EFFECTS OF TESTER-LISTENER VARIABLES ON THE SCORING ACCURACY OF SINGLE SYLLABLE WORD SPEECH DISCRIMINATION TESTS

by

Carol Lowrey Hippert
San Diego State University

Robert E. Novak
Associate Professor
San Diego State University

ABSTRACT

The effects of monitoring mode, presentation order of the monitoring modes, testing experience and lip reading ability of the tester, and sex of the speaker on the accuracy of scoring CID W-22 word discrimination tests were assessed. Significantly greater accuracy of scoring was found when testers watched the speakers mouth movements while listening to their responses under earphones. Order of presentation of the four different monitoring modes made a difference in the magnitudes of accuracy scores between similar conditions but not in the overall trend of the data. Clinical implications of these findings are discussed.

Speech audiometry is an integral part of audiologic evaluation. It provides valuable information concerning differential diagnosis of hearing impairment as well as insight into pragmatic communication abilities. Inconsistencies in the administration and recording of speech test materials exist across clinical audiological facilities. It was the purpose of this investigation to evaluate the effects that the mode of client response monitoring, experience of the tester, lipreading ability of the tester and sex of the client have on the scoring accuracy of single syllable word discrimination tests. This investigation was limited to the scoring of English speakers with general American dialects.

Although single syllable word speech discrimination tests are routinely completed in the majority of audiological clinics, little has been written about the effects of variables in tester-listener interaction on the scoring accuracy of these tests. Merril and Atkinson (1965), Lovernie, Burgi, and Curry (1968), and Nelson and Chaiklin (1970) studied differences is the scoring of single syllable word speech discrimination tests when their subjects scored tests using the auditory mode only to monitor client responses. The subject's scoring was then compared with the actual client responses and significant differences were noted. Nelson and Chaiklin (1970), Merril and Atkinson (1965) and Siegal (1962) found conflicting results concerning abilities of experienced versus inexperienced testers, and Palmer (1955) found no differences in the scoring of male versus female speakers.

MATERIALS AND METHODS

Twenty subjects scored single syllable word discrimination tests administered to a male and female speaker under four different listening-viewing conditions. Prior to their participation, each subject viewed a video taped Utley Lipreading Test and was required to pass a 20 dB Hearing Level (HL) (re: ANSI 1969) hearing screening at frequencies 250 through 4000 Hz bilaterally.

Two normal hearing caucasian subjects between the ages of 20-25 years were selected as speakers. Both subjects were considered to have General American dialects with no articulation disorders. Each was videotaped and recorded prior to the onset of data collection. The subject used to produce the videotaped recording of the Utley Lipreading Test Form A was a 22 year old Caucasian female with standard American English dialect and no articulation disorders. This speaker had previous experience in administering the Utley.

Videotaping was conducted in a two room double walled Industrial Acoustics Corporation (IAC) sound suite, model 1403 ACT. A Sony video camera model AVC32-50 with microphone mounted on a tripod was placed three feet from the subject at a 75 degree azimuth. The camera was focused so that the face of the subject was centered and the subject's full face from below the chin to approximately three inches above the head was visualized. An adjustable pole lamp was positioned at the subject's side to provide maximum illumination of facial features.

The male and female speakers wore TDH 39 earphones with MX 41/AR cushions. All earphones used by the two speakers and the 20 subjects were matched for frequency response characteristics.

The Hirsh recordings of the CID W-22 50 word lists C 1-4 and D 1-4 were presented to the male and female speakers via a Sony cassette tape recorder. Since these were normal hearing listeners the words were low pass filtered in order to induce discrimination errors. The Allison filter model 2BR operated in a low pass mode was set to provide an upper cut-off frequency of 800 Hz with a rejection rate of approximately 31 dB per octave. The filtered word lists were then channeled through a Grason-Stadler audiometer model 1701 at 40 dB HL. The male speaker responded to lists D 1-4 and the female speaker responded to lists C 1-4. The two speakers were instructed to write down and verbally repeat the words heard and to guess at any word of which they were unsure. Average resultant discrimination scores were 47 percent for the female and 51 percent for the male. The speaker used to present the Utley was videotaped in the same manner as described above. All even numbered test items were presented initially (1-30) with odd numbered items following (31-1).

The 20 subjects who scored the videotaped speakers were randomly selected within each of two different subpopulations. Group one consisted of ten subjects considered to be inexperienced, having no previous experience or familiarity with CID W-22 word lists or discrimination testing procedures. Group two consisted of ten subjects considered to be experienced, having a minimum of 25 hours of audiologic testing including the administration of single syllable word speech discrimination tests.

Subjects viewed the male and female speaker under the following four listening / viewing conditions:

Condition 1. Auditory via headphones plus vision (HP + V)
Condition 2. Auditory via free field plus vision (FF + V)
Condition 3. Auditory via headphones alone (HP)
Condition 4. Auditory via free field alone (FF)

Half the subjects were presented the conditions in an ascending order (a_1), given condition 1 first and condition 4 last, and half were given the conditions in a

descending order (a_2), with condition 4 presented first and condition 1 last. During the (HP + V) and (FF + V) conditions the subjects viewed the male and female speakers in black and white on either a 19 inch or 21 inch television screen. The lipreading test was also presented in this manner. Subjects were instructed to write down on a closed set CID W-22 word lists response form exactly what they heard from the male and female speakers.

RESULTS

Subject responses were scored by comparing the subject's written responses to the written responses previously obtained from the male / female videotaped speakers. Final tabulations were written in percent correct of what was actually said by the speaker. The Utley lipreading test results for each subject were scored by both the word and sentence method as described by Jeffers and Barley (1971). All data was analyzed utilizing a two-way analysis of variance or a t-test.

Table 1 gives the mean CID W-22 word list accuracy scores from all 20 subjects obtained under the four monitoring conditions.

Table 1. Mean CID W-22 word list scoring accuracy scores (percent correct) and standard deviations for all 20 subjects for each monitoring condition.

condition	mean	standard deviation
1 (headphones + vision)	99.85	1.04
2 (free field + vision)	98.95	.99
3 (headphones alone)	98.50	1.23
4 (free field alone)	97.25	1.92

It is evident that the overall mean for condition 4 (FF) (97.25 percent) is lower than for condition 1 (HP + V) (99.85 percent), conditions 2 (FF + V) (98.95 percent) or condition 3 (HP) (98.50 percent). Simple contrasts for each condition were completed so that condition 1 was compared with conditions 2, 3, and 4 and so on. Statistically significant differences were noted between conditions 1 and 4, 2 and 4, and 3 and 4. All F scores obtained were statistically significant at the .01 level of confidence. From these findings it was difficult to determine whether significant differences occured as a result of the use of headphones versus free field or the use of vision versus no vision. Therefore, general contrasts were applied between vision versus non vision conditions and headphones versus free field conditions. It was found that the scores obtained using vision were consistently better than those obtained using hearing alone for all subjects and that this difference was statistically significant at the .001 level of condifence.

Mean scores for subjects obtained under the two presentation orders (a_1 and a_2) for both the male and female speaker were compared and an analysis of variance was run to test for statistically significant differences between average scores obtained under these two presentation orders. A statistically significant difference at the .05 level of confidence was found between average scores obtained under order a_1 versus a_2. In comparing average scores per condition from order a_1 and a_2 it was found that scores obtained in order a_2 under condition 4 (FF) were significantly lower than scores obtained in the same condition in order a_1. This suggests that subjects under order a_1 had time to become familiar with the overall task before having to score under the most difficult monitoring mode. Those subjects under order a_2 were required to score under the most difficult monitoring mode initially without the benefit of practice. It appears that the lower average free field alone scores from order a_2 caused the significant difference between orders a_1 and a_2. Although a practice effect contributed to more accurate free field alone (condition 4) scores under order a_1, the effect of different listening / viewing conditions is still apparent with the scores obtained under condition 4 being the poorest overall in both order a_1 and a_2.

In order to assess the effect of experience of the subjects on their scoring accuracy, a t test was applied to the data obtained from these two subgroups. Data from each monitoring condition was evaluated separately with no significant differences between experienced and inexperienced subjects noted under any condition.

The presence of significant differences in the scoring of male versus female speakers was evaluated using a two-way analysis of variance. The scores from each subject for male videotape were averaged. This was also completed for the female videotape across all four monitoring conditions. Overall mean scores for the male versus female videotapes under all four monitoring conditions differed by only .38 percent which was not significant at the .05 level of confidence.

From the Utley Lipreading Test results, two sets of lipreading scores were obtained for each subject. One set of scores was based on the total number of words correctly identified by each subject. The second set of scores was based on the total number of complete sentences identified correctly. Each set of scores was evaluated separately and the abilities of experienced and inexperienced subjects were compared. A t test was applied to the data and no statistically significant differences in lipreading abilities were found between the experienced and inexperienced subjects. Examination of the Utley scores for both the experienced and inexperienced testers revealed that the majority of subjects obtained scores considered to be poor when scored by either the word or sentence method. Since it was found that the use of vision when scoring single syllable word speech discrimination test resulted in significantly better scoring abilities, it is obvious that the Utley is inappropriate for use as a predictive index of a clinician's speechreading ability as it relates to the use of visual information to enhance scoring accuracy of single syllable word speech discrimination tests.

DISCUSSION

It is obvious that the greatest accuracy in scoring single syllable word speech discrimination tests will occur when a written response is required of the client. However, for many clients this is not possible and for many clinicians this is felt

to be too time consuming and inconvenient. As a result of these conditions methods of monitoring speech discrimination tests are not consistant across audiology test facilities. In this investigation significant differences in scoring accuracy of monosyllabic word speech discrimination tests were observed when testers were asked to monitor client responses under four different listening-viewing conditions. It was found that visual observation of the client's mouth movements significantly enhanced scoring accuracy, and that monitoring of client's responses under headphones in combination with the use of visual observation of client's mouth movements allowed for the greatest scoring accuracy.

The clinical implications of these findings are that persons administering speech discrimination tests must be cognizant of their testing idiosyncracies and realize that variables in the method of monitoring client responses can significantly effect their scoring accuracy. When time constraints or client limitations prevent having the client provide a written response, the clinician must monitor the client in a way that will ensure optimal scoring accuracy. From this investigation it can be seen that the greatest scoring accuracy was obtained with the clinician listening under headphones while watching the client's mouth movements. The difference between the scoring accuracy using this monitoring method versus free field auditory monitoring alone was as great as eight percentage points. It is important to remember that these differences were found for subjects who were monitoring in distraction free, high fidelity conditions. With the addition of such typical clinical distractions as low fidelity audiometer talk-back systems, observers and extraneous background noise in the test booth, and the pressure to complete a large number of audiological evaluations in a given time period, it is very likely that even greater discrepancies in scoring accuracy would be found.

ACKNOWLEDGEMENTS

Address correspondence to: Robert E. Novak Ph.D., Department of Communicative Disorders, San Diego State University, San Diego, California 92182.

REFERENCES

American National Standard Specifications for Audiometers, ANSI S3.6-1969 (New York: American National Standards Institute, 1970).

Jeffers, J. and Barley, M., **Speechreading (Lipreading)** Springfield, Illinois, Charles C. Thomas 1971.

Lovernic, J.H., Burgi, E.J., and Curry, E.T., A comparative evaluation of five speech discrimination measures. **J. Speech Hear. Res.**, 11, 373-381, 1968.

Merrel, H.B., and Atkinson, C.J., The effect of selected variables upon discrimination scores. **J. Aud. Res.**, 5, 285-292, 1965.

Nelson, D.A., and Chailklin, J.B., Writedown versus talkback scoring and scoring bias in speech discrimination testing. **J. Speech Hear. Res.**, 13,645-654, 1970.

Palmer, J.M., The effect of speaker differences on the intelligibility of phonetically balanced word lists. **J. Speech Hear. Res.**, 20, 192-195, 1955.

Siegal, G.M., Experienced and inexperienced articulation examiners. **J. Speech Hear. Dis.**, 27, 28-35, 1962.

Utley, J., A test of lipreading ability. **J. Speech Dis.**, 11, 109-116, 1946.