0196/0202/88/0904-0198$02.00/0
EAR AND HEARING

Vol. 9, No. 4
Printed in U.S.A.

BEHAVIORAL TECHNIQUES IN AUDIOLOGY AND OTOLOGY

Use of the Connected Speech Test (CST) with Hearing-Impaired Listeners*

Robyn M. Cox, Genevieve C. Alexander, Christine Gilmore, and Kay M. Pusakulich

Department of Audiology & Speech Pathology, Memphis State University [R. M. C.], and Veterans Administration Medical Center [R. M. C., G. C. A., C. G., K. M. P.], Memphis, Tennessee

ABSTRACT

Two studies were performed in which hearing-impaired subjects responded to the Connected Speech Test (CST). In experiment 1, 40 subjects, divided into four groups according to extent and configuration of hearing loss, responded to the CST version 1 (CSTv1). This version of the test consisted of 57 passages of connected speech: 48 test passages and 9 practice passages. It was developed on the basis of data for normal-hearing listeners. Performance of hearing-impaired listeners for the CSTv1 revealed that, although the passages were equal in average intelligibility for normal hearers, they were not equally intelligible for hearing-impaired persons. Based on results of data analyses, the 57 passages were reconstituted into 28 pairs of passages: 24 test pairs and 4 practice pairs. The pairs were equal in average intelligibility for both normal and hearing-impaired listeners. This form of the test was named the CST version 2 (CSTv2). In experiment 2, an additional 23 hearing-impaired subjects responded to the CSTv2. Critical differences and the slope of the signal to babble ratio (SBR) function were determined for the CSTv2 for hearing-impaired listeners. When two CSTv2 pairs were used per score, the 95% critical difference for hearing-impaired subjects was about 15.5 rationalized arcsine units (rau). The mean SBR function slope for hearing-impaired listeners was 8.5 rau/dB. Comparing the critical difference with the SBR function slope, it may be seen that, for hearing-impaired listeners, differences in intelligibility equivalent to a 2 dB change in SBR can be detected with CST scores based on mean performance across two passage pairs.

Because the primary purpose of most hearing aid fittings is to improve communication in everyday life, the benefit received from a hearing aid is mainly determined by the extent to which it facilitates understanding of everyday connected speech. The Connected Speech Test (CST) is a test of intelligibility of everyday speech which has been developed primarily as a criterion measure for investigations of hearing aid benefit. The overall objective was to produce a test with high content validity (i.e., consisting of conversationally produced connected speech), a large number of equivalent forms, and an acceptably small error of measurement (sufficient to detect an intelligibility change equivalent to a signal to babble ratio change of 2 dB).

The background and rationale for the test, the development of the CST version 1 (CSTv1), and data for normal-hearing listeners were presented in an earlier paper (Cox, Alexander, & Gilmore, 1987). The purpose of the present paper is to report investigations of the use of the CST with hearing-impaired listeners, to describe test modifications that have been made as a result of these studies, and to present the psychometric properties of the resulting version of the test.

Description of CST Version 1

The CST version 1 (CSTv1) was formulated on the basis of data obtained from normal-hearing listeners. This test consists of 57 passages of conversationally produced connected speech: 48 test passages and 9 practice passages. The passages were recorded audiovisually but the scoring procedure has been developed using only the audio portion of the test. A six-talker babble is recorded separately for use as a competing signal. Each passage concerns a familiar topic and the listener is apprised of the topic in advance. Passages contain 10 syntactically simple sentences, 7 to 10 words in length. To control word familiarity, the basic vocabulary was derived from a children's educational reading source.

Each passage contains 25 key words for scoring, five words in each of five levels of difficulty. The distribution of key word consonants in various phonetic categories is rather similar to the corresponding distribution reported by Fletcher (1953) for conversations (see Fig. 3 in Cox et al, 1987). To administer the test, each passage is presented one sentence at a time. The listener repeats each sentence exactly as (s)he hears it. The repetition is scored for the

* This work was supported by VA Rehabilitation Research and Development funds.

number of key words correctly repeated. Percent correct scores are transformed into rationalized arcsine units (rau) to minimize relationship between mean score and variance (Studebaker, 1985). All passages are of equal intelligibility for the average normal hearer. The signal to babble ratio (SBR) function (passage score as a function of level of competing babble) for normal hearers is 12 rau/dB. In the range of scores from 20 to 80% correct, this is approximately equal to 12%/dB (see "Results & Discussion" section for further discussion of rationalized arcsine units).

It is expected that a single intelligibility score will be based on the mean performance for several passages. The number of passages used per score should be chosen on the basis of the desired score reliability. For normal hearers, the 95% critical difference (the difference between two test scores that will be exceeded by chance alone on only 5% of comparisons) for two CST scores is about 14 rau. This assumes that each CST score is based on mean performance across four randomly chosen passages.

EXPERIMENT 1: GENERATION OF CST VERSION 2

It has been suggested many times that speech items that are equally intelligible for normal hearers may not be equally intelligible for hearing-impaired listeners, principally because of the filtering action of the hearing loss. For example, if two speech passages are equally intelligible for normal hearers but passage A contains more high-frequency sounds than passage B, these two passages probably would not be equally intelligible for a listener with sloping high-frequency hearing loss because passage A would have fewer audible sounds than passage B. Because of this kind of consideration, it is necessary to construct speech intelligibility tests using data from hearing-impaired as well as normal-hearing listeners. In experiment 1, four groups of hearing-impaired subjects listened to the CST. The purpose of this investigation was to determine whether modifications to the CSTv1 would increase its usefulness with hearing-impaired persons. The research questions were:

1. Are the CST passages that were essentially equal in average intelligibility for normal hearers also equal in average intelligibility for hearing-impaired listeners?
2. Does the intelligibility of the CST passages interact with the configuration of the listener's hearing loss?
3. Is the intelligibility of the CST passages related to the proportions of fricatives, plosives, and other phonetic categories in the keywords?
4. Is the equivalence of the CST test items improved if specific combinations of passages are always administered and scored together?

METHOD

Subjects

The subjects were four groups of 10 hearing-impaired adults, 34 men and 6 women, aged 38 to 84 yr with a mean age of 67. Group assignment was based on: (1) extent of hearing loss, defined by the speech reception threshold (SRT), and (2) configuration of hearing loss, defined as the average slope of the audiogram from 500 to 4000 Hz. Average audiograms from each group for the test ear are shown in Figure 1. Group definitions were: group A2 = SRT < 40 dB, slope 6 to 14 dB/octave; group A3 = SRT < 40 dB, slope > 14 dB/octave; group B1 = SRT 40 to 60 dB, slope 0 to 5 dB/octave; group B2 = SRT 40 to 60 dB, slope 6 to 14 dB/octave. All hearing losses were essentially sensorineural (no air-bone gaps >15 dB). Of the 40 subjects, 16 reported a history of significant noise exposure. These were called the noise-induced group. It should be noted, however, that the mean age of this group, 64 years, suggests a possible presbyacusic component in addition to noise exposure. Eighteen subjects, with a mean age of 71, were unable to report any factors that may have caused their hearing loss. These were called the presbyacusic group (Fig. 2). Etiology was less clearly established for the other six subjects.

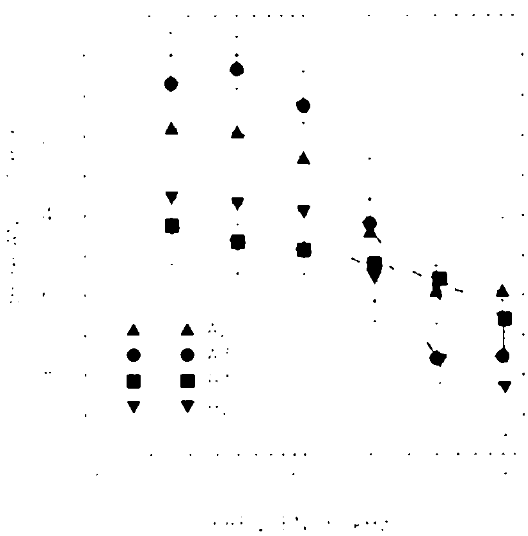

Figure 1. Average test ear audiograms for each hearing-impaired group in experiment 1. Error bars give 1 SD.

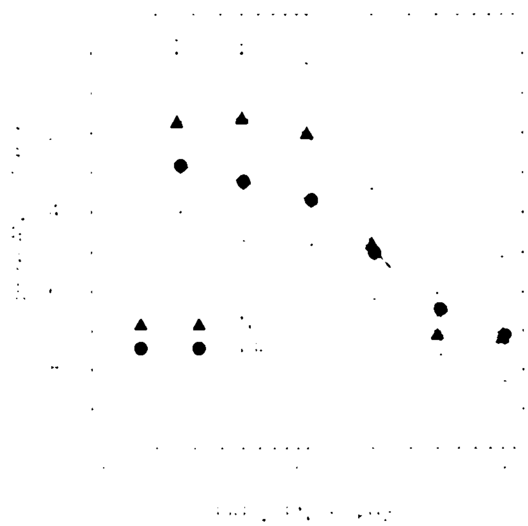

Figure 2. Average test ear audiograms for two hearing loss etiology groups in experiment 1. Error bars give 1 SD. NI, noise-induced; PR, presbyacusis.

Instrumentation for CST Presentation

The CST passages and competing babble had been recorded on three 2-channel optical disks. They were replayed by an optical disk player (Panasonic TQ 2024F) and routed to attenuators to allow independent adjustment of signal to babble ratio. The two outputs were then mixed, amplified, and delivered to an insert earphone system (Etymotic Research ER-1) that was coupled to the listener's ear using a compressible foam earplug. In the range 150 Hz to 11 kHz, this playback system delivered the same frequency response to the average eardrum as would have occurred at that location during open-ear listening in a diffuse sound field.

Calibration of the playback system was achieved with the output of the insert earphone delivered to a Zwislocki-type ear simulator coupled to a precision sound level meter. The frequency response of the system was monitored daily.

Procedure

Data collection required three test sessions per subject. The first session was used to test hearing thresholds for pure tones and speech. These data were used for group assignment. In addition, the subject's speech production was screened: subjects were required to produce standard American dialect to avoid confounding intelligibility of CST passages with speech production. Finally, the subject practiced responding to the CST test. Four to eight practice passages were used: some of the initially recorded 72 passages that were not used in CSTv1. During the practice, SBR was varied and a level was selected for subsequent testing. The SBR was adjusted for each subject with the aim of producing intelligibility scores in the 50 to 80% range. Nominal SBRs ranged from +7 to 0 dB.

The second and third test sessions were devoted to presentation and scoring of the 57 CST passages. Subjects were seated in a sound treated room viewing a video monitor that briefly displayed the passage topic before (but not during) passage presentation. They listened monaurally; the untested ear was occluded with a compressible foam earplug. Because of the large interaural attenuation associated with the ER-1 insert earphones (Killion, Wilbur, & Gudmundsen, 1985) the nontest ear was excluded from participation in the test. Delivery and scoring of the CST passages was controlled by an Apple IIe microcomputer. Each passage was presented one sentence at a time. After each sentence, both speech and babble were halted while the subject repeated the sentence or as much of it as (s)he had heard. It was emphasized that subjects were to repeat every word exactly as heard. The examiner sat across from the subject, viewing a second video monitor. The key words for the sentence were displayed on this monitor. The examiner scored the words correctly identified by entering the corresponding number on a keypad. Words containing additions, substitutions, or omissions were scored as incorrect.

Test passages were presented at the level of normal conversational speech in everyday environments plus one-half the subject's SRT. Pearsons, Bennett, and Fidell (1977) observed that the level of everyday conversational speech was 55 dB Leq outside the listener's ear (Leq = integrated, A-weighted sound level). Measurements made by the authors using a Kemar manikin revealed that this corresponded to 61 dB Leq at the average eardrum in a diffuse sound field. Hence, passages were delivered to subjects at a level that produced (61 + SRT/2) dB Leq in the Zwislocki-type ear simulator. A gain of SRT/2 was chosen because several studies have shown that, on the average, listeners with sensorineural hearing loss prefer to use an amount of hearing aid gain equal to about half their hearing loss (e.g., Byrne & Fifield, 1974). The maximum gain available was 25 dB. Hence, the 3 subjects whose SRTs exceeded 50 dB received slightly less gain than the other subjects. However, all subjects reported that the passages were presented at a comfortable loudness. Each subject heard all 57 passages at a constant SBR.

In each session, two additional practice passages were administered before data collection to refamiliarize the subject with the task and the talker. Practice passage data were not included in subsequent analyses. All experimental variables were counterbalanced or randomized to minimize order effects.

RESULTS AND DISCUSSION

The data consisted of correct/incorrect scores for 25 key words in 57 passages for 40 subjects. Percent correct scores were derived for each passage for each subject. Prior to statistical analyses, all percentage scores were transformed into rationalized arcsine units (rau) as described by Studebaker (1985). This had the effect of minimizing the relationship between mean score and variance that is characteristic of percentage scores while at the same time providing a scoring unit similar to percentages and, therefore, readily interpreted. For tests based on 50 or more words, rationalized arcsine units are within 1.3 units of the corresponding percentage value for scores in the range 12% to 88% (for example, 68% = 66.8 rau). As percentage scores increase above 88% or decrease below 12%, the corresponding rau value deviates progressively from the percentage value.

To facilitate comparison of CST results for hearing-impaired persons with those obtained for normal hearers, results for a group of 10 normal hearers were included in some statistical analyses. The data for normal hearers were obtained in a validation study of the CSTv1. Some analyses of these data for the 48 CSTv1 test passages were reported in Cox et al (1987). For the present study, analyses of data for normal hearers were performed using data from all 57 CST passages.

Equivalence of CST Passages

To evaluate the overall equivalence of the 57 CST passages, mean scores were determined for each passage in each subject group. The overall mean and standard deviation (SD) across passages for each group are shown in Table 1. Because subjects listened at different SBRs, differences in mean scores among groups are not very informative. However, the variability across passages is indicative of the overall equivalence of the passages for each group. All hearing-impaired groups had larger SDs than the normal-hearing group, indicating that the passages were more equivalent for the normal hearers than for the hearing-impaired listeners.

Variability in scores across passages could result from systematic differences in passage intelligibility (i.e., some passages always more difficult than others), from random measurement error, or from a combination of these two factors. To evaluate the extent to which systematic differences in passage intelligibility contributed to the variability

Table 1. Mean scores and standard deviations (SDs) across 57 CSTv1 passages. Results are given for four groups of hearing-impaired listeners (A2, A3, B1, B2), one group of normal hearers (NmL), and for hearing-impaired subjects grouped according to etiology of loss (NI, noise-induced; PR, presbyacusis). Data are in rationalized arcsine units (rau).

Group	Mean	SD
NmL	59.6	3.8
A2	72.9	6.2
A3	69.6	6.8
B1	72.5	5.7
B2	62.6	7.4
NI	66.5	6.6
PR	70.4	5.4

in scores, between-subject linear correlation matrices were computed for each hearing-impaired group. In this procedure, the scores for the 57 passages for each subject were correlated with the corresponding scores for each other subject in the same group. It was anticipated that if there were sizable systematic differences in passage difficulty (i.e., the ordering of passage scores was similar for many subjects), numerous significant between-subject correlations would be found. On the other hand, if passage score differences were due mostly to random error, significant between-subject correlations would not be expected.

A similar analysis for 10 normal hearers, reported in Cox et al (1987), had revealed only one significant correlation coefficient among the 45 generated, indicating that, on the whole, differences among passage scores for individual subjects could not be attributed to systematic variations in passage intelligibility. In the present study, 45 correlation coefficients were generated for each hearing impairment group. Groups A2, A3, B1, and B2 produced 7, 12, 5, and 11 significant ($p < 0.01$) correlation coefficients, respectively. Because the hearing-impaired listeners showed significant between-subject correlations in about 19% of comparisons compared to 2% in the normal hearers, this outcome suggests that the ordering of passage intelligibility was more consistent for the hearing-impaired listeners. That is, there was more of a tendency for certain passages to be more difficult than other passages for hearing-impaired listeners than for the normal hearers.

Because the four groups of hearing-impaired subjects had different audiometric configurations, it seemed possible that the ordering of passage intelligibility would be different for different groups. For example, a passage that was particularly unintelligible for listeners in group A3 might not be especially difficult for listeners in group B1. To explore this possibility, the average score was determined for each passage for each hearing-impaired group. A correlation matrix was then derived in which the average passage scores from each group were correlated with the corresponding scores from each other group. The six correlation coefficients generated ranged from 0.49 to 0.72 and all were significant ($p < 0.001$). This outcome indicates that when SBR was individually adjusted to present a moderately challenging listening condition, and the effects of random measurement error were partially removed by averaging scores across subjects, the relative intelligibility of the 57 CST passages was similar for all hearing loss groups.

In summary, analyses of the data suggested that: (1) the variability of scores for the 57 passages of the CSTv1 was greater for hearing-impaired listeners than for normal hearers, (2) some passages were systematically more difficult for hearing-impaired listeners than other passages, and (3) relative passage intelligibility did not interact with audiometric configuration.

Relationship between Intelligibility and Phonetic Categories

As noted earlier, the mean proportion of consonants in various phonetic categories in the keywords used to score the CST passages is similar to corresponding proportions in conversations. Table 2 gives the means and SDs (in rau) for each phonetic category across the 57 CSTv1 passages (the data for 24 pairs of CSTv2 passages reported in Table 2 are discussed below). Because the analyses described above showed that there was a tendency for certain passages to be less intelligible than other passages, it was of some interest to determine whether passage intelligibility was related to the proportions of keyword consonants in each phonetic category. To explore this issue, stepwise linear multiple regression analyses were performed. Using these analyses it was possible to estimate the percentage of variation in passage intelligibility scores that could be attributed to variations in proportion of keyword consonants in the different phonetic categories. Separate analyses were performed for each hearing-impaired group and for the four groups combined. In addition, analyses were performed with the subjects regrouped into the two main etiological categories represented: noise-induced loss and presbyacusis. Table 3 shows the results.

Table 3 indicates that the intelligibility of the CST passages was not significantly related ($p < 0.05$) to the proportions of nasals, voiceless plosives, or voiced plosives for any group of subjects. However, the normal hearers revealed a significant (negative) relationship between pas-

Table 2. Mean and standard deviation (SD) for keyword consonants in each phonetic category across 57 CSTv1 passages and across 24 pairs of CSTv2 passages. Data for each passage were originally derived in proportion of total consonants and then converted to rationalized arcsine units.

	57 Passages		24 Passage Pairs	
Phonetic Category	Mean	SD	Mean	SD
Nasals	16.4	6.2	16.1	3.9
Voiceless plosives	18.0	4.9	17.9	3.3
Voiced plosives	11.2	5.4	11.5	4.0
Voiceless fricatives	15.8	5.7	15.9	1.5
Voiced fricatives	6.7	5.7	7.0	1.1
Other: /l,r,j,w,h/	24.4	5.7	24.4	1.8

Table 3. Results of stepwise multiple regression analyses between mean passage intelligibility scores for 57 CSTvl passages and proportion of keyword consonants in different phonetic categories in each passage. Data are given for normal hearers (Nml), hearing-impaired listeners in four groups according to audiogram (A2, A3, B1, B2), hearing-impaired listeners in two etiological categories (NI, noise-induced; PR, presbyacusis), and for all hearing-impaired subjects combined (Comb). X indicates a phonetic category that was significantly related ($p < 0.05$) to passage intelligibility for that group. The total % is the percentage of variation in passage intelligibility score that can be attributed to the variation in the significant category(ies).

Group	Nasal	Voiceless Plosives	Voiced Plosives	Voiceless Fricatives	Voiced Fricatives	Other	Total %
Nml	—[a]	—	—	—	X	—	25
A2	—	—	—	—	—	—	
A3	—	—	—	—	X	X	17
B1	—	—	—	—	—	—	
B2	—	—	—	—	—	—	
NI	—	—	—	—	—	X	10
PR	—	—	—	X	X	—	16
Comb	—	—	—	—	—	—	

[a]Dash indicates no significant relationship.

sage intelligibility and proportion of voiced fricatives: variation in proportion of voiced fricatives accounted for 25% of the variance in normal hearers' scores. Group A3 revealed a significant relationship between intelligibility and proportions of consonants in two categories [voiced fricatives (negative) and "other" (positive)]. Together, these two categories accounted for 17% of the variation in A3 scores. The noise-induced loss group also showed a (positive) relationship between intelligibility and proportion of other consonants, accounting for 10% of the variance of passage scores in this group. Finally, the presbyacusic group revealed significant (negative) relationships between intelligibility and proportions of voiced and voiceless fricatives. Together, the fricative consonants accounted for 16% of the variance in passage scores for presbyacusic listeners.

This outcome suggested that for the A3, noise-induced, and presbyacusic hearing-impaired groups, the equivalence of the CST passages could be improved if the variation across passages in proportions of fricatives and other consonants was reduced. This was accomplished by combining the passages in designated pairs. The pairs were constructed so that the combined proportions of voiced fricatives, voiceless fricatives, and other consonants (in keywords) varied as little as possible from the overall mean of all passages. Twenty-eight passage pairs were constructed.

To evaluate the relationship between intelligibility of the passage pairs and proportions of keyword consonants in the various categories, stepwise multiple regression analyses were performed, as before, on the normal-hearing and hearing-impaired subject groups. Based on the results of these analyses, four pairs of passages were designated practice pairs and the remaining 24 pairs were considered the test items. The means and standard deviations, across the 24 test pairs, for proportions of keyword consonants are given in Table 2. The results of multiple regression analyses to examine the relationship between categories of keyword consonants and intelligibility of test pairs are shown in Table 4.

Table 4 shows that most of the significant relationships between intelligibility scores and proportion of keyword consonants disappeared when the passages were paired. However, in spite of the fact that the variation across pairs in proportion of voiced fricatives is very small (see Table 2), Table 4 shows that this variation accounted for 24% of the variability in intelligibility scores for the normal hearers. Examination of the data revealed that this outcome was due principally to the unusually high mean score obtained for one pair of passages. Because this pair of passages did not produce anomolous scores in any of the hearing-impaired groups, it was retained. Another noteworthy outcome is seen in Table 4 for the presbyacusic group. For this group of listeners, the pairing process removed the relationship between intelligibility and voiced fricatives but strengthened the (negative) relationship between intelligibility and voiceless fricatives. Examination of the data failed to reveal any unusual features that would account for this relationship. Apparently, in spite of the small variation across pairs in proportion of voiceless fricatives, this factor was significantly related to the intelligibility score for presbyacusic listeners. However, because the variation in scores across passages was small in absolute terms (see Table 5), this relationship is not likely to be a problem in the use of the test.

Mean scores and standard deviations across the 24 test pairs for each subject group are given in Table 5. Comparison of Table 5 with Table 1 indicates that the process of reducing 57 passages to 24 passage pairs had a negligible effect on the mean scores. However, calculations reveal that the use of designated passage pairs resulted in smaller variability across pairs than would have been obtained by random pairing of passages. This additional reduction of variability was seen in every group of subjects. The varia-

Table 4. Results of stepwise multiple regression analyses between mean intelligibility scores for 24 test passage pairs of CSTv2 and proportion of keyword consonants in different phonetic categories in each pair. Data are given for normal hearers (Nml), hearing-impaired listeners in four groups according to audiogram (A2, A3, B1, B2), hearing-impaired listeners in two etiological categories (NI, noise-induced; PR, presbyacusis), and for all hearing-impaired subjects combined (Comb). X indicates a phonetic category that was significantly related ($p < 0.05$) to passage intelligibility for that group. The total % is the percentage of variation in pair intelligibility scores that can be attributed to the variation in the significant category(ies).

	Phonetic Category						
Group	Nasal	Voice less Plosives	Voiced Plosives	Voice less Fricatives	Voiced Fricatives	Other	Total %
Nml	—*	—	—	—	X	—	24
A2	—	—	—	—	—	—	
A3	—	—	—	—	—	—	
B1	—	—	—	—	—	—	
B2	—	—	—	—	—	—	
NI	—	—	—	—	—	—	
PR	—	—	—	X	—	—	22
Comb	—	—	—	—	—	—	

*Dash indicates no significant relationship.

Table 5. Mean scores and standard deviations (SDs) across 24 test passage pairs of CSTv2. Results are given for four groups of hearing-impaired listeners (A2, A3, B1, B2), one group of normal hearers (Nml), and for hearing-impaired subjects grouped according to etiology of loss (NI, noise-induced; PR, presbyacusis). Data are in rationalized arcsine units (rau).

Group	Mean	SD
Nml	59.5	2.5
A2	72.5	3.8
A3	69.3	4.2
B1	71.6	3.2
B2	61.7	3.8
NI	65.9	3.5
PR	69.7	2.7

bility for hearing-impaired subjects was reduced proportionately more than the variability for normal hearers.

To evaluate the extent to which a score for one passage pair could be used to estimate a hearing-impaired individual's overall score for the entire test, linear correlation coefficients were computed for each passage pair between pair score and overall score for each hearing-impaired subject. This yielded 24 correlation coefficients ranging from 0.68 to 0.90 with an average value of 0.80. All were statistically significant ($p < 0.001$). These correlation coefficients are somewhat smaller than similar coefficients computed for normal hearers and reported in Cox et al (1987). This difference is probably largely due to the more restricted range of overall scores across subjects in the present study.

Finally, the internal consistency of the 24 passage pairs was evaluated through computation of coefficient α (Nunnelly, 1978) for the 40 hearing-impaired subjects. Although this statistic has several interpretations, its most useful interpretation in this context is as a measure of the inter-relationships among passage pairs. That is, α is a measure of the extent to which the various passage pairs are measuring the same abilities. For the 24 pairs of CST passages, $\alpha = 0.98$, indicating a high degree of interpair similarity.

EXPERIMENT 2: VALIDATION OF CST VERSION 2

On the basis of the data obtained in experiment 1, the original 57 passages of the CSTv1 were reconstituted as 28 pairs of passages: 24 test pairs and 4 practice pairs. The test pairs were essentially equivalent in intelligibility for the group of 40 hearing-impaired adults with hearing losses up to about 60 dB HL. This test was named the CST, version 2 (CSTv2). In experiment 2, another group of hearing-impaired subjects was tested using the CSTv2.

Experiment 2 had two purposes: First, to provide additional data on within-subject variability in intelligibility of passage pairs. These data were necessary for determination of critical differences for the CSTv2. Forty estimates of within-subject variability for CSTv2 pairs were available from experiment 1. However, because these were obtained retrospectively, using scores from the CSTv1, it seemed prudent to obtain additional estimates using the final form of the CSTv2. The second purpose was to determine the slope of the signal to babble ratio function for the CSTv2 for hearing-impaired listeners.

METHOD

Subjects

The subjects were 24 hearing-impaired adults, 17 men and 6 women, aged 46 to 84 yr with a mean age of 69. Six subjects were drawn from each of the four hearing impairment categories used in experiment 1 (groups A2, A3, B1, and B2). The 24 subjects were divided into two subgroups (see "Procedure"). Each subgroup contained 3 members from each hearing impairment category.

Instrumentation for CST Presentation

The CSTv2 passages, together with the competing speech babble, were redubbed onto optical laser disks. Because the original 72 passages had been reduced to 56, the final version could be accommodated on two disks instead of the three required for the initial test material. Each disk contained 12 pairs of test passages, 2 pairs of practice passages, and a segment of speech noise shaped to resemble the long-term average speech spectrum of the talker. The speech-shaped noise was used to calibrate the speech and babble levels.

The playback and calibration arrangements were identical to those described for experiment 1.

Procedure

Three sessions were required for data collection. The first session was identical to its counterpart in experiment 1 in terms of hearing testing, group assignment, and screening of speech production. During the practice session, subjects listened first to 6 to 10 "learning" passages. These learning passages were drawn from the pool of passages recorded but eliminated from the test in the earlier investigations. Next, the four practice pairs from the CSTv2 were presented as SBR was varied to grossly define the outline of the subject's signal to babble ratio (SBR) function for the test material. Finally, two SBRs were selected for presentation of the test passages. The two SBRs were 2 dB apart and were chosen with the aim of producing intelligibility scores above 30% and below the subject's SBR function asymptote. For two subjects, two SBRs satisfying these criteria could not be identified because their SBR function asymptote was relatively close to 30%. These subjects received all test passages at one SBR. Nominal SBRs ranged from +12 to +2 dB.

The second and third test sessions were devoted mainly to presentation and scoring of the 24 test passage pairs, 12 per session. Half of the subjects (subgroup 1) received the first 12 pairs at one SBR and the second 12 paris at the other SBR. These data were used to assess the within-subject variability of the passage pairs for the 12 pairs per session. Use of two SBRs per subject allowed measurement of variability at a variety of performance levels. In addition, one estimate of SBR function slope was derived per subject from the difference between the scores obtained in the two test sessions.

For the other 12 subjects (subgroup 2), the SBR condition was alternated between the two selected levels after each three passage pairs. For these subjects, SBR function slope was derived in each test session from the difference between scores at the two SBRs.

Delivery and scoring of test items was the same as described for experiment 1 except that a Zenith 181 microcomputer controlled the optical disk player. Presentation level of the test passages was determined as in experiment 1. In each session, 8 to 10 nontest passages were presented before data collection to refamiliarize the subject with the task and the talker. All experimental variables were counterbalanced or randomized to minimize order effects.

RESULTS AND DISCUSSION

Data consisted of correct/incorrect scores for 50 keywords in each of 24 pairs of passages. Percent correct scores were derived for each pair for each subject. For statistical analyses, percentage scores were again transformed into rationalized arcsine units. Because of an instrumentation problem, data for one subject in subgroup 2 were discarded, reducing this group to 11 subjects.

Determination of Critical Differences

When the CST is used to quantify intelligibility differences between conditions for groups of listeners, the significance of differences can be determined using routine statistical procedures. However, one of the major applications for the CST is to assess intelligibility under different conditions (such as aided and unaided) for a single individual. In order to evaluate the significance of differences between two scores obtained for a particular individual, it is essential to know the amount of difference likely to occur between the two scores by chance. In this context, a 95% critical difference (CD) is defined as the difference between two test scores that will be exceeded by chance alone on only 5% of comparisons. An observed difference between aided and unaided scores that is greater than the 95% CD is probably due to the effects of the hearing aid.

Two approaches to establishing critical differences for the CSTv1 were described by Cox et al (1987). In the first approach, data obtained using the CSTv1 and normal-hearing listeners were compared to performance predicted using the binomial model suggested by Thornton and Raffin (1978). As Thornton and Raffin demonstrated, the binomial model can be used to determine critical differences for monosyllabic word intelligibility scores. However, the CST does not fully comply with the assumptions necessary for application of the binomial model because the individual test words are not independent of each other (due to effects of sentence context). Nevertheless, as reported by Cox et al (1987), it was found that application of the binomial model yielded fairly accurate estimates of CD for the CSTv1 with normal-hearing listeners.

To evaluate the accuracy with which the binomial model proposed by Thornton and Raffin (1978) could predict within-subject variability for the CSTv2, data obtained for normal-hearing and hearing-impaired listeners were compared with performance predicted using the model. The within-subject variability across the 24 passage pairs (50 test words per pair) was compared with the variability that would be predicted using the binomial model. Each subject's true score was estimated using the mean score for all pairs. Figure 3 gives data for the 40 hearing-impaired subjects from experiment 1 and for the 40 normal-hearing subjects tested in the earlier study (Cox et al, 1987). Note that, as required for this comparison, data in Figure 3 reflect percentage scores. The solid curve was fitted to the 80 data points using a least squares method. The dashed curve gives the relationship predicted by the binomial model between true score and variability of CSTv2 passage pairs.

Figure 3 reveals that the within-subject variability measured for both normal and hearing-impaired subjects for the CSTv2 was greater, on the whole, than variability predicted using the binomial model. In addition, although there is considerable overlap between the normal-hearing

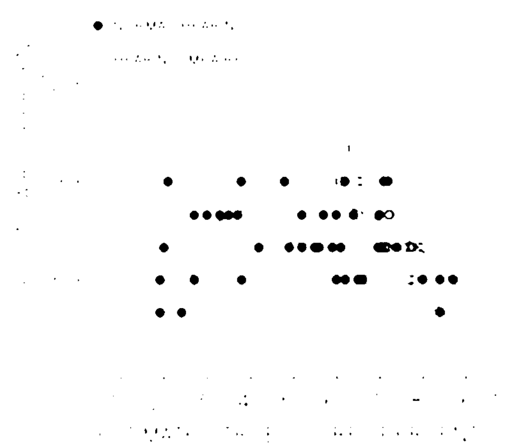

Figure 3. Performance for normal and hearing-impaired listeners on the CSTv2 compared with performance predicted using a binomial model. Data shown for normal hearers were originally obtained in an earlier investigation (Cox et al, 1987). Data for hearing-impaired subjects were obtained in experiment 1. The solid curve was fitted to all data points using a least squares method. The dashed curve gives the performance predicted by the binomial model.

and hearing-impaired groups, there is some suggestion that the within-subject variability may be greater for hearing-impaired listeners. On the basis of these data, it was concluded that an approach employing the binomial model was not appropriate for estimation of critical differences for either normal or hearing-impaired listeners for the CSTv2.

In the second approach to critical difference determination, Cox et al estimated critical differences empirically for the CSTv1 using the within-subject variability measured for 40 normal-hearing subjects. Similarly, the second approach to estimating critical differences for the CSTv2 was based on the measured variability of scores across passage pairs for each individual. Using this approach, the 95% critical difference can be calculated using the following equation:

$$\text{CD (rau)} = \frac{2.8(\text{SD})}{\sqrt{n}} \sqrt{\frac{24 - n}{23}}$$

Where n = number of CST passage pairs used per score and SD = typical within-subject variability in passage pair scores. If a 90% CD is desired, 2.3 should be substituted for 2.8 in the above equation. This equation is based on computation of the standard deviation of the distribution of differences between pairs of mean scores, each based on n CSTv2 pairs selected randomly from the population of 24 pairs (Ferguson, 1969).

Mean overall score and standard deviation (SD) across pairs were calculated for each subject (note that these SDs were for the population of passage pairs, not unbiased estimates of SDs for all possible similar pairs). For subjects from experiment 1, these were based on 24 pairs of passages. For experiment 2, calculated means and SDs were based on the 12 pair scores measured at each SBR for subgroup 1 subjects. Figure 4 illustrates these data. Inspection of Figure 4 reveals that there is no curvilinear relationship between mean score and variability (in contrast to the data seen in Fig. 3, based on percentage scores). This is an expected and desirable consequence of the conversion from percentage to rationalized arcsine scores. Because of this, it is possible to generate a critical difference that is independent of a subject's performance on the test. In addition, Figure 4 indicates that the variability for individual subjects across passage pairs was about the same for the hearing-impaired groups in both experiments. The SD for the entire group of subjects in experiment 1 was 8.3 rau. For the subjects in experiment 2, the overall SD was 7.7 rau. For both groups combined, the overall SD was 8.0 rau.

Adopting 8.0 rau as the typical within-subject variability in passage pair scores for hearing-impaired subjects, and assuming two passage pairs are used per score, the 95% CD is about 15.5 rau.

For comparison purposes, the data for 40 normal-hearing subjects obtained using CSTv1 were rescored to reflect the results that would have been obtained using CSTv2. Mean overall score and standard deviation (SD) across passage pairs were calculated for each subject as for the hearing-impaired groups. The typical within-subject SD for normal hearers was found to be 7.3 rau, a value somewhat smaller than observed for the hearing-impaired groups. Using this value and assuming two passage pairs per score, the 95% CD for normal hearers is about 14 rau.

Signal to Babble Ratio Function Slope

One method of evaluating the critical difference for a speech intelligibility test is to compare it with the SBR

Figure 4. Within-subject variability across CSTv2 passage pairs as a function of estimated true score (mean across all pairs) for hearing-impaired subjects.

function slope observed for the test materials. The slope of the SBR function denotes the rate at which intelligibility scores change as the competing babble is decreased. Using this metric, it is possible to assess the sensitivity of the test. Recall that one of the design objectives for the CST was to develop a test that would be sensitive enough to detect an intelligibility change equivalent to a change of 2 dB in SBR. For normal hearers, the SBR function slope for the CSTv1 was 12 rau/dB. When the data for normal hearers were rescored to reflect the results that would have been obtained on the CSTv2, the SBR function slope was again found to be 12 rau/dB. This outcome, together with the critical difference of 14 rau reported above, indicates that for normal hearers, intelligibility differences equivalent to 2 dB change in SBR (24 rau) can be detected with CST scores based on mean performance across two passage pairs.

Several investigators have reported that SBR function slope for intelligibility test materials may be less steep for hearing-impaired listeners than for normal hearers (see, for example, Wilson, Caley, Haenel, & Browning, 1975). Hence, it was important to assess the SBR slope for the CSTv2 using hearing-impaired persons as well as normal hearers. Estimates of SBR slope for the CSTv2 were derived using data from both subgroups in experiment 2. Subgroup 1 and subgroup 2 yielded SBR function slope estimates of 9.2 and 7.7 rau/dB, respectively. The average SBR function slope derived from these data was 8.5 rau/dB. In addition, the data suggest that, on average, the SBR function slope does not differ systematically across the four hearing impairment groups. The outcome of the present study is in agreement with previous investigations that have shown the SBR function slope for speech intelligibility scores to be less steep for hearing-impaired listeners than for normal hearers.

As described above, for hearing-impaired listeners, the 95% CD for CST scores based on mean performance across two passage pairs was 15.5 rau. With a SBR function slope of 8.5 rau/dB, it is evident that for hearing-impaired listeners, intelligibility differences equivalent to 2 dB change in SBR (17.0 rau) can be detected with CST scores based on mean performance across two passage pairs.

SUMMARY AND FINAL COMMENTS

In an earlier paper (Cox et al, 1987) we reported the development of the Connected Speech Test, version 1 (CSTv1). This test consisted of 57 passages of connected speech: 48 test passages and 9 practice passages. All passages were equally intelligible for the average normal-hearing listener. In the present paper, we describe the development of the Connected Speech Test, version 2 (CSTv2). This test consists of 28 passage pairs; 24 test pairs and 4 practice pairs. The CSTv2 was developed on the basis of data obtained from hearing-impaired subjects. The 24 test pairs are essentially equivalent in intelligibility for all tested listeners when presented with a competing speech babble. Hence, this version of the CST is appropriate for use with both normal-hearing and hearing-impaired subjects.

For normal-hearing listeners, the SBR function slope is 12 rationalized arcsine units (rau) per dB. For hearing-impaired listeners, the SBR function slope is 8.5 rau/dB. The 95% critical difference for normal hearers for scores based on mean performance across two pairs is 14 rau. For hearing-impaired listeners, the 95% critical difference for scores based on mean performance across two pairs is 15.5 rau. Thus, a change in intelligibility equivalent to a change of 2 dB in SBR can be detected using CST scores based on two passage pairs for both normal and hearing-impaired listeners. However, the test is somewhat more sensitive when used with normal hearers.

Administration of the test using designated pairs of passages instead of single passages had at least two desirable consequences: (1) the within-subject variability of intelligibility scores across designated pairs is less than it would be with randomly selected pairs, especially for hearing-impaired listeners; and, (2) it was possible to construct the pairs so that they are equivalent in intelligibility for persons with a wide range of audiometric configurations. The subjects were chosen to represent the majority of hearing-impaired persons who are likely to wear conventional hearing aids and, thus, might be exposed to the CST for measurement of hearing aid benefit. However, a few types of audiograms were not represented. These include rising configurations and severe, sharply sloping configurations. It is uncertain whether the CSTv2 would be appropriate for use with persons having these types of impairments. Finally, because the CST is composed of everyday language, it would be too difficult for most persons with congenital, profound hearing impairment.

Although the two passages of a designated passage pair are always administered together, the combinations of passage pairs that are used to compose an intelligibility score may be chosen randomly from the available 24 pairs. In practice, since there are 12 pairs per disk, it is convenient to use all the pairs on one disk before switching to the other disk. Since there are 24 passage pairs, if two pairs are used per score, intelligibility may be compared under 12 different conditions for a single listener.

It should be kept in mind that the CSTv2 is an audio test only. Evaluation of the audiovisual CST is the subject of future research. Hence, administration of the test must allow subjects to read the passage topic on the monitor screen before the passage is presented but subjects must not be permitted to watch the talker's face while listening to the passages. Finally, there is a period of rapid learning when listeners are first exposed to the CST task. Hence, it is essential to allow subjects to respond to the practice passages before data are collected. In addition, reliability of the data is improved if one practice pair is repeated each time test conditions are changed.

The CSTv2 is uniquely suited to the measurement of hearing aid benefit in that it has high content validity (conversationally produced connected speech), good sensitivity, and a large number of equivalent forms.

References

Byrne D, Fifield D. Evaluation of hearing aid fittings for infants. Br J Audiol 1974;8:47–54.

Cox RM, Alexander GC, Gilmore C. Development of the Connected Speech Test (CST). Ear Hear 1987;8(suppl):119S–126S.

Ferguson GA. Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Co., 1969: 140, 147.

Fletcher H. Speech and Hearing in Communication. New York: Van Nostrand Reinhold Co., 1953: 96.

Killion MC, Wilbur LA, Gudmundsen GI. Insert earphones for more interaural attenuation. Hear Instrum 1985;36(2):34–36.

Nunnelly JC. Psychometric Theory. New York: McGraw-Hill, 1978: 229.

Pearsons KS, Bennett RL, Fidell S. Speech levels in various noise environments. United States Environmental Protection Agency, 1977; Report EPA 600/1-77-025.

Studebaker GA. A "rationalized" arcsine transform. J Speech Hear Res 1985;28:455–462.

Thornton AR, Raffin MJM. Speech-discrimination scores modeled as a binomial variable. J Speech Hear Res 1978;21:507–518.

Wilson RH, Caley KE, Haenel JL, Browning KM. Northwestern University Auditory Test No. 6: Normative and comparative intelligibility functions. J Am Audiol Soc 1975;1:221–228.

Acknowledgments: Robert M. Joyce wrote the software for presentation and scoring of the CSTv1 and the CSTv2. Jeffrey N. Moore assisted with data collection in experiment 2.

Address reprint requests to Robyn M. Cox, Ph.D., Memphis Speech and Hearing Center, 807 Jefferson Ave., Memphis, TN 38105.

Received January 13, 1988; accepted March 8, 1988.