J Am Acad Audiol 19:496–506 (2008)

A Comparison of Recognition Performances in Speech-Spectrum Noise by Listeners with Normal Hearing on PB-50, CID W-22, NU–6, W-1 Spondaic Words, and Monosyllabic Digits Spoken by the Same Speaker

DOI: 10.3766/jaaa.19.6.5

Richard H. Wilson*
Rachel McArdle†‡
Heidi Roberts‡

Abstract

Background: So that portions of the classic Miller, Heise, and Lichten (1951) study could be replicated, new recorded versions of the words and digits were made because none of the three common monosyllabic word lists (PAL PB-50, CID W-22, and NU–6) contained the 9 monosyllabic digits (1–10, excluding 7) that were used by Miller et al. It is well established that different psychometric characteristics have been observed for different lists and even for the same materials spoken by different speakers. 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 the CID W-1 spondaic words. A professional female speaker with a General American dialect recorded the materials during four recording sessions within a 2-week interval. The recording order of the 582 words was random.

Purpose: To determine—on listeners with normal hearing—the psychometric properties of the five speech materials presented in speech-spectrum noise.

Research Design: A quasi-experimental, repeated-measures design was used.

Study Sample: Twenty-four young adult listeners ($M = 23$ years) with normal pure-tone thresholds ($\leq$20-dB HL at 250 to 8000 Hz) participated. The participants were university students who were unfamiliar with the test materials.

Data Collection and Analysis: The 582 words were presented at four signal-to-noise ratios (SNRs; −7-, −2-, 3-, and 8-dB) in speech-spectrum noise fixed at 72-dB SPL. Although the main metric of interest was the 50% point on the function for each word established with the Spearman-Kärber equation (Finney, 1952), the percentage correct on each word at each SNR was evaluated. The psychometric characteristics of the PB-50, CID W-22, and NU–6 monosyllabic word lists were compared with one another, with the CID W-1 spondaic words, and with the 9 monosyllabic digits.

Results: 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–6 word lists were equivalent, ±0.2 dB. The mean recognition performance at the 50% point with the 36 W-1 spondaic words was ~6.2 dB lower than the 50% point with the monosyllabic words. Recognition performance on the monosyllabic digits was 1–2 dB better than mean performance on the monosyllabic words.

*VA Medical Center, Mountain Home, TN, and Departments of Surgery and Communicative Disorders, East Tennessee State University, Johnson City, TN; †VA Medical Center, Bay Pines, FL; ‡Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL

Richard H. Wilson, Ph.D., VA Medical Center, Audiology (126), Mountain Home, TN 37684; Phone: 423-979-3561; Fax: 423-979-3403; E-mail: RICHARD.WILSON2@VA.GOV

Portions of this paper were presented at the American Academy of Audiology Convention, March 2006, Minneapolis, MN, and at the International Symposium on Auditory and Audiological Research (ISAAR), August 2007, Helsingør, Denmark. An abbreviated version of the manuscript was included in the ISAAR proceedings.

Conclusions: Word-recognition performances on the three sets of materials (PB-50, CID W-22, and NU–6) were equivalent, as were the performances on the four lists that make up each of the three materials. Phonetic/phonemic balance does not appear to be an important consideration in the compilation of word-recognition lists used to evaluate the ability of listeners to understand speech.

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, this issue).

Key Words: Auditory perception, hearing loss, speech perception, word recognition in multitalker babble

Abbreviations: NU–6 = Northwestern University Auditory Test No. 6; rms = root-mean-square; SNR = signal-to-noise ratio

Sumario

Antecedentes: Para poder replicar el estudio clásico de Miller, Heise y Lichten (1951), se hicieron nuevas versiones grabadas de las palabras y los dígitos, dado que ninguna de las tres listas comunes de monosilábicos (PAL PB-50, CID W22, y NU-6) contenían los 9 dígitos monosilábicos (1–10, excluyendo el 7) que usaron Miller y col. Se ha establecido que existen características psicométricas diferentes para las diferentes listas y aún para los mismos materiales expresados por diferentes hablantes. Se tomó la decisión de grabar cuatro listas de cada uno de los tres grupos de palabras monosilábicas, sin incluir los dígitos monosilábicos en los tres grupos de listas de palabras, y las palabras espondaicas CID W-1. Una hablante profesional femenina con un dialecto americano general gravó los materiales durante cuatro sesiones de grabación en un intervalo de 2 semanas. El orden de grabación de las 582 palabras fue al azar.

Propósito: Determinar – en sujetos con audición normal – las propiedades psicométricas de los cinco materiales de lenguaje presentados en ruido centrado dentro del espectro del lenguaje.

Diseño de la Investigación: Se utilizó un diseño cuasi-experimental de mediciones repetidas.

Muestra del Estudio: Participaron veinticuatro adultos jóvenes ($M = 23$ años) con umbrales tonales puros normales ($\leq$ 20 dB HL entre 250 y 8000 Hz). Los participantes eran estudiantes universitarios sin familiaridad con el material de prueba.

Recolección y Análisis de los Datos: Las 582 palabras fueron presentadas a cuatro tasas señal-ruido (SNR de -7, -2, 3 y 8 dB) en ruido centrado dentro del espectro del lenguaje, fijo a 72 dB SPL. Aunque la medida principal de interés era el punto 50 % en la función de cada palabra, establecido por la ecuación Spearman-Kärber (Finney, 1952), se evaluó el porcentaje correcto para cada palabra en cada SNR. Las características psicométricas de las listas de palabras monosilábicas en PB-50, CID W'22 y NU-6 fueron comparadas entre sí, y con las palabras espondaicas CID W-1 y con los 9 dígitos monosilábicos.

Resultados: El desempeño en el reconocimiento de las cuatro listas dentro de cada uno de los 3 materiales de palabras monosilábicas fue equivalente, $\pm$ 0.4 d B. Asimismo, el desempeño en el reconocimiento de palabras en las listas de palabras PB-50, W-22 y NU-6 fue equivalente; $\pm$ 0.2 dB. El desempeño medio de reconocimiento en el punto 50% con las 36 palabras espondaicas W-1 fue $\sim$ 6.2 dB, más bajo que el punto 50% con las palabras monosilábicas. El desempeño en el reconocimiento con los dígitos monosilábicos fue 1–2 dB mejor que el desempeño medio con las palabras monosilábicas.

Conclusiones: El desempeño en el reconocimiento de palabras en los tres grupos de materiales (PB-50, CID W-22 y NU-6) fue equivalente, así como el desempeño en las cuatro listas que se conformaron de los tres materiales. El balance fonético/fonémico no parece ser una consideración importante en la compilación de las listas de reconocimiento de palabras usadas para evaluar la habilidad de los oyentes para entender el lenguaje.

Un trabajo acompañante examina las variables acústicas, fonético-fonológicas y léxicas que puede predecir la facilidad o dificultad relativa para que estas palabras monosilábicas sean reconocibles en medio de ruido (McArdle y Wilson, en este ejemplar).

Palabras Clave: Percepción auditiva, pérdida auditiva, percepción del lenguaje, reconocimiento de palabras en balbuceo de hablantes múltiples

Abreviaturas: NU-6 = Prueba auditiva No. 6 de la Universidad Northwestern; rms = raíz media cuadrada; SNR = tasa señal-ruido

Preparation for a project involving recognition performance on the monosyllabic digits and other monosyllabic words made it necessary to record a new version of the digits and words because none of the three common monosyllabic word lists (PAL PB-50 [Egan, 1948]; CID W-22 [Hirsh et al, 1952]; Northwestern University Auditory Test No. 6 [NU–6; Tillman and Carhart, 1966]) contained the 9 monosyllabic digits (1–10, excluding the bisyllabic *7*). 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; Rintelmann et al, 1974; Wilson et al, 1990). Silverman and Hirsh (1955) and Kreul et al (1969) were among the first to recognize that the psychometric characteristics of word lists were valid only for the word lists as spoken on one occasion by a speaker.

The question that prompted the current experiment was about which of the three common monosyllabic word sets (PB-50, W-22, and NU–6) would best serve in the planned project. Because there was no basis on which the decision could be made regarding which of the three monosyllabic word sets to use and because the three monosyllabic word sets contained a number of words in common, 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 spondaic 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 (SNRs) in speech-spectrum noise. Speech-spectrum noise was selected because it provided a waveform that was consistent from one sample to the next with minimal amplitude modulations. The primary metric used to evaluate the data was the 50% point on the psychometric function for each word. The 50% point, which was calculated with the Spearman-Kärber equation (Finney, 1952; Wilson et al, 1973), provided one performance descriptor about each word that was based on data from four SNRs. Additionally, the percentage of correct performances on each word at each SNR was evaluated. Finally, a companion paper examines the acoustic, phonetic, and lexical variables that may predict the relative ease or difficulty for which these monosyllable words were recognized in noise (McArdle and Wilson, this issue).

METHODS

Materials

Four of the PB-50 lists (8, 9, 10, and 11), the four CID W-22 lists, the four NU–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 sets of materials $\times$ 4 lists $\times$ 50 words), 485 words appear only once in the 12 lists, 56 words appear in two sets, and 1 word (*have*) appears in all three sets. The PB-50s and W-22s share 30 words, the PB-50s and NU–6 share 14 words, and the W-22s and NU–6 share 12 words. Interestingly, only five of the nine monosyllabic digits appear in the 12 word lists. Digits *one*, *two*, and *three* are in the W-22 lists; digit *five* is in the NU–6 list; and digit *10* appears both in the W-22 and the PB-50 lists.

The materials were recorded by a professional female speaker during four recording sessions within a 2-week interval. The recording order of the 582 words was random. The female speaker spoke with a General American dialect and was seated in a 3 $\times$ 3 m double-wall sound booth (IAC, Model 1203) 15 cm from the microphone (Sanken, Model CU44X), which was protected by a pop screen (Stedman, Model PS-101). From the microphone, the signal was amplified (GML, Model 8304), digitized with a 24-bit analog-to-digital converter (Lucid, Model ADA 8824), and recorded at 44,100 samples per second on a computer (Macintosh, Model G4) through a recording interface (MOTU, Model 2408 MKIII).

During the recording sessions, a speech coach and two audiologists monitored and judged the spoken materials through a loudspeaker (Genelec, Model 1031A). Three trials of each word were recorded. If the speech coach and audiologists determined that the three samples were insufficient representations of the word, then the speaker repeated the word. The carrier phrase "you will cite" was used. This particular carrier phrase was used because it provided a hard stop at the end of the phrase to limit coarticulation between the carrier phrase and the target word. The level of the carrier phrase was monitored by the speaker to zero on a vu meter (Coleman, Model MB2) with the level of the words falling naturally (American National Standards Institute [ANSI], 2004).

Following the recording sessions, the carrier phrases were removed and the materials were edited to select which of the three productions of each word was to be used in the final version. One carrier phrase was selected and concatenated to each of the 582 words using a 200 msec silent interval between the end of the carrier phrase and the onset of the word. Based on pilot data that involved 200 of the monosyllabic words, four SNRs were chosen for use in the experiment (-7, -2, 3, and 8 dB). The intention was to generate psychometric functions from 10–20% correct (slightly above floor effects) to 80–90% correct (slightly below ceiling effects), with an overall target recognition performance for the four SNRs of 50% correct. The level of the carrier phrase–word segment was attenuated or amplified digitally to achieve the appropriate level for each SNR. There were 2,328 stimulus files (582 words × 4 SNRs).

Four randomizations of the 582 words were made for each of the four SNRs incorporating a 3-s interstimulus interval. The randomizations for each SNR were made by compiling the words into eleven 50-word lists and one 32-word list. After the lists were generated, the words were mixed digitally with a sample of speech-spectrum noise to achieve the appropriate SNR and were recorded on one channel, with the speech samples in quiet recorded on the second channel for use in monitoring the progression of the test.

The effective rms sound pressure levels and durations of the words were measured. The mean rms for the spondaic words was −18.7 dB (re: maximum digitization range), with the rms for the monosyllabic words and monosyllabic digits about 3 dB lower (−21.7 and −21.9 dB, respectively). The standard deviations were 0.8, 1.3, and 1.2 dB, respectively, for the three materials. The rms for the 950-msec carrier phrase was −18.6 dB. The mean duration of the spondaic words (905 msec) was about twice the duration of the monosyllabic words and digits (both 554 msec). The standard deviations for the materials ranged from 86 to 90 msec. From an acoustic perspective, then, the monosyllabic materials were fairly homogeneous.

Subjects

Twenty-four young adult listeners (18–30 years old; M = 23 years) participated. The listeners had pure-tone thresholds ≤20-dB hearing level (HL) at octave and interoctave frequencies from 250 through 8000 Hz (ANSI, 2004). All participants were university students who were unfamiliar with the test paradigm and test materials.

Procedures

Following an initial session in which the informed consent and pure-tone threshold data were obtained, the listeners subsequently participated in four 60-minute sessions. Using a counterbalanced design, each of the four randomizations of the 582 words was used an equal number of times at each SNR. For presentation, then, the 582 words were intermingled randomly. During each of the four sessions, 582 words were presented. So that session effects would be avoided, ~25% of the 582 words used in each session were presented randomly at each of the four SNRs (−7 dB, −2 dB, 3 dB, and 8 dB). 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. By the end of the four visits, each of the 582 words had been presented once to each listener at each of the four SNRs.

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 nontest ear covered by an insert phone. The right ears of the even-numbered listeners were tested, and the left ears of the odd-numbered listeners were tested. All testing was conducted in a double-wall sound booth (IAC, Model 1203). The listeners responded verbally to the stimuli, with the responses logged into a spreadsheet for analysis.

RESULTS AND DISCUSSION

The overall results indicated that the mean 50% point for the 582 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, the 50% points of four of the monosyllabic words were at less favorable SNRs (i.e., easier) than the 50% point for the easiest spondaic word (*toothbrush*). In terms of mean percentage correct for the four SNRs, the mean correct recognition for the 582 words was 50.7% (SD = 17.9%), ranging from 93.8% (*choice*) to 4.2% (*bob*). These results indicate that on average, the SNRs selected with the pilot data provided the desired range of results around the 50% point. The data from the 582 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.

Monosyllabic Words

Graphic analysis of the PB-50, W-22, and NU–6 data is presented in Figures 1 and 2 with corresponding numeric descriptive data listed in Tables 1, 2, and 3. Figure 1 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 for each of the three materials are plotted in the fourth quadrant. The numeric mean percentage correct recognition performances (and standard deviations) at each SNR for the four lists of each type of material are listed in Table 1. In Figure 1, the lines connecting the datum points are linear regressions from which the 50% points on the functions and the slopes of the functions were calculated (see right columns, Table 2).

In Table 3, 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 the overall mean percentage of correct values of 47.9%, 49.9%, and 49.6% for the

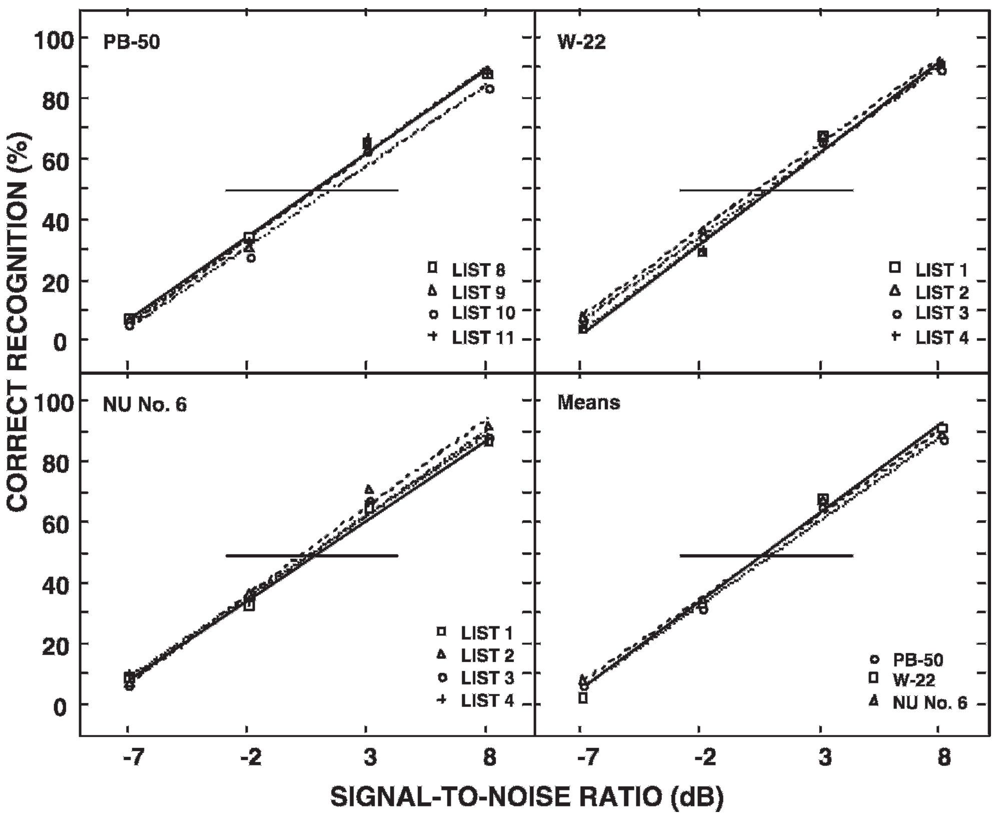

Figure 1. The mean percentage correct recognition at four signal-to-noise ratios (SNRs) for the four lists of each of the three monosyllabic-words-list materials. The mean data for each of the lists are illustrated in the fourth quadrant. The lines with each set of data are the linear regressions used to describe the data. The 50% point for each function and the slope of the function at the 50% point are listed in Table 2. The horizontal line in each panel indicates the 50% point on each function.

PB-50, W-22, and NU–6 lists, respectively. It is obvious from the data in Figure 1 and from the data in the three tables that there are minimal differences both among the four lists of each set of materials and among the three word sets (fourth quadrant).

As Table 2 shows, the variability among the four lists of each material at the 50% points essentially was the same with minimum and maximum differences between lists of 0.6 dB for NU–6 (List 1 and List 2 difference in Table 2), 0.7 dB for PB-50s (Lists 8 and 11, List 10 differences), and 0.7 dB for W-22s (List 1 and List 2 difference). The mean 50% points for the three materials (Table 2, column 6) ranged 0.4 dB with a 0.5-dB S/N mean for the W-22 and NU–6 lists and with a 0.9-dB S/N for the PB-50 lists. The slopes of the mean functions for the 12 lists (Table 2, column 7) varied 0.7%/dB, from 5.3%/dB to 6.0%/dB.

Finally in Table 2, descriptive statistics are provided for the 50% points calculated with the Spearman-Kärber equation from the data on the individual words that compose each list. As one would expect, because the Spearman-Kärber equation is based on a linear model and because the functions are linear, the 50% points based on the individual data calculated from the Spearman-Kärber equation (column 2) are essentially the same as the 50% points calculated from the linear regressions used to describe the mean data (column 6).

The data in columns 2–5 of Table 2 provide an indication of the variability at the 50% point that is associated with each list of words. The data in the "mean" rows were compiled from the 200 words in each set of materials. Overall, the variability of the 50% points is about the same for the three sets of materials in terms of the standard deviations, maximum values, and minimum values. If there were any difference, it would be between the standard deviations for the W-22 lists (3.0 dB) and for the PB-50 and NU–6 lists (3.5 dB). Accordingly, the W-22 lists also had a smaller range (maximum–minimum) of 50% points (13.8 dB) than did either the PB-50 lists (15.8 dB) or the NU–6 lists (15.6 dB).

The data in Figure 2 are bivariate plots of the recognition performances of the 24 listeners on the three combinations of the three monosyllabic word materials, which are depicted on the ordinate and abscissa. The lists shown were selected because they were most representative of the mean data for the respective materials. The different symbols depict the performances at the various SNRs with the larger, filled symbols representing the mean data at each SNR. The

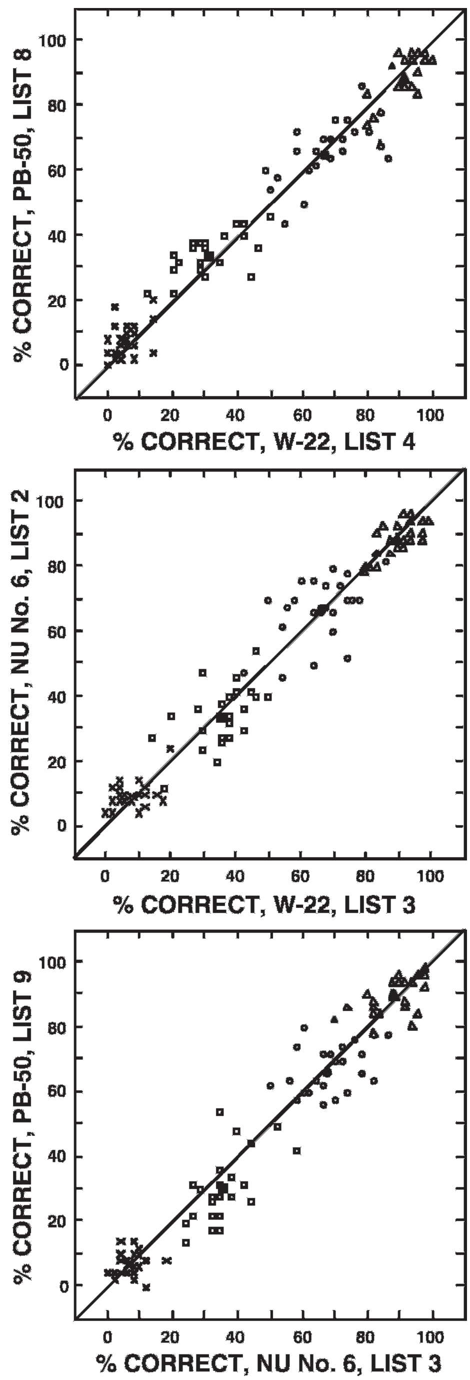

Figure 2. Bivariate plots of the percentage of correct recognition by each of the 24 listeners at −7-, −2-, 3-, and 8-dB S/N (Xs, squares, circles, and triangles, respectively). Comparisons among the three lists are made with selection of the list(s) most representative of the mean data for each type of material used.

diagonal line represents equal-recognition performance on the two lists. In all three comparisons in Figure 2, the distributions of the datum points around the diagonal lines indicate equal performance at the level of the individual listener on the two materials.

Collectively, the mean data in Figure 1 and Tables 1, 2, and 3 and the data from the individual listeners shown in Figure 2 indicate that there are no practical differences among the recognition performances obtained on the three types of materials (PB-50, W-22, and NU–6), at least for listeners with normal hearing. However, substantial variability existed among the recognition performances obtained on the 582 individual words and was randomly distributed among the various sets of word materials. These findings are examined in a companion paper (McArdle and Wilson, this issue) in terms of the acoustic, phonetic, and lexical variables that may be used to predict recognition performance in noise.

To the best of our knowledge, this study is the first direct comparison of the recognition performances achieved on the PB-50, W-22, and NU–6 materials spoken and recorded by the same speaker. Here, it is important to add that the materials were spoken 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–6 materials spoken by the same speaker (Auditec of St. Louis) but recorded during different sessions. Beattie et al found that the NU–6 materials were slightly more difficult than the W-22 materials, as evidenced by the 1–2 dB displacement of the performances on the NU–6 materials to the higher, more favorable SNRs. The NU–6 materials also had a more gradual slope, 4.2%/dB versus 4.6%/dB for the W-22 materials.

In contrast to the Beattie et al (1977) findings, Wilson and Oyler (1997) made the same comparison between the Auditec versions of the W-22 and NU–6 materials using a procedure in which the materials were presented in an interleaved paradigm to avoid possible "session" effects. Wilson and Oyler found that for listeners with normal hearing and listeners with hearing loss their recognition performances on the NU–6 materials were 4–8% better than the performances on the W-22 materials. For both the Beattie et al and the Wilson and Oyler 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, which were recorded using a random interleaved design and were administered with a similar interleaved design, revealed no appreciable differences among performances on the PB-50, W-22, and NU–6 materials that were spoken by the same speaker.

Although the characteristic of phonetic/phonemic balance in word lists was not a focus of the current study, 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

Table 1. Mean Percentage Correct Recognition (and standard deviations) at the Four Signal-to-Noise Ratios (SNRs)

	−7 dB S/N		−2 dB S/N		3 dB S/N		8 dB S/N	
List	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
PB-50								
8	7.3	10.4	34.3	29.5	65.4	27.6	88.4	12.8
9	6.9	16.6	30.9	26.1	66.4	24.0	89.6	13.5
10	6.0	13.1	28.3	24.4	63.1	26.7	83.6	21.4
11	6.4	11.9	33.8	23.8	67.1	27.7	88.5	16.4
W-22								
1	4.2	6.1	29.8	21.6	68.5	21.6	91.6	11.7
2	8.5	15.4	37.3	28.0	72.3	24.5	91.2	12.4
3	7.5	13.3	35.5	26.0	66.1	25.1	90.4	11.8
4	5.3	10.0	30.6	22.1	66.6	25.4	92.1	11.1
NU–6								
1	9.3	16.7	32.9	26.6	65.0	27.5	86.8	17.4
2	9.6	15.7	34.2	27.9	67.2	26.7	88.2	17.3
3	6.7	12.5	35.4	27.4	67.4	25.2	88.8	12.7
4	7.0	13.5	36.4	27.2	70.7	25.6	91.5	12.1
W-1 Spondaic Words								
	33.1	18.1	81.9	17.1	97.1	7.1	99.0	3.6
Digits (1–10, excluding 7)								
	13.9	16.0	49.1	28.4	78.2	24.5	96.3	7.0

Note: Data were obtained by 24 listeners with normal hearing on the PB-50, W-22, and NU–6 monosyllabic word lists; on the W-1 spondaic words; and for digits 1–10, excluding 7.

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. The data in the current study were amenable to such an evaluation.

Figure 3 shows the mean 50% points derived with the Spearman-Kärber equation for the 12 "organized" 50-word lists studied (PB-50, triangles; W-22, squares; NU–6, inverted triangles) and the twelve 50-word lists that were compiled randomly from the corpus of 540 monosyllabic words. The vertical lines are the standard deviations. The randomly selected lists demonstrate slightly more noise in the data in the form of interlist 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 from 1.5-dB S/N (PB-50, List 3) to 0.2-dB S/N (NU–6, List 4), whereas the 50% points for the lists of words selected randomly ranged 1.8 dB from 1.4-dB S/N (10th list) to −0.4-dB S/N (5th list). These sets of ranges are comparable and <2 dB, which with consideration to the slopes of the functions discussed earlier is an approximate 10% disparity (based on a slope of 5%/dB). The conclusion from this analysis is that the concept of phonetic/phonemic balance has minimal effect on the intelligibility characteristics of word-recognition materials.

Campanelli (1962) made a similar observation. Other characteristics, such as the speaker of the materials and presentation level, are much more influential on the word-recognition performance that is achieved. Perhaps, as Rosen and Corcoran (1982) indicated, "… the entire concept of phonemic balancing as it is currently practised is flawed as it deals only with the first order statistics of phoneme occurrence and no higher ones. Since coarticulation effects alter the acoustic and visual representation of nearly all sounds, the 'balancing' is only occurring at some rather abstract level" (p. 247).

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 mean function for the spondaic words is shown in Figure 4 (inverted triangles) with the numeric data (including standard deviations) listed in Table 1. The line through the datum points in the figure is the best-fit, second-degree polynomial that is used to describe the data. The 50% point calculated from the mean function in Figure 4 was −5.7-dB S/N with a slope at that point of 11.8%/dB, which is about twice as steep as the mean functions for the monosyllabic words and digits (Table 3). In contrast, the mean 50% point calculated from the 36 individual spondaic word functions 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 is 2.8%/dB steeper than the

Table 2. The Mean 50% Points, Standard Deviations (intersubject), Maximum, and Minimum Calculated with the Spearman-Kärber Equation from the Individual Word Data

List	*M* (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–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

Note: Data mentioned in the table title are listed along with the 50% points calculated with the linear regressions used to fit the mean data for each of the twelve 50-word monosyllabic lists. The slopes of the functions at the 50% points also are listed.

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).

Finally in Figure 4, the 50% point on the mean spondaic word function is 6.2 dB lower than the 50% point on the mean function for the PB-50 words (triangles) and 4.5 dB lower than the 50% point on the mean function for the digits (Xs). The 6.2-dB difference is close to the ~10-dB difference observed by Hirsh et al (1952) between the 50% points on the functions for 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 recordings (speaker) and 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 given as independent sets of materials. The Hirsh et al study with the spondaic words was closer, therefore, to a closed-set response paradigm than was the current study. Better performance is obtained in a closed-set paradigm than in an open-set paradigm (Wilson and Antablin, 1982).

Monosyllabic Digits

The mean 50% point for the nine monosyllabic digits calculated with the Spearman-Kärber equation was −1.4-dB S/N with a 3.4-dB standard deviation (Table 4, column 1). Among the nine digits, the 50% points ranged 10.2 dB, from 4.9-dB S/N (*five*) to −5.3-dB S/N (*two*). The mean function for the monosyllabic digits is shown in Figure 4 (Xs) with the numeric data (including standard deviations) listed in Table 1. The

Table 3. The Mean Percentage Correct (and standard deviations in dB) at Four Signal-to-Noise Ratios for the PB-50, W-22, and NU–6 Monosyllabic Word Materials; the Spondaic Words; and Monosyllabic-Digit Materials

	PB-50		W-22		NU–6		Spondaic Words		Digits	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*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		−5.7		−1.2	
Slope at 50% (%/dB)	5.3		5.8		5.5		11.8		5.6	

Note: Data mentioned in the table title are listed for each of the four signal-to-noise ratios (SNRs). The means of the four values also are listed. The mean 50% point (dB S/N) calculated from the mean functions in Figures 1 and 4 and the slopes (%/dB) of the mean functions also are listed.

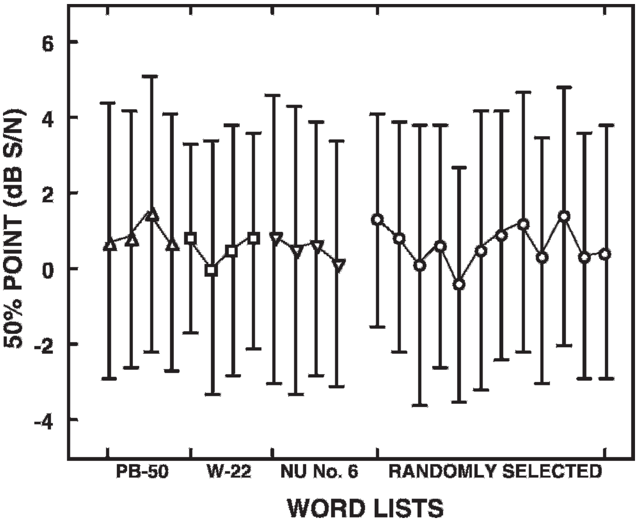

Figure 3. The mean 50% points for the individual words calculated with the Spearman-Kärber equation are shown for lists 1–4 of the PB-50 (triangles), of the CID W-22 (squares), and of the NU–6 (inverted triangles). The 12 randomly compiled lists (circles) are shown in the right half of the figure. The vertical bars indicate one standard deviation.

line through the datum points in Figure 4 is the linear regression used to describe the data. The 50% point calculated from the mean digit function in Figure 4 was −1.2-dB S/N with a slope of 5.6%/dB. As Figure 4 shows, the mean for the digits is 2-dB lower than the mean for the three monosyllabic word sets, but the slopes of the two materials are essentially the same.

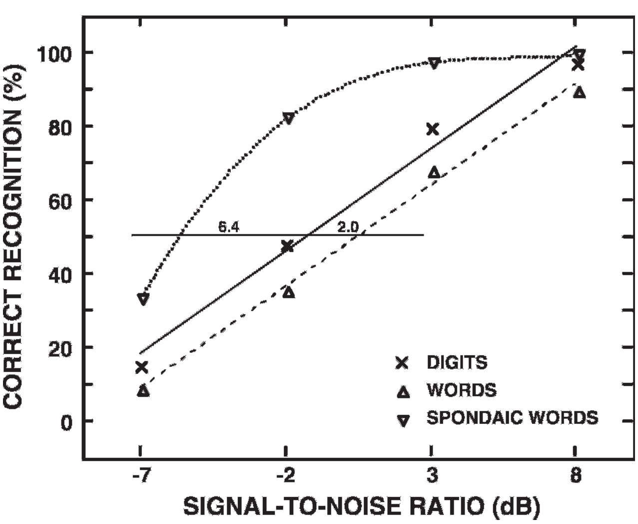

Figure 4. Mean psychometric functions for the monosyllabic digits (Xs), the W-1 spondaic words (inverted triangles), and PB-50 monosyllabic words (triangles) obtained from the 24 listeners. The lines connecting the datum points are the best-fit, linear regressions and second-degree polynomial used to describe the data. The numbers on the horizontal line are the decibel differences between adjacent functions. The 50% point for each function and the slope of the function at the 50% point are listed in Table 3. The horizontal line indicates the 50% points on the functions.

Table 4. The Mean 50% Points (in decibels signal-to-noise ratio [SNR]) Calculated with the Spearman-Kärber Equation and Other Measures of Central Tendency (in dB) for the Nine Monosyllabic Digits

	Monosyllabic Digits	9 Randomly Selected Monosyllabic Words		
		Random 1	Random 2	Random 3
Mean	−1.4	−0.7	1.8	−0.4
SD	3.4	4.0	4.3	1.7
Max	4.9	3.8	6.8	1.8
Min	−5.3	−8.0	−5.8	−3.0
Range	10.2	11.9	12.5	4.8

Note: For comparison, corresponding data are shown for three sets of nine monosyllabic words selected randomly from the corpus of monosyllabic words that comprised the PB-50, W-22, and NU–6 lists.

The psychometric functions for the nine digits are shown in Figure 5 with the wider, dark line representing the function of the mean data. Third-degree polynomials were used to describe the data for the individual digits with a linear regression used with the mean data. The heterogeneous characteristic of the family of functions depicted in Figure 5 is obvious and is typical for monosyllabic words (Wilson, 2003). As with the data compiled with the Spearman-Kärber equation, the data in Figure 5 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 0–20% correct at −7-dB S/N to 80–100% 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 nine individual functions, the results are slightly different, with a −1.5-dB S/N 50% point (*SD* = 4.1 dB) 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 in Figure 4 suggests a difference of ~2 dB in performance at the 50% point with the digits being easier (i.e., at less favorable SNRs). 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 approximate a closed-set response paradigm, the effects of which were discussed previously. In the current experiment, the monosyllabic digits were embedded randomly in the entire set of 582 words.

So one could evaluate if, in fact, the digits were easier than the other monosyllabic words, three 9-word samples were selected randomly from the list of 582 words (excluding the spondaic words and the monosyllabic digits). The descriptive statistics of the three groups of nine randomly selected monosyllabic words are listed in Table 4. The mean 50% point for the digits was −1.4-dB S/N with a 3.4-dB standard deviation.

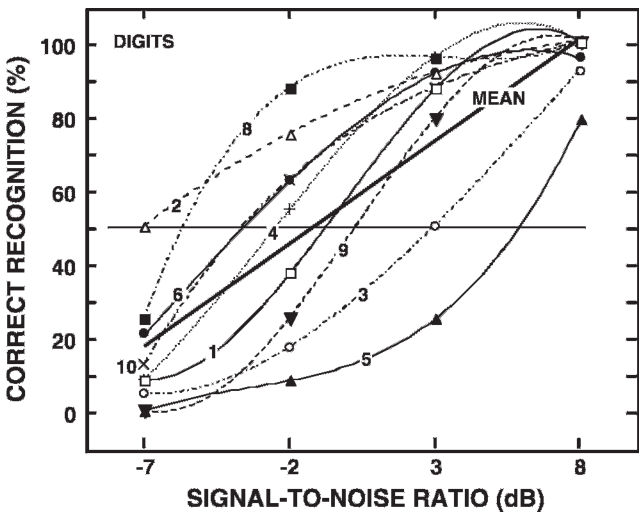

Figure 5. The mean psychometric functions are shown for the nine monosyllabic digits and the mean function. The lines through the datum points are the best-fit, third-degree polynomials used to describe the data. A linear regression was used with the mean data. The horizontal line indicates the 50% points on the functions.

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. These results are by no means conclusive, but along with the relations depicted in Figure 4, the indication is that the monosyllabic digits are slightly easier to recognize at comparable SNRs than are other randomly selected monosyllabic words.

Finally, with regard to the monosyllabic digits, the functions in Figure 5 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 variabilities were 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 5 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 (a) young listeners with normal hearing were involved, (b) the materials were spoken on one occasion by the current speaker, (c) the materials were presented in speech-spectrum noise, and (d) the design involved the materials being presented randomly in an intermingled manner:

1. Word-recognition performances on the four lists of words that make up each of the three monosyllabic word sets (PB-50, W-22, and NU–6) are equivalent, ±0.4 dB.
2. Mean word-recognition performances on the four PB-50, W-22, and NU–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 in the compilation of word-recognition lists used to evaluate the ability of listeners to understand speech.

Acknowledgments. 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.

REFERENCES

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

Beattie RC, Edgerton BJ, Svihovec DA. (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 Disord* 42: 60–64.

Campanelli PA. (1962) A measure of intra-list stability of four PAL word lists. *J Aud Res* 2:50–55.

Egan J. (1948) Articulation testing methods. *Laryngoscope* 58: 955–991.

Finney DJ. (1952) *Statistical Method in Biological Assay.* London: C. Griffen.

Heckendorf AL, Wiley TL, Wilson RH. (1997) Performance norms for the VA compact disc versions of CID W-22 (Hirsh) and PB-50 (Rush Hughes) word lists. *J Am Acad Audiol* 8:163–172.

Hirsh IJ, Davis H, Silverman SR, Reynolds EG, Eldert E, Benson RW. (1952) Development of materials for speech audiometry. *J Speech Hear Disord* 17:321–337.

Kreul EJ, Bell DW, Nixon JC. (1969) Factors affecting speech discrimination test difficulty. *J Speech Hear Res* 12:281–287.

Lehiste I, Peterson GE. (1959) Linguistic considerations in the study of speech intelligibility. *J Acoust Soc Am* 31:280–286.

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

McArdle R, Wilson RH. (This issue) Predicting word-recognition performance in noise by young listeners with normal hearing using acoustic, phonetic, and lexical variables. *J Am Acad Audiol* 19(6):507–518.

Peterson GE, Lehiste I. (1962) Revised CNC lists for auditory tests. *J Speech Hear Disord* 27:62–70.

Rintelmann WF, et al. (1974) Six experiments on speech discrimination utilizing CNC monosyllables. *J Aud Res Suppl* 2, 1–30.

Rosen S, Corcoran T. (1982) A video-recorded test of lipreading for British English. *Br J Audiol* 16:245–254.

Silverman S, Hirsh I. (1955) Problems related to the use of speech in clinical audiometry. *Ann Otol Rhinol Laryngol* 64:1234–1244.

Tillman TW, 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 RH. (2003) Development of a speech in multitalker babble paradigm to assess word-recognition performance. *J Am Acad Audiol* 14:453–470.

Wilson RH, Antablin JK. (1982) The Picture Identification Task—a reply to Dillon. *J Speech Hear Disord* 47:111–112.

Wilson RH, Coley KE, Haenel J, Browning K. (1976) Northwestern University Auditory Test No. 6: normative and comparative intelligibility functions. *J Am Audiol Soc* 1:221–228.

Wilson RH, Margolis RH. (1983) Measurement of the auditory thresholds for speech stimuli. In: Konkle DF, Rintelmann WF, eds. *Principles of Speech Audiometry.* Baltimore: University Park Press, 79–126.

Wilson RH, Morgan DE, Dirks DD. (1973) A proposed SRT procedure and its statistical precedent. *J Speech Hear Disord* 38: 184–191.

Wilson RH, 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 RH, Zizz CA, Shanks JE, Causey GD. (1990) Normative data in quiet, broadband noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker. *J Speech Hear Disord* 55:771–778.