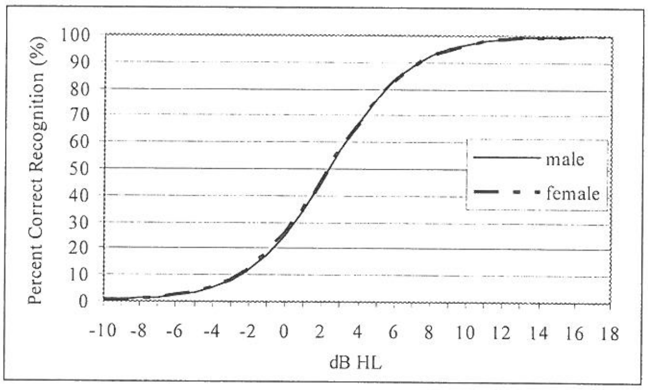

Fig. 7. Mean psychometric functions for the 25 selected male and female Polish talker bisyllabic words after intensity adjustment

Discussion

The main purpose of this research project was to develop a homogeneous subset of Polish bisyllabic words for use in measuring the SRT. This purpose was accomplished for recordings of a male and a female native Polish talker. The homogeneity of the subset of bisyllabic words can be seen by referring to Figures 5 (male) and 6 (female), which contain the predicted psychometric functions for the 25 selected bisyllabic words after intensity adjustment.

The slopes from 20 to 80% for the 25 bisyllabic words encompassed a range of 8.7 to 12.9 %/dB (M = 10.1 %/dB) for the male words and a range of 8.1 to 11.8 %/dB (M = 9.8 %/dB) for the female words. The means for the slopes from 20 to 80% for the Polish male and female bisyllabic psychometric functions are in close agreement with means for SRT materials that have been reported in other languages. The mean slope for English spondaic words has been reported to be as low as 7.2 %/dB [Wilson & Strouse 1999] to 8 %/dB [Hirsh et al. 1952] or as high as 12 %/dB [Beattie et al. 1975]. Both Hudgins et al. [1947] and Young, Dudley, and Gunter [1982] reported the mean slope for English spondaic words to be 10%/dB. The mean slope for Spanish trisyllabic SRT materials has been reported by Christensen [1995] to be 11.1 %/dB for a male talker and 9.7 %/dB for a female talker. In research involving Italian trisyllabic SRT materials, Greer [1997] reported a mean slope of 7.3 %/dB for a male talker.

A great deal of research remains to be done in the field of Polish speech audiometry materials. Future research could examine the similarities between the mean SRT obtained with the 25 adjusted bisyllabic words from this study and the mean PTA of the test subjects. Future research could also include examining bisyllabic word homogeneity and performance for hearing impaired individuals.

In addition to the research that can be conducted on the current Polish speech audiometry materials, there is also a need to develop new Polish speech materials. For example, speech materials could be created for children on the basis of word familiarity. Many of the present-day speech materials for Polish children contain words not highly familiar to children [M. Malesińska, personal communication, December 16 1999]. There is also a need to develop high-quality recordings of Polish speech materials used in aural rehabilitation for those with cochlear implants.

In summary, the bisyllabic studies resulted in the development of a homogeneous subset of 25 bisyllabic Polish words for use in measuring the SRT. These bisyllabic words are homogeneous with respect to audibility and also with respect to psychometric function slope. The bisyllabic words for both the male and female talkers are contained on the CD entitled Brigham Young University Polish Speech Audiometry Materials (Disc 1.0). These recordings can be utilized to measure the SRT in individuals whose native language is Polish.

References

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 [1991]. American National Standard maximum permissible ambient noise levels for audiometric test rooms (ANSI S3.1-1991). New York: ANSI.

American National Standards Institute [1996]. American National Standard specifications for audiometers (ANSI S3.6-1996). New York: ANSI.

American Speech-Language-Hearing Association Committee on Audiologic Evaluation. [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, (Supplement 2) 17-24.

Beattie R. C., Svihovec D. V., Edgerton B. J. [1975]. Relative intelligibility of the CID spondees as presented via monitored live voice. „Journal of Speech and Hearing Disorders" 40, 84-91.

Brandy W. T. [1966]. Reliability of voice tests of speech discrimination. „Journal of Speech and Hearing Research" 9, 461-465.

Bystrzanowska T. [1969]. Audiologia kliniczna. (Wyd. II). [Clinical audiology (2nd ed.)]. Warsaw: Politechnika Gdańska Instytut Radiokomunikacji.

Bystrzanowska T. [1978]. Audiologia kliniczna. (Wyd. III). [Clinical audiology (3rd ed.)]. Warsaw: Politechnika Gdańska Instytut Radiokomunikacji.

Cambron N. K., Wilson R. H., Shanks J. E. [1991]. Spondaic word detection and recognition functions for female and male speakers. „Ear and Hearing" 12, 64-70.

Campbell R. A. [1965]. Discrimination test word difficulty. „Journal of Speech and Hearing Research" 8, 13-22.

Cancel C. A. [1968]. Spanish speech audiometry. „International Audiology" 4, 206-208.

Carhart R. [1965]. Problems in the measurement of speech discrimination. „Archives of Otolaryngology" 82, 253-260.

Christensen L. K. [1995]. Performance Intensity Functions For Digitally Recorded Spanish Speech Audiometry. Unpublished master's thesis, Brigham Young University, Provo, Utah.

Comstock C. L., Martin F. N. [1984]. A children's Spanish speaking word discrimination test for non-Spanish-speaking clinicians. „Ear and Hearing" 5, 166-170.

Creston J. E., Gillespie M., Krohn C. [1966]. Speech audiometry: Taped vs live voice. „Archives of Otolaryngology" 83, 40-43.

Fletcher H. [1922]. The nature of speech and its interpretation. Journal of the Franklin Institute, 193, 729-747.

Greer L. F. [1997]. Performance Intensity Functions For Digitally Recorded Italian Speech Audiometry Materials. Unpublished master's thesis, Brigham Young University, Provo, Utah.

Harris R. W., Goffi M. V. S., Pedalini M. E. B., Gygi M. A. Merrill A., [2001]. Psychometrically Equivalent Brazilian Portuguese Trisyllabic Words Spoken by Male and Female Talkers. „Pro-Fono" 13(1), 37-53.

Harris R. W., Hilton L. M. [1991]. Speech Audiometry Materials. (Digital Compact Disc Recording) Provo, Utah: Department of Educational Psychology, 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. [1980]. Influence of the speaker and other factors affecting speech intelligibility. „Audiology" 19, 434-455.

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.

Kamm C., Carterette E. C., Morgan D. E., Dirks D. D. [1980]. Use of digitized speech materials in audiological research. „Journal of Speech and Hearing Research" 23, 709-721.

Kreul E. J., Bell D. W., Nixon J. C. [1969]. Factors affecting speech discrimination test difficulty. „Journal of Speech and Hearing Research" 12, 281-287.

Martin F. N., Sides D. G. [1985]. Survey of current audiometric practices. „Asha" 27, 29-36.

Nakamichi. [n.d.] Nakamichi MR-1 discrete head professional cassette deck owner's manual. Japan: Nakamichi Corporation.

Penrod J. P. [1979]. Talker effects on word-discrimination scores of adults with sensorineural hearing impairment. „Journal of Speech and Hearing Disorders" 44, 340-349.

Pruszewicz A., Demenko G., Richter L., Wika T. [1994]. Nowe listy artykulacynje do badań audiometrycznych. [New articulation lists for speech audiometry]. „Otolaryngologia Polska" 48, 50-54.

Ridgway J. [1986]. Compact disks - A revolution in the making. „Canadian Library Journal" 43(1), 23-29.

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.

Sony. [1991]. Compact disc player operating instructions. Japan: Sony Corporation.

Studio Audio and Video Limited. [1996]. Sadie Disk Editor (Version 3.0) [Computer Software].

Weisleder P., Hodgson W. R. [1989]. Evaluation of four Spanish word-recognition-ability lists. „Ear and Hearing" 10, 387-393.

Wilson R. H., Strouse A. [1999]. Psychometrically equivalent spondaic words spoken by a female speaker. „Journal of Speech, Language, and Hearing Research" 42, 1336-1346.

Young L. L., Dudley B., Gunter M. B. [1982]. Thresholds and psychometric functions of individual spondaic words. „Journal of Speech and Hearing Research" 25, 586-593.

Zakrzewski A., Jassem W., Pruszewicz A., Obrębowski A. [1975]. Identification and discrimination of speech sounds in monosyllabic meaningful words and nonsense words by children. „Audiology" 14, 21-26.

Zakrzewski A., Pruszewicz A., Rydzewski B. [1973]. W sprawie badanie słuchu audiometria mowy na terenie polski. „Otolaryngologia Polska" 27, 333-339.

Zgółkowa H. [1983]. Słownictwo współczesnej Polszczyzny mówionej. Lista frekwencyjna i rangowa. Poznań: Uniwersytet im. Adama Mickiewicza w Poznaniu.

Adres do korespondencji

David L. McPherson
Brigham Young University
129 TLRB, PO BOX 28633
Provo, Utah 84602-8633, USA
e-mail: david_mcpherson@byu.edu