Performance Norms for the VA Compact Disc Versions of CID W-22 (Hirsh) and PB-50 (Rush Hughes) Word Lists

Anastasia L. Heckendorf*
Terry L. Wiley†
Richard H. Wilson†

Abstract

The purpose of this study was to establish normative articulation functions for the Hirsh recordings of the CID W-22 word lists and the Rush Hughes recordings of the PB-50 word lists as recorded on the VA compact disc. Twenty-four young adults with normal hearing listened to both sets of materials presented in quiet at 12 levels (0–56 dB HL). Presentation levels on average needed to be 10 dB higher for the PB-50 word lists to obtain word recognition performance scores equal to those for the W-22s. Maximum word recognition scores of 95 to 100 percent were obtained for the W-22s at lower presentation levels (>40 dB HL), compared to 80 to 90 percent for the PB-50s at the maximum presentation level (56 dB HL). Mean slopes for the W-22 and PB-50 functions were 4.1 percent/dB and 1.9 percent/dB, respectively. Test-retest reliability was judged to be clinically appropriate for both sets of materials. Item analyses revealed a substantially skewed distribution toward easy items for the W-22 word lists compared to a more balanced distribution of item difficulty for the PB-50 lists. List equivalency was also judged to be clinically appropriate for the PB-50 word lists.

Key Words: Articulation functions, compact disc, speech audiometry, word recognition

Abbreviations: ANSI = American National Standards Institute, CD = compact disc, CNC = consonant-nucleus-consonant

In 1991, the Department of Veterans Affairs produced a compact disc (CD) (Speech Recognition and Identification Materials, Disc 1.1) that contained several sets of word recognition materials, including (1) a male recording of the Maryland consonant-nucleus-consonant (CNC) lists, (2) a female recording of the CID W-1 spondaic words and the Northwestern University Auditory Test No. 6 (NU-6), (3) the Rush Hughes recording of the PB-50 lists, and (4) the Hirsh recordings of the CID W-22 monosyllables (Wilson, 1993). Normative data on the CD version of the W-1 lists (Cambron et al, 1991) and the NU-6 lists (Wilson et al, 1990) are available. Normative data on the W-22 and PB-50 materials as recorded on the CD, however, are not available.

Despite a paucity of data on the characteristics of the Rush Hughes recordings of the PB-50 lists, available evidence indicates that the PB-50 materials are more sensitive than the W-22 lists to individual differences in word recognition performance for listeners with normal hearing and listeners with hearing impairment (Davis, 1948, 1978; Thurlow et al, 1948, 1949; Eldert and Davis, 1951; Hirsh et al, 1952; Silverman and Hirsh, 1955; Goldstein et al, 1956; Goetzinger et al, 1961; Carhart, 1965; Lovrinic et al, 1968; Goetzinger, 1972). An issue of primary interest in the present study was whether changes associated with the production of the CD recordings altered the performance characteristics of either the W-22 or the PB-50 material. Also, although several reports discuss differences in word recognition performance for the W-22 and PB-50 lists (Hirsh et al, 1952; Goetzinger and Rousey, 1959; Goetzinger et al, 1961; Carhart, 1965, 1970; Goetzinger, 1972; Davis, 1978), a direct performance comparison of the two tests on the same group of sub-
jects is lacking. Thus, the main purpose of the current investigation was to provide normative word recognition data on the W-22 (Hirsh) and PB-50 (Rush Hughes) word lists as recorded on the VA CD. Secondarily, test–retest measures and item analyses were examined for the two sets of materials. Finally, list equivalency was evaluated for the PB-50 word lists.

METHOD

Subjects

The volunteer subject group consisted of 12 men and 12 women (N = 24) 19 to 34 years (mean = 26 years) of age. All subjects had (1) a negative history of any recent otic disorder and normal otoscopic findings (ASHA, 1990), (2) normal middle ear function defined according to ASHA (1990) guidelines, (3) an ipsilateral acoustic reflex for a 1000-Hz tonal activator <95 dB HL, and (4) air-conduction thresholds of 0 to 10 dB HL (ANSI, 1989) at octave intervals from 250 to 8000 Hz. For each subject, the ear with the lower mean threshold at 1000, 2000, and 4000 Hz served as the test ear; an equal number of right and left ears were represented.

Instrumentation and Calibration

Tympanograms and acoustic reflexes were obtained using an acoustic admittance screening instrument (Grason-Stadler, Model GSI 27A) that met ANSI S3.39-1987 specifications. Pure-tone air-conduction audiometry was performed with a console audiometer (Grason-Stadler, Model GSI 16) in accordance with ANSI S3.6-1989. Testing was performed with the subject seated in a sound room (Industrial Acoustics Company, 1200 Series). The test signals were delivered to an earphone (TDH 50-P) mounted in a supra-aural cushion (Telephonics P/N 510C017-1). Word recognition stimuli from Speech Recognition and Identification Materials, Disc 1.1 were directed from a CD player (Sony, Model 497) to a single channel of the audiometer. The speech channel of the audiometer was calibrated to 19.7 dB SPL for a 1000-Hz tone.

Procedures

Each subject participated in three 60- to 90-minute sessions, with a 10-minute break in the middle of each session. During the first session, either the PB-50 word lists (Lists 8, 9, 10, 11) or the W-22 word lists (Lists 1, 2, 3, 4) were presented monaurally in quiet at 12 levels (0, 4, 8, 12, 16, 20, 24, 28, 32, 40, 48, 56 dB HL). The remaining set of lists was presented under the same conditions during the second session. During the third session, performed within 1 week of initial testing, each subject was retested with three lists from each set of materials at levels of 8, 20, and 32 dB HL to provide data for test–retest reliability. All four PB-50 lists also were presented during the final third session at 56 dB HL for evaluation of list equivalency.

The order of list presentation was randomized with restrictions across subjects and levels so that the same word list was never presented at consecutive levels. An ascending order of levels was used to minimize any possible learning effects. For all word recognition testing, subjects wrote their responses. Correctness was based on phonetic or homonym spelling with omissions of final consonants, plurals, or tense markers judged as incorrect.

RESULTS AND DISCUSSION

Articulation Functions

Table 1 lists and Figure 1 illustrates mean word recognition scores in percent correct obtained for the W-22 and PB-50 materials from the 24 listeners. The lines through the data points are the best-fit, second-degree polynomials. Except for the two lowest presentation levels, recognition performance on the W-22s exceeded performance on the PB-50s. As expected from the binomial model (Thornton
The results of the current investigation suggested that the PB-50 word lists recorded by Rush Hughes (recordings available on CD) continued to be more sensitive to individual differences in word recognition abilities than the CID W-22 recordings, at least for listeners with normal hearing. The PB-50 lists did not result in articulation functions that demonstrated pronounced plateau regions, as observed for the W-22 functions. Rather, the function for the PB-50s continued to rise gradually even at the highest presentation levels for the majority of subjects. The steeper slope of the W-22 function (4.1%/dB) indicated that small increases in presentation level resulted in larger changes in word recognition performance relative to the PB-50s. A single increase in presentation level of 4 to 5 dB corresponded to a 16 to 20 percent improvement in W-22 word recognition performance. In contrast, the shallower slope (1.9%/dB) associated with the PB-50 function indicated that considerably smaller changes in performance occurred with increments in presentation level. An increase of 4 to 5 dB corresponded to only an 8 to 10 percent improvement in word recognition scores. The PB-50 lists more precisely differentiated individual word recognition abilities across levels than did the W-22 lists.

Figure 1  Articulation functions for the W-22 (circles) and PB-50 (squares) lists recorded on the VA CD. The mean correct recognition (%) is plotted as a function of the presentation level (dB HL). The lines connecting the data points are the best-fit, second-degree polynomials (W-22—$y = -2.2045 + 4.5885x - 0.051125x^2$, $R^2 = 0.99$; PB-50—$y = 3.9302 = 2.9315x - 0.026023x^2$, $R^2 = 0.99$).

and Raffin, 1978), variability was greatest at the middle presentation levels (see Table 1). Ceiling effects were observed for the W-22s with many 100 percent scores obtained at the highest levels. No 100 percent scores were obtained on the PB-50s.

The differences between functions calculated from the polynomial equations progressed from 0.7 dB at 20 percent, to 5.5 dB at 50 percent, to 15.8 dB at 80 percent. Data for the PB-50s were at higher levels for each percentage. The slopes of the mean functions calculated from the same equations at the 20 percent and 80 percent correct points were 3.1 percent/dB for the W-22s and 1.7 percent/dB for the PB-50s. In addition to the slopes of the mean functions, the mean slopes of the individual functions were calculated. For each individual set of data, slopes of the articulation functions were estimated at the presentation levels that most closely approximated word recognition scores of 20 percent and 80 percent correct. The slopes of the functions for the individual subjects ranged from 2.8 percent/dB to 5.7 percent/dB (W-22s) and from 1.2 percent/dB to 2.8 percent/dB (PB-50s). The mean slope of the individual functions for the W-22s was 4.1 percent/dB (±0.9%/dB), whereas the mean slope of the individual functions for the PB-50s was 1.9 percent/dB (±0.4%/dB). The mean slopes were significantly different ($F[1,23] = 181.32; p < .05$).

The original vs Current Functions

For the past 40 years, the original articulation function for the W-22s (Hirsh et al, 1952) has been a benchmark against which practically all word recognition materials have been compared. The slope of the original W-22 function was about 4 percent/dB with the 50 percent point at 24 dB SPL. Based on this original function, the W-22s are always cited as being easier than any other comparable monosyllabic word recognition material. Until the current study, however, the original articulation function for the Hirsh version of the W-22s has not been replicated. Figure 2 illustrates the articulation functions for the original data (filled circles) and for the current data (open circles). The original W-22 data were extracted from Hirsh et al (Fig. 1; 1952) by adding 20 dB to sound pressure level to determine the corresponding presentation level (dB HL). For reference purposes, the derived data points from Hirsh et al are listed in Table 2. To simplify computational comparisons, the data were then fit with a second-degree polynomial. The two W-22 functions are
Table 2  Levels (dB HL) Corresponding to Percent Correct Performance Extracted from Second-Degree Polynomial Equations Applied to Data Obtained from the Original W-22 and PB-50 Recordings Reported by Hirsh et al (1952) and Davis (1948) and the Current Investigation (CD 95) for the CD Version of These Same Tests

<table>
<thead>
<tr>
<th>Percent Correct</th>
<th>W-22 Hirsh (1952)</th>
<th>PB-50 Davis (1948)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>12.8</td>
<td>24.8</td>
</tr>
<tr>
<td>70</td>
<td>9.3</td>
<td>20.4</td>
</tr>
<tr>
<td>60</td>
<td>6.5</td>
<td>16.6</td>
</tr>
<tr>
<td>50</td>
<td>4.0</td>
<td>13.4</td>
</tr>
<tr>
<td>40</td>
<td>1.8</td>
<td>10.4</td>
</tr>
<tr>
<td>30</td>
<td>-0.2</td>
<td>7.7</td>
</tr>
<tr>
<td>20</td>
<td>-2.2</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Mean slope 4.0%/dB 3.1%/dB 3.4%/dB 1.7%/dB

displaced about 10 dB with the original function requiring less SPLs for comparable recognition performance. The differences between the W-22 functions averaged 9.5 dB and ranged from 7.3 dB at 20 percent correct, to 9.4 dB at 50 percent correct, to 12.0 dB at 80 percent correct. The slopes derived from the 20 to 80 percent points on the polynomials fit to the mean W-22 functions were 4 percent/dB (Hirsh et al, 1952) and 3.1 percent/dB (current study).

The reasons for the differences between the functions are not entirely clear. The same recordings of the materials by the same speaker (Hirsh) were evaluated, so there was no speaker difference. Calibration, however, may have been a major contributor to the observed differences for the two functions. A difference between the levels of the calibration tone and W-22 materials on the Technisonic records has been apparent for many years (Gengel and Kupperman, 1980). The Technisonic records were derived directly from the magnetic tapes from which the original W-22 function in Figure 2 was generated. To verify the reported difference between the levels of the calibration tone and W-22 materials on the Technisonic records, the levels on the 1969 records were examined using a computer and vu meter arrangement. The calibration tone and stimulus materials were reproduced (Panasonic, Model SG-HMO9) with a 1-mil stylus (Lilly and Franzen, 1968) and digitized with a 16-bit analog-to-digital converter (Antex, Model SX-10) at 20,000 samples/sec. Subsequently, each carrier phrase and target word set were edited into individual files. Routines with a waveform editor were then used to block the carrier phrase and repetitively reproduce it with the output of the digital-to-analog converter monitored on a vu meter (Tascam, Model MU-40). After monitoring the stimulus several times on the vu meter, the peak vu reading could be estimated within ± 0.5 dB. Here and throughout, 1 vu is equal to 1 dB for a sinusoid (IEEE, 1992, p. 1475). A similar process was used with the target words except that the last word of the carrier phrase ("say") was included in the block. Because the calibration tone was about 8 dB below the vu level of the carrier phrase peaks, 8 dB were added digitally to the level of the digitized calibration tone. In this manner, all readings were accomplished around 0 vu (i.e., the lower end of the vu scale was not used).

The level measurements revealed that the 1000-Hz calibration tones on the Technisonic discs were on average 10.1, 8.9, 8.2, and 10.0 vu below the peaks of the carrier phrases in Lists 1, 2, 3, and 4, respectively. Standard deviations were approximately 1.0 vu. The target words for Lists 1, 2, 3, and 4 were 7.1, 7.9, 7.6, and 8.8 vu, respectively, above the level of the calibration tone. Similar findings have been reported using a graphic level recorder (Gengel and Kupperman, 1980). In the Gengel and Kupperman report, the carrier phrases on the Technisonic discs were about 8 dB above the level of the calibration tone, with the W-22 words being 6 to 7 dB above

![Figure 2](image-url)
the level of the tone. Further confirmation of the level difference on the Technisonic records of the W-22 materials comes from the slightly different prospective of Studebaker and Sherbecoe (1991), who reported that the calibration tone was "2.3 dB below the long term RMS level of the test words." Thus, data on the Technisonic Studio recordings of the W-22 materials from three measurement techniques indicates that the calibration tone is 8 to 10 vu (or dB) below the peak level of the carrier phrases.

In the original report, Hirsh et al. (1952, p. 188) reported that "the 1,000 cps calibration tone on the inner face of every record is at the average level of the carrier phrases." Two relevant assumptions can be posited. First, the "average level" used by Hirsh et al. (1952) to calibrate the discs is not the same as the "peak level" currently used to calibrate. Second, the same calibration procedures and levels were used on both the magnetic tape version and the record version of the W-22s. Psychometrically, this set of circumstances would result in an articulation function that is displaced to lower sound pressure levels in comparison to the articulation function that is based on the calibration tone reflecting the peaks of the carrier phrases. This is consistent with the relation observed in Figure 2 in which the W-22 function from Hirsh et al. (1952) occurs at presentation levels 10 dB lower than the W-22 function generated in the current investigation.

Perhaps an additional contributing variable to the difference between the Hirsh et al. (1952) and the current study was the familiarity that the subjects had with the stimulus materials, the listening task, the listening environment, and the speaker. In the original study, the subjects (1) were given a list of the 200 words to read at the beginning of each of eight 2.5-hour test sessions on consecutive days; (2) listened to six randomizations of each of the four lists presented at 100 dB SPL, which served to "indoctrinate" the listeners; and (3) listened to each of the 24 lists (six randomizations of four lists) at five to seven 10-dB intervals on the articulation functions. Thus, each subject heard each of the four word lists 36 to 48 times. In the current study, the subjects were not allowed to read or listen to test lists other than when they were exposed to each list four times during the experiment. Furthermore, an ascending level procedure was used so that the initial presentations would not influence performance at subsequent higher presentation levels. Based on the familiarity variable, we conclude, as did Hirsh et al. (1952), that the data from the original W-22 study should be considered "experimental" based on the use of "experienced"
listeners, whereas the data from the current study are more directly representative of the clinical situation that involves patients with “naive” listening skills.

In Figure 3, the original articulation functions for the Rush Hughes recording of the PB-50s (filled squares) are illustrated along with the PB-50s’ articulation function from the current study (open squares). The original PB-50 data based on 10 listeners were extracted from Davis (Fig. 1; 1948), again using 20 dB to convert to hearing level (see Table 2), and were fit with a second-degree polynomial. For equal intelligibility, the PB-50 materials presented from the CD required higher presentation levels than were required in the reports of 40 years ago.

For the PB-50 materials, the functions in Figure 3 differed by 0.7 dB at 20 percent correct, 5.9 dB at 50 percent, and 17.9 dB at 80 percent correct with an average difference between the 20 percent and 80 percent correct points of 7.3 dB. Additionally, the observed slope of the original function (3.4%/dB) is twice the slope of the function from the current study (1.7%/dB). The paucity of details in the Davis report (1948) about the experimental conditions, subjects, calibration, etc. makes it difficult to speculate on the differences between the two functions. We would, however, like to offer two scenarios. First, with the original Davis data, there is the suggestion (Davis et al, 1949) that the 1000-Hz calibration tone was 2.5 dB below the peaks of the carrier phrase, which would account for some of the difference between the two functions shown in Figure 3. Second, time and the copying/recording process have degraded the PB-50 materials with the substantial loss of high-frequency energy. Support for this comes from observations of the digital waveforms of the Rush Hughes recordings. The amplitudes of the initial and final consonants, especially the /s/ and /sh/, are substantially reduced in comparison to initial and final consonants in other speech materials that we have observed. This type of degradation could account for portions of the difference in the slopes of the functions shown in Figure 3.

**Comparisons across CD Functions**

Figure 4 illustrates the functions obtained in the present investigation for the CD versions of the W-22 (circles) and PB-50 (squares) word lists and in a study by Wilson et al (1990) for the CD version of the NU-6 (triangles) word lists. At the lower presentation levels (below 12 dB HL), the means for the NU-6 word lists were lower than those for both the W-22 and PB-50 lists. At 16 dB HL, the means for the NU-6 and PB-50 recordings were essentially the same (44.0% ±15.6% and 46.2% ±10.0%), respectively. Above 24 dB HL, the mean scores for the NU-6 word lists closely approached the mean scores for the W-22 word lists. Differences between mean scores for the W-22 and NU-6 recordings were less than 4 percent at presentation levels of 24, 28, and 32 dB HL. Data for the NU-6 recordings were not available for the three highest presentation levels (40, 48, and 56 dB HL) used in the current investigation.

The corresponding slopes of the mean functions for each of the three sets of materials were compared. The steepest slope occurred for the NU-6 word lists (4.5%/dB); the slope for the CID W-22 word lists (4.1%/dB) was slightly shallower. Beyond the linear portion of the articulation function (28 dB HL), however, the plateau regions of the functions were essentially indistinguishable. In contrast, the mean slope for the PB-50 word lists (1.9%/dB) was about half as steep as the functions for either the W-22 or NU-6 recordings. Not only does the linear portion vary, but the function for the PB-50s did not result in a characteristic plateau region. Performance for the PB-50s continued to improve with increased presentation levels, even at the highest levels studied.

**Test-Retest Reliability**

Retest performance scores were compared to initial test scores at levels of 8, 20, and 32 dB HL. The results of this comparison are listed in Table 3 and are illustrated as bivariate plots in Figures 5 (W-22) and 6 (PB-50). Six individual one-way, repeated measures analyses of variance (ANOVAs) were completed by level to determine if significant differences existed between initial and retest scores for both the W-22s and PB-50s. A significant difference between initial and retest conditions was found only for the CID W-22 word lists at a presentation level of 8 dB HL (F[1,23] = 14.84; p < .05); a test–retest difference of approximately 9 percent was observed for this condition, compared to 3 to 4 percent for other retest conditions. For all other analyses, there were no significant differences (p > .05) between initial and retest conditions. Mean retest scores were higher than initial mean scores for each condition, indicating the possibility of a slight learning effect. The retest increase of 1 to 4 percent, however, corresponded
Table 3  Mean Word Recognition Scores and Standard Deviations (Both in Percent Correct) Obtained at Three Levels (8, 20, and 32 dB HL) for Test and Retest Presentations of the W-22 and PB-50 Recordings

<table>
<thead>
<tr>
<th></th>
<th>8 dB HL</th>
<th>20 dB HL</th>
<th>32 dB HL</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Retest</td>
<td>Test</td>
<td>Retest</td>
<td>Test</td>
</tr>
<tr>
<td>W-22</td>
<td>Mean</td>
<td>25.1</td>
<td>33.7</td>
<td>73.3</td>
<td>75.4</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>14.3</td>
<td>13.7</td>
<td>11.1</td>
<td>11.0</td>
</tr>
<tr>
<td>PB-50</td>
<td>Mean</td>
<td>24.1</td>
<td>27.2</td>
<td>53.0</td>
<td>56.6</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>9.6</td>
<td>9.0</td>
<td>12.0</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Subjects were 24 listeners with normal hearing.

Norms for W-22 and PB-50s on CD/Heckendorf et al

Figure 5  Bivariate plots of the test (abscissa) and retest (ordinate) percent correct recognition of the W-22 materials at 8 dB HL (squares), 20 dB HL (circles), and 32 dB HL (triangles). The diagonal line represents equal test–retest performance.

List Equivalency

Although Hirsh et al (1952) demonstrated list equivalency for the W-22 word lists, similar information was not available for the PB-50 lists. Therefore, performance equivalency for the four PB-50 word lists recorded by Rush Hughes (8, 9, 10, 11) was investigated to determine if word recognition scores differed across lists; presentation was at 56 dB HL in an attempt to elicit maximum performance. Data presented by a few earlier researchers suggested that the original phonograph versions of the PB-50 word lists were not equivalent; some lists were reportedly easier or harder than others (Eldert and Davis, 1951; Wolfe and O’Connell, 1959). The four lists available on CD (8, 9, 10, 11) resulted in mean word recognition scores and standard deviations of 88.1 percent (±4.5), 89.0 percent (±4.3), 84.0 percent (±4.4), and 91.7 percent (±4.2), respectively. These results were similar to previously reported data in that List 11 was consistently easier than the other three lists (Eldert and Davis, 1951). In contrast to earlier data for phonograph versions of the PB-50s (Eldert and Davis, 1951), List 10 resulted in the lowest mean score rather than List 8. Mean scores for the current investigation varied by less than 8 percent with the standard deviations across lists essentially equivalent (approximately 4%). A list × subject repeated measures ANOVA revealed a significant main effect for list (F[3,69] = 19.09; p < .05). Post-hoc analyses were completed using Scheffe’s test (Bruning and Kintz, 1977) to further evaluate the interactions between lists; the critical difference score associated with a probability of .05 was 3.66 percent correct. Statistical analyses revealed that only interactions with List 10 were significant; scores for List 10 were consistently lower than for the other lists. Even though statistical analyses revealed a significant difference between lists, the amount of difference between means (2–8%) corresponded to one to
four items based on a 50- or 25-item test. Only for extreme word recognition scores of 0 percent and 100 percent would differences of these percentages shift an individual's score to a different confidence interval (Thornton and Raffin, 1978). Therefore, the PB-50 word lists on CD recorded by Rush Hughes were judged to be equivalent because clinical decisions would not be altered based on differences of this magnitude.

Error Analyses

Correct and incorrect responses for each word contained in the W-22 and PB-50 word lists were tallied based on 72 possible occurrences for each item (24 subjects were presented each word at three levels). Several items on the two tests were either scored correct or incorrect by a large proportion (>80%) of subjects. There were 18 such words identified correctly for the W-22s, compared to 15 for the PB-50 lists. In contrast, there were 11 words in the PB-50 word lists that were incorrect for more than 80 percent of the subjects. None of the W-22 words were incorrect 80 percent of the time. Figure 7 provides the number of items within each set of lists associated with a given number of correct responses across presentation levels. Visual inspection of the graphs in addition to the itemized error analysis revealed a highly skewed clustering of easy items for the CID W-22 word lists. Very few items were missed for more than half of the presentations across levels. In contrast, items included in the Rush Hughes lists were more evenly distributed across a wide range of difficulty just as Egan (1948) had intended when he constructed the PB-50 word lists. The majority of PB-50 items were near the center of the distribution, indicating that they were almost equally identified as correct and incorrect. Although fewer items were present near the ends of the distribution, all levels of difficulty were represented to some degree for the PB-50 words.

These error analyses are similar to those reported by Lovrinic et al (1968) for a single list of both the CID W-22 and Rush Hughes recordings available on phonograph records.
For both the current investigation and that of Lovrinic et al. (1968), more items were consistently scored correct from the W-22 word lists than from the PB-50 word lists. Similarly, the PB-50s resulted in a number of items that were perceived incorrectly by a large proportion of subjects compared to 0 and 2, for those same two studies, respectively.

The results of the current study, then, indicated that the transfer of the original tape recordings to CD did not alter the distribution of item difficulty or the pattern of item identification for either test. The item analysis provided additional support for the greater sensitivity of the Rush Hughes recordings of the PAL PB-50 word lists for differentiating word recognition abilities compared to the CID W-22 recordings. Results of the current item analysis may provide insight into the development of more sensitive word recognition tests based on reconstruction of the Rush Hughes word lists. Specifically, elimination of items that were extremely easy or difficult would provide a more balanced range of difficulty and, potentially, a more sensitive test.

Summary

The purpose of this investigation was to provide normative data for the CID W-22 and PB-50 word recognition tests available on CD. Presentation levels on average needed to be 10 dB higher for the PB-50 word lists in order to obtain word recognition performance scores equal to those for the W-22s. Maximum word recognition scores of 95 to 100 percent were obtained for the W-22s at lower presentation levels (≥40 dB HL), compared to 80 to 90 percent for the PB-50s at the maximum presentation level (56 dB HL). The slope for the W-22 articulation function (4.1%/dB) was over twice that for the PB-50 function (1.9%/dB). Error analyses demonstrated significant differences in item difficulty between W-22 and PB-50 word lists; W-22 scores were skewed toward high scores and PB-50 scores evidenced centralized distributions. Test-retest reliability was judged to be good for both the W-22 (Hirsh) and PB-50 (Rush Hughes) recordings available on CD. Similarly, the PB-50 word lists on CD were judged to be clinically equivalent.

Acknowledgments. The CD used for this study, Speech Recognition and Identification Materials, Disc 1.1, can be obtained through the Long Beach Research Foundation. Contact Richard Wilson, Ph.D., at the Veterans Affairs Medical Center, 126 Audiology, Mountain Home, TN 37684, (423) 926-1171, Ext. 7553.

The first author would like to thank her coauthors for their numerous revisions and input leading to the completion of this project.

A paper based on portions of this research was presented at the 1994 meeting of the American Speech-Language-Hearing Association in New Orleans, LA.

This work is based in large part on a master's thesis by the first author.

REFERENCES


Institute of Electrical and Electronics Engineers. (1992). *Electrical and Electronics Terms.* York: IEEE.


