A computer-based system of speech-training aids for the deaf—A progress report*

by R. S. NICKERSON, D. N. KALIKOW, and K. N. STEVENS

Bolt Beranek and Newman, Inc.
Cambridge, Massachusetts

This paper is a progress report on an effort to develop a computer-based system of speech-training aids for the deaf. The project was begun with the assumption that an attempt to design such a system would probably fail, and that a more promising approach would be to attempt to evolve one through use. Accordingly, a system incorporating some of the capabilities that it was thought would be useful for speech training was developed, and installed at the Clarke School for the Deaf where it is now being used on an experimental basis in a remedial speech-training program. The expectation was that the capabilities of the system would be modified and extended as attempts to use it provided insights concerning what features it should have. To ensure that such insights do in fact guide the system’s evolution, developers and users are engaged in a continuing dialogue concerning the desirability and feasibility of specific modifications and extensions, both in the training procedures that are used in conjunction with the system and in the characteristics of the system itself.

The general considerations that governed the initial development of the system were the following. Deaf students receive only minimal acoustic information from the speech of others and from their own vocalizations. The speech skills they acquire are based on cues they receive from their residual hearing and from visual observations of the gestures of others. Often these skills are inadequate and incorrect, and the students thus need special training in order to help them to produce intelligible speech. As a part of this training, it is customary for a teacher to produce speech-like patterns or to describe the patterns to the child and for the student to try to imitate these patterns. The student is encouraged by the teacher if he produces the correct speech gesture. Three problems arise in this kind of training situation: (1) the relevant attributes of the speech sample produced by the teacher often cannot be seen, felt, nor heard by the student; (2) the student must rely on the teacher to indicate whether or not his production is acceptable; and (3) the teacher must make a subjective judgment as to the adequacy of the student’s production. All three of these problems provide motivation for developing a set of displays for use in a speech-training situation.

The idea of using visual displays of speech parameters to aid in speech training of the deaf is, of course, very old. In recent years, numerous instruments have been developed to produce a variety of different visual patterns. Our system incorporates within a single unit some of the kinds of displays described previously by others (although usually in modified form), as well as some new displays.

The system is built around a small digital computer, the Digital Equipment Corporation PDP-8E. Speech information is obtained from a miniature accelerometer attached by thin double-stick tape either to the throat or the nose, and from a head-mounted voice microphone. The accelerometer (BBN Model 501), which is approximately .3 inches in height and diameter, and weighs about 1.8 grams, is used to simplify the extraction of certain parameters that are relatively difficult to derive from a microphone output. When the accelerometer is attached to the throat it gives a waveform that has periodic peaks at the frequency of the glottal output during voiced sounds. The output is fed to a pitch extractor circuit that measures the time between positive-going zero crossings of the waveform and reports the pitch periods to the computer. When attached to the nose, the accelerometer provides a signal that is a measure of the amount of acoustic coupling to the nasal cavity through the velarpharyngeal port. In this case, the output, which is 10-15 dB higher when the velum is lowered—during nasalized sounds—than when it is raised, is fed to a component that rectifies and low-pass filters it, and sends the result on to the computer. The use of the accelerometer for the acquisition of pitch and nasality information is described more fully by Stevens, Kalikow, and Willemain. The output of the voice microphone is fed into a filter bank that reports to the computer the energy in each of 19 frequency bands within the range 100-6560 Hz. Data from the pitch extractor or nasality circuit (only one of these components is operational at a given time in the current system) and the filter bank are sampled by the computer 100 times per second, and used to generate a variety of visual displays. Control inputs from the user are given to
the computer via a set of push-buttons and analog knobs. For further details concerning the system, see Nickerson and Stevens.4, 5

Several different types of displays have been programmed. One provides the child with a game-like situation in which he can “shoot baskets” by performing certain vocal exercises. Another represents certain speech parameters in terms of changes of features of a cartoon face. Still another provides the capability of displaying individual speech parameters (amplitude, voicing, fundamental frequency, nasality) either singly or in various combinations as time functions. The system continuously records, both digitally and on analog tape, the most recent two seconds of speech. Most displays, therefore, have freeze and replay capabilities. Some of the programs permit the teacher to produce a target pattern on the display which the child can then be asked to attempt to match. They also incorporate the capability of moving patterns about on the display so as to facilitate visual comparison of a representation of a student’s utterance against a target that he may be attempting to match. In addition, they provide the means of showing, on request, the values of some of the parameters that are displayed.

Our initial experience with the system has been encouraging; however, it seems clear that how effective any speech-training aids will prove to be in practice will be bounded above by the specifics of the ways in which they are used. As technical developments make it feasible to do increasingly complex real time analyses of speech and to generate nearly anything one wants by way of displays, it becomes more and more apparent that pedagogical uncertainties impose the real limits on what one can expect to accomplish with speech-training aids, no matter how technologically sophisticated they may be.

A thought experiment demonstrates this point. Imagine a machine that could perform in real time any type of analysis of speech that one wished, and generate any display that one might specify. The fact is that we do not really know what analyses should be performed or what displays should be developed. Moreover, even if we knew the answer to these questions, it is not clear that enough is known about speech acquisition among the deaf to provide the basis for the training procedures that would take full advantage of such capabilities. What does seem clear to us is that the flexibility of a computer-based system provides opportunities for the type of exploration that is likely to be required to make progress on these problems.

Finally, the sort of close collaboration between researchers and teachers that we have attempted to maintain in this project is essential, we believe, if efforts to evolve effective training aids are to have a reasonable chance of success. This is not a new idea. Kopp expressed the need for a greater interaction between teachers and researchers by suggesting that the field would benefit “if we could make more teachers, and more researchers teachers.” Other writers have also advocated such interaction,7, 8 but few serious attempts to collaborate seem to have been made. The strategy is a reasonable one, we feel, not only for the development of this particular system but for that of any complex system that is to involve a real time interaction between men and computers on problems for which approaches are not highly formalized and the solutions are not well understood. As David9 has pointed out, the great versatility of the computer represents both an opportunity and a challenge. The opportunity is for creativity and innovation; the challenge is to be discriminating and practical. A close coupling between a system’s developers and its users is perhaps the only way to assure a balance between innovativeness and practicality from which something both new and useful may emerge.

ACKNOWLEDGMENTS

Each of the following individuals has contributed significantly to the design, implementation, and use of the system described in this report: Robb Adams, Patricia Archambault, Arthur Boothroyd, Douglas Dodds, Ann Rollins, Robert Storm, and Thomas Willenmain.

This project was sponsored by the U.S. Office of Education Media Services and Captioned Films Branch of the Bureau of Education for the Handicapped, under Contract No. OEC-0-71-4670 (015). We are pleased to acknowledge the encouragement and helpful suggestions of Lois Elliott, who served as contract monitor during the initial stages of the project.

REFERENCES