

Northwestern University Auditory Test No. 6 in multi-talker babble: A preliminary report

Richard H. Wilson, PhD, and Anne Strouse, PhD

James H. Quillen VA Medical Center Mountain Home, TN; and Departments of Surgery and Communicative Disorders, East Tennessee State University, Johnson City, TN

Abstract—The purpose of this project was to develop a spoken word-recognition task that could be used clinically to evaluate recognition performance of individuals with hearing loss in a background noise. The test instrument incorporated monosyllabic words at seven levels over a 35-dB range presented in a background of “multi-talker” babble that was fixed in level. In Experiment 1, we established normative data on 24 young adult listeners with normal hearing and on 50 older adult listeners with high-frequency hearing loss. In Experiment 2, we examined the effects that age and hearing loss have on understanding speech in multi-talker babble by studying 15 subjects in each decade interval from 20 to 79 years.

Key words: *aging, auditory perception, compact disc, speech perception.*

This material is based upon work supported by Rehabilitation, Research and Development Service and by Medical Research Service, Department of Veterans Affairs. The first author is supported by a Senior Research Career Scientist award, and the second author is supported by an Advanced Research Career Development award. The Rehabilitation, Research and Development Service, Department of Veterans Affairs, sponsors both awards.

Address all correspondence and requests for reprints to Richard H. Wilson, PhD, VA Medical Center, Audiology (126), Mountain Home, TN 37684; 423-926-1171, ext. 7553; fax: 423-979-3403; email: richard.wilson2@med.va.gov.

INTRODUCTION

One of the most common complaints that individuals with hearing loss have is that they can hear speech (sensitivity), but they cannot understand speech (acuity), especially when the acoustic environment has some type of competing background noise. In the evaluation of spoken word-recognition abilities, most audiologists present the speech materials (typically monosyllabic words) in quiet and do not address the listener’s complaint of not being able to understand speech in a noisy environment (1). Speech-in-noise data, in addition to addressing the complaint of the patient, should be useful in (1) selecting the appropriate amplification strategy, (2) determining patient expectations with hearing aids and/or assistive listening devices, and (3) defining subjective outcome measures.

Speech recognition in noise is not routinely evaluated by audiologists, mainly because no standardized traditional word-recognition materials are available in noise. The standardization issue has been complicated by the variety of competing noises that have been studied. Several sentence materials in competing noise paradigms have been developed in research laboratories, including the Connected Speech Test (CST), the Speech-in-Noise (SIN) Test, the Speech Perception in Noise (SPIN), and the Hearing in Noise Test (HINT) (2–6). These sentence materials have not been incorporated into clinical audiology practice for several reasons. First, audiologists

historically and traditionally are schooled that monosyllabic words are the material of choice for evaluating the ability of patients to understand speech. Second, this schooling stresses the evaluation of word recognition in a quiet environment and minimizes the evaluation of how patients do or do not understand speech in a noisy environment. Third, in addition to being unfamiliar to audiologists, sentence materials are often lengthy to administer, involve psychometric procedures not commonly used clinically (e.g., adaptive techniques), and because of the shadowing response technique, are difficult for many older patients to perform.

The purpose of this report is to describe the development of a word-recognition test instrument intended for clinical use to quantify the ability of listeners with hearing loss to understand speech in a noisy background environment. For the instrument to be clinically viable, several prerequisites were involved in the design, including the use of traditional word stimuli and the length of the test procedure. Additionally, the instrument was required to evaluate recognition performance at multiple presentation levels and to provide a quantity (score) that was easy to compute and easy to interpret.

To meet these criteria, the test instrument evolved with the following characteristics:

1. The Northwestern University Auditory Test No. 6 (N.U. No. 6) monosyllabic materials recorded by the VA (Veterans Affairs) female speaker (7) were selected because of N.U. No. 6's sensitivity to the variety of word-recognition performances that individuals with hearing loss exhibited. N.U. No. 6's clinical use is widespread and is familiar to audiologists.

2. We selected a multi-talker babble as the competing background noise because "multi-talker" babble is the most common noise environment that listeners encounter in everyday life.

3. To mimic the real world in which background noises are maintained at fairly constant levels, we fixed the level of the multi-talker babble and varied the level of the speech signal.

4. Because the presentation of speech in multi-talker babble creates a complex acoustic environment, a set of practice materials was incorporated into the instrument that served to familiarize the listener with the acoustic and/or listening environment of the test paradigm and with the response task.

5. To use multiple speech presentation levels, the instrument evolved as one in which ten words were pre-

sented at each of seven levels with a quasi-randomized design. Using 70 words satisfied the criteria that the instrument must be time-efficient.

6. The instrument design was amenable to quantification by the Spearman-Kärber method that is a simple metric yielding an estimate of the 50 percent correct point in terms of the decibel signal-to-babble (S/B) ratio (8). The Spearman-Kärber method, which assumes data points above and below 50 percent correct (preferably 100 to 0 percent correct), has been shown to produce threshold estimates comparable to the threshold estimates calculated from data fit with orthogonal polynomials (9).

We performed two experiments. Experiment 1 involved the development of the materials and normalization on 24 young adults with normal hearing. Additionally, we studied 50 older adults with sensorineural hearing loss in a first step to determine the effects that mild-to-moderate hearing loss has on recognition performance in multi-talker babble (10–12). Experiment 2 involved a more detailed examination of the effects that age and hearing loss have on understanding speech in the multi-talker babble paradigm. In this experiment, we studied 15 subjects in each decade interval from 20 to 79 years (13).

METHOD

Materials

The 100 words from Lists 3 and 4 of N.U. No. 6 (VA female speaker) were interleaved into a list with 30 words arbitrarily designated practice items and 70 words arbitrarily designated test items. The 30 practice words were always used for practice, and the 70 test words were always used as test items. Based on pilot data, each word was assigned one presentation level and was always presented at that level. The 30 practice words were divided into six blocks of five words with each word in a block assigned a level between 0 and 30 dB in 5-dB steps. The 70 test words were divided into seven blocks of ten words with each word in a block assigned a level between 0 and 35 dB in 5-dB steps. Because of the minimal responses expected, the lowest level in the test sequence was not included in the practice sequence. The levels were accomplished digitally with in-house routines. Each block contained one word at each of the six practice levels or seven test levels. The recordings were made in quiet (Channel 1) and in a multi-talker babble (Channel 2). The babble, which Causey (14) recorded,

consisted of three female and three male speakers talking about various topics (15). Each block of six or seven words was time-locked to a segment of the babble, regardless of the signal-to-babble ratio. We then shuffled the segments to form randomizations of the six practice and seven test blocks. In this manner, each word maintained its temporal location in the multi-talker babble. The babble was mixed with the words digitally and recorded at a level that produced signal-to-babble ratios ranging from -5 to 20 dB (practice) and -10 to 20 dB (test) in 5 -dB steps. We constructed two randomizations of the 70 test words. For Experiment 1, the 70-word lists were recorded on DAT (Sony, Model 2500A,B). For Experiment 2, each of the 70-word lists was recorded on a compact disc (Pinnacle Micro, Model RCD-1000) as two 35-word lists, which enabled examination of abbreviated versions of the test instrument.

Procedures

The two experiments had commonalities. For the speech in multi-talker babble, we studied signal-to-babble ratios between -10 and 20 dB with the level of the babble fixed at 50 -dB hearing level (HL) (70 -dB sound-pressure level (SPL)). Each listener practiced on the listening task (30 words) before the test items were administered. The materials were presented from either a DAT (Sony, Model DTC-59ES) or a compact disc player (Sony, Model CDP-497), through an audiometer (Grason-Stadler, Model 10), to supra-aural earphones (TDH-50P). We conducted all testing in a double-wall sound booth in a 1-hour session. The listeners, who responded verbally, were paid for their participation.

Experiment 1

Psychometric functions for the words in quiet and in multi-talker babble were obtained from 24 young adults (mean age = 23.4 years) with normal hearing (≤ 20 -dB HL at 250 to 8,000 Hz) and from 50 older adults (mean age = 53.3 years) with high-frequency sensorineural hearing loss (see **Table 1**). In quiet, the practice and test functions were obtained on the young adults between 5 - and 35 -dB HL. To obtain recognition data in quiet above and below the 50 percent correct point on the older adults with hearing loss, we based the presentation levels of the quiet condition on the data obtained in the practice session, with adjustments made in 5 -dB steps to accommodate the degree of hearing loss. Of the 50 listeners, 40 listeners were evaluated with the 20 - to 50 -dB HL pre-

Table 1.

Mean pure-tone thresholds (in dB HL*) and standard deviations (SDs) (in dB) for the 50 patients with sensorineural hearing loss in Experiment 1.

	FREQUENCY (Hz)					
	250	500	1,000	2,000	4,000	8,000
Mean	18.71	15.7	16.0	22.5	51.9	56.1
SD	0.0	11.2	11.9	13.8	17.0	20.6

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

sentation range used with the listeners with normal hearing. Two subjects with the most severe hearing losses used a 40 - to 70 -dB HL range of presentation levels. The quiet and babble conditions were alternated with one ear of each listener evaluated.

Experiment 2

In this experiment, we studied 15 subjects in each of the six-decade intervals from 20 to 79 years. Pure-tone thresholds were obtained from each subject (see **Figure 1**), and word recognition was evaluated in quiet with N.U. No. 6 at 50 -dB HL (16). Following the practice set of 30 words in multi-talker babble, four lists of 35 words were administered in a random order. Both ears of each

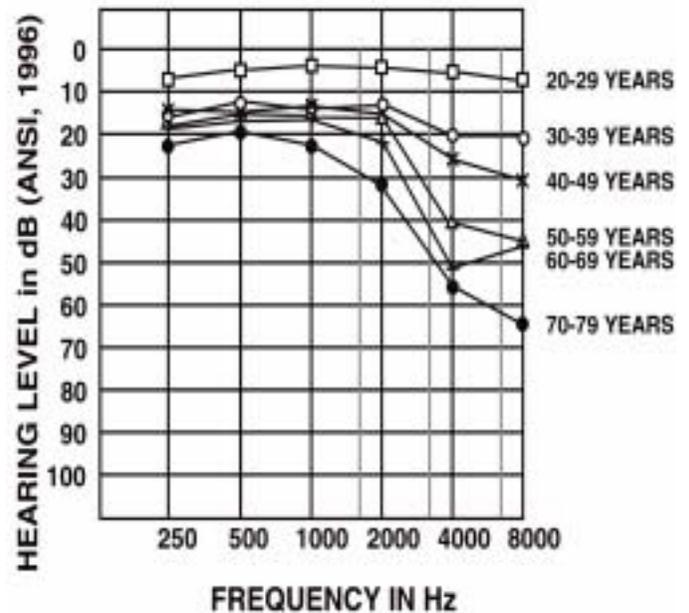


Figure 1.

Mean audiogram of six groups of 15 listeners ($n = 30$ ears). Source: American National Standards Institute (1996). Specification for audiometers (ANSI S3.6-1996). New York.

young and older listener were evaluated with the ear order alternated.

RESULTS AND DISCUSSION

Experiment 1

The mean percent correct recognition data and standard deviations (SDs) for the practice and test sessions are listed for the 24 listeners with normal hearing (**Table 2**) and for the 50 listeners with hearing loss (**Table 3**). The mean slopes of the functions between the 20 and 80 percent correct points, which were calculated from the third-degree polynomials used to fit the data, also are included in the tables. Orthogonal polynomials, which we used simply to describe the data, are convenient in that specific points on the function (e.g., the 50 percent correct point) and slopes at those points on the function can be calculated. Because of the different presentation levels used in the quiet condition with the subjects with hearing loss, we do not present those data in tabular form. Individually, the 50 percent correct points in quiet for the group with hearing loss estimated with the Spearman-Kärber equation ranged from 23-dB HL to 58-dB HL with a mean of 37.6-dB HL (SD = 8.0 dB).

Table 2.

Mean percent correct recognition (and standard deviations) for N.U. No. 6 of lists 3 and 4 combined in quiet and in babble and slopes of functions (%/dB) between 20 and 80 percent correct points. Subjects were 24 young adults with normal hearing in Experiment 1.

Level	Practice Session		Test Session	
	Mean	SD	Mean	SD
<i>Quiet (dB HL)</i>				
5	—	—	0.8	(4.1)
10	7.5	(14.2)	5.0	(8.3)
15	45.8	(18.2)	32.5	(17.8)
20	70.8	(18.2)	58.3	(19.9)
25	85.8	(19.1)	90.4	(16.3)
30	94.2	(15.0)	95.4	(6.6)
35	94.2	(9.3)	96.7	(7.0)
Slope	5.5%/dB	—	5.6%/dB	—
<i>Babble (decibel signal-to-babble ratio, re: 50-dB HL)</i>				
-10	—	—	0.4	(2.0)
-5	5.0	(13.5)	5.0	(5.1)
0	60.8	(17.2)	47.1	(14.6)
5	89.2	(14.4)	57.1	(13.3)
10	94.2	(9.3)	95.8	(5.8)
15	98.3	(5.6)	98.8	(4.5)
20	98.3	(5.6)	97.1	(4.6)
Slope	8.6%/dB	—	5.4%/dB	—

Table 3.

Mean percent correct recognition and standard deviations for N.U. No. 6 of lists 3 and 4 combined in multi-talker babble. Slopes of functions (%/dB) between 20 and 80 percent correct points also are listed. Subjects were 50 adults with sensorineural hearing loss in Experiment 1.

Level	Practice Session		Test Session	
	Mean	SD	Mean	SD
<i>Babble (decibel signal-to-babble ratio, re: 50-dB HL)</i>				
-10	—	—	0.4	2.0
-5	1.2	4.8	1.0	4.6
0	22.4	25.1	18.0	15.8
5	49.6	29.1	32.4	19.9
10	80.0	26.5	69.8	23.3
15	89.2	19.5	86.6	14.9
20	90.8	12.3	88.8	13.0
Slope	5.3%/dB	—	4.9%/dB	—

For both groups of listeners, the recognition performance on the practice items was better than the performance on the test items. This difference for the young adults with normal hearing was about 2 dB (quiet) and 3 to 4 dB (babble). For the older subjects with hearing loss, the difference in the babble condition was 2 dB. These 2 to 4 dB differences simply indicate that the practice items, which were different words from the test words, were easier than the test items. Because the practice items were always administered before the test items, possible learning effects could not be evaluated.

As one would expect, the standard deviations in **Table 2** indicate less variability in the babble condition than in the quiet condition. In effect, the multi-talker babble, which is a masking condition, equalizes the audibility differences among the subjects who are present in the quiet data. This difference in variability between the quiet and noise conditions is common in studies involving quiet and masking conditions. Finally from **Tables 2** and **3**, the slopes of the mean functions are only slightly steeper for the young listeners with normal hearing (5.5%/dB) than for the older listeners with hearing loss (4.9%/dB). This relation between the slopes of the functions for the two groups indicates that the groups have similar improvements in word-recognition ability as the presentation level is increased.

We fit the test data from each subject with third-degree polynomials from which we calculated the levels at each 10 percent correct increment. The data at each 10 percent increment from 10 to 90 percent correct then were averaged for the respective subject and listening conditions (17) and are presented in **Figure 2**. In the

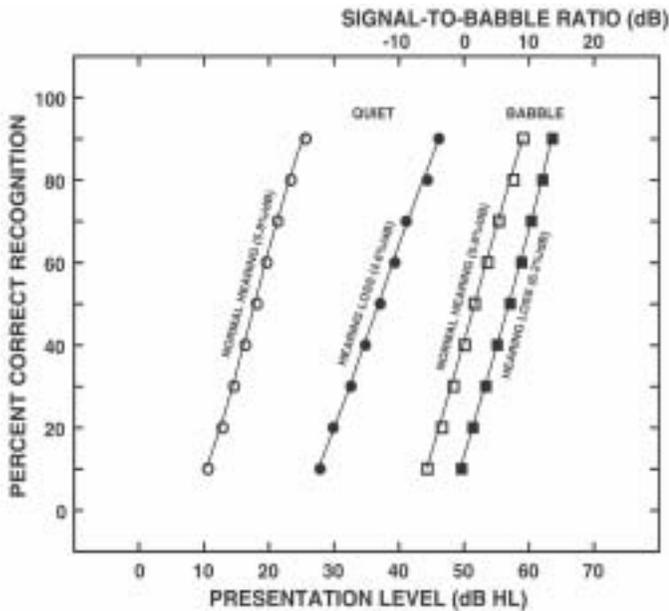


Figure 2.

Psychometric word-recognition functions in Experiment 1 for 24 young listeners with normal hearing (open symbols) and 50 older adults with high-frequency sensorineural hearing loss (filled symbols) for the quiet condition (circles) and multi-talker babble condition (squares).

quiet condition, the recognition performance of the older group with hearing loss was about 19 dB poorer than the performance of the younger group with normal hearing. Additionally, the slope of the function for the older group was about 1%/dB steeper than the function for the younger group. These performance differences between groups are for the most part attributable to differences in audibility in which the pure-tone thresholds for the older group are at levels 15 to 20 dB below (in hearing level) those for the younger group. Finally, it is of interest to examine the variability associated with the functions depicted in **Figure 2**. For the young listeners with normal hearing, the standard deviations at the 10 percent increments ranged from 3.6 to 4.1 dB in quiet and from 1.4 to 1.9 dB in babble. As one would expect, the standard deviations for the older group with hearing loss were larger than those for the young listeners, ranging from 7.9 to 8.3 dB in quiet and from 3.5 to 3.9 in babble. Thus for both groups of listeners, the babble (just as a broadband noise masker) reduced variability in comparison to the variability observed in the quiet condition.

The data in **Figure 2** for the babble condition indicate about a 5-dB difference between performances by

the two subject groups with the function for the older group slightly steeper (6.2%/dB) than the function for the younger group (5.6%/dB). In terms of masking, one would expect that the two groups' performances to be equal in the babble condition because theoretically, at least with detection, the babble masker shifts the audibility of the two groups to the same level (18). This relation did not evolve, because the older group with hearing loss required on average a 5-dB better signal-to-babble ratio to obtain the same performance as that obtained by the younger group with normal hearing. The displacement of the functions for the two subject groups can be attributed to sensitivity differences as well as to the other degradation phenomenon associated with high-frequency hearing loss and the aging process.

One design objective was a test instrument that was amenable to a simple metric to estimate the 50-percent correct point. This criterion was met with the Spearman-Kärber equation (8), which is expressed as

$$T_{50\%} = i + \frac{1}{2}(d) - (d) \left(\frac{\# \text{ words correct}}{\# \text{ words per step}} \right)$$

in which i is the highest presentation level and d is the step size. The validity of the 50-percent points established with the Spearman-Kärber equation was examined by comparing them with the 50-percent points calculated from the polynomial equations used to fit each set of data for each listener. For the 24 young adults, the mean 50-percent points for the quiet condition were 18.5-dB HL (Spearman-Kärber) and 18.2-dB HL (polynomial) with standard deviations of 2.9 dB. The mean 50-percent points for the babble condition were 2.4-dB S/B (Spearman-Kärber) and 1.8-dB S/B (polynomial), with standard deviations of 1.3 dB and 2.5 dB, respectively. Both of these differences were not significant ($p > 0.05$). For the 50 older subjects with hearing loss, **Figure 3** depicts the 50-percent points calculated with the polynomial equations on the ordinate and the corresponding 50-percent points from the Spearman-Kärber method on the abscissa. The diagonal line represents equal performance. The two methods of calculating the 50-percent correct points produced equivalent results that averaged <1 dB for both the quiet and babble conditions.

Finally, from the young adult group, we derived the 90th percentiles (quiet—21.7-dB HL, and babble—4.0-dB S/B) from the 50-percent correct points calculated with the Spearman-Kärber equation. These 90th percentiles were

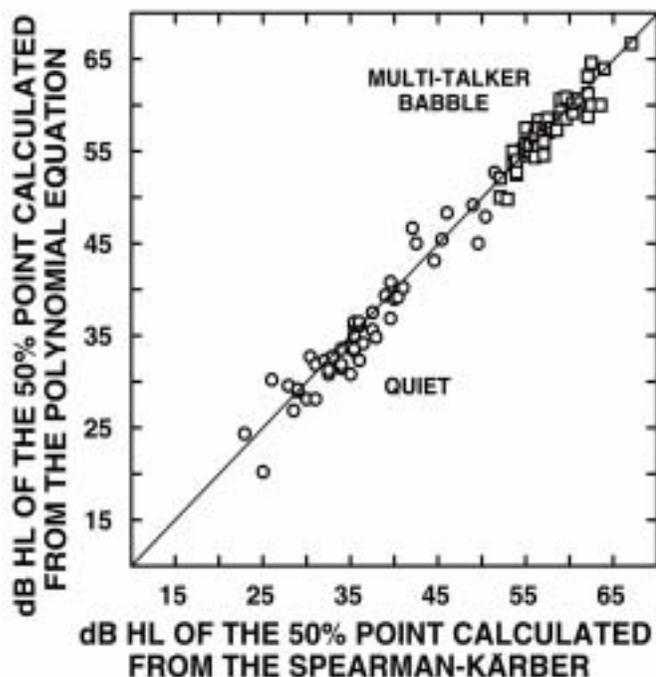


Figure 3.

A bivariate plot of hearing levels at which 50 percent correct points for 50 listeners with hearing loss in Experiment 1 were obtained from Spearman-Kärber method (abscissa) and from polynomial equation (ordinate). Data in both quiet (circles) and babble (squares) conditions are shown. Diagonal line shows equal performance.

used to define the normal ranges of recognition performances. When we applied the 90th percentile criteria to the 50 older listeners with hearing loss, all 50 subjects were beyond the 21.7-dB HL cutoff in quiet, which reflects the sensitivity differences between the two groups, and 40 subjects were beyond the 4.0-dB S/B cutoff in babble, which reflects an impaired ability to understand speech in a background of multi-talker babble.

Experiment 2

We assessed the influence of (1) the six age groups, (2) the four lists of 35 words, (3) ear, and (4) interactions of these factors using a mixed model analysis of variance (ANOVA) with word list and ear as within-subjects factors and age group as the between-subjects factor. The results of the ANOVA revealed that the main effects for word list and ear did not reach statistical significance [word list: $F(3,252) = 0.73$, $p > 0.05$; ear: $F(1,84) = 3.2$, $p > 0.05$]. Thus, the ear and word list data were combined for the remainder

of the analysis. The main effect of age group was significant [$F(5,84) = 27.0$, $p < 0.0001$]. To examine the effect of hearing loss, we used the four-frequency pure-tone average (500, 1,000, 2,000, and 4,000 Hz) for the right and left ears as a covariate in separate analyses of covariance (ANCOVA) that examined the same variables. The results of the ANCOVA [$F(5,82) = 4.0$, $p < 0.01$] were not different from the ANOVA results, indicating that the differences in performance between age groups were not owing solely to differences in hearing sensitivity.

The mean psychometric functions for the words in 50-dB HL multi-talker babble for the six age groups are presented in **Figure 4**. The descriptive lines are the best-fit linear regressions whose r^2 values ranged from 0.95 to 0.97. The regressions were fit only over the dynamic portion of the data with the use of a datum point in the 0 to 10 percent correct region as the lower anchor and when possible with the use of a >90-percent correct datum point as the upper boundary. Obviously, for the two older groups of listeners, the criterion for the upper boundary was modified. The 50-percent correct points in signal-to-babble ratio and the slopes of the functions at the 50-percent points are presented in **Table 4**. Between the 20- and 50-year groups, there is 1- to 1.5-dB/decade

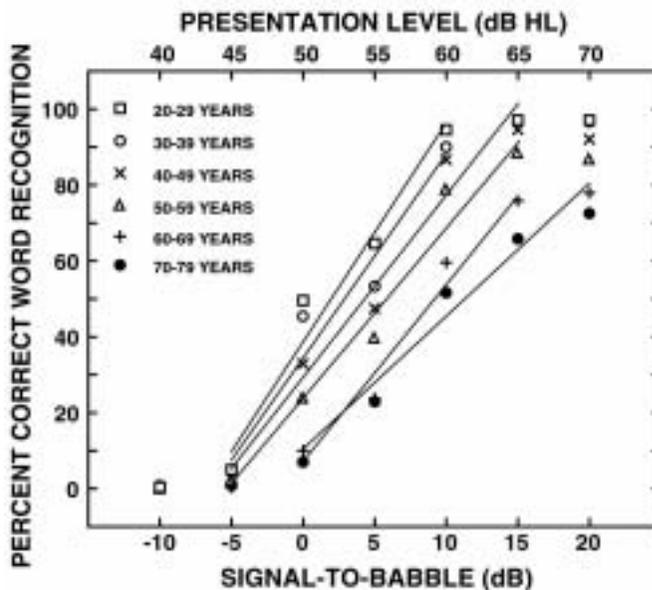


Figure 4.

Mean psychometric functions for N.U. No. 6 materials presented in 50-dB HL multi-talker babble to 6 age groups of 15 listeners ($n = 30$ ears) in Experiment 2. Lines are best-fit linear regressions whose r^2 values ranged from 0.95 to 0.97.

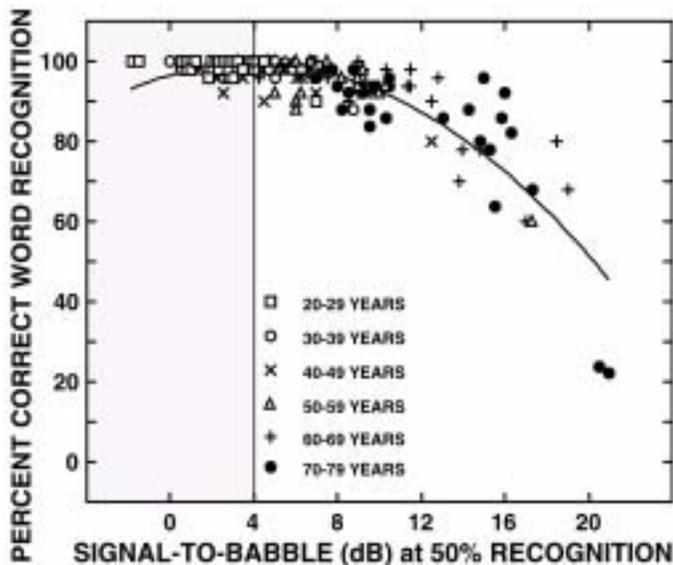
Table 4.

Signal-to-babble (S/B) ratios at which 50-percent correct points occurred on mean psychometric functions and on slopes of functions. Data are from two ears of 15 listeners in Experiment 2.

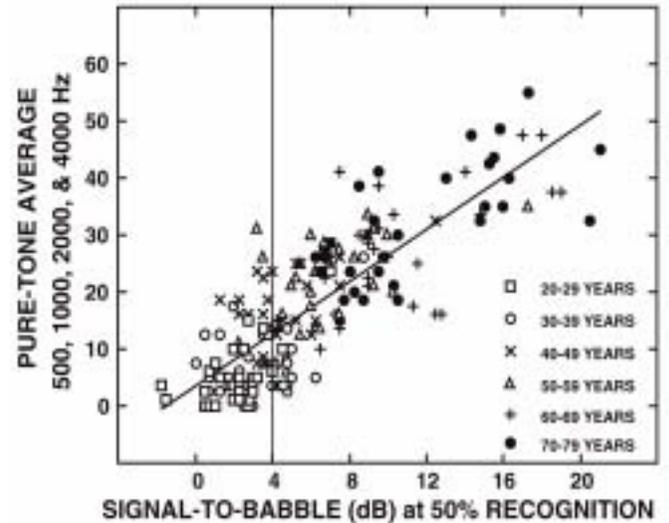
Age Group (years)	50% Correct (dB S/B)	Slope (%/dB)
20–29	1.9	5.7
30–39	2.8	5.3
40–49	4.4	4.7
50–59	5.8	4.6
60–69	9.2	4.7
70–79	11.8	3.5

decrement in the level at which we observed the 50-percent points. The slopes for the 20- through 50-year groups, however, decrease only about 1%/dB over the four decades. For the 60- and 70-year groups, performance is appreciably poorer and the slopes are substantially more gradual.

Figures 5, 6, and 7 are bivariate plots of the 50-percent correct points in signal-to-babble ratio (abscissa) and the percent correct word recognition in quiet (ordinate, Figure 5), the pure-tone average (ordinate, Figure 6), and age (ordinate, Figure 7). The shaded

**Figure 5.**

A bivariate plot of 50% correct on test materials (abscissa) and correct word-recognition performance in quiet at 50-dB HL (ordinate) from each of 15 subjects in 6 age groupings. Shaded region represents the 90th percentile of 50% point based on 20- to 29-year listeners in Experiment 1.

**Figure 6.**

A bivariate plot of the 50% correct points on test materials (abscissa) and the pure-tone average (dB HL) at 500, 1,000, 2,000, and 4,000 Hz (ordinate) from each of 15 subjects in 6 age groupings. Shaded represents 90th percentile of 50% point based on the 20- to 29-year listeners in Experiment 1.

region represents the 90th percentile for the 50-percent point derived from the 24 listeners with normal hearing in Experiment 1. The data in Figure 5 illustrate that many of the individuals with good word recognition in quiet have difficulty understanding in a background of noise. The majority of listeners in the two older groups had word-recognition performance in quiet above 80-percent correct. In the multi-talker babble, however, the recognition performance for the two older groups was reduced substantially with 50-percent correct points in the 8- to 16-dB S/B range, which is 4 to 12 dB above the performance range of the young listeners with normal hearing. The data also demonstrate a direct relation between both degrees of hearing loss (Figure 6) and age (Figure 7) and the ability to understand speech in background noise. The data in Figure 6 indicate that as the degree of hearing loss increases, the S/B ratio required for 50-percent correct recognition also increases. A similar relationship is observed as a function of age (Figure 7). As age increases, the S/B ratio required for 50 percent correct recognition also increases.

A substantial difference in performance variability for older listeners as compared to younger listeners is evident in Figures 6 and 7. In the figures, the majority of data points for the 20- through 50-year groups are clustered in the same general area on the plot. In contrast, the

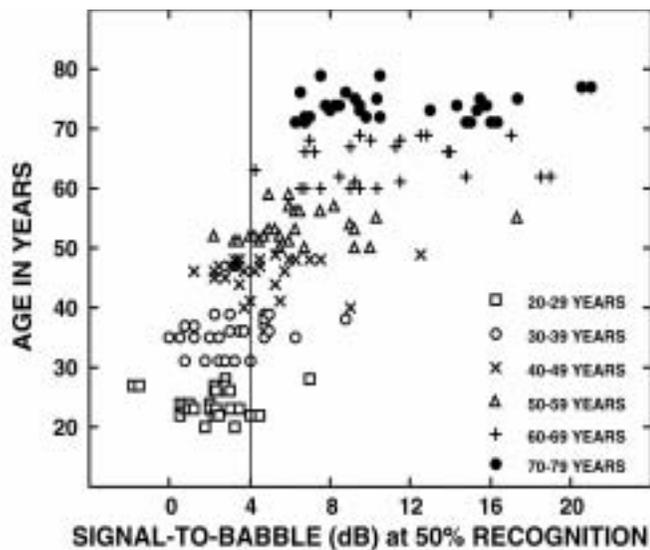


Figure 7.

A bivariate plot 50% correct points on test materials (abscissa) and age of individual subjects (ordinate) from each of six age groupings. Shaded region represents 90th percentile of the 50% point based on 20- to 29-year listeners in Experiment 1.

data points for the 60- and 70-year groups are scattered widely and indicate a less consistent pattern of performance. These differences in performances for the older groups reflect the extreme variability in word-recognition performance in older subjects with hearing loss.

CONCLUSIONS

The speech recognition in multi-talker babble paradigm described in this report demonstrates its utility in the auditory evaluation of individuals with hearing loss. The current data indicate that the speech in multi-talker babble paradigm provides a quick and easy procedure that can be used clinically to assess the ability of patients to understand speech in a competing message background. Our ultimate interest in the multi-talker babble paradigm is to provide the audiologist with information that can be used in hearing aid selection, fitting, verification, and counseling. The materials described in this report, which are available on audio compact disc (16), are currently undergoing refinements to make the test instrument more efficient by eliminating the lowest two S/B ratios, increasing the range of the highest S/B ratio and decreasing the step size from 5 dB to 4 dB.

REFERENCES

1. Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C. Standardization of a test of speech perception in noise. *J Speech Hear Res* 1984;27:32-48.
2. Cox RM, Alexander GC, Gilmore C. Development of the connected speech test (CST). *Ear Hear* 1987;8:119S-26S.
3. Cox RM, Alexander GC, Gilmore C, Pusakulich KM. Use of the connected speech test (CST) with hearing-impaired listeners. *Ear Hear* 1988;9:198-207.
4. Killion MC, Villchur E. Kessler was right—partly: But SIN test shows some aids improve hearing in noise. *Hear J* 1993;46:31-5.
5. Nilsson M, Soli SD, Sullivan, JA. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 1994; 95:1085-99.
6. Department of Veterans Affairs. Speech recognition and identification materials [computer program]. Version 1.1. Long Beach (CA): VA Medical Center; 1991.
7. Finney DJ. *Statistical method in biological assay*. London: C. Griffen; 1952.
8. Wilson, RH, Strouse, AL. Psychometrically equivalent spondaic words spoken by a female speaker. *J Speech Hear Lang Res* 1999;42:1336-46.
9. Noe CM, Wilson RH. Evaluation of NU 6 lists 3 and 4 in multi-talker babble for subjects with normal hearing and hearing impairment tests at 2.5 and 5 dB steps. The NIH/VA Hearing Aid Research and Development Conference; 1997 September; Bethesda (MD).
10. Norwood-Chapman L, Wilson RH, Thelin JW. Clinical evaluation of word recognition in multi-talker babble under four listening conditions. American Academy of Audiology Convention; 1997; Fort Lauderdale (FL).
11. Wilson RH, Oyler AL, Sumrall R. Psychometric functions for Northwestern University Auditory Test No. 6 in quiet and multi-talker babble. American Academy of Audiology Convention; 1996 April; Salt Lake City (UT).
12. Wilson RH, Strouse A. Word recognition in multi-talker babble. American Speech Language-Hearing Association Convention; 1999 November; San Francisco (CA).
13. Causey GD. Personal communication. 1988.
14. Sperry JL, Wiley TL, Chial MR. Word recognition performance in various background competitors. *J Am Acad Audiol* 1997;8:71-80.
15. Department of Veterans Affairs. Speech recognition and identification materials [computer program]. Version 2.0. Mountain Home (TN): VA Medical Center; 1998.
16. Wilson RH, Margolis RH. Measurements of auditory thresholds for speech stimuli. In: Konkle DF, Rintelmann WF,

editors. Principles of speech audiometry. Baltimore: University Park Press; 1983. p. 79–126.

17. Hawkins D, Stevens SS. Masking of pure tones and of speech by white noise. *J Acoust Soc Am* 1950;22:6–13.

Submitted for publication June 30, 2000. Accepted in revised form March 29, 2001.