J Am Acad Audiol 23:92–96 (2012)

The Words-in-Noise Test (WIN), List 3: A Practice List

DOI: 10.3766/jaaa.23.2.3

Richard H. Wilson*†
Kelly L. Watts*†

Abstract

Background: The Words-in-Noise Test (WIN) was developed as an instrument to quantify the ability of listeners to understand monosyllabic words in background noise using multitalker babble (Wilson, 2003). The 50% point, which is calculated with the Spearman-Kärber equation (Finney, 1952), is used as the evaluative metric with the WIN materials. Initially, the WIN was designed as a 70-word instrument that presented ten unique words at each of seven signal-to-noise ratios from 24 to 0 dB in 4 dB decrements. Subsequently, the 70-word list was parsed into two 35-word lists that achieved equivalent recognition performances (Wilson and Burks, 2005). This report involves the development of a third list (WIN List 3) that was developed to serve as a practice list to familiarize the participant with listening to words presented in background babble.

Purpose: To determine—on young listeners with normal hearing and on older listeners with sensorineural hearing loss—the psychometric properties of the WIN List 3 materials.

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

Study Sample: Twenty-four young adult listeners ($M = 21.6$ yr) with normal pure-tone thresholds ($\leq$20 dB HL at 250 to 8000 Hz) and 24 older listeners ($M = 65.9$ yr) with sensorineural hearing loss participated.

Data Collection and Analysis: The level of the babble was fixed at 80 dB SPL with the level of the words varied from 104 to 80 dB SPL in 4 dB decrements.

Results: For listeners with normal hearing, the 50% points for Lists 1 and 2 were similar (4.3 and 5.1 dB S/N, respectively), both of which were lower than the 50% point for List 3 (7.4 dB S/N). A similar relation was observed with the listeners with hearing loss, 50% points for Lists 1 and 2 of 12.2 and 12.4 dB S/N, respectively, compared to 15.8 dB S/N for List 3. The differences between Lists 1 and 2 and List 3 were significant. The relations among the psychometric functions and the relations among the individual data both reflected these differences.

Conclusions: The significant ~3 dB difference between performances on WIN Lists 1 and 2 and on WIN List 3 by the listeners with normal hearing and the listeners with hearing loss dictates caution with the use of List 3. The use of WIN List 3 should be reserved for ancillary purposes in which equivalent recognition performances are not required, for example, as a practice list or a stand alone measure.

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

Abbreviations: S/N, SNR = signal-to-noise ratio; WIN = Words-in-Noise Test

The Words-in-Noise Test (WIN) was developed to evaluate word-recognition performance in a fixed level multitalker babble (Wilson, 2003; Wilson and McArdle, 2007). Initially, the test paradigm consisted of ten monosyllabic words at each of seven signal-to-noise ratios (SNR, S/N) from 24 to 0 dB in 4 dB decrements. Each word was time-locked with a unique segment of babble and was always presented at the same SNR. The 70 words were from the Northwestern University Auditory Test No. 6 (NU-6; Tillman and Carhart,

*James H. Quillen VA Medical Center, Mountain Home, TN; Department of Audiology and Speech-Language Pathology, East Tennessee State University, Johnson City, TN

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

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. The contents of this paper do not represent the views of the Department of Veterans Affairs or the United States Government.

1966) and were spoken by a female speaker (Department of Veterans Affairs, 1998). For clinic use, the list of 70 words was divided into two comparable lists of 35 words each (Wilson and Burks, 2005). The quantification metric of the WIN is the 50% correct point that is calculated with the Spearman-Kärber equation (Finney, 1952; Wilson et al, 1973). The 90th percentile of the 50% points (6.0 dB S/N) is the upper limit of performance by listeners with normal hearing (Wilson et al, 2003); 50% points at SNRs higher than 6 dB are considered abnormal. As shown by Wilson et al (2007), the WIN provides data that are comparable to the QuickSIN (Killion et al, 2004). The WIN increasingly has been incorporated into audiological protocols, both in clinics and in research laboratories. Over the years, several audiologists have inquired about an additional WIN list that could be used as a practice list to familiarize the listener with the task of listening to speech in a background of multitalker babble. This note describes the development of WIN List 3, which is intended for that purpose.

Originally, the 150 words from Lists 2–4 of NU-6 were evaluated in four experiments by developing psychometric functions on each word embedded in a unique segment of babble (Wilson, 2003). After Experiment 1, 48 words were eliminated because of irregularities of one sort or another, leaving 102 words. Following Experiment 2, the 70 WIN words were finalized leaving 32 unused words. Collectively, then, there were 80 potential words that could be considered in constructing the 35-word WIN List 3. Devising the words for List 3, therefore, was an imperfect task as the word pool consisted of the undesirable words that had been discarded during formulation of the original WIN list. The same rules used with the original WIN list were used to compile List 3. First, target performances by young listeners with normal hearing were 30, 50, 70, and 90% correct at SNRs of 0, 4, 8, and 12 dB, gradually increasing toward 100% correct at 24 dB S/N. Second, for listeners with hearing loss (pure-tone average <40 dB HL; ANSI, 2004) a systematic change across the SNRs was sought. The words for WIN List 3 are listed in Appendix A. Supplemental to the online version of this article is a WAV file (Audio 1) of WIN List 3 that includes a 10 sec calibration tone at the end of the file. The WIN list is on the left channel, and the same words in quiet are on the right channel for monitoring purposes.

The 24 young adult listeners (mean = 21.6 yr, SD = 3.3 yr) had pure-tone thresholds at the octave frequencies that averaged <5 dB HL. The 24 older listeners (mean age = 65.9 yr, SD = 4.7 yr) had mean thresholds at the 250–8000 Hz octaves of 17.3, 19.0, 22.1, 40.0, 66.0, and 65.7 dB HL with a traditional three-frequency pure-tone average of 27.0 dB HL (SD = 8.3 dB) and a 1000, 2000, and 4000 Hz average of 42.9 dB HL (SD = 8.9). A randomization of each of the three 35-word WIN lists was presented to each listener with the presentation order counterbalanced so that each list was given an equal number of times in each of the three order positions. The level of the babble was fixed at 80 dB SPL with the level of the words varied from 104 to 80 dB SPL in 4 dB decrements. The materials were reproduced by a compact disc player (Sony, Model CDP-CE375) and fed through an audiometer (Grason-Stadler, Model 61) to an insert earphone (Etymōtic Research, Model ER-3A). The testing was conducted in a sound booth with the verbal responses of the listeners recorded into a spreadsheet.

The results are listed in Table 1 and illustrated in Figures 1 and 2 (the percents correct for the individual words are listed in Appendix A). Consider first in Table 1 the mean 50% points derived with the Spearman-Kärber equation. For both groups of listeners the data indicate that List 3 is more difficult than List 1 and List 2. For the listeners with normal hearing, Lists 1 and 2

Table 1. Mean Percent Correct (and SDs) for Each List at Each SNR for the Two Groups of Listeners

	List 1		List 2		List 3	
dB S/N	Mean	SD	Mean	SD	Mean	SD
Normal Hearing						
24	97.5	6.8	100.0	0.0	97.5	6.8
20	97.5	6.8	100.0	0.0	86.7	14.0
16	97.5	6.8	95.8	8.3	94.2	9.3
12	98.3	5.6	97.5	6.8	89.2	13.2
8	76.7	18.3	71.7	22.0	58.3	21.2
4	59.2	18.2	49.2	19.5	21.7	16.6
0	16.7	16.3	8.3	13.1	16.7	19.3
Spearman-Kärber 50%						
Mean (dB S/N)	4.3		5.1		7.4	
SD (dB)	1.3		1.1		2.3	
90th Percentile	6.0		6.0		10.6	
Polynomial						
50% (dB S/N)	3.2		4.1		6.8	
Slope (%/dB)	8.7		8.2		5.8	
Hearing Loss						
24	85.8	20.8	88.3	15.5	85.0	20.6
20	86.7	20.1	86.7	19.3	75.0	22.3
16	76.7	21.8	67.5	22.7	64.2	27.6
12	64.2	33.4	55.8	27.6	20.0	20.4
8	27.5	25.6	31.7	26.3	5.8	9.3
4	5.0	10.6	10.8	15.6	4.2	10.2
0	0.0	0.0	0.0	0.0	1.7	5.6
Spearman-Kärber 50%						
Mean (dB S/N)	12.2		12.4		15.8	
SD (dB)	4.0		3.9		3.1	
Polynomial						
50% (dB S/N)	11.1		11.5		15.2	
Slope (%/dB)	6.5		5.2		6.3	

Note: Also included are the mean 50% points calculated with the Spearman-Kärber equation and calculated from the polynomials used to describe the data in Figure 2.

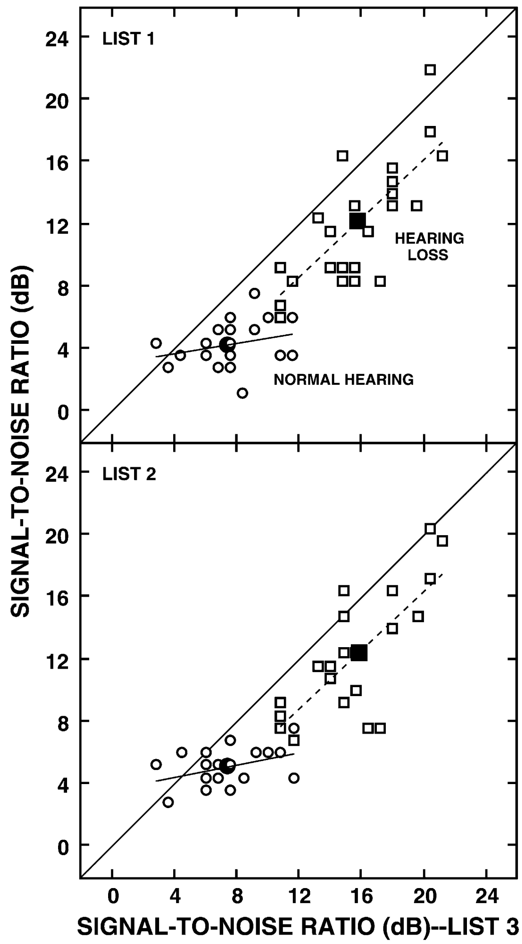

Figure 1. Bivariate plots of the 50% points calculated with the Spearman-Kärber equation for the individual listeners with normal hearing (circles) and with hearing loss (squares). The filled symbols represent the respective mean values. Linear regressions for each set of data also are shown.

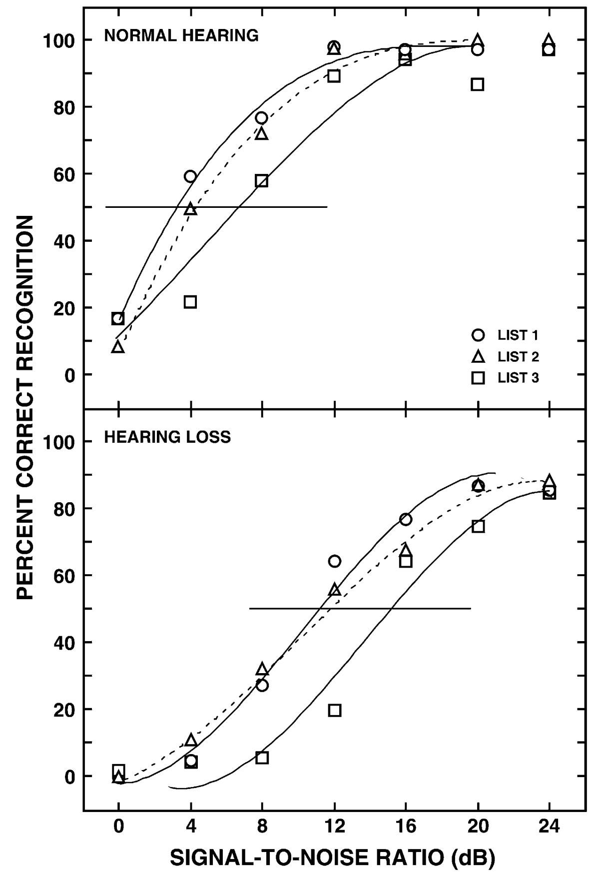

Figure 2. Psychometric functions for WIN List 1 (circles), List 2 (triangles), and List 3 (squares) for the two groups of listeners. The lines connecting the datum points are the best-fit, third-degree polynomials used to describe the data.

(50% points = 4.3 and 5.1 dB S/N) are about 2.5 dB easier than List 3 (50% point = 7.4 dB S/N), whereas for the listeners with hearing loss, the difference between Lists 1 and 2 and List 3 increased to 3.5 dB (12.2 and 12.4 dB S/N vs. 15.8 dB S/N). Repeated measures analysis of variance was used to examine the list differences within each group of listeners. The main effect of list was significant for both the listeners with normal hearing [$F(2,46) = 34.5$; $p < .001$] and the listeners with hearing loss [$F(2,46) = 32.4$; $p < .001$]. Post hoc comparisons using Bonferroni corrections for multiple comparisons showed that List 3 was significantly more difficult ($p < .001$) than List 1 or List 2 for both groups of listeners. Performances on List 1 and List 2 were not statistically different for either group of listeners ($p > .01$). The list differences for the individual listeners are reinforced by the bivariate plots in Figure 1 in which the data for List 1 (top panel) and List 2 (bottom panel) are on the ordinate with the data for List 3 on the abscissa. Almost all the data points for both groups of listeners are below the diagonal line that represents equal performance, which indicates poorer performance on List 3 than on Lists 1 or 2. Also noteworthy in the figure for the listeners with hearing loss are the linear regressions with slopes that approach unity. The unity slopes indicate a consistent relation between Lists 1 and 3 and between Lists 2 and 3. The only difference between the data sets is the displacement of the List 3 data to the higher SNRs. The R^2 values were 0.53 and 0.56 for the List 1 and List 2 comparisons with List 3, respectively.

The mean 50% points for the three WIN lists combined in each of the three presentation order positions were the same. For the listeners with normal hearing the means were 5.7, 5.7, and 5.4 dB S/N for presentation order positions 1, 2, and 3, respectively, and for the listeners with hearing loss the means were 13.5, 13.1, and 13.7 dB S/N. Repeated measures ANOVAs for the listeners with normal hearing [$F(2,46) = 0.11$; $p = .90$] and for the listeners with hearing loss [$F(2,46) = 0.27$; $p = .77$] showed no significant differences for presentation order.

The psychometric functions for the three lists and two groups of listeners are shown in Figure 2 with the mean percent correct data and standard deviations at each SNR listed in Table 1. Again, the displacement of the List 3 functions from the List 1 and 2 functions is obvious. For the listeners with normal hearing, the slopes of the functions are different, ranging from 8.7 and 8.2%/dB for Lists 1 and 2, respectively, to 5.8%/dB for List 3. Interestingly, the slopes for the three lists are similar for the listeners with hearing loss. The other obvious relation in Figure 2 is the similarity of the List 1 and 2 functions, especially when contrasted to the List 3 functions. As one would expect, at the extremes of the psychometric functions the data for the three lists are fairly equivalent indicating a commonality among the three lists where floor and ceiling effects are influential. The percent correct for each word by the two groups of listeners are listed in the supplemental materials.

Finally from Table 1, it is noteworthy that the 90th percentiles for Lists 1 and 2 for the listeners with normal hearing were 6.0 dB S/N, which is exactly the 90th percentile established with the original WIN (Wilson et al, 2003). In contrast, the 90th percentile for List 3 was 10.6 dB S/N.

In conclusion, for equal performances on WIN Lists 1, 2, and 3, List 3 requires a 2.5 to 3.5 dB more favorable SNR than are required by Lists 1 and 2. As indicated in the introduction, after the original 70 WIN words had been taken from the 150-word pool, the pool was populated with many words with psychometric properties that limited their usefulness in the compilation of words that were homogeneous with respect to recognition performance at selected SNRs. The significant ~3 dB differences between performances on WIN Lists 1 and 2 and on WIN List 3 for the listeners with normal hearing and the listeners with hearing loss preclude inclusion of List 3 in the same role as Lists 1 and 2. There is no reason, however, that a clinician or an investigator could not use List 3 as a practice list or for other stand alone purposes.

REFERENCES

American National Standards Institute. (2004) *Specification for Audiometers (ANSI S3.6 2004)*. New York: American National Standards Institute.

Department of Veterans Affairs. (1998) *Speech Recognition and Identification Materials. Disc 2.0*. Mountain Home, TN: VA Medical Center.

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

Killion MC, Niquette PA, Gudmundsen 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* 116:2395–2405.

Tillman TW, Carhart R. (1966) *An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words. Northwestern University Auditory Test No. 6*. USAF School of Aerospace Medicine Technical Report SAM-TR-66-55. Brooks Air Force Base, TX: USAF School of Aerospace Medicine.

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, 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* 40:321–328. http://www.rehab.research.va.gov/jour/03/40/4/pdf/Wilson-B.pdf.

Wilson RH, Burks CA. (2005) The use of 35 words for evaluation of hearing loss in signal-to-babble ratio: a clinic protocol. *J Rehabil Res Dev* 42:839–852. http://www.rehab.research.va.gov/jour/05/42/6/pdf/wilson.pdf.

Wilson RH, McArdle R. (2007) Intra- and inter-session test, retest reliability of the Words-in-Noise (WIN) test. *J Am Acad Audiol* 18:813–825.

Wilson RH, McArdle R, Smith SL. (2007) An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 50:844–856.

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

Appendix A. Recognition Performances on Each Word of WIN List 3 by the 24 Listeners with Normal Hearing and the 24 Listeners with Hearing Loss

			% Correct	
dB S/N	Word #	Word	Normals	Hearing Loss
24	1	WALK	100.0	91.7
24	2	PEG	95.8	87.5
24	3	ROSE	95.8	79.2
24	4	FIT	95.8	70.8
24	5	WHEN	100.0	95.8
20	6	SHIRT	100.0	75.0
20	7	BONE	100.0	75.0
20	8	VOTE	95.8	83.3
20	9	CHECK	70.8	83.3
20	10	NEAT	66.7	58.3
16	11	THIN	91.7	62.5
16	12	JUG	100.0	70.8
16	13	NEAR	87.5	33.3
16	14	PHONE	91.7	79.2
16	15	HALL	100.0	75.0
12	16	CAUSE	87.5	20.8
12	17	SEIZE	95.8	12.5
12	18	SHOULD	91.7	12.5
12	19	LEAN	91.7	8.3
12	20	CHAT	79.2	45.8
8	21	LID	54.2	0.0
8	22	HOLE	50.0	25.0
8	23	LEASE	83.3	4.2
8	24	YEARN	66.7	0.0
8	25	PEARL	37.5	0.0
4	26	PERCH	12.5	8.3
4	27	DIP	58.3	8.3
4	28	SHALL	25.0	0.0
4	29	TEAM	8.3	4.2
4	30	CHAIN	4.2	0.0
0	31	CHEEK	25.0	8.3
0	32	GERM	29.2	0.0
0	33	FIVE	8.3	0.0
0	34	HIT	12.5	0.0
0	35	NAME	8.3	0.0