AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION

Journal of Speech and Hearing Research, Volume 38, 234–243, February 1995

Accuracy of Speech Intelligibility Index Predictions for Noise-Masked Young Listeners With Normal Hearing and for Elderly Listeners With Hearing Impairment

Sarah E. Hargus*
Sandra Gordon-Salant
University of Maryland
College Park

This study examined whether the accuracy of Speech Intelligibility Index (SII) predictions is affected by subject age when between-groups auditory sensitivity differences are controlled. SII predictive accuracy was assessed for elderly listeners with hearing impairment (EHI) and for young noise-masked listeners with normal hearing (NMN). SII predictive accuracy was poorer for the EHI subjects than for the NMN subjects across a range of test conditions and stimuli. Speech test redundancy, speech presentation level, signal-to-babble ratio, and babble level also affected SII predictive accuracy. The results suggest that the speech recognition difficulties experienced in noise by elderly listeners do not result solely from reduced auditory sensitivity.

KEY WORDS: Speech Intelligibility Index, elderly (age effects), noise masking, speech recognition, hearing loss

The study of the speech recognition difficulties of older adults with hearing impairment has had a long history (e.g., Gaeth, 1948; Goetzinger, Proud, Dirks, & Embrey, 1961). In quiet conditions, young and elderly listeners tend to perform similarly (Dubno, Dirks, & Morgan, 1984; Dubno & Schaefer, 1992; Gordon-Salant, 1987; Helfer & Wilber, 1990). When noise or other distortion is added to the listening situation, poorer performance is often observed for older subjects than for younger subjects (CHABA, 1988; Dubno et al., 1984; Gelfand, Piper, & Silman, 1986; Helfer & Wilber, 1990; Kalikow, Stevens, & Elliott, 1977; Plomp & Mimpen, 1979). Some authors (Gelfand et al., 1986; Jerger, Jerger, Oliver, & Pirozzolo, 1989; Patterson, Nimmo-Smith, Weber, & Milroy, 1982) have postulated that this difference results from age-related auditory processing changes in the peripheral or central auditory nervous system. In contrast, recent studies (Humes & Roberts, 1990; Humes, Watson, Christensen, Cokely, Halling, & Lee, 1994; Jerger, Jerger, & Pirozzolo, 1991; van Rooij & Plomp, 1990, 1991, 1992) have suggested that much of the variance in speech recognition performance between young and elderly subjects can be explained by differences in speech signal audibility.

One means of studying the effects of speech signal audibility on speech recognition performance involves application of the Articulation Index (AI; French & Steinberg, 1947). Most studies of the AI have found that it overestimates the performance of listeners with hearing impairment (Dirks, Bell, Rossman, & Kincaid, 1986; Dubno &

*Currently affiliated with the Medical University of South Carolina.

© 1995, American Speech-Language-Hearing Association

0022-4685/95/3801-0234

Dirks, 1989; Dubno, Dirks, & Schaefer, 1989; Kamm, Dirks, & Bell, 1985; Kamm, Dirks, & Carterette, 1982; Pavlovic, 1984; Pavlovic, Studebaker, & Sherbecoe, 1986; Schum, Matthews, & Lee, 1991). This suggests that the AI is not sufficiently accurate to predict absolute speech recognition ability. The few studies employing the AI with elderly listeners have found that it is accurate in quiet situations, but overestimates performance in noisy conditions (Dubno et al., 1984; Schum et al., 1991). This suggests that in quiet, speech signal audibility is the main factor affecting speech recognition in elderly listeners, whereas in noise other factors become important.

It is often difficult to match young and elderly subject groups for signal audibility. When subjects with normal hearing are used, the elderly listeners tend to have poorer hearing than the young listeners. When subjects with hearing impairment are used, the young and elderly listeners often differ in degree, configuration, and etiology of hearing loss. One way to control for threshold differences is to use noise-masked listeners with normal hearing whose masked thresholds equal the quiet thresholds of the subjects with hearing impairment. Several authors (e.g., Dubno & Schaefer, 1992; Humes & Christopherson, 1991; Humes, Dirks, Bell, & Kincaid, 1987; Zurek & Delhorne, 1987) have found that this method is effective for equating audibility between subjects with normal hearing and subjects with hearing impairment.

Combining AI procedures with the use of noise-masked subjects cancels out audibility differences between young and elderly listeners, allowing possible age effects to be examined independently of sensitivity differences. AI calculation methods affect how precisely the model accounts for audibility differences. Most previous studies that have used the AI noted limitations with the original calculation procedures (ANSI, 1969) and modified those procedures in various ways, including use of locally developed frequency importance functions, locally developed transfer functions, and the actual speech spectrum (Dirks et al., 1986; Dubno & Dirks, 1989; Dubno et al., 1989; Kamm et al., 1982, Kamm et al., 1985; Pavlovic, 1984). In response to limitations observed with the original ANSI (1969) standard methods for calculating the AI, a new standard has been developed. This proposed standard suggests a new name for the AI: the Speech Intelligibility Index (SII).

The proposed SII procedures embody several improvements over ANSI S3.5-1969. One change is that the proposed standard provides frequency importance functions for a variety of speech materials in an Appendix, as well as one function for "average speech." The Appendix recommends that the SII be calculated using the importance function characteristic of the chosen speech material. In addition, the proposed standard instructs the user to develop transfer functions (functions depicting the relationship between calculated SII scores and actual performance) rather than providing these functions. The user is instructed to develop functions that reflect the speech materials to be used and the proficiency of the talkers and listeners that will be using the system. These changes should allow speech signal audibility to be accounted for more precisely.

In the current study, the accuracy of the SII for predicting observed speech recognition scores was compared for elderly subjects with hearing impairment and noise-masked young listeners with normal hearing. The use of noise-masked normal subjects as the comparison group for the elderly listeners allowed SII accuracy to be assessed in conditions of essentially equal signal audibility. The overall goal of the study was to determine whether or not the speech recognition difficulties of elderly listeners in noise can be explained by audibility. Because the SII predicts speech recognition scores exclusively on the basis of signal audibility, an absence of a group effect on SII predictive accuracy would support the hypothesis that decreased audibility is the primary problem for older adults listening to speech in noise. Alternatively, a finding of significantly better SII predictive accuracy for noise-masked younger subjects compared to older subjects would suggest that factors other than audibility are involved in the speech understanding difficulties of older listeners in noise.

The study employed both nonsense syllables and sentences of varying predictability to evaluate whether SII accuracy is affected by speech material redundancy. Various test conditions also were employed, to permit examination of the effects of signal level, overall babble level, and signal-to-babble (S/B) ratio on SII predictive accuracy. Although previous studies have used both AI procedures and noise-masked subjects with normal hearing to examine differences in performance between young and elderly listeners, none have used the SII or employed speech stimuli with varying redundancy.

Method

Subjects

Three subject groups participated in this study. Group 1 consisted of 6 young listeners (22–26 years) with normal hearing (YN). Group 2 consisted of 10 young listeners (22–27 years) with normal hearing who were tested using noise masking (NMN). All subjects with normal hearing had pure tone thresholds $\leq$15 dB HL re: ANSI (1989) for 250–8000 Hz. Group 3 consisted of 10 elderly listeners (65–83 years) with mild-to-moderate sensorineural hearing impairment (EHI). All EHI subjects reported difficulty understanding speech in noise. The primary etiology of hearing loss was presbycusis. The mean audiogram for the EHI subjects, compared to an average audiogram for young listeners with normal hearing, is displayed in Figure 1.

All subjects exhibited word recognition scores (Northwestern University Test #6, Tillman & Carhart, 1966) in quiet of 88% or better. Immittance measures confirmed that middle ear function was normal in each subject. The subjects were native speakers of English, and none had participated in hearing experiments before volunteering for this study.

Stimuli and Apparatus

For threshold testing, pulsed swept pure tones were presented by a Grason-Stadler Type E800-4 Bekesy Audiome-

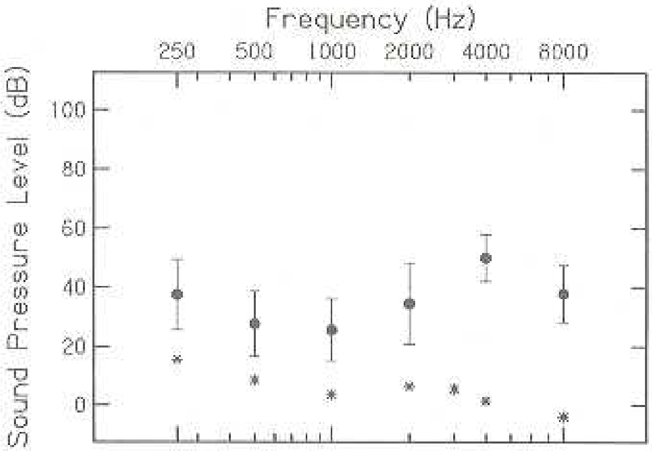

FIGURE 1. Mean thresholds at octave frequencies for the 10 EHI subjects (filled circles). The bars at each frequency represent ± 2 *SD*. Also shown are reference equivalent threshold sound pressure levels (in dB) for the Etymotic ER-3A Insert Earphone (asterisks), as presented in Appendix G of the standard specification for audiometers (ANSI S3.6-1989).

ter and routed to an insert earphone (Etymotic ER-3A). The output of the audiometer was calibrated daily.

The masker for each NMN subject was produced by a white noise generator (Colbourn S81-02), routed through both channels of a ⅓-octave equalizer (Ashly Model 3102), amplified (Crown D150A), attenuated (Hewlett-Packard 350D), and delivered to the test earphone. The equalizer settings for each NMN subject were determined following procedures similar to those described by Humes et al. (1987).

Speech stimuli were selected with a goal of sampling a range of speech material redundancy. The test featuring the least redundancy was the Nonsense Syllable Test (NST) (Resnick, Dubno, Hoffnung, & Levitt, 1975). The five consonant-vowel (CV) subtests of the NST were used. The high-predictability (PH) items of the Revised Speech Perception in Noise Test (R-SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) served as the material featuring the highest redundancy. The low-predictability (PL) items of the R-SPIN sampled an intermediate level of redundancy. The R-SPIN 12-talker babble was presented with each of the three types of speech signals during all conditions.

The speech stimuli and their respective calibration tones were low-pass filtered (5 kHz) and digitized onto a laboratory computer (10 kHz sampling rate). A waveform editing program was used to edit the stimuli. After editing, the stimuli were converted back to analog form, filtered, and recorded onto one channel of a digital audiotape (DAT) using a DAT player/recorder (Sony PCM 2500A). The 12-talker babble was recorded directly from audiocassette to the second channel of the DAT. Tapes were prepared of the 8 R-SPIN lists and of 8 versions of the NST CVs. Each R-SPIN list contained 25 PH and 25 PL items. In each version of the NST CVs, both the order of the subtests and the order of syllables within each subtest were randomized. Each version contained a total of 48 syllables to be identified.

Procedures

The output of the two channels of the DAT player/recorder was selectively amplified (Crown D-75), attenuated (HP 350D), and then mixed (Colbourn S82-24). The mixer output was amplified (Colbourn S78-03) and routed to a single insert earphone (Etymotic ER-3A). For the NMN subjects, the masker was mixed with the speech and babble before amplification and routing to the earphone. Speech, babble, and masking noise levels were calibrated daily. For all subjects, the test ear was the ear that most closely matched the hearing criteria or the right ear if the two ears were the same. In cases where there were asymmetrical audiometric thresholds, the test ear was the ear with better hearing sensitivity.

Bekesy audiometry was used to determine pure tone thresholds at the center frequencies of the 18 ⅓-octave SII calculation bands: 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600, 2000, 2500, 3150, 4000, 5000, 6300, and 8000 Hz. For the YN and EHI subjects, thresholds were measured once before speech testing. For each NMN subject, thresholds were measured first in quiet and then in the presence of the noise masker. Following adjustments to the masker, the Bekesy procedure was repeated until the NMN subject's masked thresholds were within 5 dB of the quiet thresholds of the EHI subject to whom the NMN subject was randomly assigned.

Speech testing took place over two test sessions, one test per session. For half of the subjects, the R-SPIN sentences were presented during the first session and the NST CVs were presented during the second session; for the other half of the subjects, the order was reversed. Subjects responded to the R-SPIN sentences by writing the final word of each sentence on an answer sheet. For the NST CVs, subjects responded by selecting the appropriate syllable from a set of choices on an answer sheet.

For the YN subjects, each stimulus was presented at 60 dB SPL in quiet and at 7 S/B ratios (−12, −8, −4, 0, +4, +8, and +12 dB). These conditions were selected to sample a broad range of SII and performance scores. If a subject scored 0% in a condition, then any lower S/B ratio conditions were omitted.

For the NMN and EHI subjects, each stimulus was presented at three levels to assess the effect of presentation level on predictive accuracy. Babble levels were chosen so that accuracy could be investigated for various S/B ratios, as well as with constant S/B ratios for the different signal levels. The combination of babble and speech levels resulted in six listening conditions. These conditions are displayed in Table 1.

The order of the test conditions and the assignment of test condition to stimulus list were randomized among the subjects. Performance was scored in percent correct for all materials.

Measurement of Speech and Babble Spectra

Additional tapes containing either R-SPIN key words or NST CV syllables were created for speech and babble

TABLE 1. The six listening conditions (in dB S/B ratio) resulting from combinations of speech levels and babble levels, used for NMN and EHI subjects.

Speech level (dB SPL)	Babble level (dB SPL) 52	67	82
60	+8	—	—
75	+23	+8	—
90	+38	+23	+8

Note. Minimum S/B ratio = +8 dB. — indicates condition not used in this study.

spectral measurements. The digitized speech stimuli were edited to remove the carrier sentence or phrase, and the key words or syllables were concatenated. The concatenated lists were converted back to analog form, filtered, and recorded onto one channel of the DAT. The R-SPIN 12-talker babble was recorded on the second channel. This procedure was carried out for Lists 1 and 3 of the R-SPIN and for one version of the NST.

The speech and babble were then played through the test apparatus and insert earphone to a sound-level meter with a ⅓-octave band filter via a 2-cm^3 coupler. One-third-octave band levels were measured for the key words, the CVs, and the babble at each presentation level used in the study. R-SPIN Lists 1 and 3 were measured separately, and the results for the two lists were averaged.

Calculation of the SII

The SII was calculated for each subject in each test condition using the methods detailed in Draft V3.0 of the proposed new standard. The ⅓-octave procedure was employed. For each subject and test condition, a data file was created that included speech and babble spectral information as measured in this experiment, the subject's pure tone thresholds, and the frequency importance function either for the SPIN (in the case of the two types of R-SPIN items) or for "various nonsense syllable tests" (in the case of the NST CVs). These functions were provided in an appendix of the proposed standard.

Generation of Transfer Functions

The proposed SII procedure instructs the user to develop normative transfer functions for the test conditions and speech materials of interest. To that end, normal transfer functions relating calculated SIIs to percent correct performance scores were generated for the NST CVs and for the PH and PL items of the R-SPIN using the Stata statistical package (Stata Corporation, 1993). The YN data were entered, and a regression analysis was performed. The regression equation providing the best degree of fit was selected for each speech stimulus. The resulting curves were used to convert the SIIs for the NMN and EHI subjects to predicted percent correct scores.

Results

Threshold Matching

The criterion of a 5-dB match at each frequency between the quiet thresholds of each EHI subject and the masked thresholds of the matched NMN subject was met for 3 of the 10 pairs. In each of the remaining 7 cases, no more than three bands ever exceeded the 5-dB goal. The average RMS error between matched thresholds was 3.1 dB. Figure 2 shows paired audiograms representing the best and worst threshold matches.

Transfer Functions

Normal transfer functions relating SII to performance for the R-SPIN (PH), R-SPIN (PL), and NST-CVs are displayed in Figures 3, 4, and 5, respectively. These figures include the data from the YN subjects that were used to generate these functions. For each material, the regression equation describing the curve was a third-order polynomial that accounted for at least 92% of the variance in performance.

Predictive Accuracy of the SII

SII predictive accuracy first was examined by plotting the calculated SII score against the performance score for each subject and test condition on a graph of the normal transfer

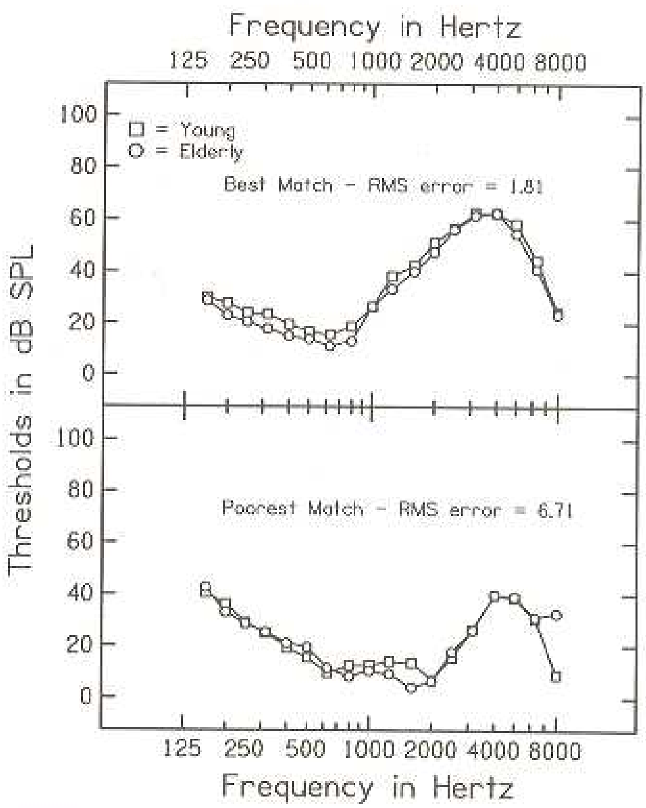

FIGURE 2. Paired audiograms representing the best (top panel) and worst (bottom panel) threshold matches between EHI and NMN subjects.

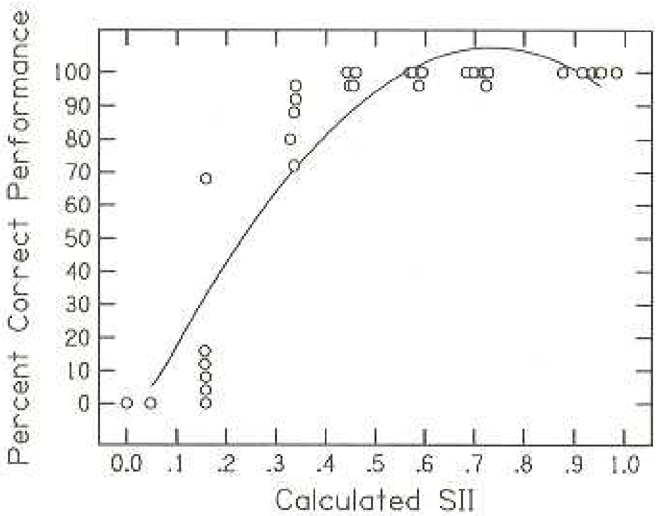

FIGURE 3. Normal transfer function relating SII to performance for the R-SPIN (PH) stimuli, including data from YN subjects.

function ± 2 *SDs*, as derived by the above-referenced statistical package. The normal functions and the data for NMN and EHI subjects are shown for the R-SPIN (PH) items, the R-SPIN (PL) items, and the NST CVs in Figures 6, 7, and 8, respectively.

Visual inspection of these figures suggests that there are differences in SII predictive accuracy among the three speech stimuli. For the R-SPIN (PH), most of the data are clustered around the normal function. The data for the R-SPIN (PL) and the NST CVs are more scattered; approximately 25 of 120 data points fall more than 2 *SD* below the normal function for each test. This suggests that the SII predicted performance more accurately for the R-SPIN (PH) than for the R-SPIN (PL) or the NST CVs.

Figures 6, 7, 8 also suggest that there are group effects on SII predictive accuracy. For each speech material, approximately ⅔ of the data points falling more than 2 *SD* from the normal function represent the performance of EHI subjects. This suggests that the SII was more accurate for the NMN subjects than for the EHI subjects.

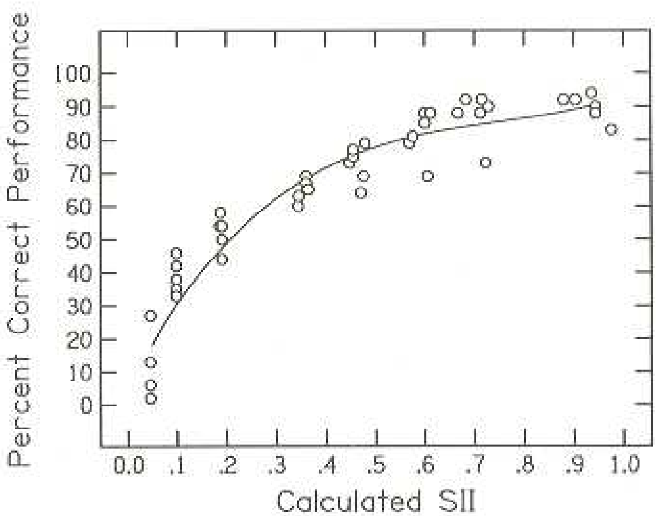

FIGURE 5. Normal transfer function relating SII to performance for the NST CV stimuli, including data from YN subjects.

SIIs were converted to predicted percent-correct scores using the transfer function curves shown in Figures 3, 4, and 5. Difference scores for each subject and condition were obtained by subtracting the predicted score from the observed score.

The absolute value of each difference score indicates the accuracy of the prediction: numbers closer to 0 indicate better accuracy. The sign of the difference score indicates the direction of the difference: a positive number signifies that the observed score was higher than the predicted score (performance was underpredicted), and a negative number shows that actual performance was lower than predicted performance (performance was overpredicted).

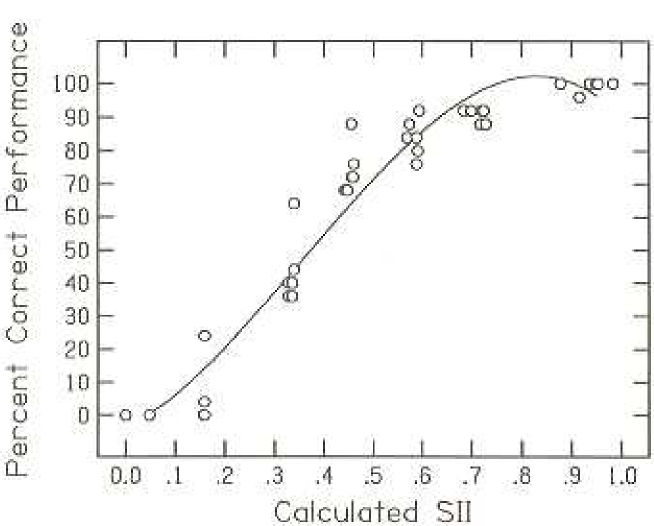

FIGURE 4. Normal transfer function relating SII to performance for the R-SPIN (PL) stimuli, including data from YN subjects.

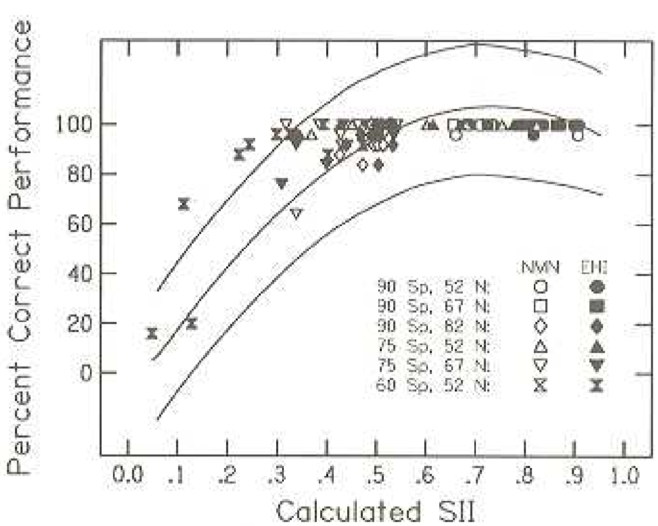

FIGURE 6. Normal transfer function ±2 *SD* for R-SPIN PH items, including data for EHI and NMN subjects.

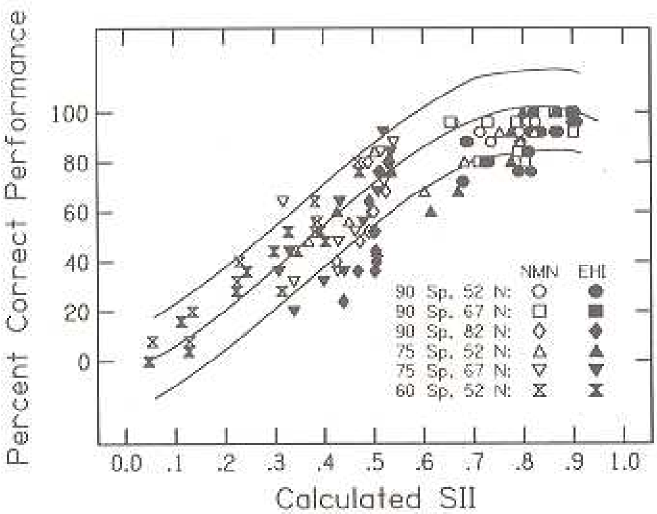

FIGURE 7. Normal transfer function ±2 *SD* for R-SPIN PL items, including data for EHI and NMN subjects.

Means and *SD*s of the difference scores in percent correct were calculated as a function of subject group, speech material, and one of three factors: (a) S/B ratio, (b) speech level, and (c) babble level. The means and *SD*s as a function of S/B ratio, speech level, and babble level are displayed in Tables 2, 3, and 4, respectively.

Individual predicted and observed scores in percent correct were then converted to rationalized arcsine units (RAUs) following procedures described by Studebaker (1991). Difference scores in RAUs were determined by subtracting the predicted score in RAUs from the observed score in RAUs. The difference scores in RAUs were analyzed in three separate multivariate analyses of variance (MANOVA) corresponding to each of the three factors of speech level, S/B ratio, and babble level. The design for each MANOVA included one between-subjects factor (subject group) and two within-subjects factors (speech stimulus and speech level or S/B ratio or babble level). Multiple comparison tests

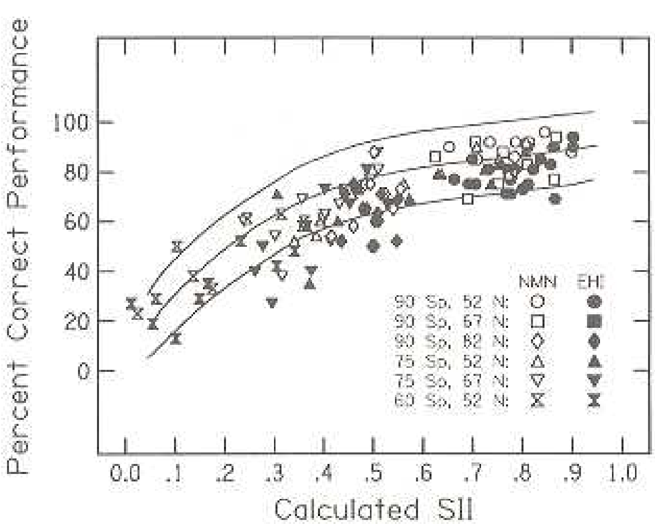

FIGURE 8. Normal transfer function ±2 *SD* for NST CV subtests, including data for EHI and NMN subjects.

TABLE 2. Means and standard deviations of observed − predicted difference scores (in percent correct) as a function of S/B ratio.

S/B ratio (dB)		Speech material		
		PH	PL	NST
Group 2 (NMN)				
+8	*M*	0.00	−8.10	−10.20
	SD	4.16	10.52	10.66
+23	*M*	0.00	−7.10	−2.40
	SD	0.00	8.20	8.04
+38	*M*	−0.80	−4.50	4.20
	SD	1.69	4.01	3.29
Group 3 (EHI)				
+8	*M*	0.60	−17.80	−17.40
	SD	5.27	17.38	9.40
+23	*M*	−1.40	−11.20	−8.30
	SD	1.26	7.44	5.77
+38	*M*	0.00	−8.30	−2.50
	SD	0.00	6.86	4.77

Note. Speech presentation level: 90 dB SPL.

were performed after each MANOVA, when necessary, using the Student-Newman-Keuls procedure (Kirk, 1968).

The first MANOVA analyzed the effects of subject group, speech stimulus, and S/B ratio. Significant main effects were

TABLE 3. Means and standard deviations of observed − predicted difference scores (in percent correct) as a function of speech level.

Speech level (dB SPL)		Speech material		
		PH	PL	NST
Group 2 (NMN)				
60	*M*	17.20	5.50	−0.44
	SD	6.89	8.13	9.91
75	*M*	3.10	−3.50	−5.10
	SD	6.52	11.23	7.52
90	*M*	−0.80	−4.50	4.20
	SD	1.69	4.01	3.29
Group 3 (EHI)				
60	*M*	20.10	3.60	−10.89
	SD	17.81	6.08	7.29
75	*M*	4.30	−11.00	−7.50
	SD	7.89	12.71	11.20
90	*M*	0.00	−8.30	−2.50
	SD	0.00	6.86	4.77

Note. Babble level: 52 dB SPL.

TABLE 4. Means and standard deviations of observed – predicted difference scores (in percent correct) as a function of babble level.

Babble level (dB SPL)		Speech material		
		PH	PL	NST
		Group 2 (NMN)		
52	*M*	17.20	5.50	−0.44
	SD	6.89	8.13	9.91
67	*M*	8.90	−0.40	−5.70
	SD	11.44	13.20	9.06
82	*M*	0.00	−8.10	−10.20
	SD	4.16	10.52	10.66
		Group 3 (EHI)		
52	*M*	20.10	3.60	−10.89
	SD	17.81	6.08	7.29
67	*M*	8.20	−5.90	−11.50
	SD	7.60	14.77	14.46
82	*M*	0.60	−17.80	−17.40
	SD	5.27	17.38	9.40

Note. S/B ratio: +8 dB.

observed for subject group [$F(1, 18) = 4.87$, $p < .05$] and stimulus [$F(2, 36) = 28.73$, $p < .001$], and a significant interaction was observed between stimulus and S/B ratio [$F(4, 72) = 8.30$, $p < .001$]. The following main effects and interactions were not significant: S/B ratio [$F(2, 36) = 2.14$, $p > .05$], group × stimulus [$F(2, 36) = 2.39$, $p > .05$], group × S/B ratio [$F(2, 36) = 0.12$, $p > .05$], and group × stimulus × S/B ratio [$F(4, 72) = 0.36$, $p > .05$].

The main effect of group corresponded to generally better SII predictive accuracy for the NMN subjects than for the EHI subjects. A simple main effects analysis of the test × S/B ratio interaction revealed that the effect of S/B ratio was significant for the NST CVs [$F(2, 57) = 13.28$, $p < .001$] but not for the R-SPIN (PH) [$F(2, 57) = 2.60$, $p > .05$] or the R-SPIN (PL) [$F(2, 57) = 0.65$, $p > .05$].

Multiple comparison tests for the NST CVs showed that differences were significant at the .05 level between the mean difference scores observed at all three S/B ratios. SII accuracy was poorest in the +8 dB S/B condition and improved as the S/B ratio improved. Accuracy for the two R-SPIN materials did not change as a function of S/B ratio.

The second MANOVA evaluated the significance of the subject group, stimulus, and speech presentation level factors. Results revealed significant main effects of subject group [$F(1, 16) = 4.67$, $p < .05$], stimulus [$F(2, 32) = 35.91$, $p < .001$], and speech level [$F(2, 32) = 22.25$, $p < .001$]. The SII was more accurate for the NMN subjects than for the EHI subjects, accounting for the main effect of subject group.

A significant interaction between stimulus and speech presentation level [$F(4, 64) = 13.88$, $p < .001$] also was observed. A simple main effects analysis of this interaction revealed that the effect of speech level was significant for all three stimuli [PH: $F(2, 57) = 23.96$, $p < .001$; PL: $F(2, 57) = 16.29$, $p < .001$; NST: $F(2, 55) = 4.37$, $p < .05$], and that the effect of stimulus was significant for all three speech levels [60 dB SPL: $F(2, 55) = 30.76$, $p < .001$; 75 dB SPL: $F(2, 57) = 12.21$, $p < .001$; 90 dB SPL: $F(2, 57) = 24.70$, $p < .001$]. The following interactions were not significant: group × stimulus [$F(2, 32) = 2.93$, $p > .05$], group × speech level [$F(2, 32) = 0.03$, $p > .05$], and group × test × speech level [$F(4, 64) = 0.63$, $p > .05$].

Multiple comparison tests were performed to determine the source of the stimulus × speech level interaction. A criterion of .05 was used to test significance. For both types of R-SPIN sentences, predictive accuracy was significantly different for 60 dB SPL than for the two higher presentation levels. For the R-SPIN (PH), SII accuracy was better at 75 and 90 dB SPL than at 60 dB SPL. In contrast, SII accuracy for the R-SPIN (PL) was better at 60 dB SPL than at the other two levels. For the NST CVs, significant differences were observed between 90 dB SPL and the other two levels, with better accuracy at the highest speech presentation level.

The effect of speech stimulus was significant at all three speech presentation levels. At 60 dB SPL, difference scores were significantly different at the .05 level for all three materials. Performance for the R-SPIN (PH) was considerably underpredicted, R-SPIN (PL) performance was slightly underpredicted, and NST CV performance was overpredicted. For 75 dB SPL, SII accuracy for the R-SPIN (PH) was significantly different from that observed for the NST CVs and the R-SPIN (PL). R-SPIN (PH) performance was underpredicted, whereas performance for the other stimuli was overpredicted. Difference scores for the NST CVs and the R-SPIN (PL) were not significantly different. At 90 dB SPL, predictive accuracy was significantly poorer for the R-SPIN (PL) than for the other two tests.

In the third MANOVA, the significance of the subject group, stimulus, and babble level factors was assessed. Significant main effects of subject group [$F(1, 16) = 5.60$, $p < .05$], stimulus [$F(2, 32) = 43.66$, $p < .001$], and babble level [$F(2, 32) = 14.13$, $p < .001$] were observed. Interactions between group and stimulus [$F(2, 32) = 1.12$, $p > .05$]; group and babble level [$F(2, 32) = 0.19$, $p > .05$]; stimulus and babble level [$F(4, 64) = 2.06$, $p > .05$]; and group, stimulus, and babble level [$F(4, 64) = 0.51$, $p > .05$] were not significant.

The main effect of subject group again corresponded to generally smaller difference scores for NMN subjects than for EHI subjects. Multiple comparison tests were performed to reveal the exact nature of the effects of test and babble level. Mean difference scores for the R-SPIN (PH) were significantly different at the .05 level from those observed for the other two tests. For the R-SPIN (PH), the SII underpredicted performance, whereas the SII overpredicted performance for the other two stimuli. SII accuracy was significantly different for all three babble levels. Performance was underpredicted for 52 dB SPL, underpredicted only very slightly for 67 dB SPL, and overpredicted for 82 dB SPL.

Discussion

Subject Age and Predictive Accuracy

The first question addressed in this study was whether SII predictive accuracy was affected by subject age. A clear effect of subject group was observed in all three data analyses, with better SII accuracy for the NMN subjects than for the EHI subjects across a range of test conditions and materials. This result is consistent with the limited literature assessing AI accuracy for young and elderly subjects. For example, Dubno et al. (1984) found that in conditions of equal performance, the AI predicted higher scores for elderly listeners than for younger listeners. Schum et al. (1991) also found that the AI overpredicted speech recognition in babble for elderly subjects with hearing loss. Dubno and colleagues concluded that the AI did not adequately account for age-related decrements in speech recognition. This observation appears to be true for the SII, despite the use of new procedures.

In some earlier studies, speech recognition performance differences between young and elderly subjects were related to differences in hearing sensitivity (e.g., Gelfand et al., 1986; Kalikow et al., 1977; Plomp & Mimpen, 1979). In the current study, sensitivity differences were controlled by simulating the hearing losses of the EHI subjects in young listeners with normal hearing. Although the threshold matching procedures did not produce perfect matches for all EHI subjects, the average difference across frequency (1.81–6.71 dB) was considered adequate for this study, because all speech stimuli were presented at suprathreshold levels. The RMS error observed in the current study between EHI thresholds and NMN thresholds was similar to that observed by Humes et al. (1987).

The SII calculation procedures employed in the current study used individual subject thresholds, thus accounting for any signal audibility differences that may have resulted from imprecise threshold matching. As a result, differences in SII predictive accuracy can be attributed to factors other than sensitivity differences.

One plausible explanation for the group effect observed here is that elderly listeners find babble noise more distracting than do younger listeners. This hypothesis is supported by McDowd and Filion (1992), who reported that the ability to ignore irrelevant stimuli seems to decrease with advancing age.

Another interpretation of the group effect derives from the comparison of EHI subjects with cochlear hearing losses to NMN subjects with simulated hearing losses. That is, cochlear damage may impose changes in signal processing, in addition to loss of sensitivity, that may not be simulated adequately by noise masking of subjects with normal hearing. Additional comparisons between EHI subjects and younger subjects with hearing loss may be necessary to clarify the source of the group effect in the present study.

The EHI subjects in the current experiment had mild-to-moderate hearing losses, and good or excellent word recognition abilities in quiet. Thus, the findings may not generalize to elderly listeners with other auditory characteristics. Kamm et al. (1985) found that for a subject with poor word recognition in quiet, the AI greatly overpredicted performance. Other studies have shown that AI accuracy is reduced for listeners with hearing losses that are moderately severe or poorer (Dubno et al., 1989; Pavlovic, 1984).

Speech Material and Predictive Accuracy

The effect of speech material was significant in all three data analyses, although it interacted significantly with the effects of S/B ratio and speech presentation level. In general, performance tended to be underpredicted for the R-SPIN (PH) and overpredicted for the other two materials. This trend is consistent with the results of earlier studies involving the speech materials used in the current experiment (Dirks et al., 1986; Kamm et al., 1985).

The finding of underprediction for the R-SPIN (PH) items was not surprising, because recognition of these sentences involves the use of contextual cues. AI procedures traditionally have not accounted for context effects, and this remains true for the SII. For the R-SPIN (PL) and the NST CVs, where performance depends primarily on audibility, the SII should have been more accurate than it was for the R-SPIN (PH). In fact, accuracy tended to be poorer for these two materials.

This result may be attributed to the use of frequency importance functions that were not developed specifically for the materials to which they were applied. For the NST CVs, the importance function was based on materials in which the phonemes of English are represented equally often. This does not characterize the group of CV subtests used in the current study. For the R-SPIN, one importance function was provided in the SII for both sentence types. This single function was derived from two functions, one developed for high-context sentence materials and one developed for low-context sentence materials (Bell, Dirks, & Trine, 1992). In light of the difference in the recognition task for the two types of items, it is not surprising that performance would be underpredicted for the PH items and overpredicted for the PL items when using one combined function for these two types of sentences.

Effect of Test Condition on Predictive Accuracy

The effect of speech presentation level on SII accuracy varied among the three speech stimuli. At higher speech levels, SII accuracy improved for the R-SPIN (PH), decreased for the R-SPIN (PL), and improved slightly for the NST CVs. In this analysis, the babble level was constant, and therefore the S/B ratio improved at higher levels. The pattern of results could be associated with self-masking effects of speech (e.g., spread of masking) at high levels, coupled with the differing effects of contextual cues available with the different speech materials. At high levels, the listeners appeared to have more difficulty accurately perceiving the speech signal than the SII predicted. This observation applies primarily to the R-SPIN (PL) items. For the NST CVs, possible level effects may have been reduced by the closed-set nature of the NST. Similarly, the availability of contextual cues for the R-SPIN (PH) items probably served to minimize any level effects that may have occurred.

There was an interaction between babble level and speech stimulus. For the R-SPIN (PH), SII accuracy improved with increasing babble level, whereas accuracy for the other two stimuli decreased. It should be noted that as babble levels increased, speech levels also increased. Thus, at higher babble levels, it appears that the self-masking effects of the speech and the masking effects of the babble were greater than the SII model predicts. For the R-SPIN (PH), this overmasking may have reduced the effects of context, causing performance to be closer to the predicted scores.

The effect of S/B ratio on SII accuracy also depended on the speech stimulus. For the NST CVs, S/B ratio significantly affected SII accuracy. For both R-SPIN materials, in contrast, accuracy did not vary with S/B ratio. The high presentation level (90 dB SPL) used for all three S/B ratios may have canceled out the effects of background noise for the R-SPIN. This did not occur for the NST CVs, perhaps because of the difference in the phonemic content of the two tests. Unlike the R-SPIN lists, the group of NST CV subtests is not phonetically balanced. More importantly, these subtests were developed to maximize difficult phonemic contrasts for listeners with hearing impairment (Dubno, Dirks, & Langhofer, 1982) and contain many weak, high-frequency consonants. Thus, the NST CVs contain a higher proportion of easily masked sounds compared to the R-SPIN lists.

Conclusions

The results show reduced predictive accuracy of the SII for the EHI listeners as compared to young listeners with simulated hearing losses that were almost identical to those of the older listeners. This effect was observed over a range of listening conditions for three different speech materials. These findings tentatively suggest that there is an age-related decrease in speech recognition abilities in noise that is separate from the absolute sensitivity loss. Recent studies (Gordon-Salant & Fitzgibbons, 1993; Humes & Christopherson, 1991; Klein, Mills, & Adkins, 1990) have shown further that age-related decline in various aspects of auditory processing, such as duration discrimination, frequency discrimination, and upward spread of masking, may contribute to the speech recognition problems of elderly people.

Procedural factors examined in this study also affected SII predictive accuracy, suggesting that the proposed new SII method continues to have some limitations. The effects of different speech materials, speech presentation levels, levels of background noise, and subject groups need to be explored further. Adjustment to the SII procedures, such as the incorporation of more importance functions, may be necessary to improve its predictive accuracy. At present, it appears that the proposed SII may be useful for predicting relative performance across a set of conditions, but it is somewhat limited for predicting absolute speech recognition performance, particularly for elderly listeners with hearing loss.

Acknowledgment

This research was supported in part by a grant from the National Institute on Aging awarded to the second author. The authors thank Chaslav Pavlovic for providing a draft of the proposed SII calculation procedures and a copy of the computer program for calculating the SII. We are also grateful to Peter Fitzgibbons, Patricia Balfour, and Grace Yeni-Komshian for their assistance and advice regarding research design, instrumentation, and statistical analysis, and to Judy Dubno, Jayne Ahlstrom, Lois Matthews, Lenore Holte, and two anonymous reviewers for their helpful comments on earlier drafts of this manuscript. Finally, we thank the subjects who participated in this experiment for their generous contributions of time and effort.

References

American National Standards Institute (1969). *American national standard method for the calculation of the articulation index* (ANSI S3.5-1969). New York: ANSI.

American National Standards Institute (1989). *Specifications for audiometers* (ANSI S3.6-1989). New York: ANSI.

Bell, T. S., Dirks, D. D., & Trine, T. D. (1992). Frequency-importance functions for words in high- and low-context sentences. *Journal of Speech and Hearing Research, 27,* 32–48.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32–48.

Committee on Hearing, Bioacoustics, and Biomechanics (CHABA)—Working Group on Speech Understanding and Aging. (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83,* 859–893.

Dirks, D. D., Bell, T. S., Rossman, R. N., & Kincaid, G. E. (1986). Articulation index predictions of contextually dependent words. *Journal of the Acoustical Society of America, 80,* 82–92.

Dubno, J. R., & Dirks, D. D. (1989). Auditory filter characteristics and consonant recognition for hearing-impaired listeners. *Journal of the Acoustical Society of America, 85,* 1666–1675.

Dubno, J. R., Dirks, D. D., & Langhofer, L. R. (1982). Evaluation of hearing-impaired listeners using a nonsense-syllable test. II. Syllable recognition and consonant confusion patterns. *Journal of Speech and Hearing Research, 25,* 141–148.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America, 76,* 87–96.

Dubno, J. R., Dirks, D. D., & Schaefer, A. B. (1989). Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II: Articulation index predictions. *Journal of the Acoustical Society of America, 85,* 355–364.

Dubno, J. R., & Schaefer, A. B. (1992). Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *Journal of the Acoustical Society of America, 91,* 2110–2121.

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19,* 90–119.

Gaeth, J. (1948). *A study of phonemic regression associated with hearing loss.* Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Gelfand, S. A., Piper, N., & Silman, S. (1986). Consonant recognition in quiet and in noise with aging among normal hearing listeners. *Journal of the Acoustical Society of America, 80,* 1589–1598.

Goetzinger, C., Proud, D., Dirks, D., & Embrey, J. (1961). A study of hearing in advanced age. *Archives of Otolaryngology, 73,* 662–674.

Gordon-Salant, S. (1987). Age-related differences in speech recognition performance as a function of test format and paradigm. *Ear and Hearing, 8,* 277–282.

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36,* 1276–1285.

Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging and speech perception in reverberation and noise. *Journal of Speech and Hearing Research, 3,* 149–155.

Humes, L. E., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. *Journal of Speech and Hearing Research, 34*, 686–693.

Humes, L. E., Dirks, D. D., Bell, T. S., & Kincaid, G. E. (1987). Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers. *Journal of the Acoustical Society of America, 81*, 765–773.

Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research, 33*, 726–735.

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. *Journal of Speech and Hearing Research, 37*, 465–474.

Jerger, J., Jerger, S., Oliver, T., & Pirozzolo, F. (1989). Speech understanding in the elderly. *Ear and Hearing, 10*, 79–89.

Jerger, J., Jerger, S., & Pirozzolo, F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age and cognitive abilities in the elderly. *Ear and Hearing, 12*, 103–109.

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. *Journal of the Acoustical Society of America, 61*, 1337–1351.

Kamm, C. A., Dirks, D. D., & Bell, T. S. (1985). Speech recognition and the articulation index for normal and hearing-impaired listeners. *Journal of the Acoustical Society of America, 77*, 281–288.

Kamm, C. A., Dirks, D. D., & Carterette, E. C. (1982). Some effects of spectral shaping on recognition of speech by hearing-impaired listeners. *Journal of the Acoustical Society of America, 71*, 1211–1224.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks-Cole.

Klein, A. J., Mills, J., & Adkins, W. Y. (1990). Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners. *Journal of the Acoustical Society of America, 87*, 1266–1271.

McDowd, J. M., & Filion, D. (1992). Aging, selective attention and inhibitory processes: A psychophysiological approach. *Psychology and Aging, 7*, 65–71.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America, 72*, 1788–1803.

Pavlovic, C. V. (1984). Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment. *Journal of the Acoustical Society of America, 75*, 1253–1258.

Pavlovic, C. V., Studebaker, G. A., & Sherbecoe, R. L. (1986). An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. *Journal of the Acoustical Society of America, 80*, 50–57.

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America, 66*, 1333–1342.

Resnick, S. B., Dubno, J. R., Hoffnung, S., & Levitt, H. (1975). Phoneme errors on a nonsense syllable test. *Journal of the Acoustical Society of America, 58* (Suppl. 1), 114.

Schum, D. J., Matthews, L. J., & Lee, F. (1991). Actual and predicted word-recognition performance of elderly hearing-impaired listeners. *Journal of Speech and Hearing Research, 34*, 636–642.

Stata Corporation. (1993). *Stata: Release 3.1* [Computer program]. College Station, TX: Author.

Studebaker, G. A. (1991). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28*, 455–462.

Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6* (Technical Report, SAM-TR-66-55). Brooks AFB, TX: USAF School of Aerospace Medicine.

van Rooij, J. C. G. M., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analysis. *Journal of the Acoustical Society of America, 88*, 2611–2624.

van Rooij, J. C. G. M., & Plomp, R. (1991). The effect of linguistic entropy on speech perception in noise in young and elderly listeners. *Journal of the Acoustical Society of America, 90*, 2985–2991.

van Rooij, J. C. G. M., & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion. *Journal of the Acoustical Society of America, 91*, 1028–1033.

Zurek, P. A., & Delhorne, L. A. (1987). Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *Journal of the Acoustical Society of America, 82*, 1548–1559.

Received April 22, 1994
Accepted September 14, 1994

Contact author: Sandra Gordon-Salant, PhD, University of Maryland at College Park, Department of Hearing and Speech Sciences, Lefrak Hall, College Park, MD 20742.