International Journal of Audiology 2012; 51: 708–712

Technical Report

Speech-in-noise measures: Variable versus fixed speech and noise levels

Richard H. Wilson* & Rachel McArdle†

*VA Medical Center, Mountain Home, TN, USA, and East Tennessee State University, Johnson City, USA, and †Bay Pines VA Healthcare System, Bay Pines, FL, USA, and University of South Florida, Tampa, USA

Abstract

Objective: The purpose was to determine if speech-recognition performances were the same when the speech level was fixed and the noise level varied as when the noise level was fixed and the speech level varied. *Design:* A descriptive/quasi-experimental experiment was conducted with Lists 3 and 4 of the revised speech perception in noise (R-SPIN) test, which involves high predictability (HP) and low predictability (LP) words. The R-SPIN was modified into a multiple signal-to-noise paradigm (23- to −1-dB in 3-dB decrements) from which the 50% points were calculated with the Spearman-Kärber equation. *Study sample:* Sixteen young listeners with normal hearing and 48 older listeners with pure-tone hearing losses participated. *Results:* The listeners with normal hearing performed better than the listeners with hearing loss on both the HP and LP conditions. For both groups of listeners, (1) performance on the HP sentences was better than on the LP sentences, and (2) the mean 50% points were 0.1 to 0.4 dB lower (better) on the speech-variable, babble-fixed condition than on the speech-fixed, babble-variable condition. *Conclusions:* For practical purposes the ≤ 0.4-dB differences are not considered noteworthy as the differences are smaller than the decibel value of one word on the test (0.6 dB).

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

This technical report addresses the following question: Is speech-recognition performance in a masking paradigm the same when the level of the speech is fixed and the level of the noise varied as it is when the level of the speech is varied and the level of the noise fixed? Intuitively, within certain limits, *viz*. on the linear segment of the masking function (Hawkins & Stevens, 1950) the answer is *yes*. Although as discussed below there are previous reports in the literature that address this very issue using non-speech, detection paradigms, skeptics remain when the signals involved are speech and noise. If one is generating a psychometric function for which recognition performance is the dependent variable and signal-to-noise ratio (SNR)[1] is the independent variable, then functionally a series of performance measures is obtained at multiple SNRs. The measures made at each SNR are independent. The intuitive part is that as long as the measurement is on the linear segment of the masking function, the auditory system and hence the listener is oblivious to the method used to vary the SNR (i.e. which of the two signals is fixed and which is varied). This generalization for the linear segment of the masking function does not necessarily apply to the extremes of the masking function at which the function becomes non-linear, owing to audibility issues at the lower end of the function and to distortion at the higher end of the function (e.g. Studebaker et al, 1999).

Yost and Soderquist (1981) used a 30-ms, 1000-Hz signal to establish forward masking functions with 500-ms tonal maskers at 800, 1000, and 1100 Hz. In one condition the levels of the signal were fixed and the levels of the masker varied, whereas in the second condition the levels of the masker were fixed and the levels of the signal varied. An adaptive technique (Levitt, 1971) was used in 2-dB steps with two listeners. The results indicated that the same results were obtained when the signal was fixed and the masker varied and when the masker was fixed and the signal varied. Subsequently, a similar forward masking experiment with four participants was reported by Weber (1986) in which the signal was a 10-ms, 1000-Hz signal and the maskers were 60-Hz, narrow-band maskers at 600, 800, 1000, 1150, and 1250 Hz. In one condition, the signal level was fixed and the masker level varied, whereas in the second condition the manipulation of the variables was just reversed. The main result was that the signal-fixed and the masker-fixed conditions produced equivalent thresholds. The second finding was that the functions for the 1000 Hz masker was linear throughout the range of masker levels (20- to 80-dB SPL), whereas the remaining four sets of functions were non-linear at the lowest masker levels (20- to 40-dB SPL).

To the best of our knowledge, there are no published speech-recognition data that have examined the effects of which variable is fixed in level and which variable is varied in level. One unpublished report

Correspondence: Richard H. Wilson, VA Medical Center, Audiology (126), Mountain Home, TN 37684, USA. E-mail: richard.wilson2@va.gov

(Received 25 January 2012; accepted 9 April 2012)

ISSN 1499-2027 print/ISSN 1708-8186 online © 2012 British Society of Audiology, International Society of Audiology, and Nordic Audiological Society
DOI: 10.3109/14992027.2012.684407

Abbreviations

ANSI	American National Standards Institute
CD	Compact disc (audio)
HFPTA	High-frequency, pure-tone average (1000, 2000, and 4000 Hz)
HP	High predictability
LP	Low predictability
PTA	Pure-tone average (500, 1000, and 2000 Hz)
R-SPIN	Revised speech perception in noise test
S/N, SNR	Signal-to-noise ratio
SPL	Sound-pressure level

(Trine, 1995) did examine this question in two experiments using the speech intelligibility rating (SIR) test materials (Cox & McDaniel, 1989) presented in broadband noise that was steady state (continuous) and interrupted at rates of 2, 4, 8, 16, 32, 64, and 128 Hz. The SIR is a series of segments read by a male speaker from an encyclopedia for children. The task of the listeners with normal hearing was to adjust the level of the masker or speech sample until they could "just understand 50%" of the SIR message. Multiple estimates of the 50% points were obtained for each condition. In Experiment 1 with 10 listeners, the SIR message was fixed at 60-dB SPL and the level of the masker varied. In Experiment 1b with five listeners who were new to the study, the level of the masker was fixed at 85-dB SPL and the level of the SIR message varied. For each of the eight masker conditions there was little difference between performances on the two experiments. The differences of the 50% points ranged from − 2.7 dB (8 Hz) to 2.9 dB (64 Hz) with overall performance 0.6 dB better when the level of the speech was varied and the level of the masker fixed. Because the number of participants in the two experiments was small and unequal and because different participants were involved in the two experiments, it is difficult to determine if the small differences are real and, if real, owing to what variable (i.e. participants, conditions, or a combination of the two).

The purpose of the experiment described herein was to determine whether or not the same speech-recognition performances were obtained: (1) when the level of the speech was fixed and the level of the noise varied, and (2) when the level of the noise was fixed and the level of the speech varied. A second purpose was to compare performances on these two paradigms by listeners with normal hearing and by listeners with hearing loss for pure tones. The revised speech perception in noise (R-SPIN)[2] materials (Kalikow et al, 1977; Bilger, 1984; Bilger et al, 1984) in a multiple SNR paradigm (Wilson et al, 2012) were selected for study. The R-SPIN, which is a word-recognition test in multitalker babble that uses a sentence paradigm in which the last word in the sentence is the target word, has two types of sentences: (1) high-predictability (HP) sentences or words[3], which provide syntactic, semantic, and prosodic cues that help predict the target word, and (2) low-predictability (LP) sentences, which provide little, if any, cues in the sentence that help predict the target word. Complementary pairs of 50-sentence lists have the same 50 target words, each of which is presented in an HP and LP sentence. For example, the target word *spoon* appears in R-SPIN Lists 1 and 2, which is the first list pair. In List 1, *Stir your coffee with a* <u>*spoon*</u> is the HP sentence and in List 2, *Bob could have known about the* <u>*spoon*</u> is the LP sentence. By administering both lists of a list pair, data on 50 words under both (HP and LP) conditions are obtained. Typically when administered at an 8-dB S/N to listeners with hearing loss, recognition performance on the HP sentences is about 40% better than performance on the LP sentences (Bilger, 1984; Schum & Matthews, 1992; Humes et al, 1994).

Methods

Materials

The R-SPIN paradigm used in the current study was a modification of the traditional R-SPIN that over the course of a list pair (45 target words in both HP and LP sentences) involved the presentation of 10 words at each of nine SNRs from 23 to − 1 dB (Wilson et al, 2012). (Note: because nine SNRs were involved, only 45 of the 50 sentences of each R-SPIN list were used.) From the masking functions generated from each listener, the 50% points were calculated with the Spearman-Kärber equation (Finney, 1952) and the morphology of the functions examined. For this experiment, R-SPIN Lists 3 and 4 were used. For each list of the list pair at each SNR, two HP and three LP words or three HP and two LP words were administered. Collectively over the two lists, at each SNR the five HP words and the five LP words were the same, only the sentences were different. Four randomizations of each of the two lists were prepared. For two of the randomizations, the level of the multitalker babble was fixed and the level of the sentences was varied to one of the nine SNRs, whereas for the remaining two randomizations, the level of the sentences was fixed and the level of the multitalker babble segments was varied to one of the nine SNRs. When the lists were compiled, the carrier phrases (*Number* __) that had accompanied each R-SPIN sentence were removed. Each list administration took about five minutes.

Subjects

Two groups of listeners participated. The first group included 16 young adults (mean = 23.5 years, SD = 2.4 years) with normal pure-tone thresholds (≤ 20-dB HL, ANSI, 2004). The PTA (500, 1000, and 2000 Hz) was 5.0-dB HL (SD = 5.2 dB) and the HFPTA (1000, 2000, and 4000 Hz) was 4.4-dB HL (SD = 5.1 dB). The second group was composed of 48 older listeners with hearing loss (mean age = 68.1 years; SD = 7.9 years). For the group with hearing loss, the PTA was 32.2-dB HL (SD = 6.6 dB) and the HFPTA was

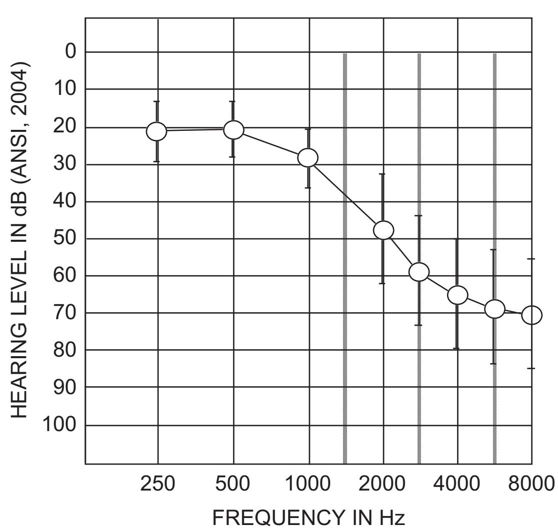

Figure 1. The mean audiogram of the test ear of the 48 listeners with hearing loss who participated in the experiment. The vertical bars represent ±1 standard deviation.

Table 1. The percent correct recognition (and standard deviations) for the various combinations of conditions at the nine SNRs (dB). The 50% points calculated from the individual data with the Spearman-Kärber equation and from the polynomial equation used to describe the data are listed along with the slopes of the functions at the 50% points.

	Listeners with normal hearing				Listeners with hearing loss			
	Speech fixed		Babble fixed		Speech fixed		Babble fixed	
Condition and dB S/N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
High predictability (HP)								
23	100.0	0.0	100.0	0.0	98.3	6.9	100.0	0.0
20	100.0	0.0	100.0	0.0	99.2	4.0	99.6	2.9
17	100.0	0.0	100.0	0.0	98.3	5.6	97.5	6.7
14	100.0	0.0	100.0	0.0	96.7	8.6	95.8	10.1
11	100.0	0.0	97.5	6.8	93.8	13.1	92.9	11.3
8	97.5	6.8	98.8	5.0	73.8	28.9	80.8	26.1
5	97.5	6.8	95.0	8.9	71.3	28.0	74.2	26.4
2	87.5	16.1	91.3	12.6	29.6	30.0	30.8	29.7
− 1	45.0	18.6	58.8	18.6	4.2	11.6	5.8	14.3
SK 50% (dB S/N)	− 0.3	0.9	− 0.7	0.9	4.6	3.0	4.2	2.8
Polynomial								
50% Point (dB S/N)	− 0.8		− 1.6		3.4		3.0	
Slope @ 50% (%/dB)	19.7		17.1		8.5		9.1	
Low predictability (LP)								
23	98.8	5.0	100.0	0.0	92.9	15.2	93.3	13.3
20	100.0	0.0	98.8	5.0	88.3	22.2	87.9	19.2
17	95.0	8.9	93.8	9.6	78.3	20.6	80.8	19.8
14	96.3	8.1	96.3	10.9	77.1	24.8	80.4	24.6
11	90.0	16.3	88.8	12.6	60.4	32.4	59.6	33.4
8	80.0	14.6	82.5	16.1	37.5	29.4	36.3	29.9
5	85.0	17.1	86.3	15.9	28.3	27.0	27.5	27.2
2	48.8	27.3	52.5	25.2	7.9	15.3	10.0	16.5
− 1	26.3	12.0	28.8	14.5	1.7	5.6	2.1	6.2
SK 50% (dB S/N)	2.9	1.8	2.7	2.0	10.3	4.2	10.2	3.8
Polynomial								
50% point (dB S/N)	1.5		1.1		9.6		9.5	
Slope @ 50% (%/dB)	8.5		8.9		5.3		5.5	

46.9-dB HL (SD = 8.5 dB). The mean audiogram for the listeners with hearing loss is shown in Figure 1.

Procedures

For the *babble-fixed* condition the level of the multitalker babble was fixed at 80-dB SPL and the sentences were delivered at nine SNRs from 23 to − 1 dB in 3-dB decrements, which corresponded to sentence levels of 103- to 79-dB SPL. For the *sentence-fixed* condition, the level of the sentences was fixed at 90-dB SPL and the babble was delivered at nine SNRs from 23 to − 1 dB in 3-dB decrements, which corresponded to babble levels of 67- to 91-dB SPL. When the babble was fixed at 80-dB SPL, the average level of the sentences was 91-dB SPL across the nine SNRs. When the sentences were fixed at 90-dB SPL, the average level of the babble was 79-dB SPL. The most comparable SNR between the babble-fixed and sentence-fixed conditions was 11 dB at which the sentence and babble levels were within 1 dB.

Each listener was presented different randomizations of List 3 and of List 4 for the babble-fixed and sentence-fixed conditions. The order of the two conditions was alternated and counterbalanced among the listeners. The materials were reproduced by a CD player (Sony, Model CDP-CE375) and fed through an audiometer (Grason-Stadler, Model 61) to an insert earphone (ER-3A). Testing was monaural with the right ears tested on the even numbered listeners and the left ears tested on the odd numbered listeners. All testing was conducted in a sound booth with the verbal responses of the listener recorded by the experimenter into a spreadsheet. Data collection took < 1 hour.

Results and Discussion

The mean 50% points calculated from the individual data with the Spearman-Kärber equation (and standard deviations) are listed in Table 1 along with the mean percent correct recognition performances at each SNR. Consider first the mean 50% point data (SK 50%) for the HP conditions in the upper panel of Table 1. For both groups of listeners on the HP materials, the mean 50% points were 0.4 dB lower (better) on the babble-fixed condition than on the sentence-fixed condition (− 0.3 and − 0.7 dB S/N, listeners with normal hearing; 4.6 and 4.2 dB S/N, listeners with hearing loss). For the LP materials these differences were in the same direction but reduced to 0.2 dB (listeners with normal hearing) and 0.1 dB (listeners with hearing loss). The 50% points for the sentence-fixed and babble-fixed conditions within each hearing group and for each

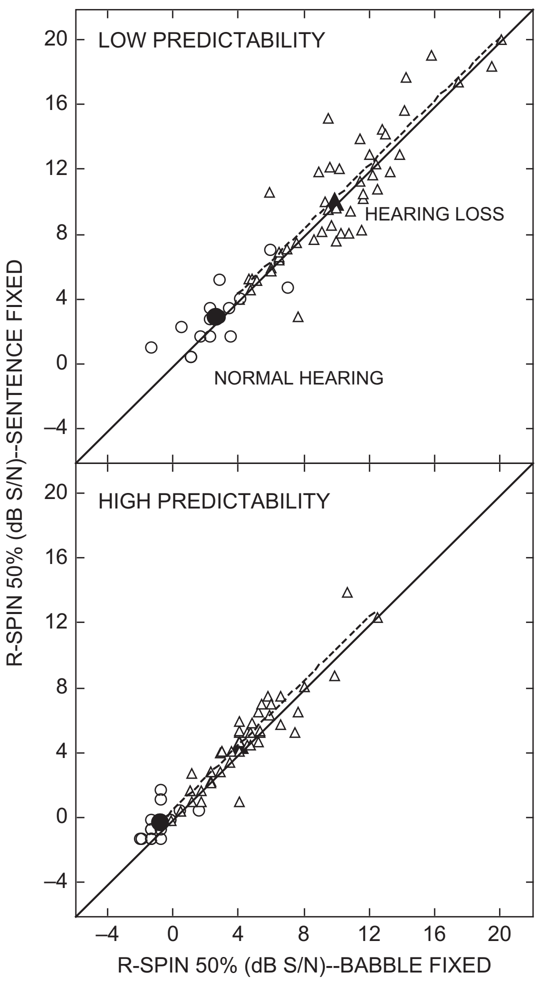

Figure 2. Bivariate plots of the 50% points on the R-SPIN with the babble level fixed (abscissa) and the speech level fixed (ordinate) are illustrated. The data for the 16 listeners with normal hearing (circles) and for the 48 listeners with hearing loss (triangles) are shown. The diagonal solid line represents equal performance and the dashed line is the linear regressions used to describe the data for the listeners with hearing loss.

predictability condition (HP and LP) were examined using paired t-tests. The results indicated that the slight performance improvement seen when the level of the babble was fixed and the level of the sentences varied for the HP words (0.4 dB) was a significant difference for the listeners with hearing loss [$t(47) = 2.4$, $p < .05$], but not for the listeners with normal hearing [$t(15) = 1.9$, $p > .05$]. The number of participants in the analysis was much smaller for the listeners with normal hearing ($n = 16$) than for the listeners with hearing loss ($n = 48$), thus decreasing the power of the statistical analysis. In essence, the 0.4 mean differences between conditions for both groups of listeners with the HP materials were irrelevant and unobtainable when considering an individual listener. With the presentation paradigm used that involved a 3-dB step size and five words at each SNR, the Spearman-Kärber equation values each word as "worth" 0.6 dB (3 dB / 5 words). Thus, a difference of 0.4 dB is less than the value of a single token and is not considered important clinically. The mean 50% point data for the LP conditions in the lower panel of Table 1 (SK 50%) indicate 0.1 to 0.2 dB differences between performances on the sentence-fixed and babble-fixed conditions, neither of which was significant.

The similarities between the sentence-fixed and babble-fixed conditions are emphasized when the individual data are examined. In Figure 2, the 50% points for the individual listeners with normal hearing (circles) and with hearing loss (triangles) are presented as bivariate plots with the data for the sentence-fixed condition on the ordinate and the data for the babble-fixed condition on the abscissa. The solid diagonal lines represent equal performance and the dashed lines are the linear regressions used to fit the data for the listeners with hearing loss. The slopes of the two linear functions were both 0.99 dB/dB, which indicates a one-to-one relation between performances on the babble-fixed and sentence-fixed conditions. Defining equal performance as ± 0.6 dB (± 1 token), collectively on the HP task, 42 of the 64 listeners (65.6%) had equal performances on the sentence-fixed and babble-fixed conditions with 17 (26.6%) having better performance on the babble-fixed condition and 5 (7.8%) having better performance on the sentence-fixed condition. On the LP task, 33 of the 64 listeners (51.6%) had equal performances on the sentence-fixed and babble-fixed conditions, with the remaining 31 listeners almost evenly divided between better performance on the sentence-fixed condition (21.9%) and on the babble-fixed condition (26.6%).

The mean percent correct recognition data from the two groups of listeners are shown in Figure 3 and are listed along with the standard deviations in Table 1. The data for the babble-fixed conditions are depicted with filled symbols and the data for the sentence-fixed conditions are shown with the open symbols. Several relations are of note. First, recognition performances at the 50% points on the functions by the listeners with normal hearing were substantially better than performances by the listeners with hearing loss with a 4.4-dB difference on the HP sentences and an 8.3-dB difference on the LP sentences. Second, the differences between the functions for the babble-fixed and sentence-fixed conditions are practically nil. Overall on the dynamic portions of the psychometric functions in Figure 3 and Table 1, when the babble level was fixed and the sentence level varied, recognition performances were slightly better than when the sentence level was fixed and the babble level varied. For the listeners with normal hearing these differences were 5.0% (HP) and 1.8% (LP), and for the listeners with hearing loss these differences were 2.4% (HP) and 0.6% (LP). In terms of decibels, the differences are tempered somewhat when consideration is given to the slopes of the respective functions. For example, the 5.0% difference for the HP condition with the listeners with normal hearing is only about 0.3 dB when the ~18%/dB slope is considered (Table 1). The largest single difference at any of the SNRs was 13.8% for the listeners with normal hearing in the HP condition at − 1dB S/N, which translates to a 0.8 dB difference. For the remaining three comparisons (LP, listeners with normal hearing; HP and LP, listeners with hearing loss), the differences between performances with the sentence level fixed and the babble level fixed are substantially less, both in terms of percent correct and decibels. The slopes of the functions for the paired conditions in Figure 3 and Table 1 are in very good agreement, ranging from 17.1 to 19.7 %/dB for the listeners with normal hearing in the HP condition to 5.3 to 5.5 %/dB for the listeners with hearing loss in the LP condition.

Finally from Table 1, consider the recognition performances obtained at 8-dB S/N, which is the SNR typically used with the R-SPIN protocol. At 8-dB S/N, the differences between performances on the HP and LP materials were 16.9% (17.5% sentence fixed; 16.3% babble fixed) for the listeners with normal hearing, and 40.4% (36.3% sentence fixed; 44.5% babble fixed) for the listeners with hearing loss. The 16.9% difference was minimized because of

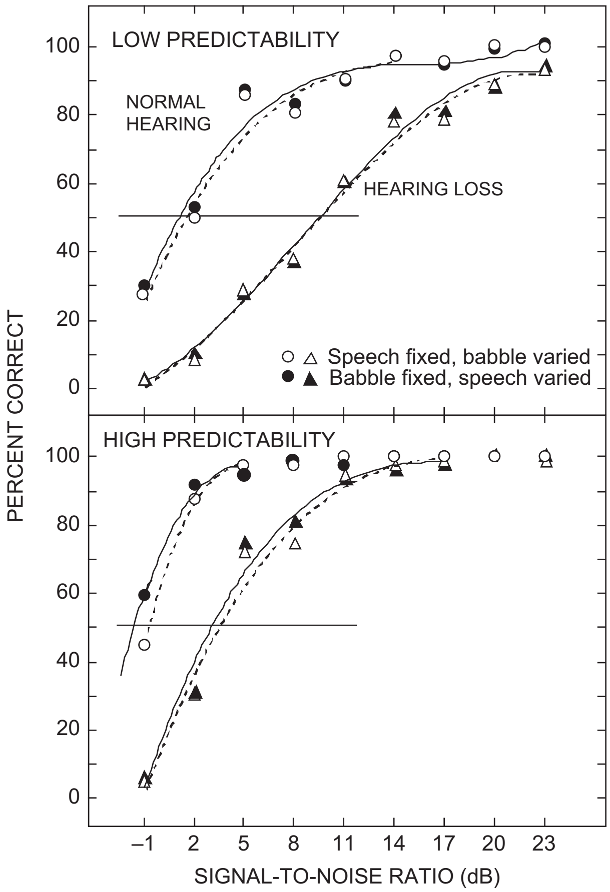

Figure 3. The mean percent correct recognition for the low-predictability and high-predictability R-SPIN sentences from List Pair 3 and 4 are depicted. The data from the 16 listeners with normal hearing are shown with circles and from the 48 listeners with hearing loss are shown with triangles. The filled symbols are the data for the babble-level-fixed condition and the open symbols are the data for the sentence-level-fixed condition. The lines through the datum points are the best-fit, third-degree polynomials used to describe the data.

ceiling effects caused by the convergence of the two functions in the range of maximum performance. The 40.4% difference is almost identical to the 41.3% HP, LP difference reported for 128 listeners with sensorineural hearing loss by Bilger (1984, Table 1), to the 41.5% HP, LP difference reported for 98 listeners with hearing loss by Schum and Matthews (1992), and to the 41.1% HP, LP difference observed at two presentation levels on 50 listeners with sensorineural hearing loss by Humes et al (1994), thereby providing concurrent validity to the current study.

Conclusions

The relations among the data and distributions of the data in the current study reinforce the equality of recognition performances that were obtained on the sentence-fixed and babble-fixed conditions. The current data agree with the findings reported earlier with non-speech stimuli by Yost and Soderquist (1981) and by Weber (1986) and with speech stimuli by Trine (1995) that indicate on the linear segment of the masking function, equivalent results are obtained with either the level of the speech fixed and the level of the noise varied or the level of the noise is varied and the level of the speech fixed.

Acknowledgements

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), a Senior Research Career Scientist Award to the first author, and a Career Development Award to the second author. The authors acknowledge the contributions made by to the project by Kelly Watts. Portions of this work were presented at the annual American Auditory Society Conference (Scottsdale, March, 2009). The contents of this paper do not represent the views of the Department of Veterans Affairs or the United States Government.

Notes

1. In this paper, signal-to-noise ratio is abbreviated two ways. When used in conjunction with a decibel value, S/N is used. When used in text, SNR is used.
2. The R-SPIN is available on CD for a nominal charge through the Department of Speech and Hearing Science, the University of Illinois, Champaign, IL 61820, USA.
3. Throughout the manuscript the terms *sentences*, *speech*, and *words* are used interchangeably within the context of the R-SPIN.

Declaration of interest: The authors report no conflicts of interest.

References

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

Bilger R.C., Nuetzel J.M., Rabinowitz W.M. & Rzeczkowski C. 1984. Standardization of a test of speech perception in noise. *J Speech Hear Res*, 27, 32–48.

Bilger R.C. 1984. Speech recognition test development. In: E. Elkins (ed.), *Speech Recognition by the Hearing Impaired*. *ASHA Reports*, 14, 2–15.

Cox R.M. & McDaniel D.M. 1989. Development of the speech intelligibility rating (SIR) test for hearing aid comparisons. *J Speech Hear Res*, 32, 347–352.

Finney D.J. *Statistical Method in Biological Assay*. London: C. Griffen; 1952.

Hawkins J.E. & Stevens S.S. 1950. The masking of pure tones and of speech by white noise. *J Acoust Soc Am*, 22, 6–13.

Humes L.E., Watson B.U., Christensen L.A., Cokely C.G., Halling D.C. & Lee L. 1994. Factors associated with individual differences in clinical measures of speech recognition among the elderly. *J Speech Hear Res*, 37, 465–474.

Kalikow D.N., Stevens K.N. & Elliott L.L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am*, 61, 1337–1351.

Levitt H. 1971. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49, 467–477.

Schum D.J. & Matthews L.J. 1992. SPIN test performance of elderly hearing-impaired listeners. *J Am Acad Audiol*, 3, 303–307.

Studebaker G.A., Sherbecoe R.L., McDaniel D.M. & Gwaltney C.A. 1999. Monosyllabic word recognition at higher-than-normal speech and noise levels. *J Acoust Soc Am*, 105, 2431–2444.

Trine T.D. 1995. *Speech recognition in modulated noise and temporal resolution: Effects of listening bandwidth*. Unpublished dissertation. University of Minnesota.

Weber D.L. 1986. Procedure-independent growth-of-masking functions. *J Acoust Soc Am*, 79, 2082–2084.

Wilson R.H., McArdle R., Watts K.L. & Smith S.L. 2012. The revised speech perception in noise test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *J Am Acad Audiol*, (in press).

Yost W.A. & Soderquist D.R. 1981. Varying the masker or the signal level in forward masking: A procedural comparison. *Acoustics Letters*, 4, 219–223.