Comparison of Objective and Subjective Measures of Speech Intelligibility in Elderly Hearing-Impaired Listeners

Robyn M. Cox
Department of Audiology and Speech Pathology
Memphis State University, and
Veteran's Administration Medical Center
Memphis, TN

Genevieve C. Alexander
Izel M. Rivera
Veteran's Administration Medical Center
Memphis, TN

Three experiments were performed to evaluate the use of subjective intelligibility estimations as a method for measuring hearing aid benefit. Subjective and objective speech intelligibility scores were compared for young normal-hearing and elderly hearing-impaired listeners. Objective intelligibility scores were obtained using the Connected Speech Test (CST). This test consists of conversationally produced passages of speech that the listener repeats sentence by sentence. To provide subjective intelligibility scores, listeners estimated the percentage that they understood of each CST passage. Comparison of the two types of scores revealed that they were closely related in both groups of subjects ($r = 82-.92$). Although the two types of scores were essentially equal for normal-hearing subjects, the hearing-impaired listeners tended to produce subjective estimations of intelligibility that were significantly lower than their objective scores. Manipulation of visual cues and amplification, in an attempt to influence the hearing-impaired listeners' expectation of understanding speech, had no effect on the subjective-objective score differential. The difference between subjective and objective scores in the hearing-impaired group was not related to audiometric variables such as speech reception threshold, audiogram, or duration of hearing loss. It was concluded that comparative hearing aid evaluations using subjective intelligibility estimates would usually produce the same *relative* outcome as evaluations using the objective intelligibility measurement procedure. However, scores obtained with the objective procedure had smaller critical differences. Thus, when both types of scores are based on the same number of passages, the objective measurement procedure would be the more sensitive to differences among hearing aids.

KEY WORDS: hearing aid, speech intelligibility, subjective measures

Quantification of the intelligibility of connected speech by asking the listener "How much do you understand?" has been used in hearing assessment and hearing aid fitting for many years. Formal use of subjective intelligibility estimation was reported as early as 1947 (Falconer & Davis), and other investigations using variants of this approach have appeared regularly in the audiological literature ever since. For example, recent applications of subjective intelligibility estimates used them to compare hearing aids (Cox & McDaniel, 1984), measure speech reception threshold (Walker & Byrne, 1985), and determine a frequency importance function for connected speech (Studebaker, Pavlovic, & Sherbecoe, 1987).

Subjective intelligibility tests elicit a judgment of understanding from the listener. Objective procedures, in contrast, rely on the tester to evaluate the listener's understanding: They typically require recognition or repetition of speech test material. Subjective estimation of speech intelligibility has several advantages over most objective measurement methods. First, the subjective method produces data much more rapidly: A speech passage may be sampled and judged in 30 s in contrast to the

several minutes required to obtain a score with an objective test. Second, the subjective method allows ready quantification of the intelligibility of speech passages that are similar to everyday connected speech, whereas most objective speech tests use a type of speech (e.g., single monosyllables) not often encountered in everyday listening. Third, the subjective approach may afford a more valid measure of speech understanding because a subjective estimate quantifies the proportion of speech that is understood, whereas an objective method relies on repetition of words without considering comprehension.

Objective approaches to intelligibility measurement also appear to have certain advantages. Because they elicit repetition or identification of the words understood, these tests provide data that lend themselves to analysis of specific error patterns. Moreover, objective test data provide definite assurance about the accuracy with which the target speech was received. In contrast, when a subjective estimate of intelligibility is obtained, the tester has no guarantee that the listener actually understood the proportion of the test material that he or she reports understanding. The listener may have misinterpreted the test material or may have inaccurately evaluated the proportion missed. Also, speech intelligibility estimates may be affected by nonauditory factors such as listener expectations and personality traits.

Given the widespread use of both subjective and objective methods of speech intelligibility measurement in clinical and research applications, it is surprising that relatively little information has been reported on the relationship between the two types of measures. Peters (1965) measured intelligibility for each of 24 talkers using (a) a multiple choice objective test (not described) yielding a percentage score, and (b) a subjective intelligibility rating of connected speech on a 7-point equal-appearing interval scale. The listeners presumably had normal hearing. The correlation between objective and subjective intelligibility scores was only .58. Because this outcome indicates that only about one third of the variation in subjective scores could be attributed to the variation in objective scores, it suggests that the two types of test may be measuring fundamentally different aspects of speech understanding. Factors that could have limited the outcome in this study include relatively poor test-retest reliability and the use of different speech samples to generate the subjective and objective intelligibility scores.

Speaks, Parker, Harris, and Kuhl (1972) reported subjective and objective intelligibility data obtained from normal hearers listening to the CID sentences (Silverman & Hirsh, 1955). In each condition, each listener provided percentage intelligibility estimates for 10 sentences and repeated the words from 10 different sentences. The correlation between mean subjective and objective intelligibility scores was .93. However, when extreme scores were excluded, the correlation decreased to .84.

These two investigations agree that there is a positive correlation between subjective and objective intelligibility scores for connected speech, at least for normal hearers. However, the strength of the relationship between the two types of data was different in the two studies. There appear to be no reported investigations in which subjective and objective speech intelligibility data were compared for hear-

ing-impaired listeners. Moreover, the influence of nonauditory factors on intelligibility estimates has not been assessed. Studebaker, Bisset, Van Ort, and Hoffnung (1982) reported data suggesting that hearing-impaired listeners are less accurate than normal hearers in judging intelligibility of connected discourse. However, work by Yanz, Carlstrom, and Thibodeau (1985) and Gordon-Salant (1986) indicates that hearing-impaired listeners are as accurate as normal hearers in judging their own understanding of speech.

The investigations reported in this paper were undertaken to examine the relationship between subjective and objective intelligibility measures of connected speech for elderly hearing-impaired listeners. The major impetus for these studies was to evaluate the utility of subjective intelligibility measures as a basis for assessing hearing aid benefit. Because most hearing aids are worn by elderly individuals with adventitious hearing loss, the hearing-impaired subjects were chosen to represent this group.

EXPERIMENT 1: NORMAL-HEARING LISTENERS

It cannot be assumed a priori that an objective intelligibility score for a particular speech sample should be equal to the subjectively estimated intelligibility score of the same sample. First, unless the objective score is based on repetition of every word in the sample, the chosen scoring words might not accurately reflect the overall intelligibility of the entire sample. Second, because of the redundancy that characterizes connected speech, perfect subjective intelligibility may be attained without comprehension of every word in the sample. Finally, when the speech is presented to hearing-impaired listeners, the acoustic filtering imparted by the hearing loss may affect the relationship between subjective and objective intelligibility scores by reducing one type of score more than the other. As a result of these considerations, the intrinsic relationship between scores for the subjective and objective intelligibility test procedures was empirically established using data obtained from normal-hearing subjects.

Method

Objective Speech Intelligibility Test

Speech intelligibility was objectively quantified using the Connected Speech Test (CST; Cox, Alexander, & Gilmore, 1987). This test comprises 48 passages of conversationally produced connected speech on everyday topics. A multi-talker babble serves as a competing stimulus. Each passage contains 10 syntactically simple sentences, 7–10 words in length. A passage is presented one sentence at a time. After each sentence, both speech and babble are halted while the subject repeats the sentence or as much of it as he or she understood. Subjects are instructed to repeat every word exactly as heard. Each passage contains 25 scoring words. The intelligibility score for each passage is based on the percentage of scoring words that are correctly repeated. The

test is recorded on optical disk, and passages may be administered either audiovisually or in audio only. Before a passage begins, subjects are shown a word describing the topic of the passage. During audiovisual administration, a front view of the talker's head and shoulders is shown on a monitor.

In the present study, a set of six CST passages was administered for each objective score. The sets have been empirically determined to be equal in average intelligibility for normal hearers and hearing-impaired listeners similar to those serving in Experiments 2 and 3 (Cox, Alexander, Gilmore, & Pusakulich, 1988; 1989). A score, in percent correct, reflected the mean performance across all the passages in a particular set. Thus, each score was based on 150 scoring words.

Subjective Speech Intelligibility Test

Subjective estimations of speech intelligibility were obtained using an intelligibility rating procedure similar to the one used by Cox and McDaniel (1989) for the Speech Intelligibility Rating (SIR) test. Subjects were shown an equal-appearing interval scale, numbered from 0 to 100, with the ends defined as "no words understood" and "all words understood." The scale is shown in Figure 1.

The CST passages served as stimuli for subjective intelligibility estimations. A passage was first administered as described above to produce the objective score. After responding to all 10 sentences of a passage, the listener consulted the rating scale in Figure 1 and supplied a number that reflected the percentage of the preceding passage that he or she had understood. Six passages were presented, and their intelligibility estimated, for each subjective score. A score, in percent, reflected the mean estimate across all six passages. Thus, each subjective score was based on six intelligibility estimates. Use of the same set of six CST passages to produce both the objective score and the corresponding subjective score established without question the equivalence of the speech stimuli for the two types of scores.

Subjects

Adults who reported normal hearing served as subjects. Group 1 consisted of 28 individuals, aged 15–41 years, with

an average age of 25. Two had previous brief experience with the subjective and objective intelligibility measurement tasks. Group 2 comprised 17 of the members of Group 1. This subgroup was aged 15–41 years, with a mean age of 24.

Procedure

The CST passages were replayed using a two-channel optical disk player (Panasonic TQ 2024F). The test passages and competing babble were mixed and delivered monaurally using an insert earphone (Etymotic Research ER-1) coupled to the ear via a compressible foam earplug. The nontest ear was plugged. The signal was delivered at 61 dB L_{eq} (L_{eq} = equivalent continuous dBA level) calibrated in a Zwislocki-type ear simulator (approximately normal conversational level). The passages were presented in audio only (without the visual image of the talker). Each subject listened at a single signal-to-babble ratio (SBR), but the SBR was varied across subjects to produce a range of scores. Normal SBRs ranged from −5 dB to +4 dB.

A microcomputer (Zenith 181 or 184) controlled the sentence-by-sentence delivery of each passage. After the subject responded to each sentence, the examiner keyed in the scoring words that were correctly repeated. At the end of each passage, the subject was asked to generate a number on the 0–100 scale that reflected the percentage of the preceding passage that was understood. Several practice passages were administered before data collection.

Each Group 1 subject provided two objective intelligibility scores and two corresponding subjective intelligibility estimates. Each subjective-objective score pair was based on a different set of six CST passages.

In addition to providing two subjective-objective score pairs for the original CST material, Group 2 subjects also provided two subjective-objective score pairs for CST passages that had been filtered to simulate the acoustic effect of a high-frequency sloping hearing loss.[1]

---

[1]Four Group 2 subjects served twice: once to provide the comparison of subjective and objective data shown in Figure 2, and again about a year later to provide the comparison of data for filtered and unfiltered speech shown in Figure 3 For these subjects, the unfiltered speech data in Figure 3 are different from those in Figure 2.

WORDS UNDERSTOOD
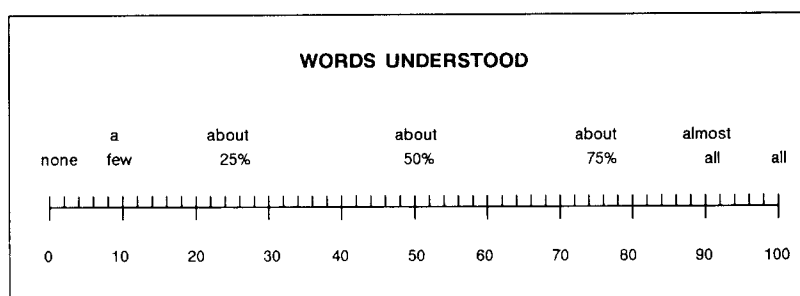
FIGURE 1. Facsimile of the equal-appearing interval scale used by listeners for the subjective intelligibility estimation task.

In the filtered conditions, the CST passages and competing babble, after mixing, were low-pass filtered using two cascaded graphic equalizers (Yamaha, GQ1031BII) before delivery to the insert earphone. Above the cut-off frequency of 400 Hz, the simulated loss sloped at 22 dB/octave to 1.2 kHz and was 40 ± 3 dB in the range 1.2 kHz to 8.0 kHz. Both the SBR and the presentation level of the filtered speech-plus-babble stimulus were adjusted to achieve approximately equal loudness and intelligibility with the unfiltered counterpart. The order of presentation of filtered/unfiltered conditions was counterbalanced across subjects in Group 2. When the condition was changed from filtered to unfiltered, or vice versa, several practice passages were administered prior to data collection in the new condition.

Results

Group 1

The data comprised two objective intelligibility scores and two corresponding subjective intelligibility estimates for 28 subjects. To homogenize the variances of these percentage data, all values were transformed into rationalized arcsine units (raus) before analysis as Studebaker (1985) described. The scale for rationalized arcsine units extends from −23 to 123. Values in the range from 20 to 80 are within about one unit of the corresponding percentage score.

Figure 2 illustrates the relationship between the objective and subjective intelligibility data for normal hearers. Each symbol depicts one pair of scores. There are two pairs of scores per subject. Despite some individual variation, and one aberrant subject shown by the open squares, these data are well described by the diagonal line, suggesting that
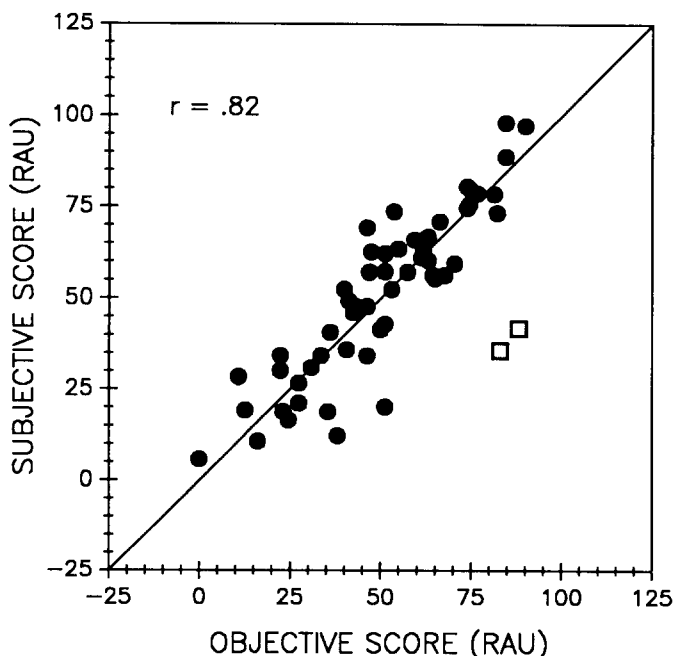

FIGURE 2. Objective and subjective intelligibility data for 28 normal-hearing listeners. Each symbol depicts one pair of scores. There are two pairs of scores per subject.

objective and subjective intelligibility scores were essentially equal for these listeners. The linear correlation coefficient between subjective and objective scores was .82 (this correlation was .87 if the aberrant subject was excluded from the analysis).

To determine whether there were systematic differences between subjective and objective scores, these data were subjected to a repeated measures analysis of variance. Variables were score (objective and subjective) and passage set (first and second). The results indicated that there was no significant difference between the two types of scores. The mean objective and subjective scores were 51.6 rau and 50.6 rau, respectively.

This study with normal hearers provided an opportunity to estimate and compare critical differences for subjective and objective intelligibility scores obtained with these measurement procedures. Critical differences are used to evaluate the significance of differences between scores of the same type (e.g., both subjective) obtained from the same listener. If the difference between scores under two conditions exceeds the 95% critical difference, there is only a 5% likelihood that this outcome occurred by chance. The size of the critical difference varies inversely with the reliability of the test scores.

For both objective and subjective measurement procedures, critical differences (CDs) were computed using the standard deviation of the distribution of differences (SD_{diff}) between scores for the first and second set of six CST passages (95% CD = 1.96 × SD_{diff}). For the objective scores, this resulted in a 95% CD of 10.3 rau. For the subjective scores, the 95% CD was found to be 15.7 rau.

Group 2

Figure 3 shows the subjective and objective scores for filtered and unfiltered CST passages for the 17 subjects in Group 2. Each symbol depicts one pair of scores. There are four pairs of scores per subject, two in each filter condition. The linear regression lines are shown for unfiltered (dashed line) and filtered (solid line) data. The main issue of interest in these data centered on the possible effect of low-pass filtering on the relationship between subjective and objective scores. If filtering resulted in a change in the relationship, we would expect the pattern of the open triangles to be different from the pattern of the filled triangles. This would be substantiated by different regression lines for filtered and unfiltered data. In fact, the two sets of data are actually very similar. Data are intermingled around the figure's diagonal (not shown), and the two regression lines are almost identical.

To explore any systematic interaction between scoring method and filtering, these data were subjected to repeated measures analysis of variance. Variables were filtering (filtered and unfiltered), score (subjective and objective), and passage set (first and second). None of the main effects or interactions approached statistical significance. The mean objective score was 51.4 rau, and the mean subjective score was 46.0 rau. This pattern in which the mean subjective score was somewhat (but not significantly) less than the
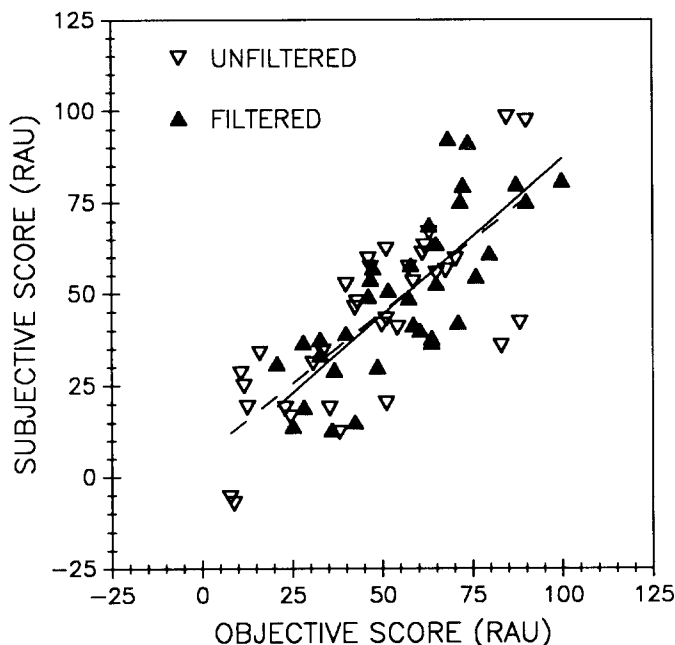
FIGURE 3. Objective and subjective intelligibility data for filtered and unfiltered speech from 17 normal-hearing listeners. Each symbol depicts one pair of scores. There are four pairs of scores per subject, two for filtered speech and two for unfiltered speech.

mean objective score was seen in both the filtered and unfiltered data.

Discussion

The data depicted in Figure 2 suggest that subjective and objective intelligibility scores, as measured in this study, were essentially equal for normal hearers. There is no evidence that differences between the two types of scores had any basis beyond measurement error. This outcome is consistent with the data reported by Speaks et al. (1972).

The data shown in Figure 3 indicate that the relationship between subjective and objective intelligibility scores was not affected by low-pass filtering the CST passages to simulate the filtering effects of a high-frequency hearing loss.

The CDs derived for objective and subjective scores were similar to analogous CDs found in previous studies. The value of 10.3 rau derived for the objective scores was quite similar to the 11.2 rau reported by Cox et al. (1989) for six-passage CST scores for normal-hearing listeners. The critical difference of 15.7 rau computed for the subjective scores was also similar to the analogous CD of 13.5 rau that can be estimated for the Speech Intelligibility Rating test reported by Cox and McDaniel (1989).[2]

Overall, these data from normal-hearing listeners were consistent with a conclusion that the subjective and objective

measures of speech intelligibility used in this study are measuring the same quantity and result in equal average values. Furthermore, the apparent equality of the subjective and objective measurement procedures persisted even when high-frequency components of the speech passages were reduced in a manner similar to the acoustic effects of high-frequency hearing loss.

EXPERIMENT 2: HEARING-IMPAIRED LISTENERS

Results of Experiment 1 suggested that, at least for young normal-hearing listeners, the main difference between the subjective and objective intelligibility assessment procedures was in the reliability of the scores, as reflected by the CDs. In the second experiment, the two measurement procedures were used to quantify speech intelligibility for elderly hearing-impaired persons. To explore the potential effect on subjective scores of variables that might be encountered in determining hearing aid benefit, measurements were made with and without hearing aids and with and without visual cues. It seemed possible, for example, that hearing-impaired listeners might underestimate their speech understanding in unaided conditions because they were accustomed to hearing poorly in this circumstance. Similarly, they might overestimate their speech understanding when visual cues were available because they expected to understand better in this condition.

METHOD

Subjects

Twenty-four individuals served as subjects. Their mean age was 68 years (*SD* = 5.2). All had bilateral sensorineural hearing impairment. The extent of their impairments ranged from mild to moderately severe. Average audiometric thresholds at octave frequencies from 250 Hz to 4000 Hz, in the test ears, were 28, 28, 38, 56, and 72 dB HL (mean *SD* = 14 6) (ANSI, 1969). Individual speech reception thresholds ranged from 10 to 65 dB HL with a mean of 38 dB. Mean audiometric slopes from 500 Hz to 4000 Hz ranged from 0 dB/octave to 22 dB/octave. All of the hearing losses were adventitious. Although precise etiology was difficult to establish, most of the subjects attributed their impairment to aging and/or noise exposure. Sixteen subjects were hearing aid wearers, and 8 did not use amplification. Corrected near and far vision ranged from 20/20 to 20/40.

Procedure

The CST passages were replayed using a two-channel optical disk player (Panasonic TQ 2024F or Sony LDP-1500). The video output was routed to a 33-cm diagonal color monitor (Panasonic CT1330M). The speech passage output was delivered to a small loudspeaker (Realistic Minimus-7) that was located directly below and slightly in front of the video image. The multitalker babble was delivered to four

---

[2]In the Speech Intelligibility Rating (SIR) test, the listener estimates the intelligibility of a spoken passage by providing a rating on a scale from 0 (no words understood) to 10 (all words understood). The CD reported here was determined with the assumption that values on the SIR scale could be translated into percentages with no loss of accuracy

identical small loudspeakers placed symmetrically around the listener's position at azimuths of 45°, 135°, 225° and 315°. In each of the eight experimental conditions the CST was presented and scored, and subjective intelligibility estimates were elicited, in the same manner as for Experiment 1.

Each subject provided both objective and subjective intelligibility scores in each of two stimulus presentation modes: audiovisual and audio only. In both of these conditions, subjects were tested unaided and while wearing each of three monaural hearing aids that differed somewhat from each other but were all potentially appropriate for the hearing loss. Each hearing aid was worn at a comfortable volume control setting selected by the subject. The hearing aids modified the intelligibility of speech delivered to the listeners and created conditions with the potential for altering the subjects' expectations of understanding speech. However, differences among the hearing aids were not of interest in this investigation.

Each subject was tested in one of three ordinary rooms (without acoustical treatment). Room A was used to simulate a living room, Room B simulated a large classroom, and Room C simulated a social event with numerous people present. In each room, the speech passages and competing multitalker babble were presented at levels that are found in everyday listening situations (Pearsons, Bennett, & Fidell, 1977). Each subject was tested in only one room. Assignment of subjects to rooms was random. Ten subjects listened in Room A, and 7 listened in each of the other two rooms. On average, the subjects assigned to each room had about equal hearing impairment; mean speech reception thresholds were 39, 37, and 38 dB HL in rooms A, B, and C respectively. Differences among the rooms were not of interest in this study. However, the use of realistic listening settings resulted in subjects performing at a range of intelligibility levels without imposition of artificial control methods.

Several CST practice passages were presented before each new experimental condition. Stimulus presentation modes and listening conditions were counterbalanced across subjects.

Results

The data comprised eight objective and eight corresponding subjective intelligibility scores for 21 subjects. Some data were missing for 3 subjects, resulting in 4–6 pairs of scores for these individuals. As with the data from normal-hearing listeners, all values were transformed into raus before analysis.

Figure 4 illustrates the relationship between the objective and subjective intelligibility scores. The upper panel shows data for the audio-only presentation mode, and the lower panel gives data for the audiovisual mode. Both sets of data are well fitted by linear regression lines, as Figure 4 shows. The linear correlation coefficients between objective and subjective scores were .92 (audio only) and .90 (audiovisual).

These data were subjected to a repeated measures analysis of variance to explore any systematic differences between objective and subjective scores and to examine the effects of the two factors that were hypothesized to influence
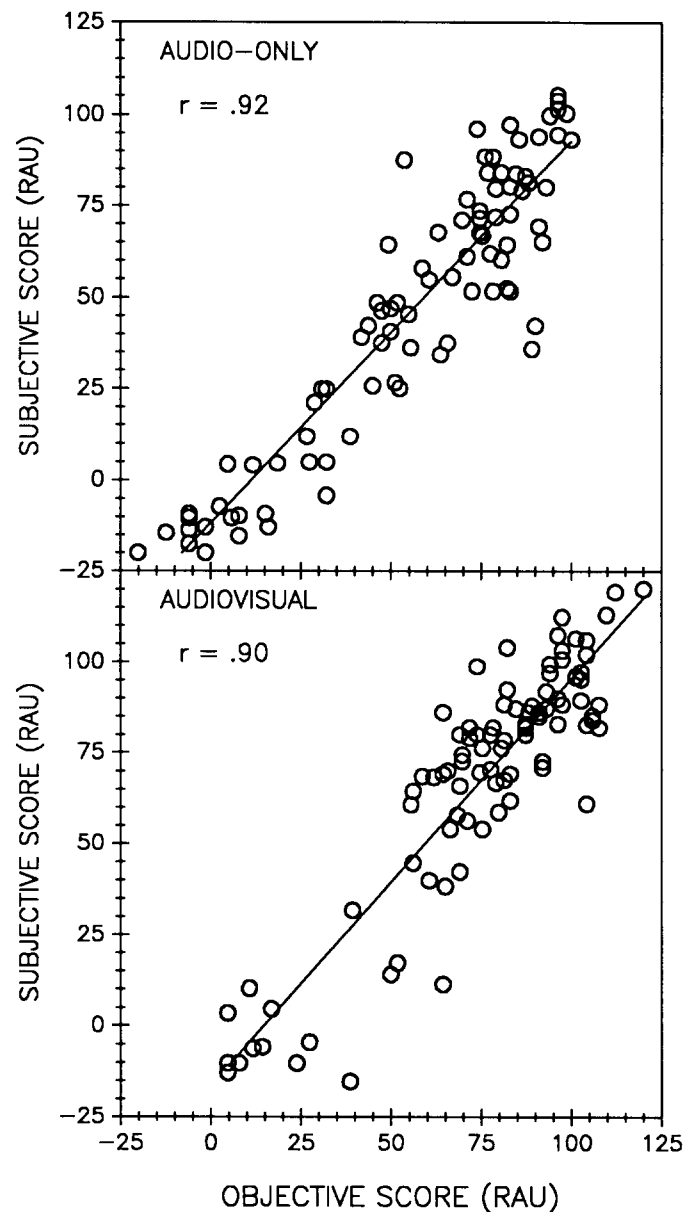
FIGURE 4. Objective and subjective intelligibility data for 24 hearing-impaired listeners. Each symbol depicts one pair of scores. The upper panel shows data for the audio-only presentation mode. The lower panel shows data for the audiovisual mode.

subjects' expectations of understanding speech: visual cues and amplification. Variables were scoring method (objective and subjective), stimulus presentation mode (audio and audiovisual), and listening condition (unaided and three aided). The subjects with missing data were omitted from this analysis. Results revealed a significant difference between the two types of score [$F(1, 20) = 14.9$, $p < .01$], with the mean objective score (66.5 rau) higher than the mean subjective score (57.8 rau). Not surprisingly, there was also a significant difference between the mean audiovisual score of 72.6 rau and the mean audio-only score of 51.6 rau [$F(1, 20) = 36.4$, $p < .01$]. Furthermore, there was a significant difference for the main effect of listening condition [$F(3, 60) =$

6.9, p < .01]. Post hoc testing using the Student-Neuman-Keuls procedure (α = .05) revealed that the mean unaided score was significantly poorer than the mean scores in the three aided conditions, but the scores did not differ across the three aided conditions.

No significant interactions emerged from this analysis. This indicated that the difference observed between subjective and objective intelligibility scores was about the same for both audio-only and audiovisual presentation modes. Furthermore, the finding that subjective scores were generally lower than objective scores was not affected when the subjects wore hearing aids. These results are illustrated in Figures 5 (presentation mode) and 6 (listening condition). Both figures show that the mean subjective score was about 9 rau less than the mean objective score, regardless of experimental condition.

Figure 7 summarizes the data for each subject by illustrating the mean objective score and mean subjective-objective score difference across all conditions. A negative subjective-objective difference indicates that, on average, the subject's objective score was higher than his or her subjective score. To explore a possible relationship between subjective-objective score differences and subject characteristics, correlation coefficients were computed between the average subjective-objective differences and SRTs, audiogram slopes, monosyllabic word discrimination scores, and ages. A modest correlation was found between age of the subject and the subjective-objective score differences [$r(22)$ = −.45, p < .05]. None of the other variables produced significant correlations. Figure 8 illustrates the relationship between subject age and score difference. Although the significant correlation suggests that as subject age increased the subjective-objective score difference became more negative, this outcome was largely due to the fact that the one subject who produced a positive difference score was also the youngest subject. If this subject is omitted from the analysis, the correlation between age and difference score becomes a nonsignificant .28.
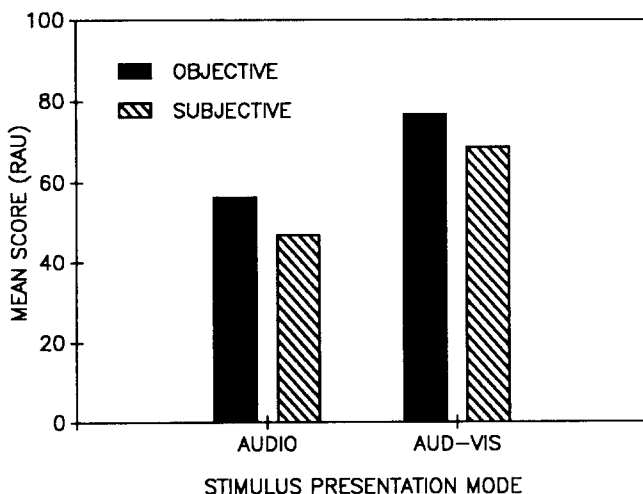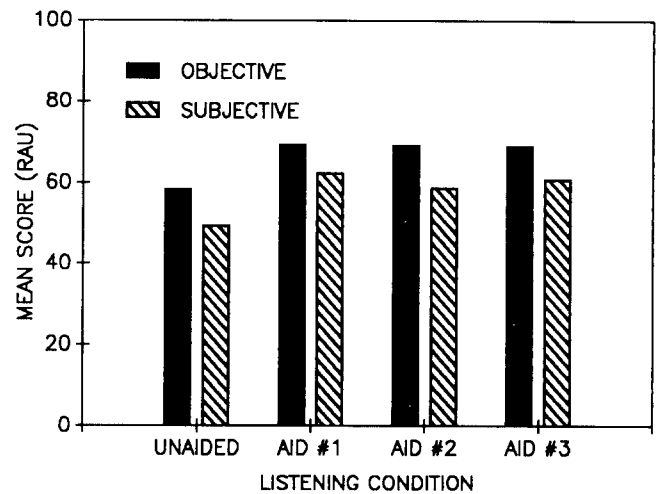

FIGURE 6. Mean objective and subjective intelligibility scores in the unaided condition and in three different hearing aid conditions.

The 24 subjects were separated into categories according to their mean subjective-objective difference. Eleven subjects fell into the defined small-difference category with mean differences of ±5 rau. Nine subjects fell into the defined large-negative-difference category with differences > −10 rau. Four subjects, designated "other," were in neither group. Data were compiled for each subject group in terms of duration of hearing loss, etiology of loss, and hearing aid experience. Results are summarized in Table 1. None of these variables seemed to distinguish among subjects in the various groups. Composite audiograms for the subjects in the small-difference and large-difference categories are shown in Figure 9. There were no significant differences between the two composite audiograms.
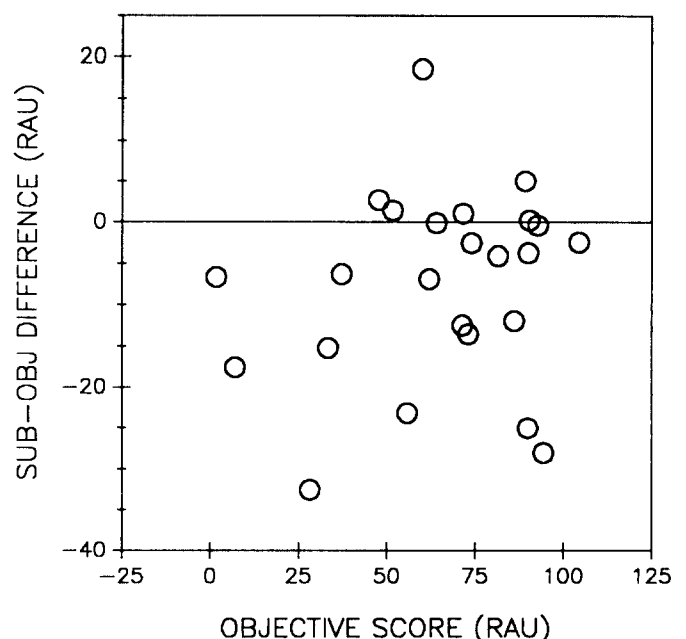

FIGURE 5. Mean objective and subjective intelligibility scores in audio-only and audiovisual stimulus presentation modes.

FIGURE 7. Mean difference between subjective and objective scores as a function of mean objective score for hearing-impaired listeners.
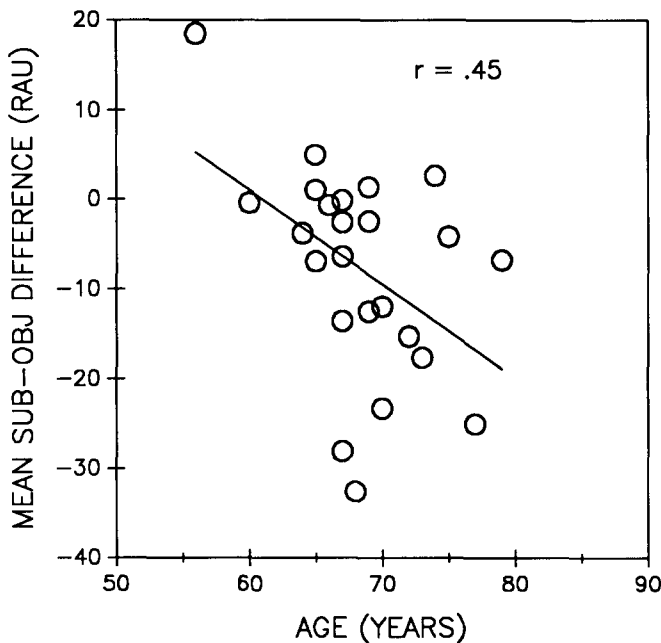
FIGURE 8. Relationship between mean subjective-objective score difference and age of hearing-impaired listeners. The linear regression line is shown.

Discussion

These results indicate that, for hearing-impaired listeners, there is a strong relationship between objective intelligibility and subjective intelligibility for the continuous discourse passages of the CST: About 83% of the variance on one measure can be attributed to the variance on the other measure. However, the mean objective and subjective intelligibility scores were not equal, as they were for the normal-hearing group. Instead, the subjective scores were systematically lower than the objective scores.

Manipulation of visual cues and amplification in an attempt to modify the hearing-impaired listener's expectations of understanding speech had no effect on the subjective-objective differential. This suggests that subjective intelligibility ratings were not highly susceptible to changing test conditions. This outcome gives some grounds for confidence that comparisons of subjectively estimated intelligibility across

TABLE 1. Statistics on etiology, duration of hearing loss, and hearing aid experience for subjects with "small," "large," and "other" mean subjective-objective differences.

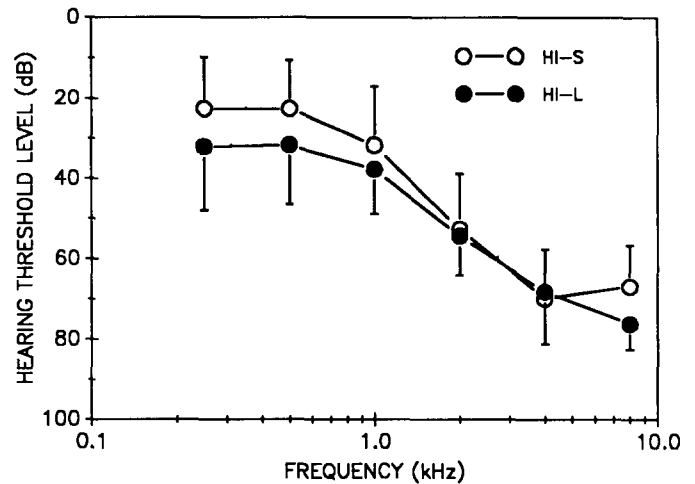| Difference category	N	Mean duration (years)	Nominal etiology (%)			Hearing aid experience (%)	
			NI	PR	Other	Yes	No
Small ±5 rau	11	18.4	55	18	27	55	45
Large >−10 rau	9	22.4	45	33	22	78	22
Other	4	33.3	75		25	75	25

Note. NI = noise-induced; PR = presbyacusis.

FIGURE 9. Composite audiograms for 9 hearing-impaired listeners with large negative discrepancy between objective and subjective scores (HI-L) and for 11 hearing-impaired listeners with small discrepancy between objective and subjective scores (HI-S).

different listening conditions would not be compromised by test factors unrelated to intelligibility.

The difference between objective and subjective intelligibility scores could not be predicted from a consideration of audiometric variables such as SRT, word discrimination score, or audiogram slope. Furthermore, this difference did not seem related to duration of hearing loss, etiology of loss, or hearing aid experience. As Figure 9 reveals, the typical subject who produced a large subjective-objective difference did not have a significantly different audiogram from the typical subject who produced a small difference. This suggests, in agreement with the filtered-speech data from Experiment 1, that the difference between objective and subjective scores was not precipitated by the acoustic filtering of high frequencies imposed by the hearing loss.

Despite the overall trend for subjective scores to be lower than objective scores, Figure 7 reveals that not all hearing-impaired subjects gave this result: Almost half of the 24 subjects had mean subjective-objective differences between plus and minus 5 rau, indicating good overall agreement between the two types of measures. In addition, as was noted above, one listener's subjective estimates were actually considerably larger than his objective scores. It was also of interest to note that there were no instances of close agreement between subjective and objective scores for subjects whose overall objective intelligibility scores were less than about 45 rau. When the mean objective score was lower than this, the mean subjective-objective differences were always negative.

These results clearly indicated an overall trend for subjective scores to be lower than objective scores for the elderly hearing-impaired listeners. Thus, despite the overall equality of the two intelligibility measures for young normal-hearing listeners, objective and subjective measurements of hearing aid performance could not be expected to give the same results for many elderly hearing-impaired persons.

However, hearing aid comparisons using objective and subjective procedures could give similar results for this group

of individuals if (a) differences between subjective and objective measurements were consistent for a given listener (so that different hearing aid conditions received similar relative scores by both measures), and (b) the difference between subjective and objective scores was independent of intelligibility performance (so that the subjective-objective difference did not change when intelligibility improved because amplification was used). To evaluate these issues, an additional study was undertaken.

EXPERIMENT 3: HEARING-IMPAIRED LISTENERS, RETEST

In the third experiment, several of the 24 subjects from Experiment 2 were retested. Time between first and second tests ranged from 1 to 16 months, with a mean test-retest interval of 10 months. This study explored the consistency of the subjective-objective differences given by individual subjects and the relationship of this difference to absolute performance level.

In addition, subjective estimates were obtained in a task involving visual stimuli. The purpose of this procedure was to assess whether subjects who produced large subjective-objective intelligibility differences would also produce significantly low subjective estimations in a task involving a different sensory modality.

Method

Subjects

Fifteen subjects were tested. Their mean age was 69 years (SD = 5.7). Average audiometric thresholds at octave frequencies from 250 Hz to 4000 Hz, in the test ears, were 23, 23, 34, 54, and 73 dB HL (mean SD = 11.7 dB). Individual speech reception thresholds ranged from 10 to 55 dB HL with a mean of 35 dB. Thirteen subjects were from the group of 24 hearing-impaired subjects who served in Experiment 2. These subjects included 8 individuals from the small-mean-difference category (called the HI-S subgroup), 4 individuals from the large-mean-difference category (called the HI-L subgroup), and the single individual who produced a large positive mean subjective-objective difference in Experiment 2. The 2 new subjects were elderly hearing-impaired individuals similar to the members of the original group for Experiment 2.

Procedure

Instrumentation and procedures for replaying and scoring the CST were similar to those used in Experiment 1. Subjects listened through an insert earphone (Etymotic ER-1) and stimuli were presented auditorily only.[3] Speech was deliv-

---

[3]Because neither hearing aids nor visual cues affected the subjective-objective score relationship in Experiment 2, there was no reason to continue their use in Experiment 3

ered to each subject at a level of (61 + SRT/2) dB L_{eq}, calibrated in a Zwislocki-type ear simulator.

Thirteen subjects listened at three different signal-to-babble ratios, and 2 subjects listened at two different SBRs. SBRs were varied so that a range of scores would be obtained from each individual. At each SBR, the subject responded to two different sets of six CST passages, producing an objective score and corresponding subjective score for each set. Nominal SBRs ranged from −6 dB to +12 dB.

For the visual estimation task, the subject was shown a grid composed of 100 squares, some of which contained a large dot. He or she was informed that there were 100 squares in the grid and asked to estimate the number of squares that contained a dot. The grid was viewed for 5 s, a length of time insufficient to count the dots. Fifteen different grids were used, three representing each of 22, 37, 49, 66, and 81 dots. They were viewed in random order. The visual estimation task always followed the auditory task. For comparison purposes, this task was also performed by a group of 10 normal-hearing subjects.

Results

The speech intelligibility data consisted of 4–6 pairs of objective and subjective scores per subject. Figure 10 illustrates the relationship between subjective and objective scores. Each symbol depicts one pair of scores. The correlation between the two types of scores was .85; the regression line is shown.

To explore any systematic differences between subjective and objective scores, these data were entered into a repeated measures analysis of variance (this analysis used data from the 13 subjects with six pairs of scores). Results
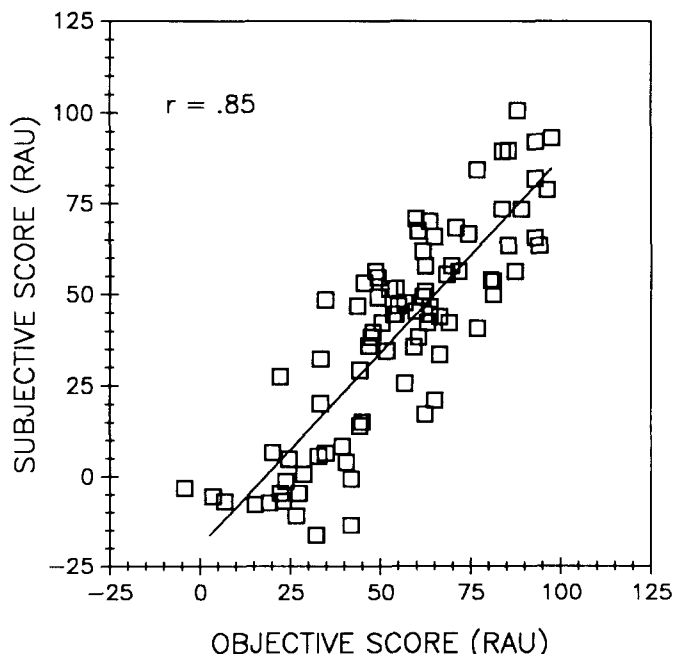

FIGURE 10. Objective and subjective intelligibility data for 15 hearing-impaired listeners, 13 of them from the group depicted in Figure 4. Each symbol depicts one pair of scores.

indicated a significant difference between mean subjective and mean objective scores [$F(1, 12) = 17.2$, $p < .01$], with the mean objective score (56.4 rau) higher than the mean subjective score (42.0 rau).

To determine whether subjects who produced larger subjective-objective score differences in Experiment 2 also produced larger differences in Experiment 3, two analyses were performed. First, the differences between subjective-objective score pairs were obtained for each subject in the HI-S and HI-L subgroups. These differences were entered into an analysis of variance to determine whether the two subgroups produced different subjective-objective differences in Experiment 3. Results indicated that the mean difference score of −11.8 rau for the HI-S subgroup was significantly smaller than the mean difference score of −30.4 rau for the HI-L subgroup [$F(1, 10) = 6.8$, $p < = .02$]. Second, a correlation coefficient was computed between each subject's mean difference scores in Experiments 2 and 3. The correlation coefficient was .72 ($df = 11$, $p < .01$), indicating a moderately strong relationship between the two sets of differences.

To evaluate the relationship between intelligibility and subjective-objective difference, two analyses were performed. First, a correlation coefficient was computed between the objective score and the corresponding subjective-objective difference for each pair of scores. The resulting correlation coefficient was .10. Second, in case a relationship between objective score and subjective-objective difference was being masked by intersubject differences, each subject's objective scores were ranked from low to high and the corresponding difference scores were assigned to low, medium, and high categories. Analysis of variance failed to reveal a significant difference between the scores in the three categories. Both of these outcomes were consistent with a conclusion that there was no overall relationship between intelligibility and score difference.

Figure 11 shows the results of the visual estimation task. These data reveal the relationship between the actual number of dots and the estimated number of dots for 10 normal-hearing subjects, 8 subjects in the HI-S subgroup, and 4 subjects in the HI-L subgroup. The figure suggests that the estimations provided by the HI-L subjects were lower than those from the two other groups for the two conditions with the largest numbers of dots. Analysis of variance of these data revealed a significant interaction between subject group and estimated number of dots [$F(8, 76) = 3.4$, $p < .01$]. Post hoc tests with the Student-Neuman-Keuls procedure ($\alpha = .05$) indicated that the estimates produced by the HI-L subjects in the 66-dots condition were significantly lower than those from the two other groups.

Finally, critical differences for subjective and objective intelligibility scores were determined from these data for hearing-impaired listeners, using the same approach as described in Experiment 1 for the normal-hearing listeners. For each type of score, differences were derived for each subject between the two sets of passages administered in each SBR condition. The 95% CD for the objective scores was 12.5 rau. For the subjective scores, the 95% CD was 22.5 rau.
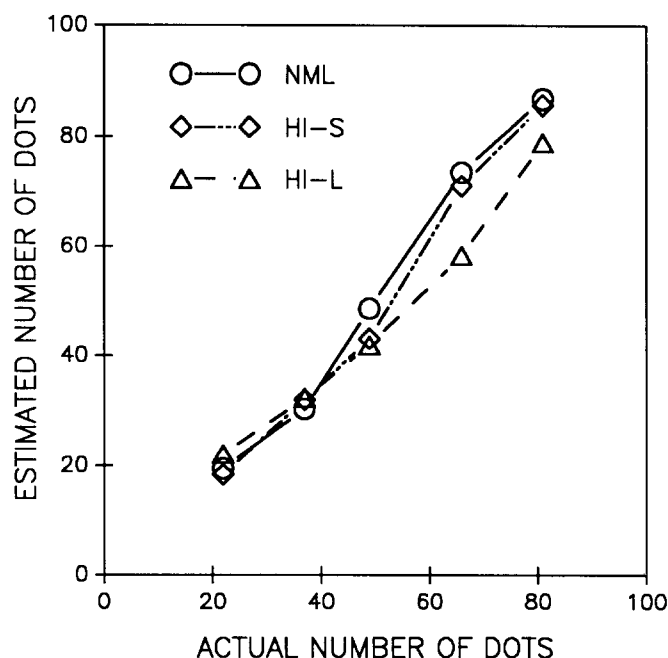
FIGURE 11. Relationship between actual and estimated number of dots in a 100-square grid. Data are given for three groups of subjects: 10 normal-hearing individuals (NML); 8 hearing-impaired individuals who produced small subjective-objective difference scores in an auditory task (HI-S); and 4 hearing-impaired individuals who produced large negative subjective-objective difference scores in an auditory task (HI-L).

Discussion

In this second set of data from the group of hearing-impaired listeners, the finding of a strong relationship between objective and subjective intelligibility scores was replicated. However, the average difference between the two types of scores was even more pronounced than in Experiment 2. This outcome suggests that the magnitude of subjective-objective differences may be sensitive to the test situation. In the realistic listening environments used in Experiment 2, the interscore differences were smaller than in the laboratory-type setting used for Experiment 3.

Subjects who produced a relatively large mean subjective-objective difference on the first occasion were likely to produce a relatively large mean subjective-objective difference on the second occasion also, despite the fact that the two measurements were obtained several months apart. In fact, about 50% of the variance in subjective-objective differences could be attributed to a consistent intersubject difference. The implication is that certain subjects are more likely than others to assign low scores to their own ability to understand speech. However, for a given subject in a particular test situation, these relatively low scores should be consistently produced. Thus, the subjectively estimated difference among several hearing aid conditions should be similar to the objectively measured difference.

The results of Experiment 2 had suggested a hypothesis that the magnitude of a subjective-objective difference was related to the magnitude of the objective score. When each subject performed at several different intelligibility levels in

the third experiment, this hypothesis was not substantiated: Better objective scores were not accompanied by smaller subjective-objective differences. This outcome implies that the magnitude of the subjective-objective difference produced by a particular listener would not be affected by differences in intelligibility that might be obtained, for example, between an aided condition and an unaided condition. Again, the result should be that relative differences among tested conditions would be similar for both objective and subjective measurements.

Hearing-impaired listeners who produced relatively low subjective intelligibility scores compared to their objective scores also tended to produce relatively low estimations in the visual task. This effect was statistically significant in one condition despite the small number of subjects. This outcome intimates that subjects who produced large discrepancies between their objective and subjective intelligibility scores were manifesting a tendency towards underestimation that was not limited to, or determined by, the auditory task.

The critical difference of 12.5 rau for objective scores was almost identical to the corresponding CD of 12.2 rau reported by Cox et al. (1989) for hearing-impaired subjects responding to the CST. The CD of 22.5 rau for the subjective estimates was similar to the 21.4 rau that can be derived from data reported by McDaniel (1988)[4] for hearing-impaired subjects responding to the SIR test. However, McDaniel's data were derived from subjective scores based on five separate estimates, whereas the present study used six separate estimates to determine the subjective score in each condition.

General Discussion

In this work, we undertook to evaluate the relationship between an objective measure of speech intelligibility and a corresponding subjective intelligibility estimation procedure. The overall purpose was to determine whether hearing aid benefit assessments for elderly hearing-impaired listeners would be likely to give the same results when the two measurement procedures were used. The results suggested that, despite differences in absolute scores, a comparative hearing aid evaluation using subjective intelligibility estimates would usually produce essentially the same result as a comparative evaluation using the objective intelligibility measurement procedure.

This is an auspicious outcome because subjective measurements have many practical advantages over objective ones. However, it is important to note that the critical differences for subjective estimates were substantially larger than those for objective scores. Furthermore, the CDs obtained in these studies for both normal-hearing and hearing-impaired listeners were in close agreement with analogous CDs drawn from several other investigations. Thus, it appears that the objective measurement procedure used here is considerably more reliable that the subjective procedure. The result would be that the subjective procedure is less sensitive to benefit

differences across hearing aid conditions, and larger differences between conditions would be required to produce a significant outcome. This situation could be remedied by obtaining more subjective estimates to generate the final score in each tested condition. However, this would diminish the administration-time advantage associated with the subjective procedure.

The most intriguing outcome of these studies was the finding that a substantial proportion of the elderly hearing-impaired listeners awarded themselves much lower subjective intelligibility scores than seemed warranted by their objective performance on the same material. That this was not a chance observation was established by the finding that, several months later, these same subjects again produced relatively low subjective scores compared to their objective scores. An overall trend towards relatively low subjective scores was not seen in the younger normal-hearing listeners.

Why should the two groups be different in the relationship between subjective and objective scores? The most obvious explanations involve the differences in hearing loss and/or age. Several factors suggest that these results cannot be attributed to the effects of hearing loss on the acoustic signal. First, the normal-hearing group did not reveal a parallel effect when the speech passages were filtered to simulate the effects of hearing loss (see Figure 3). Second, within the hearing-impaired group, those with large subjective-objective differences had about the same hearing loss, on average, as those with small subjective-objective differences (see Figure 9). Finally, the subjects who produced large subjective-objective auditory score differences also tended to generate lower subjective estimates in a visual task that was quite different from the auditory task and certainly was independent of hearing status (see Figure 11).

It is possible that the differences between subjective and objective intelligibility scores in the hearing-impaired group were related to the age of these subjects, as suggested by the correlation illustrated in Figure 8. A tendency to produce relatively low subjective estimates of intelligibility might be seen as a manifestation of the cautious response style that is mentioned frequently in the literature on aging (e.g., Okun, 1976). The fact that the relationship between subjective and objective data tended to be consistent across both auditory and visual tasks and the fact that the relative size of the auditory subjective-objective differential was consistent over a period of several months support the view that subjective intelligibility estimates are influenced to a substantial degree by the individual's overall response style.

Several investigators have applied the principles of signal-detection theory to determine whether older listeners are more cautious than younger listeners in responding to speech understanding tasks. With this approach, the subject provides both a response to a speech recognition test item and a rating of confidence in the correctness or incorrectness of the response. These data are used to assess the subject's criterion for judging the accuracy of his or her responses. Yanz and Anderson (1984) were unable to detect a difference between normal-hearing young and normal-hearing elderly listeners in their criteria for responding to a monosyllabic word recognition test. Gordon-Salant (1986) found that the two age groups did differ in response criteria when

---

[4]McDaniel reported a 95% CD of 2 14 scale intervals for hearing-impaired subjects responding to the SIR test. This would be roughly equivalent to 21 4 rau

responding to speech intelligibility tests, with older listeners displaying a *less cautious* criterion than younger listeners. Finally, Jerger, Johnson, and Jerger (1988) reported that, although elderly listeners displayed the gamut of lax, neutral, and strict response criteria, an individual's score on speech intelligibility tests was independent of his or her response criterion. Taken together, these studies to not support a hypothesis that elderly listeners as a whole are unusually cautious in their responses to speech intelligibility tasks or that response criterion is related to the outcome of an *objective* intelligibility test. However, the relationship between response criterion and subjective intelligibility scores has not been reported. It seems possible that response criterion might affect subjective scores even though it does not affect objective scores. In that case, response criterion would be related to the difference between objective and subjective intelligibility scores. Further research is needed to explore this possibility.

Acknowledgments

References

American National Standards Institute. (1969) American national standard specification for audiometers (ANSI S3.6-1969). New York: Author.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing, 8* (Suppl.) 119S–126S.

Cox, R. M., Alexander, G. C., Gilmore, C., & Pusakulich, K. M. (1988). Use of the Connected Speech Test (CST) with hearing-impaired listeners. *Ear and Hearing, 9,* 198–207.

Cox, R. M., Alexander, C. G., Gilmore, C., & Pusakulich, K. M. (1989). The Connected Speech Test Version 3: Audiovisual administration. *Ear and Hearing, 10,* 29–33.

Cox, R. M., & McDaniel, D. M. (1984). Intelligibility ratings of continuous discourse: Application to hearing aid selection. *Journal of the Acoustical Society of America, 76,* 758–766.

Cox, R. M., & McDaniel, D. M. (1989). Development of the Speech Intelligibility Rating (SIR) Test for hearing aid comparisons.

Journal of Speech and Hearing Research, 32, 347–352.

Falconer, G., & Davis, H. (1947). The intelligibility of connected discourse as a test for the threshold of speech. *Laryngoscope, 57,* 581–595.

Gordon-Salant, S. (1986). Effects of aging on response criteria in speech-recognition tasks. *Journal of Speech and Hearing Research, 29,* 155–162.

Jerger, J., Johnson, K., & Jerger, S. (1988). Effect of response criterion on measures of speech understanding in the elderly. *Ear and Hearing, 9,* 49–56.

McDaniel, D. M. (1988). *Evaluation of the speech intelligibility rating (SIR) test in hearing aid selection.* Unpublished doctoral dissertation, Memphis State University, Memphis, TN.

Okun, M. A. (1976). Adult age and cautiousness in decision: A review of the literature. *Human Development, 19,* 220–233.

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). *Speech levels in various noise environments.* EPA Rep No 600/1-77-025.

Peters, R. W. (1965). *A rating scale technique for the measurement of speech intelligibility* (AD-629 308). Springfield, VA: Clearinghouse for Federal and Scientific Technical Information, U.S. Department of Commerce.

Silverman, S. R., & Hirsh, I. J. (1955). Problems related to the use of speech in clinical audiometry. *Annals of Otology, Rhinology & Laryngology, 64,* 1234–1245.

Speaks, C., Parker, B., Harris, C., & Kuhl, P. (1972). Intelligibility of connected discourse. *Journal of Speech and Hearing Research, 15,* 590–602.

Studebaker, G. A. (1985). A 'rationalized' arcsine transform. *Journal of Speech and Hearing Research, 28,* 455–462.

Studebaker, G. A., Bisset, J. D., Van Ort, D. M., & Hoffnung, S. (1982). Paired comparison judgments of relative intelligibility in noise. *Journal of the Acoustical Society of America, 72,* 80–92.

Studebaker, G. A., Pavlovic, C. V., & Sherbecoe, R. L. (1987). A frequency importance function for continuous discourse. *Journal of the Acoustical Society of America, 81,* 1130–1138.

Walker, G., & Byrne, D. (1985). Reliability of speech intelligibility estimation for measuring speech reception thresholds in quiet and in noise. *Australian Journal of Audiology, 7,* 23–31.

Yanz, J. L., & Anderson, S. M. (1984). Comparison of speech perception skills in young and old listeners. *Ear and Hearing, 5,* 134–137.

Yanz, J. L., Carlstrom, J. E., & Thibodeau, L. M. (1985). Self-assessment of communication skills: Toward the development of a new speech audiometric tool. *Ear and Hearing, 6,* 211–215.