Recognition performance on single-speaker recordings of W-22, NU6, and PB-50 by listeners with normal hearing

RICHARD H. WILSON[1] AND RACHEL MCARDLE[2]

[1] *Auditory Research Laboratory (126), James H. Quillen VA Medical Center, Mountain Home, Tennessee, 37684, USA and Departments of Surgery and Communicative Disorders, East Tennessee State University, Johnson City, Tennessee, 37614, USA*

[2] *Auditory Research Laboratory (126), Bay Pines VA Healthcare System, Bay Pines, Florida, 33744, USA and Department of Communicative Disorders and Sciences, University of South Florida, Tampa, Florida, 33620, USA*

The psychometric characteristics of the PB-50, CID W-22, and NU No. 6 monosyllabic word lists were compared with one another, with the CID W-1 spondaic words, and with the nine monosyllabic digits. The 583 words were spoken by the same speaker and were presented at 4 levels (−7-, −2-, 3-, and 8-dB S/N) in speech-spectrum noise fixed at 72-dB SPL. Twenty-four young adults with normal hearing participated in four sessions. Recognition performance on the four lists within each of the three monosyllabic word materials were equivalent, ±0.4 dB. Likewise, word-recognition performance on the PB-50, W-22, and NU No. 6 word lists were equivalent, ±0.2 dB. The mean recognition performance at the 50% point with the 36 W-1 spondaic words was ~6-dB better than mean performance on monosyllabic words. Recognition performance on the monosyllabic digits was 1-2 dB better than mean performance on the monosyllabic words.

INTRODUCTION

Throughout the literature, different psychometric properties have been observed for different lists of word-recognition materials (e.g., Silverman and Hirsh, 1955; Lovrinic et al, 1968; Wilson et al, 1976; Beattie et al, 1977; Heckendorf et al, 1997; Wilson and Oyler, 1997) and even for the same materials spoken by different speakers (e.g., Tillman and Carhart, 1966; Wilson et al, 1990). Kruel et al (1969) were among the first to recognize that the psychometric characteristics of word lists were only valid for the word lists as spoken on one occasion by a speaker. The question that prompted the current experiment was which of the three common monosyllabic sets (PB-50, W-22, and NU No. 6) would best serve in the replication of portions of the classic Miller, Heise, Lichten (1951) study involving monosyllabic words and digits? The decision was made to record four lists of each of the three monosyllabic word sets, the monosyllabic digits not included in the three sets of word lists, and, as a point of reference, the CID W-1 spondiac words (Hirsh et al, 1952). The purpose of the current study was to determine on listeners with normal hearing the psychometric properties of the five speech materials when presented at four signal-to-noise ratios (SNR) in speech-spectrum noise using the 50% point for each word established with the Spearman-Kärber

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

equation (Finney, 1952) and the percent correct on each word at each SNR. A companion paper examines the acoustic, phonetic/phonological, and lexical variables that may predict the relative ease or difficulty for which these monosyllable words were recognized in noise (McArdle and Wilson, 2007).

METHODS

Four of the PB-50 lists (8, 9, 10, and 11), the 4 CID W-22 lists, the 4 NU No. 6 lists, the nine monosyllabic digits, and the 36 CID W-1 spondaic words were included in the study. Of the 600 possible monosyllabic words (3 materials x 4 lists x 50 words), 485 words appear only once in the 12 lists, 56 words appear in 2 sets, and one word (*have*) appears in all three sets. Only 4 of the 9 monosyllabic digits appear in the 12 word lists.

The 583 words were recorded digitally (Macintosh, G4) in random order by a professional female speaker during four recording sessions within a two-week interval. The recording microphone (Sanken, Model CU44X) was protected by a pop screen (Stedman, Model PS-101). A speech coach and two audiologists monitored and judged the spoken materials with three trials of each word recorded using the carrier phrase, "you will cite", which ends in a hard stop, was monitored to zero on a vu meter with the level of the words falling naturally. The materials were edited to select which of the three productions of each word were to be used. One carrier phrase was selected and concatenated to each of the 583 words using a 200-ms silent interval between the carrier phrase and the word.

Four signal-to-noise ratios (SNRs) were used (−7, −2, 3, and 8 dB), which were selected to generate psychometric functions from 10-20% correct to 80-90% correct. There were 2332 stimulus files (583 words x 4 SNRs). Four randomizations of the 583 words were made for each of the four SNRs incorporating a 3-s inter-stimulus interval. The words were mixed digitally with a sample of speech-spectrum noise to achieve the appropriate SNR and were compiled into 11, 50-word lists and one 33-word list.

Twenty-four, young adult listeners (18 to 30 years; mean = 23 years) participated. The listeners had pure-tone thresholds ≤20-dB HL at octave and inter-octave frequencies from 250 to 8000 Hz (ANSI, 1996). All participants were university students who were unfamiliar with the test paradigm and test materials.

The listeners subsequently participated in four, 60-minute sessions. Using a counterbalanced design, each of the four randomizations of the 583 words was presented an equal number of times at each SNR. In this manner the 583 words randomly were intermingled. During each of the four sessions 583 words were presented with ~25% of the words presented at each of the four SNRs, which was a further step to avoid any order effects. The level of the speech-spectrum noise was fixed at 72-dB SPL with the level of the speech varied from 65-dB SPL (−7-dB S/N) to 80-dB SPL (8-dB S/N) in 5-dB steps.

The word-recognition materials were reproduced by a CD player (Sony, Model G150), fed through an audiometer (Interacoustics, Model AC40), and delivered to the test ear through an ER-3A insert phone with the non-test ear covered by an insert phone. All testing was conducted in a double-wall sound booth and the responses logged into a spreadsheet for analysis.

RESULTS AND DISCUSSION

The overall results indicated that the mean 50% point for the 583 words across the four SNRs was 0.4-dB S/N with a standard deviation of 3.6 dB. The 50% points ranged over a 17.9 dB interval from −8.3-dB S/N (*choice*) to 9.7-dB S/N (*bob*). Interestingly, four of the monosyllabic words were easier than the easiest spondaic word (*toothbrush*). The overall mean correct recognition for the 583 words was 50.7% (SD = 17.9%) ranging from 93.8% (*choice*) to 4.2% (*bob*). The data from the 583 words were parsed into the four lists of the three monosyllabic word sets, the spondaic word list, and the monosyllabic digit list for subsequent evaluations.

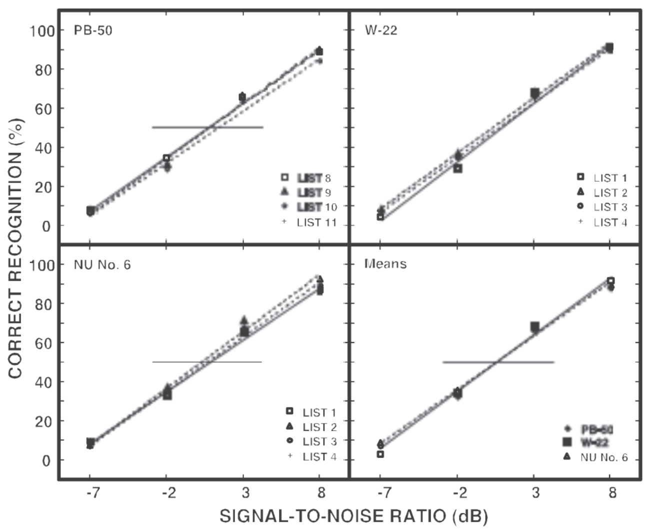

Fig. 1: The mean percent correct recognition at four signal-to-noise ratios for the four lists of each of the three monosyllabic words list materials.

Monosyllabic words

Graphic analysis of the PB-50, W-22, and NU No. 6 data are presented in Figure 1, which depicts the mean functions for each of the four lists that comprise each of the three monosyllabic word sets (quadrants 1-3). The mean data are plotted in the 4th quadrant. The 50% points on the linear regressions and the slopes are listed in the right columns of Table 1. The mean data for the four lists of each of the three word sets at each SNR are in close agreement with one another with overall mean percent correct values of 47.9%, 49.9%, and 49.6% for the PB-50, W-22, and NU No. 6 lists, respectively. From the data in Figure 1 and the table, there are minimal differences both among the four lists of each set of material and among the three word sets (4th quadrant). As can be seen in Table 1, the variability among the four lists of each material at the 50% points essentially were the same with minimum, maximum differences between lists of 0.6 to 0.7 dB. The mean 50% points for the three materials (column 6) ranged 0.4 dB, whereas the slopes of the mean functions varied 0.7%/dB from 5.3%/dB to 6.0%/dB.

List	Mean (dB S/N)	SD (dB)	Max (dB)	Min (dB)	50% Point (dB S/N)	Slope (%/dB)
PB-50						
8	0.7	3.6	8.2	−5.3	0.7	5.5
9	0.8	3.4	8.2	−8.0	0.8	5.7
10	1.5	3.6	9.7	−7.4	1.4	5.4
11	0.7	3.4	9.3	−7.0	0.7	5.6
Mean	0.9	3.5	8.9	−6.9	0.9	5.3
W-22						
1	0.8	2.5	8.6	−4.7	0.8	6.0
2	0.0	3.4	6.3	−8.0	0.1	5.7
3	0.5	3.3	6.8	−7.4	0.5	5.6
4	0.8	2.9	7.2	−6.2	0.7	5.9
Mean	0.5	3.0	7.2	−6.6	0.5	5.8
NU No. 6						
1	0.8	3.8	9.3	−8.3	0.8	5.3
2	0.5	3.8	7.2	−7.8	0.5	5.8
3	0.6	3.4	8.6	−6.8	0.6	5.6
4	0.2	3.2	8.2	−6.2	0.2	5.4
Mean	0.5	3.6	8.3	−7.3	0.5	5.5

Table 1: *The mean 50% points and standard deviations (inter-subject) calculated with the Spearman-Kärber equation from the individual word data are listed along with the 50% points and slopes calculated with the linear regressions.*

To the best of our knowledge, this is the first direct comparison of the recognition performances achieved on the PB-50, W-22, and NU No. 6 materials spoken and recorded by the same speaker during the same recording sessions. Two previous studies (Beattie et al, 1977; Wilson and Oyler, 1997) compared performances on the W-22 and NU No. 6 materials spoken by the same speaker (Auditec of St. Louis) but recorded during different sessions. Beattie et al found the NU No. 6 materials were slightly more difficult than the W-22 materials, whereas Wilson and Oyler found that recognition performances on the NU No. 6 materials were 4 to 8% better than the performances on the W-22 materials. For both studies, the small differences between the recognition performances on the two materials, which were in opposite directions, were significant. The data from the current study observed no appreciable differences among performances on the PB-50, W-22, and NU No. 6 materials that were spoken by the same speaker.

The data generated made possible an evaluation of word lists with respect to the phonetic/phonemic balance incorporated into the lists (Egan, 1948; Lehiste and Peterson, 1959; Peterson and Lehiste, 1962). The concept of phonetic/phonemic balance of a word list is attractive in that it is reasonable to evaluate the ability of a listener to understand speech using a variety of speech sounds. We are unaware, however, of a study that has evaluated the (un)importance of including phonetic/phonemic balance as a characteristic of word lists used for word-recognition testing. Figure 2 shows the mean 50% points derived with the Spearman-Kärber equation for the 12 "organized" 50-word lists studied (PB-50, W-22, NU No. 6) and 12, 50-word lists that were compiled randomly from the 540 monosyllabic words.

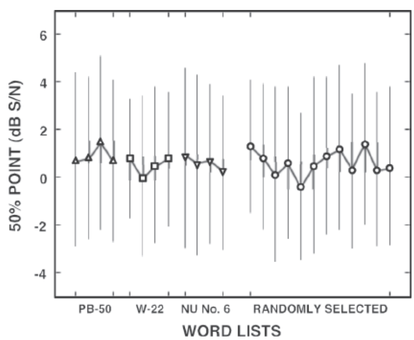

Fig. 2: The mean 50% points for the individual words calculated with the Spearman-Kärber equation are shown for lists 1-4 of the three word lists and 12 randomly compiled lists (circles). The vertical bars indicate one standard deviation.

The vertical lines are the standard deviations. The randomly selected lists demonstrate slightly more noise in the data (i.e., inter-list variability), but the absolute magnitude of the differences, both within and between the organized and random groups of words, is small. For the organized word lists, the 50% points ranged 1.3 dB, whereas the 50% points for the lists of words selected randomly ranged 1.8 dB. These sets of ranges are comparable and <2-dB, which with consideration to the slopes of the functions discussed earlier is an ~10% disparity. The conclusion from this analysis is that the concept of phonetic/phonemic balance has minimal impact on the intelligibility characteristics of word-recognition materials. Other characteristics, such as the speaker of the materials and presentation level, are much more influential on the word-recognition performance.

Spondaic Words

The mean 50% point for the 36 spondaic words calculated with the Spearman-Kärber equation was −5.3-dB S/N with a 1.4-dB standard deviation. The 50% points ranged 5.4-dB from −2.2-dB S/N (farewell) to −7.6-dB S/N (toothbrush). The 50% point calculated from the mean linear regression was −5.5-dB S/N with a slope of 9.9%/dB. In contrast, the mean 50% point calculated from the functions of the individual 36 spondaic words was −6.0-dB S/N (SD = 4.1 dB) and the mean slope was 12.7%/dB (SD = 4.2 dB). Because of the dynamics involved, the mean slope of the 36 functions was 2.8%/dB steeper than the slope of the mean function of the 36 words and is a better predictor of the slope of the functions for the individual spondaic words (Wilson and Margolis, 1983). The 50% point on the mean spondaic word function was 6.0 to 6.5 dB lower than the 50% point on the mean function for the monosyllabic words and 4.3-dB lower than the 50% point on the mean function for the digits. These differences are only slightly smaller than the ~10-dB difference observed by Hirsh et al (1952) between the W-1 spondaic words and the W-22 monosyllabic words. The larger difference observed by Hirsh et al may be attributable to differences in the experimental designs. In the current study, the spondaic words were intermingled with the monosyllabic words whereas with the Hirsh et al study, the spondiac and monosyllabic words were each given as a set of materials.

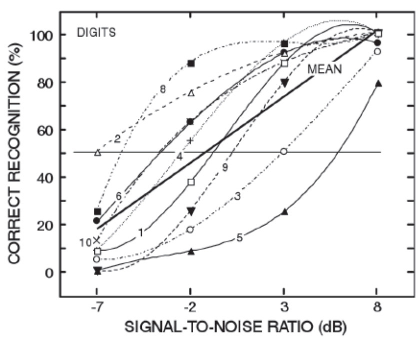

Fig. 3: The mean psychometric functions are shown for the nine monosyllabic digits and the mean function. The lines through the datum points are the best-fit, 3rd degree polynomials used to describe the data. A linear regression was used with the mean data.

Monosyllabic Digits

The mean 50% point for the 9 monosyllabic digits calculated with the Spearman-Kärber equation was −1.4-dB S/N with a 3.4-dB standard deviation. Among the 9 digits, the 50% points ranged 10.2-dB, from 4.9-dB S/N (five) to −5.3-dB S/N (two). The psychometric functions for the nine digits are shown in Figure 3 with the wider, dark line representing the mean data. The 50% point calculated from the mean digit function in Figure 3 was −1.2-dB S/N with a slope of 5.6%/dB (Table 2). The mean for the digits is 2-dB lower than the mean for the three monosyllabic word sets, but the mean slopes of the two materials are essentially the same. The heterogeneous characteristic of the family of functions depicted in Figure 3 is obvious and is typical for monosyllabic words (Wilson, 2003). As with the data complied with the Spearman-Kärber equation, the data in Figure 3 exhibit a 12.9-dB range of scores at the 50% point from −7.0-dB S/N (two) to 5.9-dB S/N (five). Interestingly, all of the functions for the digits, except two, ranged from 0 to 20% percent correct at −7-dB S/N to 80 to 100% percent correct at 8-dB S/N with the only differences being the morphologies of the functions. Again, when the mean 50% point and the mean slope are calculated from the 9 individual functions, the results are slightly different with −1.5-dB S/N (SD = 4.1 dB) 50% point and a mean slope of 10.0%/dB (SD = 2.9 dB).

The initial observation of the comparison of the monosyllabic word and monosyllabic digit mean data suggests that the digits are ~2-dB easier. Differences of this magnitude and slightly larger magnitudes between the word and digit functions are noted in the literature but usually in paradigms in which the digits are presented as unique lists that in many ways approximates a closed-set response. In the current experiment, the monosyllabic digits were embedded randomly in the entire set of 583 words. To evaluate if, in fact, the digits were easier than the other monosyllabic words, three, nine-word samples were selected randomly from the list of 583 words (excluding the spondaic words and the monosyllabic digits). The mean 50% point for the digits was −1.4-dB S/N with a 3.4-dB standard deviation. The three random samples, which represented a variety of performances with mean 50% points at −0.7, 1.8, and −0.4-dB S/N, required more favorable SNRs for 50% correct recognition than did the digits.

	PB-50		W-22		NU No. 6		Spondaic Words		Digits	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
−7 dB S/N	6.7	13.1	6.4	11.8	8.1	14.7	33.1	18.1	14.4	15.9
−2 dB S/N	31.8	26.0	33.3	24.6	34.7	27.1	81.9	17.1	47.7	27.3
3 dB S/N	65.5	26.4	68.4	24.1	66.7	26.2	97.1	7.1	79.2	24.5
8 dB S/N	87.5	16.4	91.3	11.7	88.8	15.1	99.0	3.6	96.7	7.0
Mean	47.9		49.9		49.6		77.8		59.5	
50% point dB S/N	0.9		0.5		0.5		−5.5		−1.2	
Slope at 50% (%/dB)	5.3			5.8				5.5		

Table 2: *The mean percent correct in decibels signal-to-noise ratio (and standard deviations in dB) for the word materials are listed for each of the four SNRs. The slopes (%/dB) at the 50% point for each material also are listed.*

These results are by no means conclusive but the indication is that the monosyllabic digits are slightly easier to recognize than are other monosyllabic words. Finally with regard to the monosyllabic digits, the functions in Figure 3 indicate a 10.2 dB range of the 50% points from −5.3 dB S/N (two) to 4.9-dB S/N (five). Similar variably was observed with the three random selections of nine monosyllabic words that had ranges of 9.6 dB, 8.3 dB, and 16.7 dB respectively. Thus, the variability observed in Figure 3 is viewed as representative of the variability inherent in measures of word recognition with monosyllabic words.

CONCLUSIONS

The following conclusions can be drawn from the data in the present study when (1) young listeners with normal hearing are involved, (2) the materials are spoken on one occasion by a speaker, (3) the materials are presented in speech-spectrum noise, and (4) the design involves the materials presented randomly in an intermingled manner:

1. Word-recognition performance on the four lists of words that comprise each of the three monosyllabic word sets (PB-50, W-22, and NU No. 6) are equivalent, ±0.4 dB.
2. Mean word-recognition performances on the four PB-50, W-22, and NU No. 6 word lists are equivalent, ±0.2 dB.
3. Mean recognition performance at the 50% point with the 36 W-1 spondaic words is ~6-dB lower than the 50% point for the monosyllabic words.
4. Mean recognition performance on the monosyllabic digits appears to be 1-2 dB better than mean performance on the monosyllabic words from the three sets of word materials.
5. Phonetic/phonemic balance does not appear to be an important consideration when compiling word-recognition lists used to evaluate the ability of listeners to understand speech.

ACKNOWLEDGMENT

The Rehabilitation Research and Development Service, Department of Veterans Affairs supported this work through a Merit Review, the Auditory and Vestibular Dysfunction Research Enhancement Award Program (REAP), and a Senior Research Career Scientist award to the first author and a Research Career Development award to the second author. The authors acknowledge the contributions made by Christopher A. Burks during the recording process and Heidi Roberts during data collection. A more complete version of this report will appear in a forthcoming issue of the Journal of the American Academy of Audiology.

REFERENCES

American National Standards Institute. (**1996**). “Specification for audiometers (ANSI S3.6-1996),” New York, Author.

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

Egan, J. (**1948**). “Articulation testing methods,” The Laryngoscope, **58**, 955-991.

Finney, D. J. (**1952**). “Statistical method in biological assay,” London, C. Griffen.

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

Heckendorf, A. L., Wiley, T. L., and Wilson, R. H. (**1997**). "Performance Norms for the

VA Compact Disc Versions of CID W-22 (Hirsh) and PB-50 (Rush Hughes) Word Lists," J. Am. Acad. Audio. **8**, 163-172.

Kreul, E. J., Bell, D. W., and Nixon, J. C. (**1969**). “Factors affecting speech discrimination test difficulty,” J. Speech Hear. Res. **12**, 281-287.

Lehiste, I., and Peterson, G. E. (**1959**). “Linguistic considerations in the study of speech intelligibility,” J. Acoust. Soc. Am. **31**, 280-286.

Lovrinic, J., Burgi, E., and Curry, E. (**1968**). “A comparative evaluation of five speech discrimination measures,” J. Speech Hear. Res. **11**, 372-381.

McArdle, R., and Wilson, R. H. (**2007**). “Interpreting Word-Recognition Data from the PB-50, W-22, and NU No. 6 Lists using the Neighborhood Activation Model and Phonemic Features of the Materials,” J, Am, Acad, Audio (under review).

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). “The intelligibility of speech as a function of the context of the test materials,” J. Exp. Psychol. **41**, 329-335.

Peterson, G. E. and Lehiste, I. (**1962**). “Revised CNC lists for Auditory tests,” J. Speech Hear Dis. **27**, 62-70.

Silverman S, and Hirsh, I. (**1955**). “Problems related to the use of speech in clinical audiometry,” Ann. Otol. Rhinol. Laryngo. **64**, 1234-1244.

Tillman, T. W., and Carhart, R. (**1966**). “An expanded test for speech discrimination utilizing CNC monosyllabic words,” Northwestern University Auditory Test No. 6. Brooks Air Force Base, TX USAF School of Aerospace Medicine Technical Report.

Wilson, R. H., Coley, K. E., Haenel, J., and Browning, K. (**1976**). “Northwestern University Auditory Test No. 6: Normative and Comparative Intelligibility Functions,” J. Am. Audio. Soc. **1**, 221-228.

Wilson, R. H. and Margolis, R. H. (**1983**). “Measurement of the auditory thresholds for speech stimuli”. in Principles of speech audiometry. Baltimore: University Park Press, 79-126.

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

Wilson, R. H., Zizz, C. A., Shanks, J. E., and Causey, G. D. (**1990**). "Normative data in quiet, broadband noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker," J. Speech Hear. Dis. **55**, 771-778.