Improving the Efficiency of Speech Audiometry: Computer-based Approach

Brad A. Stach*1
M. Leighanne Davis-Thaxton*2
James Jerger*

Abstract

Speech audiometric measures were carried out on 981 patients using a computer-based system that permitted a comparison between two modes of signal presentation, manual and automatic. The manual mode, designed to simulate the pace of live-voice testing, allowed the presentation of speech signals at a rate dictated by the patient's response. The automatic mode, designed to simulate tape-player testing, fixed the interstimulus interval. A comparison of elapsed time for the completion of word recognition testing with phonetically balanced (PB) word lists showed that use of the manual mode resulted in increased efficiency. Testing time was reduced by an average of 22 percent. These results suggest that the clinical efficiency of speech audiometric testing can be enhanced by using a computer-based manual approach. The combination of digitally recorded speech stimuli with a computer-based manual approach may provide a useful compromise between the efficiency of live-voice testing and the signal consistency of magnetic-tape or compact-disc recordings.

Key Words: Microcomputer, speech audiometry, word recognition

Despite the audiologist's substantial reliance on measures of speech understanding and progress in sound reproduction, clinical speech audiometric practices in use today suffer from important, technologically related shortcomings. The best evidence supporting this notion is the fact that the majority of clinicians still carry out speech audiometry by live-voice testing. Recently, Martin et al (1994) reported results of a survey showing that 80 percent of respondents still used monitored live voice for speech intelligibility testing, and 90 percent of respondents still used live-voice testing for establishing speech thresholds. These are remarkable findings, in view of the fact that live-voice testing lacks standardization and requires carefully controlled vocal effort to ensure consistent intensity of signal presentation level.

One reason that live-voice testing is still used, despite these well-known and significant limitations, is that conventional recording playback systems are not able to provide the flexibility of live-voice methodology that is so desirable for routine clinical assessment. That is, whether using magnetic-tape or CD players, the clinician does not have the flexibility to control presentation, change the rate of presentation, randomize lists or items within a list, or easily repeat a stimulus. In testing some patients, this is a trivial concern. In others, it is not. Clinicians are familiar with the need to repeat signal presentations. Repetition is often necessary in patients who cannot respond quickly enough to a speech signal before the next one begins. The opposite can also occur. Clinicians are often faced with the child who requires more rapid signal presentation to remain on task. This lack of control over signal presentation rate often leads to abandonment of recorded materials in favor of the more flexible live-voice testing strategy.

One alternative to conventional recordings is the use of computer-based speech audiometry. The application of microcomputer technology to speech audiometry has the potential to provide a number of enhancements over conventional techniques. On the one hand, a computer-based system can provide all of the benefits of compact-disc technology by serving as a digital recorder. On the other hand, such a system can also allow control of stimulus presentation, control of inter-stimulus interval, randomization of stimulus presentation, and control of stimulus repetition. Thus, a computer-based system combines the
consistency of CD-quality speech signals with the advantages of the flexibility inherent in live-voice testing.

We have recently developed a computer-based speech audiometric system that addresses many of these flexibility issues. The purpose of the present study was to compare the relative efficiency of a computer-based manual presentation method with a conventional automatic playback method of speech audiometry. The evaluation was completed during routine clinical speech audiometric testing of 981 patients. Results showed an overall reduction in testing time with the manual-presentation strategy.

METHOD

Subjects

Subjects were drawn from the routine clinical caseload of patients evaluated at The Methodist Hospital Audiology Service, the Neurosensory Center of Houston. The subjects were chosen consecutively from among those who could complete a phonetically balanced (PB) list of monosyllabic words in the English language. Testing was carried out on 981 patients, 498 males and 487 females. They ranged in age from 7 to 97 years, with a mean age of 51 years. Hearing sensitivity, represented as the pure-tone average (PTA) of hearing threshold levels (HTLs) at 500, 1000, and 2000 Hz, ranged from -8 to 97 dB HTL, with a mean PTA of 23.8 dB HTL. Only those patients who could not complete routine speech audiometric measures were excluded, including patients who were not proficient in the English language, patients who were not cognitively capable of completing the task, and patients whose responses were deemed invalid due to a suspicion of malingering.

Computer-based Speech Audiometer

The computer-based speech audiometer was developed on the Macintosh II microcomputer. Digitally recorded speech stimuli were stored on a hard-disk drive and accessed under program control. The program was designed to retrieve speech stimuli from memory and present them through the Macintosh sound output system. The Macintosh output was routed to a commercial audiometer (either Grason Stadler GSI 10 or GSI 16) for presentation through earphones.

The speech signals were digitally recorded from magnetic tape via the Farallon\textsuperscript{TM} MacRecorder\textsuperscript{TM} analog-to-digital (A/D) converter system and stored on the computer's hard-disk drive. The speech signals were sampled at a rate of 11,000/sec, with 8-bit amplitude resolution. The Farallon waveform editing program was used to edit the sampled signal to remove pre- and post-speech silent intervals. The speech materials were four lists of 50 words from the PAL PB-50 test. A 1000-Hz sinusoid, recorded at the average of frequent peaks of the words, served as the calibration tone. To assure equivalency of amplitude input to the computer across speech materials, the voltage output of the tape player was monitored. All taped calibration tones were set to an identical voltage, providing equivalent signal output levels to the A/D converter.

Two-channel D/A output of speech signals was achieved by program calls to Farallon HyperCard\textsuperscript{®}X commands. Signals were routed through the Macintosh stereo output. Operator interface software was written in Object Pascal within a MacApp\textsuperscript{™} user-interface framework.

For all test types, the interface was designed to display the speech signals that were being presented, to allow designation of a response as correct or incorrect, to show elapsed time from the onset of testing, and to calculate a percent correct score. An example of the user interface is shown in Figure 1. The interface had four control "buttons," one for starting the test, one for scoring a test item as correct, one for scoring an item as incorrect, and one for repeating the previous stimulus. Selection of a speech stimulus from any given list of words was programmed to be random without replacement.

Procedures

Two modes of testing were used with each subject. The manual mode was designed to simulate the live-voice paradigm. Testing was initiated by a mouse-click or function-key press. When the patient responded, the audiologist chose either the response-correct or response-incorrect button on the computer interface. Immediately following this selection, the next word was presented to the patient. Thus, the pace of the test was dictated by the patient's response. If a patient responded quickly to the stimulus, the next signal would be presented rapidly. If the patient responded more slowly, then the rate of stimulus presentation would be reduced accordingly. The automatic mode was designed to mimic tape-recorded, automatic-playback testing. Stimuli were presented with a fixed interstimulus interval of 2 sec, which is comparable to the intervals found on commercially recorded tapes and compact discs. The
audiologist again identified whether or not the response was appropriate, but that identification had no impact on the rate of signal presentation.

Lists of 25 words were presented to 981 subjects under the two conditions. Lists were presented using both modes to each subject at the same signal intensity. One ear was tested in the automatic mode and the other in the manual mode. Ears were assigned to the two modes randomly.

The computer was programmed to quantify elapsed time between the onset of the initial speech signal and the completion of the 25-word list. Although interstimulus interval in the automatic mode was fixed, slight variations occurred in elapsed time across lists as a result of differences in word length. Records were kept of elapsed times and of difference in testing time as a function of presentation mode under all conditions for all subjects.

RESULTS

Mean elapsed time for the manual mode was 87.8 seconds (SD = 12.0 sec), while mean elapsed time for the automatic mode was 107.3 seconds (SD = 2.7 sec). Thus, elapsed time for word-recognition testing in the manual mode was 22 percent less than elapsed time in the automatic mode. This difference reached statistical significance (t = 49.3; df = 980; p < .05). Variability in the manual mode was substantially greater than in the automatic mode. The range of times in the automatic mode varied from 95 sec to 115 sec. In the manual mode, elapsed time ranged from 60 sec to 160 sec. The distribution of elapsed times for both presentation modes is shown in Figure 2.

In terms of individual performance, 812 (83%) subjects had elapsed times that were at least 10 sec shorter for the manual mode, and 26 (2%) had elapsed times that were at least 10 sec longer for the manual mode. The remaining 143 (15%) subjects had times within 10 sec for both presentation modes.

Because of the wide range of elapsed times in the manual mode, we asked whether any factors could be identified that might have influenced the amount of time necessary to complete word-recognition testing. Data analysis of the entire group showed no substantive correlation between elapsed time and age ($r^2 = 0.09$) or between elapsed time and degree of hearing loss (PTA) on the right ear ($r^2 = 0.14$) or the left ear ($r^2 = 0.17$). To further evaluate this question, we chose two subgroups of 24 subjects each from the extremes of the distribution. One subgroup, the fast responders, had elapsed times of less than 72 sec. The other subgroup, the slow responders,
had elapsed times of more than 117 sec. The mean age of the fast responders was 48 years, while the mean age of the slow responders was 69 years. This difference was statistically significant \((t = 3.86; \text{df} = 46; p < .05)\). In addition, and probably as a consequence, the fast-responder subgroup had less hearing loss, with a mean PTA of 16.9 dB, than the slow-responder subgroup, with a mean PTA of 45.0 dB. The difference was also statistically significant \((t = 5.48; \text{df} = 46; p < .05)\).

**DISCUSSION**

One of the reasons that live-voice, speech-audiometric testing is so popular, despite its obvious shortcomings, is that it is perceived to make word-recognition testing more efficient in the clinic. Results of the present study suggest that this perception is probably correct. We designed a computer-based approach to speech audiometry that resembled live-voice testing in terms of the pace at which stimuli could be presented. When compared to an approach that resembled the automatic-playback system of magnetic-tape or CD players, the live-voice, or manual, mode of presentation was shown to be 22 percent faster, on the average. In addition, in individual cases, over 80 percent of patients completed testing faster in the manual mode than in the automatic mode. To the extent that the manual mode accurately simulates live-voice testing, the perception that live-voice testing is more clinically efficient appears to be a valid one.

To be sure, the actual difference in elapsed time between manual and automatic modes of presentation was relatively small (19.5 sec). While such differences may not be clinically meaningful in routine audiometric assessment, the trend toward a more comprehensive speech audiometric evaluation may render such differences clinically significant when several different lists are administered in the course of an evaluation.

In the best case, the manual mode saved 49 sec of testing time for a PB-word list. In the worst case, the manual mode resulted in an elapsed time that was 65 sec longer than the conventional automatic mode. Analysis of the data extremes showed the faster responders to be significantly younger and have significantly less hearing loss than the slower responders. This finding is congruent with the clinical impression that conventional speech audiometric methods are too slow for some patients and too fast for others. In the majority of cases, the manual presentation mode resulted in increased clinical efficiency. In other cases, however, additional testing time was necessary. It seems likely that, for this latter group of patients, the allowance of more testing time should result in increased accuracy of speech audiometric measures, although no data are available from this study to support that notion.

The question of accuracy of these measures as a function of test modes is an interesting one and should be pursued. In the present study, we did not assess whether scores were different as a function of test mode. We tested one ear with the manual mode and the other with the automatic mode, and ears were assigned randomly. No meaningful differences in word-recognition scores were observed as a function of test mode. It would be interesting to know whether or not scores in the same ear would differ with test mode, especially in the slow responder and fast responder subgroups.

We were encouraged by the results of this study for at least two reasons. First, by completing the testing on nearly 1000 patients, we were able to conclude that the computer-based system was feasible for routine use in a busy clinical practice. Second, results of the study showed that efficiency could be increased by using a manual mode, much in the same way that efficiency is increased by live-voice testing. Unlike live-voice testing, however, this enhanced efficiency was accomplished using digitally recorded signals. Recent attempts by other laboratories to enhance speech audiometric techniques have included the use of compact disc technology (Wilson and Preece, 1990) to provide consistent, high-fidelity stimulus presentation. The problem with such systems, however, is that the manner of signal presentation, the automatic presentation of stimuli at fixed intervals, can be as inefficient as the conventional analog, tape-player concept. By combining a manual mode of presentation with digitally recorded speech signals, we were able to demonstrate the effectiveness of a system that took advantage of the best aspects of both live-voice testing and recorded speech materials.

**Acknowledgment.** This work was supported in part by a special-project grant from the American Speech-Language-Hearing Foundation and Apple Computer, Inc.

**REFERENCES**
