Towards the Automated Assessment of Speech Intelligibility: Some Results from the Amplification Research Group

Donald G. Jamieson, Dino Dell'Orletta, and Ketan Ramji
Speech Communication Laboratory
Department of Communicative Disorders
University of Western Ontario

The measurement of a listener's ability to understand speech is widely recognized to be an important aspect of audiological assessment. While numerous procedures have been proposed for such measurement, there remains a need for cost- and time-effective procedures which provide valid and reliable assessments of an individual listener's capacity to understand speech in various listening situations. This brief report outlines an approach being developed at the University of Western Ontario by members of the Amplification Research Group.

Background

Speech intelligibility tests measure the capacity of an individual listener to understand speech in a particular listening situation. The speech intelligibility measure which is used most frequently in current clinical practice is the phonetically-balanced monosyllabic word test (e.g., CID W-22, Hirsh et al., 1952; Harvard PB-50, Egan, 1948; NU-6, Tillman and Carhart, 1966). Typically, these tests are administered as audiotaped sequences of words, with the client attempting to repeat each word to the examiner. The examiner typically scores each response by hand. In some instances, "live voice" is used — the list of words is read to the client, who then repeats what is heard.

One can understand why these tests have enjoyed wide acceptance: they are easy to administer and score, they are relatively economical of the clinician's time, and they provide a number which can be used to rank order intelligibility. However, the tests have been widely criticized for their relative insensitivity, poor test-retest reliability, large word familiarity effects, and strong practice effects (e.g., CHABA, 1988; Edgerton and Danhauer, 1979; McCarthy, 1983; Ostergard, 1983; Schwartz and Surr, 1979; Sher and Owen, 1974; Silverman and Hirsh, 1955; Thornton and Raffin, 1978; Walden, Schwartz, Williams, Holum-Hardegen, and Crowley, 1983). Testing with taped stimuli also causes difficulties when the rate of presentation is inappropriate for a given listener (see McLennon and Knox, 1975). Testing using live voice permits the presentation rate to be varied to accommodate the abilities of different listeners, but the testing conditions are nonstandard and they cannot be reproduced from test to test (Brandy, 1966). Finally, hand scoring of spoken responses is both tedious for the examiner and is subject to the examiner's own perceptual errors (Nelson and Chaiklin, 1970). In an attempt to deal with these difficulties, we have been exploring the possibilities for automated, computer-based testing and scoring using tests which may be less vulnerable to criticism on psychometric grounds.

Basic Philosophy

In developing our system, we were concerned with several factors: (1) whenever possible, we used general-purpose, "off the shelf" items, which were available in volume —to reduce costs and to increase both the reliability of the system components and the level of maintenance support which was available; (2) the acoustic properties of the speech tests were precisely defined and were developed to facilitate control and calibration; (3) the pace of testing was designed to be adjustable, to accommodate the needs of different listeners; and (4) testing was designed so that the listener's responses could be collected and scored automatically, to minimize the demands on the examiner's time and the possibilities for scoring errors. The system we developed is based on an IBM/AT-compatible microcomputer (Zenith Model 248). The computer has a 40 megabyte hard disk, 512K of memory, an EGA graphics card, two EGA color monitors connected in parallel, a mouse (Logitech "bus" mouse, which is used by the subject to select the desired response), and a digital-to-analog converter board (DataTranslation 2801A, which is used to output the signals from the computer's disk to an analog speech signal). The computer runs specially-written programs to control testing, collect the listener's responses, and score the data. The same computer also is used as a general-purpose computer to do word processing, maintain database records, plot data, do statistical analyses, analyze acoustic signals,[1] and so forth.

When required by the testing program, the desired speech signal is converted from a digital signal (stored on the computer's disk) to an electrical signal by the DataTranslation board, then conditioned electroacoustically by: (1) a low-pass filter (Kemo model VBF 25 MD, which is used to ensure that the speech signal is free from any aliasing effects); (2) a simple programmable attenuator and mixer (so that the computer can control the level of the signal and, where used, the noise); (3) a general-purpose amplifier (Crown D-75, which is used to set the overall level of the signal to that required for the particular test); (4) a general-purpose multimeter (which is used to check that the levels are precisely as specified for the particular test);

1. The system described here is identical to that being used to develop CSRE: the Canadian Speech Research Environment. The CSRE software permits detailed speech recording, editing, analysis, synthesis, replay, and testing, on an inexpensive, general-purpose microcomputer-based facility (Jamieson and Nearey, 1988).

and (5) the sound transducer (e.g., Etymotic Research ER-3A insert earphone).

Our total cost for the entire system, including the general-purpose computer, was about CDN$12,000. With a less expensive filter, this cost could be reduced by as much as $5,000. In the next two sections, we describe our use of the system with two automated testing procedures.

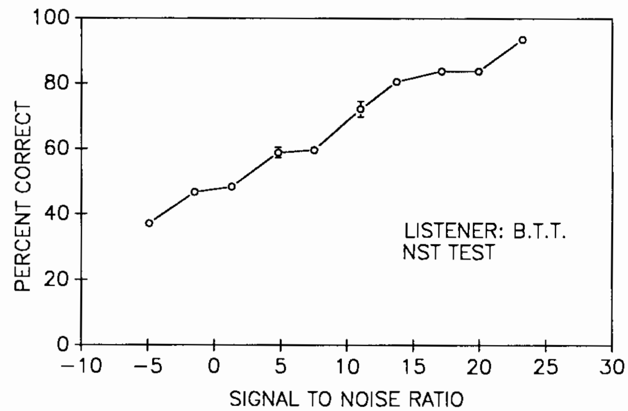
Specific Applications: I. Nonsense Syllable Testing

Levitt and Resnick (1978) introduced the Nonsense Syllable Test as a way to obtain both an estimate of a listener's relative ability to understand speech under different listening conditions, and to analyze the listener's specific difficulties in understanding speech. The test has been examined under a variety of listening conditions (e.g., Dubno, Dirks, and Langhofer, 1982; Dubno and Dirks, 1982; Dubno and Levitt, 1981; Edgerton and Danhauer, 1979). For our version of the NST, we recorded, on computer disk, 55 separate consonant-vowel (CV) and vowel-consonant (VC) syllables[2] spoken by a single Canadian talker. These signals were edited, using a general-purpose waveform editor (Jamieson and Nearey, 1988), to ensure that they met our acoustic criteria and, then, judged by a panel of normally-hearing listeners to ensure that they sounded natural and were readily identifiable.

The 55 syllables are arranged into seven closed-set modules, which are presented to listeners individually. Each test session consists of one presentation of each of the seven modules. Within a module, syllables are presented in a random order. During each trial of the test, the listener first sees the set of possible responses displayed on a computer screen. Next, the target sound is presented to the listener by the computer within the carrier phrase, "Point to the sound ..." The listener uses the mouse to move a visible pointer on the computer's monitor to indicate the desired response, then, presses the mouse button so that the response is recorded.[3] At that point, the computer reorganizes the set of possible alternative responses on the computer monitor and, then, begins the next trial by displaying a new screen of possible responses.

One run through the full NST (one presentation of each module) requires approximately eight minutes. At the conclusion of testing, the computer scores the data and permits several possible analyses. Figure 1 displays the results of one such analysis—an example of data collected with the NST under various signal to noise ratio conditions. It can be seen that the NST is sensitive to small changes in the listening conditions over a wide range of variation and that the data obtained are highly reproducible from test occasion to occasion.

Figure 1. NST results for an individual listener under several listening conditions.



The panel displays a performance intensity function for the NST, showing the overall score (percentage of test items identified correctly) which was obtained by the listener during a number of test sessions at various signal-to-noise ratios. The error bars show the standard deviation found when repeated measures are taken at a particular signal-to-noise ratio. These values are very small, indicating the very high reproducibility of test results with the NST.

Specific Applications: II. Continuous Discourse Rating

Tests such as the NST can provide detailed analytical information concerning the specific types of difficulties which a hearing impaired listener may have with the reception of speech. A separate clinical need is to obtain a rapid measure of a listener's overall level of understanding of speech in a particular listening situation — for example, with a particular hearing aid. In an attempt to meet this need, we have been investigating the possibility that listeners can provide direct ratings of the intelligibility of continuous discourse samples (following Speaks, Parker, Harris, and Kuhl, 1972). Continuous discourse contains different acoustical cues to the identity of individual words (i.e., dynamic, coarticulation effects, both spectral and temporal), as well as lexical, semantic, and syntactic cues to word identity, which cannot be examined in tests using syllables or isolated words. While a variety of alternative rating approaches have been considered (e.g., Cox, et al., 1987; Cox and McDaniel, 1984; Giolas, 1966; Speaks, et al., 1972), our approach has been to play listeners a 40 to 60 second passage of a story and have them judge the percentage of words in the passage that they were able to understand. They respond by using the mouse to move a pointer on the computer screen to indicate this judgement on a continuous scale (from 0 to 100). While the results of such testing are less orderly than with the

2. There is evidence that consonant clusters (CCVC, CVCC, CCCVC) can improve the prediction of listeners' speech understanding abilities (Bilger and Matthies, 1985). Such clusters may be added to future versions of this NST.

3. Since this response is under the subject's control, the procedure permits the listener to control the pace of the test.

NST (for example, performance-intensity functions are considerably steeper, and test-retest reliability is less good), listeners can make such judgements quickly and easily, and the test can provide some clinically useful information very rapidly indeed.

Future Work

We are continuing to refine these tests and to develop additional tests for rapid, reliable assessment of listeners' capacities to understand speech in various situations. Because the system is general-purpose and flexible, it also lends itself quite readily to the rapid development of other types of tests, including ones using languages other than English. An important application of our testing will continue to involve the evaluation of the effects of alternative hearing aid gain functions on speech understanding (e.g., Jamieson and Miller, 1987; Jamieson, Miller, and Raftery, 1988), and the evaluation of various "speech enhancement" and "noise reduction" processing schemes.

Acknowledgements

This work has been supported by grants from Health and Welfare Canada under the National Health Research Development Program, and by the Natural Sciences and Engineering Research Council of Canada under the Strategic Research Program. We are indebted to Barb Bentley, John Booth, Meg Cheesman, Shane Moodie, Lynne Powell, Emmet Raftery, and Richard Seewald for assistance and advice at various stages of the project.

Address all correspondence to:
Dr. Donald G. Jamieson
Speech Communication Laboratory
Department of Communicative Disorders
University of Western Ontario
London, Ontario, Canada
N6G 1H1.

References

Bilger, R.C., & Matthies, M.L. (1985). The role of complexity in tests of word recognition. *ASHA, 27,* 117.

Brandy, W.T. (1966). Reliability of voice tests of speech discrimination. *Journal of Speech and Hearing Research, 9,* 461-465.

CHABA Working Group of Speech Understanding and Aging, (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83,* 859-895.

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

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

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

Dubno, J.R., & Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *Journal of the Acoustical Society of America, 69,* 249-261.

Edgerton, B.J., & Danhauer, J.L. (1979). *Clinical Implications of Speech Discrimination Testing Using Nonsense Stimuli.* Baltimore, Maryland: University Park Press.

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope, 58,* 955-991.

Giolas, , T.G. (1966). Comparative intelligibility scores of sentence lists and continuous discourse. *The Journal of Auditory Research, 6,* 31-38.

Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E., Eldert, E., & Benson, R.W. (1952). Development of materials for speech and audiometry. *Journal of Speech and Hearing Disorders, 17,* 321-337.

Jamieson, D.G., & Miller, D. (1987). Optimizing procedures for prescribing and fitting hearing aids. Canadian Association of Speech-Language Pathologists and Audiologists, Halifax.

Jamieson, D.G., Miller, D., & Raftery, E. (1988). Evaluation of alternative hearing aid prescription rules. Canadian Association of Speech-Language Pathologists and Audiologists, Banff.

Jamieson, D.G., & Nearey, T.M. (1988). CSRE: A speech research environment. *Abstracts of the Association for Research in Otolaryngology.*

Levitt, H., & Resnick, S.B. (1978). Speech reception by the hearing-impaired: Methods of testing and the development of new tests. *Scandinavian Audiology Supplement, 6,* 107-130.

McCarthy, P.A. (1983). The use of speech audiometry in aural rehabilitation. *Seminars in Hearing, 4,* 273-284.

McLennon, R.O., Jr., & Knox, A.W. (1975). Patient-controlled delivery of monosyllabic words in a test of auditory discrimination. *Journal of Speech and Hearing Disorders, 40,* 538-543.

Nelson, D.A., & Chaiklin, J.B. (1970). Writedown versus talkback scoring bias in speech discrimination testing. *Journal of Speech and Hearing Research, 13,* 645-654.

Ostergard, C.A. (1983). Factors influencing speech audiometry. *Seminars in Hearing, 4,* 221-239.

Schwartz, D.M., & Surr, R.K. (1979). Three experiments on the California Consonant Test. *Journal of Speech and Hearing Disorders, 44,* 61-72.

Sher, A.E., & Owens, E. (1974). Consonant confusions associated with hearing loss above 2000 Hz. *Journal of Speech and Hearing Research, 17,* 669-680.

Silverman, S., & Hirsh, I. (1955). Problems related to the use of speech in clinical audiometry. *Annals of Otolaryngology, 64,* 1234-1244.

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

Thorton, A., & Raffin, M. (1978). Speech discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research, 21,* 507-518.

Tillman, T.W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CVC monsyllabic words.* (Northwestern University Auditory Test No.6). Technical Report, SAM-TR-66-55, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base: Texas.

Walden, B.E., Schwartz, D.M., Williams, D.L., Holum-Hardegen, L.L., & Crowley, J.M. (1983). Test of the assumptions underlying comparative hearing aid evaluations. *Journal of Speech and Hearing Disorders, 48,* 264-273.