0196/0202/89/1001-0029$02.00/0
EAR AND HEARING
Copyright © 1989 by The Williams & Wilkins Co.

Vol. 10, No. 1
Printed in U.S.A.

BEHAVIORAL TECHNIQUES IN AUDIOLOGY AND OTOLOGY

The Connected Speech Test Version 3: Audiovisual Administration*

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

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

ABSTRACT

The Connected Speech Test (CST) is used to measure the intelligibility of everyday speech; it is intended primarily for quantifying hearing aid benefit. The test consists of 48 passages of conversationally produced connected speech, each passage concerning a familiar topic and comprising 10 sentences. Listeners are apprised of the passage topic in advance and are required to repeat the sentences one at a time. Each passage contains 25 scoring words. The test is recorded audiovisually. In previous papers, we have reported the development of the test materials and investigations of the use of the audio portion with normally hearing and hearing-impaired listeners. Audio versions of the test have been developed for use with normal hearers (CST version 1), and for hearing-impaired listeners (CST version 2). In the present paper, we report a study of the test administered audiovisually. Twenty-six normally hearing subjects responded to audiovisual presentation of all 48 test passages. On the basis of the results, a new version of the test (CST version 3) was generated. In this new version, the test passages are presented in designated sets of four or six passages. The critical difference between two scores is estimated to be the same for both audio and audiovisual administration.

The Connected Speech Test (CST) is a test of the intelligibility of everyday speech. It has been developed primarily for use as a criterion measure in investigations of hearing aid benefit. Two previous papers have reported: (1) the background and rationale for the test, data for normally hearing listeners, and the development of the CST version 1 (CSTv1) (Cox, Alexander, & Gilmore, 1987a); and (2) investigations of the use of the CSTv1 with hearing-impaired listeners and subsequent modifications of the test, leading to the generation of the Connected Speech Test version 2 (CSTv2) (Cox, Alexander, Gilmore, & Pusakulich, 1988).

In brief, the test consists of 48 passages of connected speech, each with 25 scoring words. Each passage is about a familiar topic and consists of 10 sentences. A word describing the topic is presented to the subject before the passage is presented. A multitalker babble is provided as a competing signal; its level may be adjusted to meet the needs of the evaluation (e.g., to simulate a particular type of listening environment). The sentences are played one at a time and the listener is required to repeat each sentence exactly as heard. It is recommended that at least four passages be administered, and the results averaged, for each listening condition.

Performance is quantified in terms of the percentage of scoring words correctly repeated and this number is transformed into rationalized arcsine units (Studebaker, 1985). Although 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. The transformation from percentage into rationalized arcsine units minimizes the relationship between overall score and variability. As a result, the critical difference between two scores remains the same, regardless of score magnitude. (The rationale for determination of critical differences for the CST is fully discussed in the previous papers and, for this reason, is not repeated here.)

The CSTv1 is intended for presentation to normally hearing subjects. Each passage is essentially equivalent in intelligibility to each other passage for normal hearers. The 4 (or more) passages presented in each condition may be selected randomly (within replacement) from the corpus of 48. The 95% critical difference between two scores (each based on four passages) is about 14 rationalized arcsine units (rau). The CSTv2 is appropriate for use with either normally hearing or hearing-impaired listeners. In order to more precisely equate the forms administered to hearing-impaired subjects, the 48 test passages are sorted into 24 pairs of passages. For each pair a more difficult passage is matched with a less difficult passage so that all pairs are equal in average intelligibility for both normal

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

and hearing-impaired listeners. The four passages presented in each condition should be two randomly chosen pairs of passages. The 95% critical difference between two scores (each based on four passages) is about 15.5 rau for hearing-impaired subjects.

The connected speech passages were recorded audiovisually. The talker was a female who was selected because her everyday speech, presented auditorily only, was empirically determined to be of average intelligibility (Cox, Alexander, & Gilmore, 1987b). In addition, she was judged to be average in generation of speechreading cues. Factors considered in this determination included: absence of distracting mannerisms while taking, no unusual assymetries in mouth or jaw, normal lip mobility, and some visibility of teeth and tongue during speech. However, no formal testing was undertaken to quantify this talker's speechreading cues. It is important to note that the CST versions 1 and 2 utilize only the audio portion of the test recordings, except that the listeners are permitted to view the topic word on a monitor screen before auditing each passage. No speechreading cues are provided during administration of these tests.

Because speechreading cues are available to some extent in many communicative situations, there are obvious reasons to assess hearing aid benefit using an audiovisual test. The optimal hearing aid frequency response may differ; depending on the presence or absence of speechreading cues. The important cues for place of articulation are often missing from a degraded auditory signal. However,. speechreading cues may supplement auditory cues to significantly reduce deficits in the perception of place of articulation. This type of consideration suggests that hearing aid benefit may interact with presence/absence of visual cues.

The present paper reports a study of the Connected Speech Test presented with visual as well as audio cues. The purposes were: (1) to investigate the equivalence of the 24 passage pairs of the CSTv2 when administered audiovisually and (2) to generate a version of the test that was appropriate for audiovisual administration.

METHOD

Subjects

Twenty-six individuals with thresholds <20 dB HL from 250 Hz through 8000 Hz served as subjects (exception: one subject's threshold at 8000 Hz was 45 dB). Their ages ranged from 23 to 50 with a mean of 31 years. They included students, clerical workers, technicians, and maintenance personnel. None had training in speechreading. All reported normal or corrected-normal vision at 1 m (the test distance).

In order to generalize the speechreading results from these essentially normally hearing subjects to postlingually hearing-impaired persons, it is only necessary to assume that the normally hearing listeners utilize visual cues in essentially the same way as the hearing-impaired persons. Published evidence supports this assumption. Erber (1972), Benguerel and Pichora-Fuller (1982), and Owens and Blazek (1985) have all reported data indicating that there are no differences between normal and hearing-impaired groups in viseme perception. Many other studies also support this conclusion.

Recordings

The Connected Speech Test passages were recorded on videotape using a broadcast quality camera (Sony, model DXCM3A with Fujinon lens). Lighting consisted of a 1000 watt back light, a 1500 watt key light, and 1000 watt diffused fill light. The film was made in color against a grey background. The talker used light, everyday, makeup. These conditions were chosen to provide a clear but not excessively detailed picture, similar to typical everyday experience. For information on audio recording, see Cox et al (1987a). The talker's head, neck, and top of shoulders were photographed from a 0° azimuth. When replayed on a 33 cm diagonal monitor, the image is slightly smaller than life-sized. The edited master tape was dubbed to optical laser disk (Panasonic recorder, model TQ2026F).

Procedure

The test passages were replayed using a 2-channel optical disk player (Panasonic, model TQ2024F). The video output was routed to a 33 cm diagonal color monitor (Panasonic, model CT-1330M). The audio outputs (passages and babble) were attenuated, mixed, amplified, and presented to an insert earphone (Etymotic ER-1) that was coupled to the test ear using a compressible foam earplug. This playback system delivered the same frequency response to the average eardrum as would have occurred there during open-ear listening in a diffuse sound field. The nontest ear was plugged.

The audio passages were delivered at a level of 41 dB Leq (equivalent continuous A-weighted level), calibrated in a Zwislocki-type ear simulator. This was 20 dB below the level of normal conversation speech in a quiet environment (Pearsons, Bennett, & Fidell, 1977). For 15 subjects, the signal to babble ratio (SBR) was set at −5 dB. The remaining 11 subjects listened at an SBR of −7 dB. The combination of presentation level and SBR conditions were selected on the basis of pilot data with the intention of eliciting a wide range of overall scores but avoiding scores near 0 and 100%.

Subjects were seated in a single-walled sound room, 1 m in front of the monitor. They were instructed to watch the monitor and listen to the audio signal and to repeat each sentence exactly as they perceived it. Six to eight practice passages were presented before the test passages to familiarize the listener with the task and to allow learning effects of asymptote. All 24 pairs of test passages were then presented with order controlled to minimize order effects. Delivery and scoring of the passages were controlled by microcomputer (Zenith, model 181).

RESULTS

The data consisted of scores, in rationalized arcsine units, for each of 24 pairs of passages (CSTv2) for each subject. Mean scores of passage pairs across subjects ranged from 56.4 to 71.4 rau. To evaluate the within-subject equivalence of the passage pairs presented audiovisually, the standard deviation of passage pair scores was computed for each subject. These data were compared with analogous standard deviations determined for the same passage pairs presented audio-only to 40 normal hearers in previously reported studies (Cox et al, 1987a). Figure 1 presents the data for both conditions. In this figure, each *filled circle* depicts one listener's within-subject standard deviation as a function of estimated true score (mean across all pairs) for audio-only presentation. Each *open circle* depicts a within-subject standard deviation as a function of estimated true score for audiovisual presentation.

The figure suggests that, on the whole, the within-subject variability of passage pair scores was somewhat greater in the audiovisual condition than in the audio-only condition. Calculations confirmed this observation. As noted by Cox et al (1988), the typical within-subject variability of CSTv2 (audio) passage pairs was 7.3 rau for normal hearers. In the present study, the typical within-subject variability for the same passage pairs presented audiovisually was 8.2 rau. This outcome indicates that, not surprisingly, the speechreading cues generated by the CST talker were not exactly equivalent across passages.

In an attempt to more precisely equate the test forms, the 24 CSTv2 passage pairs were reconstituted into sets of four passages, each consisting of two CSTv2 pairs. In this process, passage pairs having higher audiovisual scores were joined with passage pairs having lower audiovisual scores. This resulted in 12 sets of passages with mean scores across subjects from 63.4 to 65.5 rau. Within-subject standard deviations were again computed for both audio and audiovisual data using the new sets of four passages. The results are shown in Figure 2. As this figure shows, the distributions of within-subject standard deviations were essentially overlapping in the audio and audiovisual conditions. The typical within subject standard deviation for the audio condition was 5.05 rau. In the audiovisual condition, the typical within-subject standard deviation was 5.17 rau.

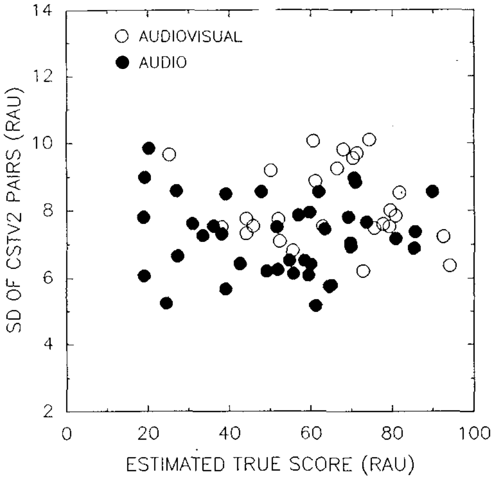

Figure 1. Within-subject standard deviation of passage pair scores as a function of estimated true score (mean score across all passage pairs). *Filled circles* depict data for 40 subjects from previous studies who received only the audio portion of the test. *Open circles* depict data for the 26 subjects in this study who received the test audiovisually.

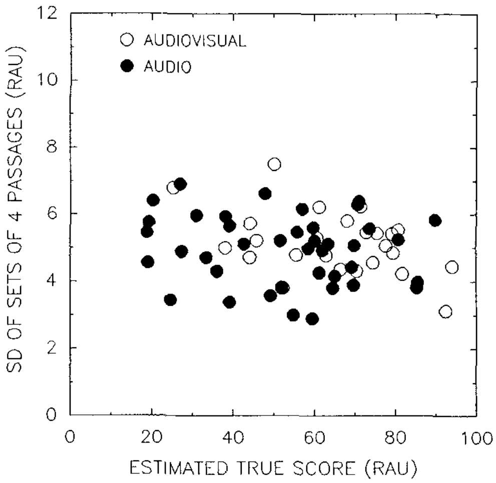

Figure 2. Within-subject standard deviation of scores for sets of four passages as a function of estimated true score (mean score across all sets). *Filled circles* depict data for 40 subjects from previous studies who received only the audio portion of the test. *Open circles* depict data for the 26 subjects in this study who received the test audiovisually.

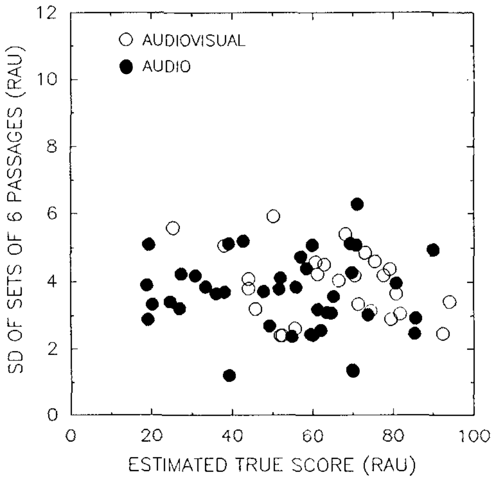

Figure 3. Within-subject standard deviation of scores for sets of six passages as a function of estimated true score (mean score across all sets). *Filled circles* depict data for 40 subjects from previous studies who received only the audio portion of the test. *Open circles* depict data for the 26 subjects in this study who received the test audiovisually.

Finally, even greater equality of audio and audiovisual variability was attained by subdividing the 24 CSTv2 passage pairs into sets of six passages (three pairs each). This resulted in eight sets of passages with mean scores across subjects from 64.3 to 64.6 rau. Figure 3 shows the relevant data on within-subject variability. Note that the distributions of audio and audiovisual data are almost identical. The typical within-subject standard deviations were 4.0 rau in both conditions.

This new version of the test, in which the 48 test passages are subdivided into sets of four and sets of six was called the Connected Speech Test, version 3 (CSTv3). Table 1 gives the passage titles for the various sets.

Table 1. Topic words for each Connected Speech Test (version 2) passage pair and designated sets of four and six passages comprising the Connected Speech Test, version 3.

Passage	Sets		Passage	Sets	
Pair	4	6	Pair	4	6
window/glove	}	}	cabbage/gold	}	}
umbrella/giraffe			weed/chimney		
lung/dove	}		lead/calendar	}	
carrot/grass		}	lion/zebra		}
nail/woodpecker	}		lizard/wolf	}	
owl/vegetable			orange/oyster		
lemon/violin	}	}	dice/eagle	}	}
wheat/ice			ear/liver		
donkey/guitar	}		leopard/eye	}	
envelope/grasshopper		}	zipper/egg		}
lettuce/dictionary	}		clock/kangaroo	}	
lawn/cactus			camel/goose		

DISCUSSION

Because the within-subject variability across CSTv3 test forms (sets of four or six passages) is essentially the same for both audio and audiovisual presentations, it is appropriate to use the same critical difference between two scores for both types of administration. Previous investigations have indicated that the 95% critical difference between two scores when each is based on a set of four passages (two CSTv2 pairs) is about 14 rau for normal hearers and 15.5 rau for hearing-impaired listeners. Using the same calculation scheme [described in Cox et al (1988)], the 95% critical difference for two scores each based on a set of six passages (three CSTv2 pairs) is estimated as 11.2 rau for normal hearers and 12.2 rau for hearing-impaired listeners. If a smaller critical difference is desired, sets of four can be combined at random into sets of eight, and the new critical difference calculated using the equation given in Cox et al (1988), bearing in mind that this equation requires entry of the number of pairs of passages used per score.

Caveats are in order. First, the outcome of this study of normal hearers may not validly generalize to prelingually impaired individuals. Because congenitally or prelingually hearing impaired individuals acquire language using primarily nonauditory cures, they probably use speechreading cues differently from persons who acquired language through the normal, primarily auditory mode. As a result, they may combine auditory and visual information in an anomalous manner.

Second, although the results of this investigation suggest that the CSTv3 portends to be a useful test of audiovisual intelligibility for postlingually impaired persons, they have not established that the test is especially useful if administered by vision alone. It should be kept in mind that the passages were never presented by vision only. Thus, it cannot necessarily be concluded that the CSTv3 sets would be equivalent if administered in this way. Furthermore, Kricos and Lesner (1982) demonstrated that there is considerable variability in the generation of speechreading cues across talkers (from their study it would appear that a talker of average visual intelligibility probably generates about six separate viseme categories). The location of the CST talker on the visual intelligibility dimension is not known. Additional investigations would be necessary to resolve these issues.

References

Benguerel A and Pichora-Fuller MK. Coarticulation effects in lipreading. J Speech Hear Res 1982;25:600–607.

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

Cox RM, Alexander GC, and Gilmore C. Intelligibility of average talkers in typical listening environments. J Acoust Soc Am 1987b;81:1598–1608.

Cox RM, Alexander GC, Gilmore C, and Pusakulich KM. Use of the Connected Speech Test (CST) with hearing-impaired listeners. Ear Hear 1988;9:198–207.

Erber NP. Auditory, visual and auditory-visual recognition of consonants by children with normal and impaired hearing. J Speech Hear Res 1972;15:413–422.

Kricos PB and Lesner SA. Differences in visual intelligibility across talkers. Volta Rev 1982;84:219–225.

Owens E and Blazek B. Visemes observed by hearing-impaired and normal-hearing adult viewers. J Speech Hear Res 1985;28:381–393.

Pearsons KS, Bennett RL, and 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.

Acknowledgment: Robert M. Joyce wrote the software for presentation and scoring of the CST passages.

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

Received June 30, 1988; accepted October 7, 1988.