Evaluation of hearing aids from the standpoint of the user

Henry W Boger*

Consultant, Senior Member of the International Society of Automation, Birmingham, Hoover, USA

Background

Hearing aids are widely used, and perhaps misused, daily by a large group of people with perceived hearing difficulty. The main reason for the use of these hearing aids is exposure to loud noise. In some cases, the exposure was caused by the nature of employment, for example gunfire on the part of servicemen.

Hearing aids are dispensed (sold) by groups offering audiology tests to determine the degree of impairment. Because income is based on the sale of hearing aids, the outcome of testing and the ensuing recommendations are skewed toward the purchase of high ticket hearing aids.

The most glaring deficiency in this process is the lack of testing to determine speech recognition.

The work of Dr. Wilson and Mr. Burks

For many years, investigators have discussed the difficulty that individuals with hearing loss have understanding speech, especially in background noise [1]. Carhart and Tillman [2], based on their data and data from Groen [3], stated that ". . . by the time background talk reaches a level where it is just mildly disruptive to intelligibility for normal hearers it can become a serious masker for the sensorineural . . . " (p. 279). Carhart and Tillman suggested that communication handicap should be quantified not only by measures of pure tone sensitivity and word recognition in quiet but also by word recognition in a background of competing speech. Much of this reasoning was based on two types of hearing loss that Carhart described in 1951: "a loss of acuity" and "a deficiency in the clarity with which speech is received". Loss of acuity was displacement of the "articulation function" for the listener with hearing loss to the right of that for a listener with normal hearing. When a "deficiency in clarity" was involved, Carhart suggested that, regardless of the presentation level, the clarity of the speech signal could not be substantially changed.

Twenty-five years later, Stephens made a similar observation but called the two hearing loss components "attenuation" and "distortion". Subsequently, Plomp formalized Carhart's and Stephen's observations that hearing loss had audibility (acuity) and distortion (clarity) components . Most individuals with audibility-related hearing loss show improved (even to 100% correct) word recognition in quiet as the level of the signal increases. Individuals with distortion-related hearing loss, however, can have reduced word-recognition performance in quiet, regardless of presentation level. Degraded speech-recognition tasks, like listening in background noise, highlight the detrimental effect of this distortion component on everyday speech listening. The introduction of background noise into the word-recognition paradigm more often than not results in substantially poorer recognition performance than when speech is presented in quiet. The relation between speech understanding in quiet and in noise led Plomp and

Duquesnoy to state that "a hearing loss for speech in noise of 3 dB is more disturbing than a hearing loss for speech in quiet of 21 dB" [17, p. 101], and Killion to note that if you want to know how well an individual understands speech in background noise, then you must measure it [18].

The audibility component of hearing loss can be overcome simply with increased signal level. The distortion component, however, is more difficult to overcome. Thus, the hearing aid industry has focused primarily on this component of hearing loss. Simply stated, for listeners to overcome the distortion component of hearing loss, devices such as traditional hearing aids or technology like FM (frequency modulation) systems must improve the signal-to-noise ratio at their ear.

In a series of experiments, we developed a Words in Noise (WIN) test for measuring hearing loss in signal-to-babble (S/B) ratio [7, 19-21]. The WIN test evaluates speech understanding in a background of multitalker babble at several S/B ratios. The test has the following characteristics [19]:

1. Seventy monosyllabic words from the Northwestern University Auditory Test No. 6 (NU-6) [22] spoken by a female on a Department of Veterans Affairs (VA) compact disc [23].

2. Ten unique words at each of seven S/B ratios from 24 to 0 dB S/B in 4 dB decrements.

3. Words time-locked to a unique segment of babble for reduced variability.

4. Continuous, fixed-level babble and varied word level.

5. A 2.7 s interval between words.

6. A 50 percent correct word-recognition point as determined by the Spearman-Kärber equation [24].

7. A 6 to 12 dB or more separation in S/B ratio between listeners with normal hearing and listeners with hearing loss.

Initially, the 70 words were presented at random S/B ratios, which was appropriate for experimental purposes. Subsequently, a test that was more appropriate for clinical purposes was developed in which the listener was exposed to easy listening conditions (24 dB S/B) that progressed to more difficult listening conditions (0 dB S/B). A stopping rule terminated the test sequence when 10 words at one level were incorrectly identified [19,25]. The approximately 5-min test required 10 min when both ears were evaluated. The immediate feedback from clinicians was that the addition of 10 min to an audiological evaluation

was difficult to justify. Although we prefer the 70-word test for evaluating each ear, our conclusion was that to ensure that clinicians implemented speech-in-noise testing on a routine, widespread basis, the WIN test would have to be made more attractive time-wise, especially in light of the other tests and activities involved in an audiological evaluation. So, as with most tests that are transferred from the laboratory to the clinic, administration time is a major consideration. This issue prompted our efforts to halve the test to 2.5 min by establishing two equivalent 35-word lists with 5 words presented at each of 7 S/B ratios in 4 dB decrements from 24 to 0 dB. The use of 35 versus 70 words involves the same issues as when earlier word tests were converted from 50 to 25 words [26]. List equivalency is an elusive concept, particularly because the auditory characteristics of listeners with and without hearing loss are heterogeneous. Equivalent word-recognition lists must produce the same results from a given ear under the same listening conditions, preferably with repeated measures.

Most efforts for establishing an instrument that measured speech in background noise involved speech in sentences [27-33]. However, the use of sentences has not been widely accepted because clinical audiologists prefer monosyllabic words [34]. The WIN test was developed with the hope that it would be accepted into routine clinical use. The evaluation of word-recognition performance in quiet and in background noise with the same materials spoken by the same speaker is attractive, especially when performances in the two listening conditions are compared.

A recent study demonstrated that words and sentences presented in background multitalker babble resulted in equivalent recognition performances [35]. McArdle et al. compared the recognition performances of 36 listeners with normal hearing and 72 listeners with sensorineural hearing loss on the WIN test and the Quick Speech-In-Noise Test (QuickSIN™), a sentence-recognition task [36]. Listeners were presented with two lists from the WIN and two lists from the QuickSIN™. The listeners with normal hearing had essentially the same mean 50 percent correct recognition points on both lists of each test (WIN = 4.4 and 5.0 dB S/B ratio, QuickSIN™ = 3.9 and 4.3 dB S/B ratio). The listeners with hearing loss had mean 50 percent correct recognition points on both protocols that were 7 to 8 dB higher than those of the listeners with normal hearing (WIN = 12.3 and 12.4 dB S/B, QuickSIN™ = 10.1 and 13.3 dB S/B). The difference between the WIN and QuickSIN™ performances was not significant within either listener group. However, the approximately 8 dB difference between listeners with normal hearing and listeners with hearing loss was significant. These results indicate that words and sentences in multitalker babble provide the same differentiation between performances by listeners with normal hearing and listeners with hearing loss. For the most part, results from word-recognition tasks in quiet have not differentiated between these groups.

This report focuses on the development of a clinical word-recognition task in multitalker babble for quantifying speech understanding in background noise (and therefore one aspect of the distortion component of hearing loss). Individual word-recognition data for the 24 to 8 dB S/B conditions were compiled from four previous studies that evaluated 573 baseline listeners with sensorineural hearing loss on the 70 WIN words [7,25,37-38]. The descriptive characteristics of these listeners, who ranged from 38 to 89 yr, are presented in Table 1 (Preliminary column). Their mean audiogram can be characterized as a mild-to-moderate high-frequency hearing loss with a mean word-recognition performance of 75.6 percent correct on the female-speaker version of the NU-6 [23]. In the four studies, multitalker babble was

Table 1. Mean ± standard deviation of ages, pure-tone thresholds,* and word-recognition scores for three listener groups.

Characteristics	Preliminary	Experiment 1	Experiment 2
	(n = 573)	(n = 72)	(n = 48)
Age (yr)	66.5 ± 10.6	64.9 ± 9.5	67.9 ± 9.3
Pure-Tone Thresholds (decibel hearing level [Hz])			
250	22.0 ± 9.0	22.2 ± 9.1	21.0 ± 8.1
500	22.8 ± 9.3	22.1 ± 8.9	22.2 ± 8.9
1,000	26.6 ± 11.0	25.3 ± 9.4	25.8 ± 10.8
2,000	46.7 ± 15.0	44.9 ± 17.8	44.6 ± 18.0
3,000	62.3 ± 13.4	61.2 ± 13.9	57.9 ± 17.5
4,000	69.2 ± 14.1	68.4 ± 13.9	64.0 ± 17.0
8,000	68.3 ± 15.9	64.2 ± 16.9	62.8 ± 18.3
Word Recognition (% correct)	75.6 ± 19.7	83.6 ± 12.3	85.9 ± 13.5

*Source: American National Standards Institute. Specification for audiometers (ANSI S3.6 1996). New York (NY): American National Standards Institute; 1996.

presented at 80 dB sound pressure level (SPL) and the words were presented in 4 dB decrements from 104 dB SPL (84 dB hearing level [HL] or 24 dB S/B) to 80 dB SPL (60 dB HL or 0 dB S/B).

The 70 words are sorted in Table 2 by mean recognition performance (percent correct) of the 573 baseline listeners. The superscripted notations beside each percent correct value designate in which list each word appeared (A = Experiment 1, List 1; B = Experiment 1, List 2; 1 = Experiment 2, List 1; 2 = Experiment 2, List 2). For example, *road* was in List 1 of Experiment 1 and List 2 of Experiment 2, whereas *pain* was in List 2 of Experiment 1 and List 1 of Experiment 2. Notably, the 10 words that produced the best performances (*road* through *ditch*) included 5 words from the 24 dB S/B condition, 4 words from the 20 dB S/B condition, and 1 word from the 16 dB S/B condition (Table 2). The heterogeneity of the recognition performances, which was most apparent for the first 30 words, demonstrates that S/B ratio is not the only determinant of word-recognition performance. Furthermore, the original 70-word list was developed from data on listeners with normal hearing [19]. At the highest S/B ratios (24 to 12 dB), listeners with normal hearing had the highest performances, which makes decisions about which word to use at the highest S/B ratios arbitrary. As became apparent in the initial studies, list equivalence differs for listeners with normal hearing and listeners with hearing loss. Our focus was to establish lists that were equivalent for listeners with sensorineural hearing loss.

We performed two subsequent experiments with these 70 words (see text). Experiment 1 created two 35-word lists based on recognition performance and S/B ratio shown. Experiment 2 created two 35-word lists based only on recognition performance. Superscript notations next to % correct value indicates in which list word appeared in each experiment (i.e., Experiment 1: List 1 = A, Experiment 1: List 2 = B, Experiment 2: List 1 = 1, Experiment 2: List 2 = 2).

We initially evaluated the data in Table 2 for the expected differences between 35-word lists in 50 percent correct recognition points as determined by the Spearman-Kärber equation. The worst-case scenario involved one group at each S/B ratio that had the five words with the best performances and a second group that had the five words with the worst performances. When grouped this way, the 50 percent correct recognition points for the two lists were 14.0 and 10.8 dB S/B, respectively, which is a 3.2 dB difference. Twenty random sorts of the data at each S/B ratio revealed absolute differences between the 50 percent correct recognition points that ranged from 0.1 to 1.2

Table 2. Seventy test words (beginning in first column and continuing in fourth column) and corresponding signal-to-babble (S/B) ratios. Words rank-ordered by % correct word recognition of baseline listeners with high-frequency sensorineural hearing loss (n = 573).

Word	Decibel S/B	% Correct*	Word	Decibel S/B	% Correct*
Road	24	98.8A,2	Tool	12	56.7A,1
Gun	20	95.8A,2	Voice	12	55.5B,1
Food	24	95.1B,2	Search	12	53.2A,2
Pain	24	94.4B,1	Rush	12	51.5A,1
Juice	24	94.2B,2	Deep	8	36.8B,2
Dodge	24	93.9B,1	Pick	8	31.2A,1
Tire	20	90.8A,2	Chief	12	30.9B,2
Dog	16	90.1A,2	Young	8	29.3A,1
Ring	20	89.2B,1	Sour	8	23.2B,2
Ditch	20	88.8A,1	Half	8	21.5B,1
Chair	20	87.8A,1	Soap	8	19.9A,2
Haze	20	86.7B,2	Turn	8	19.9B,1
Youth	24	86.2A,1	Doll	8	19.0A,2
Wheat	24	85.9B,1	Bite	8	17.3B,1
Date	16	85.3B,2	Make	8	16.1A,2
Red	16	84.6A,1	Sheep	4	10.3A,2
Luck	20	84.1B,1	Long	4	7.5A,2
Shawl	20	81.8B,2	Learn	4	5.9B,1
Hate	12	81.7A,1	Mess	4	5.2A,2
Witch	12	81.7A,2	Talk	4	4.4B,1
Kick	20	81.0A,1	Beg	4	2.1B,2
Wire	16	79.1B,1	Mood	4	1.9B,1
Late	24	78.9A,2	Note	4	1.9A,1
Judge	16	78.9B,1	Far	4	1.7B,1
Good	12	78.7B,2	Mouse	4	1.6A,2
Cool	24	77.0A,1	Bath	0	0.9A,1
Such	20	75.9B,2	Nice	0	0.9B,2
Hire	24	73.6A,2	Back	0	0.7B,2
Live	16	72.3B,2	Calm	0	0.5B,2
Base	16	71.6A,1	Get	0	0.5B,1
Have	16	67.9B,2	Life	0	0.3A,1
Shack	12	66.3B,1	Kill	0	0.2A,2
Time	16	66.1A,1	Dab	0	0.0A,2
Pass	12	64.7B,2	Gaze	0	0.0A,1
Gas	16	61.6A,2	Read	0	0.0B,1

dB and a mean ± standard deviation (SD) absolute difference of 0.5 ± 0.3 dB. Based on these simulations, then, we could reasonably expect performance differences of close to 0 dB on two 35-word lists that were compiled for minimization of differences.

We developed the two 35-word lists using two strategies based on error analysis of recognition performance data from the 573 baseline listeners [39]. The first, more traditional strategy (Experiment 1) focused on recognition performance of 10 words at each of the 7 S/B ratios. Each group of words at each S/B ratio was sorted by mean recognition performance. Then, the words ranked 1 and 10 were paired and put in List 1, the words ranked 2 and 9 were paired and put in List 2, etc. Once the two lists were compiled, we made slight adjustments to equalize the means and minimize the SDs for the two sets of words at each S/B ratio. Although the data at the various S/B ratios determined the lists, the ultimate metric of interest was the 50 percent correct recognition point on the function. Figure 1 is a bivariate plot of the 50 percent correct recognition points on List 1 (abscissa) and List 2 (ordinate) for the 573 listeners. Although 258 listeners performed better on List 2 than List 1 and 216 performed better on List 1 than List 2, the mean 50 percent correct recognition point was the same for both lists (12.4 dB S/B). Finally, 25 listeners with hearing loss performed in

the normal range on List 1, 21 on List 2, and 14 on both lists. Thus, 2.5 percent of the 573 listeners with hearing loss had normal recognition performance. Test-retest data over a 12-month period were available for 315 of the 573 listeners. In the original analyses, an intraclass correlation coefficient of 0.88 was observed for the test-retest data [40]. The test-retest difference between the 50 percent correct recognition points was 0.3 dB with a 95 percent confidence interval of ±3.6 dB, which indicates that performance did not change. For the current report, the test-retest data from the 315 listeners with hearing loss were sorted into two 35-word lists (Appendix).

The second strategy (Experiment 2) focused on the performance of the 573 baseline listeners on all 70 words. The words were rank-ordered by recognition performance without regard to S/B ratio (Table 2), with groups of 10 words identified for each of 7 performance levels. For example, Performance Level 7 had the 10 words with the best overall performances (words *road* through *ditch*, Table 2), whereas Performance Level 6 had the 10 words with the second best performances (*chair* through *witch*, Table 2). Then alternate words were sorted into the two lists with minor adjustments for equalization of the means and minimization of the SDs. Finally, we made slight adjustments to ensure that each list contained five words at each of the seven S/B ratios so that we could use the Spearman-Kärber equation to calculate the 50 percent correct recognition point [24]. Although not included in this article, a bivariate plot of the 50 percent correct recognition points from these two lists was almost identical to Figure 1, with better performance on List 2 by 250 listeners, better performance on List 1 by 234 listeners, and equal performance on both lists by 89
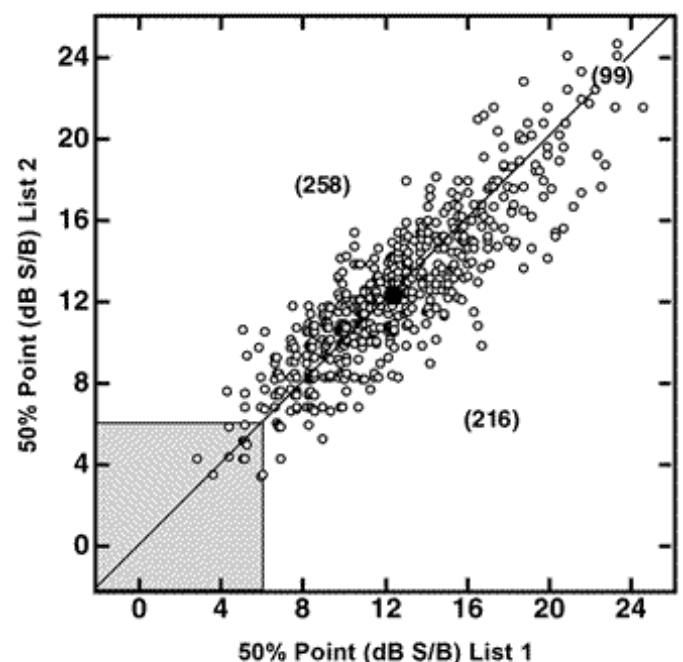

Figure 1. Bivariate plot of 50% correct recognition points (determined with Spearman-Karber equation) on List 1 (abscissa) and List 2 (ordinate) of words in Noise protocol for 573 listeners with mild-to-moderate sensorineural hearing loss. Diagonal line is equal word-recognition performance on both lists, and parenthetical numbers in plot are number of listeners with 50% points above, on, and below line. Large, filled circle is mean datum point. To minimize overlapping data points, we used a jittered algorithm that randomly multiplied x and y values by 1.02 to 0.98 in 0.01 steps. Shaded region in lower left in 90th percentile for listeners with normal hearing. (Source: Wilson RH, Abrams HB, Pillion AL. A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble [S/B]. J Rehabil Res Dev. 2003: 40(4): 321-327.)

listeners. Again, the mean 50 percent correct recognition point for the two lists was 12.4 dB S/B.

The majority of the 573 baseline listeners with hearing loss had minimal word-recognition performance (<6% correct) at the two lowest S/B ratios (4 and 0 dB). Therefore, we used data from 49 listeners with normal hearing (£20 dB HL at the 250-8,000 Hz octaves [41]) to determine equivalence at these two S/B ratios, which might have been arbitrary for listeners with hearing loss [38,42].

Experiment 1

Experiment 1 examined the word-recognition performances of 72 listeners with hearing loss on two 35-word lists in multitalker babble. The two lists were formed based on the recognition performances for each of the 10 words at each of the 7 S/B ratios from 24 to 0 dB. Groups of five words at each S/B ratio were then combined to form the two lists. A pairwise *t*-test, which compared the 50 percent correct recognition points of the 573 baseline listeners on the two lists, indicated that List 1 and List 2 were not significantly different, which corroborates Figure 1.

Methods

Materials

Two randomizations of the two 35-word lists were compiled with the WIN stimulus files from the original experiment [19]. Each word was mixed with a time-locked, unique segment of babble with five unique words at each S/B ratio from 24 to 0 dB in 4 dB decrements (e.g., the five words presented at 24 dB S/B were always presented at 24 dB S/B). Because the word-to-babble segments were concatenated at the negative-going zero crossings (the boundaries between words), babble segments were acoustically and perceptually transparent to the listener. The materials with the speech and babble mixed were recorded on a compact disc (Hewlett-Packard, Model DVD200i).

Subjects

This experiment studied 72 older listeners (mean = 66.7 yr, range = 43 to 84 yr) with sensorineural hearing loss. Inclusion criteria were-

1. £30 dB HL threshold at 500 Hz [41].

2. £40 dB HL threshold at 1,000 Hz.

3. £35 dB HL threshold above 1,000 Hz.

4. >50 percent correct maximum word recognition in quiet [23].

The descriptive data for the 72 listeners are listed in Table 1 (Experiment 1 column). Similar to the 573 baseline listeners, the 72 listeners in Experiment 1 had mild-to-moderate hearing loss with a mean 83.6 percent word-recognition ability in quiet. The listeners were recruited from audiology clinics at the VA Medical Center in Mountain Home, Tennessee. Participants signed informed consent forms before participation.

Procedures

For all conditions, the multitalker babble level was fixed at 80 dB SPL and the speech level varied from 104 to 80 dB SPL in 4 dB decrements. The conditions were designed so that half of the listeners received List 1 followed by List 2. The words in multitalker babble were reproduced on a compact disc player (Sony, Model CDP-497), routed through an audiometer (Grason-Stadler, Model 10) to a TDH-50P earphone encased in a cushion (P/N 510C017-1, Telephonics Corporation). The right ears of even-numbered listeners and the left ears of odd-numbered listeners were tested. The nontest ear was covered with a dummy earphone. All testing was conducted in a double-wall sound booth, and the listeners' verbal responses were recorded into a spreadsheet.

Results

Figure 2 depicts the mean psychometric functions for the two lists based on data from the 573 baseline listeners with hearing loss (Figure 2(a)) and the 72 listeners with hearing loss in Experiment 1 (Figure 2(b)). Figure 2 also compares the performances of these two listener groups on List 1 (Figure 2(c)) and List 2 (Figure 2(d)). The mean ± SD for the two groups with hearing loss and the group with normal hearing are listed in Table 3. As indicated earlier, the 573 baseline listeners performed the same on both lists (Figure 2(a)). The 72 listeners with hearing loss, however, performed 0.5 dB better on List 2 than List 1 (Figure 2(b)). A pairwise *t*-test indicated that this 0.5 dB difference was significant ($t_{(71)}$ = 2.199, $p < 0.05$). The 72 listeners performed about 2 dB poorer on both lists than the 573 listeners, a group effect likely due to the different inclusion criteria. The effect of list presentation order was also evaluated. A pairwise *t*-test indicated that recognition performance on the list given first was significantly better (0.5 dB) than the list given second ($t_{(71)}$ = 2.199, $p < 0.05$). Finally, intersubject variability (SD) was greater with List 2 than List 1, which is attributed to the random nature of the test design.

The slopes of the mean functions, calculated at the 50 percent correct recognition points for the two lists (Figure 2), were 6.6 to 6.7 percent/dB for both groups of listeners. When the mean slopes were computed for the individual data from the 72 listeners, the slopes were slightly steeper: 9.2 percent/dB for List 1 and 8.9 percent/dB for List 2 (Table 3). Slopes calculated from individual data better predict slopes for individual listeners than slopes of the mean function [43]. The slopes of these functions are essentially identical to those for listeners with hearing loss in similar studies [19,23]. In these earlier studies, the slopes of the functions were steeper for listeners with normal hearing than for listeners with hearing loss.

Experiment 1, based on performance equivalency at each of the seven S/B ratios, produced two lists with psychometric functions that had similar morphologies but were significantly displaced by 0.5

Table 3. Mean ± standard deviation percent correct word recognition for 5 words at each signal-to-babble (S/B) ratio on two 35-word lists. Based on data from listeners with hearing loss (573 listeners in baseline studies and 72 listeners in Experiment 1) and from 49 listeners with normal hearing at lowest S/B ratios. Mean 50 percent correct recognition points established with Spearman-Kärber equation (SK 50%), mean slopes (slope at 50%), and mean overall percent correct (mean %) also shown for listeners with hearing loss.

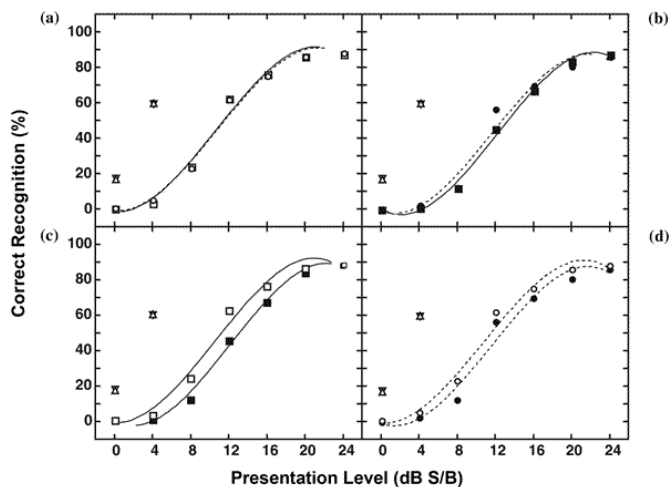| Decibel S/B	List 1			List 2		
	Hearing Loss		Normal Hearing	Hearing Loss		Normal Hearing
	(n = 573)	(n = 72)	(n = 49)	(n = 573)	(n = 72)	(n = 49)
24	87.5 ± 7.1	87.2 ± 11.1	-	88.1 ± 11.1	86.1 ± 10.8	-
20	86.2 ± 3.5	83.6 ± 5.5	-	86.2 ± 7.7	80.8 ± 14.1	-
16	76.1 ± 7.2	66.9 ± 7.2	-	75.4 ± 11.9	69.7 ± 12.6	-
12	62.3 ± 12.1	45.3 ± 9.6	-	61.8 ± 20.7	56.4 ± 24.1	-
8	23.8 ± 6.1	11.7 ± 7.2	-	23.0 ± 8.1	12.2 ± 5.7	-
4	3.2 ± 1.9	0.6 ± 1.2	63.1 ± 18.9	5.3 ± 3.7	2.5 ± 3.5	62.5 ± 11.0
0	0.3 ± 0.4	0	22.5 ± 16.5	0.5 ± 0.4	0	23.2 ± 11.9
SK 50%	12.4 ± 3.9	14.2 ± 3.7	-	12.4 ± 3.7	13.7 ± 3.3	-
Slope at 50% (%/dB)	-	9.2 ± 4.5	-	-	8.9 ± 4.1	-
Mean %	48.5	42.2	-	48.6	44	-

Figure 2. Psychometric functions for word-recognition performance on two 35-word lists sorted by signal-to-babble (S/B) ratio for (a) 573 baseline listeners with hearing loss. (b) 72 listeners with hearing loss in Experiment 1. (c). both listener groups on List 1. And (d) both listener groups on List 2. Throughout figure, List 1 depicted with soild lines and squares (open = 573-listener group, closed = 72-listener group), and List 2 depicted with dashed lines and circles (open = 573-listener group, closed = 72-listener group). Triangles are mean performance on both lists for 49 listeners with normal hearing. (Sources: Wilson RH, Weakley DG. The 500 Hz masking-level difference and word recognition in multitalker babble for 40- to 89-year old listeners with symmetrical sensorineural hearing loss. J Am Acad Audiol 2005: 16(6): 367-82, Wilson RH, Weakley DG. The use of digit triplets to evaluate word-recognition abilities in multitalker babble. Sem Hear. 2004: 25: 93-111)
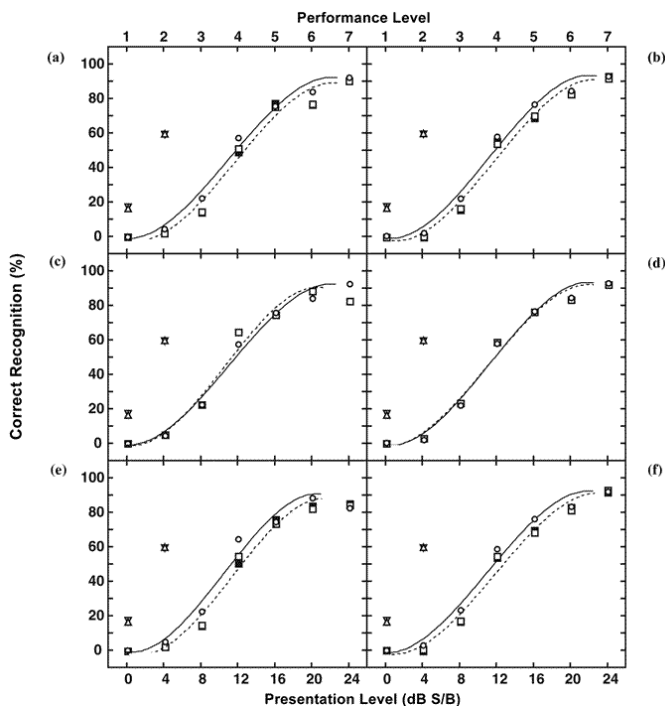

Figure 3. Top row: Psychometric functions for word-recognition performance on two 35-word lists sorted by performance level listeners with hearing loss (573 listeners in baseline studies and 48 listeners in Experiment 2) on (a) List 1 and (b) List 2. Middle row: Psychometric functions based on performance level (solid line) and signal-to-babble (S/B) ratio (dashed line) for same 573 baseline listeners on (c) List 1 and (d) List 2. Bottom row: Same word-recognition performance data as shown in (a) and (b) but sorted by S/B ratio for (e) List 1 and (f) List 2. Triangles are mean performances on both lists for 49 listeners with normal hearing. (Sources: Wilson RH, Weakley DG. The 500 Hz masking-level difference and word recognition in multitalker babble for 40- to 89-year old listeners with symmetrical sensorineural hearing loss. J Am Acad Audiol. 2005: 16(6): 367-382. Wilson RH, Weakley DG. The use of digit triplets to evaluate word-recognition abilities in multitalker babble. Sem Hear. 2004: 25: 93-111)

dB when 72 listeners with hearing loss were evaluated. Given that the listeners with hearing loss performed 8 to 12 dB poorer than the listeners with normal hearing, the 0.5 dB difference between the two lists is probably not clinically important. This difference, however, did prompt a second experiment in which a different strategy was used for devising the word lists.

Experiment 2

Experiment 2 examined the word-recognition performances of 48 listeners with hearing loss on two 35-word lists presented in multitalker babble at 24 to 0 dB S/B ratios. The two lists were formed based on the recognition performances for the 70 words regardless of S/B ratio. Again, a pairwise *t*-test on the data from the 573 baseline listeners did not demonstrate a significant difference between the 50 percent correct recognition points for List 1 and List 2.

Methods

Materials

We devised two randomizations of the two 35-word lists using the procedures described in the materials section of Experiment 1.

Subjects

The inclusion criteria for Experiment 2 were the same as for Experiment 1. The descriptive data for the 48 listeners, who ranged from 45 to 83 yr, are listed in Table 1 (Experiment 2 column). As with the two previous groups of listeners, the 48 listeners in Experiment 2 had mild-to-moderate hearing loss and a mean 85.9 percent word-recognition ability in quiet.

Procedures

Recognition performance on two randomizations of the 35-word lists in multitalker babble was determined. List 1 was presented first to the odd-numbered listeners and List 2 was presented first to the even-numbered listeners. The two lists and two randomizations were presented so that the four possible combinations were given an equal number of times. This design not only allowed evaluation of performances on List 1 and List 2 but also examination of order effects (first vs second) and randomization effects (randomization 1 vs randomization 2). All other procedures were the same as Experiment 1.

Results

The data for Experiment 2 are presented in two ways. First, the results are presented according to the word groupings based on recognition performance (Figures 3(a) and 3(b), Table 4). Performance level is the independent variable. Second, the same data are grouped in the more traditional manner with S/B ratio as the independent variable (Figures 3(e) and 3(f)). The functions from these two formats are compared in Figures 3(c) and 3(d). Data for List 1 (Figure 3(a)) and List 2 (Figure 3(b)) are sorted by recognition performance. The function for the 573 baseline listeners and the two randomizations for the 48 listeners in Experiment 2 are also shown. The 573 baseline listeners performed approximately 1 dB better than the 48 listeners, which again is a group effect. The 48 listeners performed the same on the two list randomizations. The 50 percent correct recognition points were calculated with the Spearman-Kärber equation. The 50 percent correct recognition points for the 48 listeners were 13.5 and 13.3 dB S/B for the first randomization of Lists 1 and 2, respectively (Table 4). A two-way repeated measures analysis of variance (ANOVA) on the 50 percent correct recognition points failed to demonstrate a statistically significant main effect of list or randomization.

Table 4. Mean ± standard deviation % correct word recognition for 5 words at each performance level on two 35-word lists. Based on data from listeners with hearing loss (573 listeners in baseline studies and 48 listeners in Experiment 2). Mean 50% correct recognition points established with Spearman-Kärber equation (SK 50%), mean slopes (slope at 50%), and mean overall % correct (mean %) also shown.

Performance Level	List 1		List 2	
	n = 573	n = 48	n = 573	n = 48
7	92.8 ± 4.3	90.8 ± 4.3	93.4 ± 2.4	92.1 ± 6.1
6	84.4 ± 2.7	77.5 ± 10.4	84.8 ± 1.9	82.9 ± 5.8
5	76.4 ± 3.8	75.8 ± 10.3	77.0 ± 2.9	70.4 ± 7.3
4	57.9 ± 6.0	51.7 ± 13.6	58.3 ± 12.9	54.2 ± 19.0
3	23.1 ± 6.7	15.0 ± 10.4	22.5 ± 5.1	16.7 ± 6.6
2	5.1 ± 4.0	2.5 ± 2.7	2.7 ± 2.4	0.4 ± 0.9
1	0.5 ± 0.8	0.4 ± 0.9	1.0 ± 0.8	0
SK 50%	12.4 ± 3.8	13.5 ± 4.0	12.4 ± 3.8	13.3 ± 3.8
Slope at 50% (%/dB)	-	9.7 ± 5.4	-	9.2 ± 4.8
Mean %	48.6	44.8	48.5	45.2

The main difference between Table 3 and 4 is the smaller SDs for the 573-baseline listener data obtained with the second randomization strategy. Smaller SDs were expected because the word groups created by the second strategy resulted in more homogenous recognition performances than the first strategy. Because the number of listeners in Experiments 1 and 2 differed, comparison of the SDs for these two groups was difficult.

Figures 3(c) and 3(d) depict the functions of the two lists for the 573 baseline listeners. The functions based on performance level and on S/B ratio are shown for comparison. The differences between the functions are seen at the two highest levels for List 1. Figures 3(e) and 3(f) recast the data from Figures 3(a) and 3(b) with S/B ratio as the independent variable. The same relations among the functions in Figures 3(e) and 3(f) are seen among the functions in Figures 3(a) and 3(b).

The effect of presentation order was evaluated across the four possible presentation positions (two randomizations by two lists). Although recognition performance on the fourth presentation was 0.4 dB better than on the first presentation, a one-way repeated measures ANOVA did not demonstrate a statistically significant main effect of order, which suggests that learning effects associated with the listening task are minimal and not of clinical concern.

The slopes of the mean functions calculated at the 50 percent correct recognition points for the two lists (Figure 3) were 6.5 percent/dB for both listener groups. When the mean slopes were computed for individual data from the 48-listener group, the slopes were steeper: 9.7 percent/dB for List 1 and 9.2 percent/dB for List 2 (Table 4). Again, these slopes are almost identical to the slopes from earlier WIN studies [19,25,37].

Discussion

The purpose of this study was the development of two 35-word lists in multitalker babble that clinical audiologists could use to quickly measure patients' understanding of speech in background noise. We used two randomization strategies to devise the two lists from an experimental 70-word test [19]. Listener performances were essentially equivalent on the lists from both randomization strategies. In our first experiment, based on the first randomization strategy, the recognition performance of 72 listeners with hearing loss was approximately 0.5 dB poorer on one list than the other. In our second experiment, based on the second randomization strategy, 48 listeners with hearing loss had the same recognition performance on both lists. These two

experiments and the previous baseline studies indicate that most 35-word groupings of the 70 words produce equivalent results. Because recognition performances on the two lists from the two randomization strategies showed minimal differences, either set of lists is appropriate for clinical use.

Although the 35-word lists developed in Experiment 2 use five words at each of the seven S/B ratios, we suggest that clinicians present the words from highest to lowest recognition performance. In this way, the words are presented more or less randomly with regard to their S/B ratio. Recall that earlier data indicated that listeners had equivalent recognition performances on random and descending presentation level protocols [25]. Furthermore, presentation of words in order of performance level, rather than S/B ratio, provides the listener with a listening experience that progresses from easiest to most difficult. The independent variable can be plotted as either performance level (1 to 7) or S/B ratio (24 to 0 dB S/B).

The 50 percent correct recognition point, as determined by the Spearman-Kärber equation, is the primary metric of a patient's ability to understand speech in background noise. In the audiology clinic at the VA Medical Center in Mountain Home, Tennessee, S/B-ratio hearing loss is defined as-

1. Normal hearing £6.0 dB.

2. Mild hearing loss 6.8 to 10.0 dB.

3. Moderate hearing loss 10.8 to 14.8 dB.

4. Severe hearing loss 15.6 to 19.6 dB.

5. Profound hearing loss >20 dB.

Additionally, a plot of the performances at the various presentation levels according to the independent variable is insightful because both the 50 percent correct recognition point and the function morphology are considered. Figure 4 illustrates the graphic format clinicians use to plot patient responses to a WIN test. The shaded regions define the 90th percentile performances of listeners with normal hearing [7]. The data in Figure 4 are from one listener in Experiment 2 on List 2. The independent variable was plotted as either performance level (Figure 4(a)) or S/B ratio (Figure 4(b)). Regardless of the plotting method, the extent to which a given listener's recognition performance departs from normal recognition performance is immediately obvious. In this case, the 50 percent correct recognition point was 15.6 dB S/B.

Throughout this article, we consider word-recognition performance in multitalker babble for determining S/B hearing loss. Hearing loss
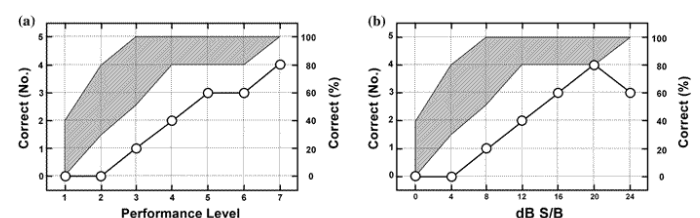

Figure 4. Sample words in Noise data plotted for clinical use. Data shown from one listener in Experiment 2. Shaded region of graphs are 10th and 90th percentile ranges of recognition performance by young listeners with normal hearing. (a) Data plotted according to word groupings defined by performance. Regardless of signal-to-babble (S/B) ration. (b) Same data plotted in traditional manner of recognition performance of each word grouping at each S/B ratio. (Source: Wilson RH, Abrans HB, Pillion AI, A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. J Rehabil Res Dev. 2003: 40(4): 321-327)

was specified by the 50 percent correct point with consideration given to the morphology of the psychometric function. Because multiple presentation levels are characteristic of the WIN test, reports of an overall percent correct score would be inappropriate. As a point of reference, however, the overall percent correct performance on the WIN test was about 80 percent for listeners with normal hearing [38] and about 50 percent for the 573 listeners with hearing loss.

Conclusions

Our goal was to shorten the WIN test from 5 to 2.5 min by halving the original 70-word list. This goal was accomplished. For most listeners, however, the test took less than 2.5 min because a stopping rule usually terminated the protocol following the 4 dB S/B level, and often, following the 8 dB S/B level. These early terminations reduced the test time by 20 and 40 s, respectively. Although the focus of this article was the development of a 35-word list for evaluating speech understanding in background noise, use of the two lists as a composite 70-word list is not precluded.

References

1. Hirsh IJ (1950) Binaural hearing aids: a review of some experiments. *J Speech Disord* 15: 114-23.

2. Carhart R, Tillman TW (1970) Interaction of competing speech signals with hearing loss. *Arch Otolaryngol* 91: 273-79.

3. Groen JJ (1969) Social hearing handicap: its measurement by speech audiometry in noise. *Int J Audiol*. 8: 82-183.

4. CARHART R (1951) Basic principles of speech audiometry. *Acta Otolaryngol* 40: 62-71. [Crossref]

5. Stephens SDG (1976) The input for a damaged cochlea: a brief review. *British J Audiol* 10: 97-101.

6. Plomp R (1978) Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am* 63: 533-549. [Crossref]

7. Wilson RH, Abrams HB, Pillion AL (2003) A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. *J Rehabil Res Dev* 40: 321-27. [Crossref]

8. Dubno JR, Dirks DD, Morgan DE (1984) Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am* 76: 87-96. [Crossref]

9. Gordon-Salant S (1987) Age-related differences in speech recognition performance as a function of test format and paradigm. *Ear Hear* 8: 277-82. [Crossref]

10. Beattie RC1 (1989) Word recognition functions for the CID W-22 test in multitalker noise for normally hearing and hearing-impaired subjects. *J Speech Hear Disord* 54: 20-32. [Crossref]

11. Pekkarinen E, Salmivalli A, Suonpaa J (1990) Effect of noise on word discrimination by subjects with impaired hearing, compared with those with normal hearing. *Scand Audiol* 19: 31-36. [Crossref]

12. Souza PE, Turner CW (1994) Masking of speech in young and elderly listeners with hearing loss. *J Speech Hear Res*. 37: 655-61.

13. Divenyi PL, Haupt KM (1997) Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects. *Ear Hear* 18: 42-61.

14. Divenyi PL, Haupt KM (1997) Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. II. Correlation analysis. *Ear Hear* 18: 100-13.

15. Divenyi PL, Haupt KM (1997) Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation. *Ear Hear* 18: 189-201.

16. Wiley TL, Cruickshanks KJ, Nondahl DM, Tweed TS, Klein R, Klein BE (1998) Aging and word recognition in competing message. *J Am Acad Audiol* 9: 191-98.

17. Plomp R, Duquesnoy AJ (1982) A model for the speech-reception threshold in noise without and with a hearing aid. *Scand Audiol Suppl* 15: 95-111.

18. Killion MC (2002) New thinking on hearing in noise: a generalized articulation index. *Sem Hear* 23: 57-75.

19. Wilson RH (2003) Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *J Am Acad Audiol* 14: 453-70. [Crossref]

20. Wilson RH, Oyler AL, Sumrall R (1996) Psychometric functions for Northwestern University Auditory Test No. 6 in quiet and multi-talker babble. In: American Academy of Audiology Convention; 1996 Apr; Salt Lake City, Utah.

21. Wilson RH, Strouse A (2002) Northwestern University Auditory Test No. 6 in multi-talker babble: a preliminary report. *J Rehabil Res Dev* 39: 105-14. [Crossref]

22. Tillman TW, Carhart R (1966) An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6. SAM-TR-66-55 (Technical Report). Brooks Air Force Base (TX): USAF School of Aerospace Medicine.

23. Department of Veterans Affairs. Speech recognition and identification materials. Disc 2.0 (compact disk). Mountain Home (TN): Department of Veterans Affairs Medical Center; 1998.

24. Finney D (1952) Statistical method in biological assay. London (England): C. Griffen.

25. Wilson RH, Burks CA, Weakley DG (2005) Word recognition in multitalker babble measured with two psychophysical methods. *J Am Acad Audiol* 16: 622-30. [Crossref]

26. Thornton AR, Raffin MJ (1978) Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res*. 21: 507-18. [Crossref]

27. Speaks C, Jerger J, Jerger S (1966) Performance-intensity characteristics of synthetic sentences. *J Speech Hear Res*. 9: 305-12. [Crossref]

28. Kalikow DN, Stevens KN, Elliot LL (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 61: 1337-51.

29. Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C (1984) Standardization of a test of speech perception in noise. *J Speech Hear Res* 27: 32-48. [Crossref]

30. Cox RM, Alexander GC, Gilmore C (1987) Development of the Connected Speech Test (CST). *Ear Hear* 8(Suppl 5): S119-26.

31. Cox RM, Alexander GC, Gilmore C, Pusakulich KM (1988) Use of the Connected Speech Test (CST) with hearing-impaired listeners. *Ear Hear* 9: 198-207.

32. Killion MC, Villchur E (1993) Kessler was right-partly: but SIN test shows some aids improve hearing in noise. *Hear J*. 46: 31-35.

33. Nilsson M, Soli SD, Sullivan JA (1994) Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95: 1085-99. [Crossref]

34. Martin FN, Champlin CA, Chambers JA (1998) Seventh survey of audiometric practices in the United States. *J Am Acad Audiol* 9: 95-104.

35. McArdle RA, Wilson RH, Burks CA (2005) Speech recognition in multitalker babble using digits, words, and sentences. *J Am Acad Audiol* 16: 726-39.

36. Killion MC, Niquette PA, Gundmundsen GI, Revit LJ, Banerjee S (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116(4 Pt 1): 2395-2405.

37. Wilson RH, Burks CA, Weakley DG (2006) Word recognition of digit triplets and monosyllabic words in multitalker babble by listeners with sensorineural hearing loss. *J Am Acad Audiol* In press 2006.

38. Wilson RH, Weakley DG (2005) The 500 Hz masking-level difference and word recognition in multitalker babble for 40- to 89-year-old listeners with symmetrical sensorineural hearing loss. *J Am Acad Audiol* 16: 367-82. [Crossref]

39. Wiley TL, Stoppenbach DT, Feldhake LJ, Moss KA, Thordardottir ET (1995) Audiologic practices: what is popular versus what is supported by evidence. *Am J Audiol* 4: 26-34.

40. McArdle RA, Chisolm TH, Abrams HB, Wilson RH, Doyle PJ (2005) The WHO-DAS II: measuring outcomes of hearing aid intervention for adults. *Trends Amplif* 9: 127-43. [Crossref]

41. American National Standards Institute. Specification for audiometers (ANSI S3.6 1996). New York (NY): American National Standards Institute; 1996.

42. Wilson RH, Weakley DG (2004) The use of digit triplets to evaluate word-recognition abilities in multitalker babble. *Sem Hear* 25: 93-111.

43. Wilson RH, Margolis RH (1983) Measurement of the auditory thresholds for speech stimuli. In: Konkle DF, Rintelmann WF, editors. Principles of speech audiometry. Baltimore (MD): University Park Press; p. 79-126.