Chapter 13

Speech Technology and Speech Training for the Hearing Impaired

Diane Kewley-Port
Indiana University
Communication Disorders Technology, Inc.

Speech technology is being implemented in a variety of digital devices to improve communication by persons with hearing impairment. Recent advances in digital signal processing algorithms for speech permit development of powerful features that are implemented in the microprocessors of wearable sensory aids. Computer-based trainers for both speech perception and speech production incorporate many kinds of speech processing algorithms as the basis of feedback. This chapter provides a brief overview of speech processing in relation to the development of various digital devices for persons with hearing impairment. The primary focus is on speech-production training aids, in particular on research to validate feedback and to demonstrate the clinical efficacy of these systems. Future research on human judgements of disordered speech is essential for the development of computer-based trainers that can either enhance speech training or substitute for humans in the clinical process.

Advances in speech technology have the potential to make significant improvements in communication abilities by persons with hearing impairment. Speech technology generally refers to the analysis, synthesis, coding, or recognition of speech by computers. Applications of speech technology are incorporated in

devices for speech reception, speech production training, and speech perception training. One reason for the increased use of speech technology for persons with hearing impairment is the maturation of the field of *digital signal processing* (DSP). DSP is an area of applied mathematics that is normally taught in electrical engineering departments. DSP provides the tools for sampling and manipulating audio (and video) signals by computers, allowing replacement of analog hardware with flexible algorithms implemented in microprocessors, such as those found in computers and digital hearing aids (Engebretson, Morley, & Popelka, 1987). The purpose of this chapter is to examine how speech technology is applied in contemporary audiological rehabilitation and speech remediation today, and what are the future directions of this effort. The application area that will be discussed in depth is that of speech-production training aids. The chapter will emphasize the relation between devices employing speech technology and human users, including the evaluation of the clinical efficacy of such systems.

Speech technology involves both computer software and audio hardware such as microphones and loudspeakers. However, the more important and rapidly changing aspects of speech technology are the digital signal processing *algorithms* implemented in computers. Since algorithms for speech processing are usually developed for applications with normal speech, it is not always clear how these algorithms are selected and modified for applications with disordered speech. In addition, although speech is generally the signal of greatest importance for persons with hearing impairment, many assistive listening devices are developed primarily from knowledge about the processing of nonspeech signals, such as pure tones or noise. In recent years a great deal more research on speech perception abilities by individuals with a hearing impairment, with or without hearing aids, has been conducted (Leek, Dorman, & Summerfield, 1987). In the future we can expect that research on algorithms for standard speech technology applications and research on speech perception through impaired audiological systems will come together to solve many of the problems associated with the digital devices surveyed below.

The migration of speech technology from hardware systems, such as traditional spectrographs, to computer-based systems (e.g., the Computerized Speech Laboratory System by Kay Elemetrics), has taken place rather slowly over the past 10 years, except in specialized research laboratories. One reason for the slow penetration of these systems into classrooms or clinics was the expense of the digital-to-analog convertors (DACs) and, of course, the computer itself. Another reason was that a sufficient body of knowledge about the application of DSP algorithms to speech was not generally available to teachers and clinicians in the disciplines of speech and hearing.

This has now changed and the effect will be far-reaching. Computer manufacturers, starting with the developers of the NEXT computers in about 1989, believe that computer manipulation of audio signals is the new frontier for software applications. Thus inexpensive DACs (under $500) are being sold for PC-compatible computers, and engineering workstations are now routinely equipped

with high quality DACs (e.g., IRIS computers by Silicon Graphics). Interest in low-cost audio capabilities is also seen in the promotion of multimedia computer applications. Multimedia refers to computers which can integrate audio with video or animation displays. Although multimedia software has a great deal of potential to improve educational software, it is by no means clear that the high cost of developing such applications will make multimedia software inexpensive or widely available very soon.

It is likely that the use of speech technology in communication devices for persons with hearing impairment will become more commonplace in 5 to 10 years. It is incumbent upon clinicians, as well as researchers, to give careful thought to the role that these systems can play to enhance communication abilities. Primary issues to keep in mind are whether the new technology represents an improvement over existing systems or methods, and whether the technology has been shown to be clinically effective. The present chapter will focus on these issues primarily as they relate to speech training aids. To put that discussion in a larger perspective, first let us briefly review current speech technology applications for persons with hearing impairments.

OVERVIEW OF SPEECH TECHNOLOGY APPLICATIONS FOR INDIVIDUALS WITH A HEARING IMPAIRMENT

Speech technology has been implemented in a variety of devices to assist communication by persons with hearing impairment. The two major categories of such devices are aids intended primarily for speech *reception* (e.g., hearing aids) and aids for speech *training* (see Figure 1 for an example speech production trainer). Aids can be classified according to the acoustic-phonetic properties they analyze from the speech signal as follows. Some aids extract global properties of speech, for example an energy contour. Others analyze information only about the segmental aspects of speech in terms of spectral-temporal properties, for example vowel quality. Another approach extracts prosodic information, especially fundamental frequency contours. Any wearable device more than 10 years old relies exclusively on analog circuitry to analyze speech. More recently, however, digital devices have become the standard since weight and power requirements are reduced while the flexibility of software code enhances performance.

Efforts to develop devices for speech reception have always been far more numerous than those for speech training. The primary effort has, of course, been devoted to the development of acoustic hearing aids. Advancements in miniaturized, integrated analog circuits permitted the development of hearing aids with a wide variety of capabilities. However, the development of digital hearing aids on microcomputer chips has the potential for an even larger and quantitatively different set of possibilities. The development of a multiprocessor system in the 1980s (Engebretson et al., 1987; Morley, Engebretson, & Trotta, 1986) has now resulted in commercially available devices (e.g., Model 80 by 3M). Currently, most digital aids are programmed to meet individual needs

following standard audiological fitting procedures. In the future, however, digital algorithms will begin to solve difficult problems, such as noise suppression or enhanced speech processing for the profoundly deaf (Faulkner, Ball, Rosen, Moore, & Fourcin, 1992). Progress on these problems will result from both current and future research on the digital signal processing of speech, as well as behavioral studies of how humans process speech through hearing aids (see Van Tassel, 1993 for a review).

Applications of speech technology, related to speech coding, have been implemented in speech processors for multichannel cochlear implants from their inception. The original processing schemes, filter banks (McDermott, McKay, & Vandali, 1992) or F0/F1/F2 coding schemes (Blamey, Dowell, Clark, & Seligman, 1987) used analog circuits. More recently Finley et al. (1991) have implemented a digital speech processor that permits more flexibility in the development and evaluation of algorithms to improve speech coding. Since multichannel implants are presently limited to fewer than 22 channels to code all the multidimensional complexity of speech, flexible software algorithms are needed to determine how to enhance and optimize performance given these limitations.

Tactile aids have been investigated for decades as a way to provide an alternate sensory channel for persons with hearing impairment (for a review see Levitt, 1988). Tactile aids have been developed that present global (single-channel vibrators), segmental (Queen's University vocoder: Brooks, Frost, Mason, & Gibson, 1986; Blamey & Alcántara, 1994) or prosodic information (Boothroyd & Hnath-Chisholm, 1988). In the laboratory, multichannel, tactile transducers are typically controlled by computers in order to evaluate different types of speech processing algorithms (Weisenberger, Broadstone, & Saunders, 1989). Clearly the migration of these algorithms to wearable tactile aids will need to rely on microprocessors. Already, the series of wearable tactile aids developed by Boothroyd (1972, 1985) for presenting fundamental frequency contours have been redesigned as digital systems (Yeung, Boothroyd, & Remond, 1988). Research continues to demonstrate improved performance based on new approaches for processing speech (e.g., a principle-components analysis of spectral-temporal information: Weisenberger, Craig, & Abbott, 1991). Thus we can anticipate wearable tactile aids based on digital techniques to become more available and less expensive.

Computer-based methods of speech perception training are now available in addition to a variety of traditional methods in use. Gagné's (1994) chapter in this volume has thoroughly reviewed the theory, development, and clinical evaluation of speech perception training programs. In particular, Gagné's comment that most problems encountered in the development of speech perception trainers deserve a great deal more research applies equally to speech production trainers. The remainder of this paper will focus on a few of the critical problems associated with the development of speech-production training aids.

SPEECH-PRODUCTION TRAINING AIDS

Since persons with profound deafness have so much difficulty in acquiring intelligible speech, for many persons with a profound hearing loss the clear choice is manual rather than oral communication. Nonetheless, for those persons who want to improve their oral speech skills, a computer-based speech training aid can provide visual feedback and speech drill that supplements the traditional methods that require intensive training by human speech teachers. Speech training aids designed to improve the naturalness and intelligibility of the speech produced by persons with a hearing impairment have existed as laboratory systems since Gault developed a tactile aid in 1926. While it has not been demonstrated that the overall communication of deaf persons with profound deafness improves with the use of such devices (Nickerson, Kalikow, & Stevens, 1976), improvement has been clinically demonstrated for more limited speech production goals, such as improvement in the intelligibility of the particular phonemes trained (Arends et al., 1991; Fletcher, 1989; Kewley-Port, Watson, Elbert, Maki, & Reed, 1991; Youdelman & Levitt, 1991).

It is helpful to consider what is meant by a computer-based speech training (CBST) aid and what are some of the more important properties of such systems. A CBST system can process one or more signals derived from speech, either acoustic or physiological, for the purpose of providing feedback about speech production. Obviously the information that persons with a hearing impairment receive through the auditory system is severely degraded. Thus, a primary purpose of a CBST system is to provide visual feedback on computer monitors which can, in some sense, replace auditory feedback and assist the teacher or student in improving speech production.

An example of a computer-based speech training aid is shown in Figure 1. This is the *I*ndiana *S*peech *T*raining *A*id (ISTRA) system (by Communication Disorders Technology) that was originally developed through NIH and NSF funding at the Boy's Town National Research Hospital and Indiana University (Kewley-Port et al., 1991; Kewley-Port, Watson, Maki, & Reed, 1987; Lippmann & Watson, 1979). The components of the trainer shown in Figure 1 are typical of other CBST systems. First, the child speaks into a microphone attached to the computer. The microcomputer is most likely an IBM PC compatible computer, although occasionally trainers have been implemented on MacIntosh computers. In addition to the computer, a speech technology device is required to capture the speech for the computer, such as the speech recognition board used in ISTRA. Computer algorithms then process the speech, sometimes in comparison to a model (here the model is a "template"), and produce a visual display with information about the speech (called *feedback*) on the computer monitor. The model in ISTRA is a template that is made from the child's own best productions of a word spoken under the supervision of a speech teacher. ISTRA feedback displays the *quality* or goodness of the speech just produced in comparison to the template of the previous best efforts as stored in the computer.

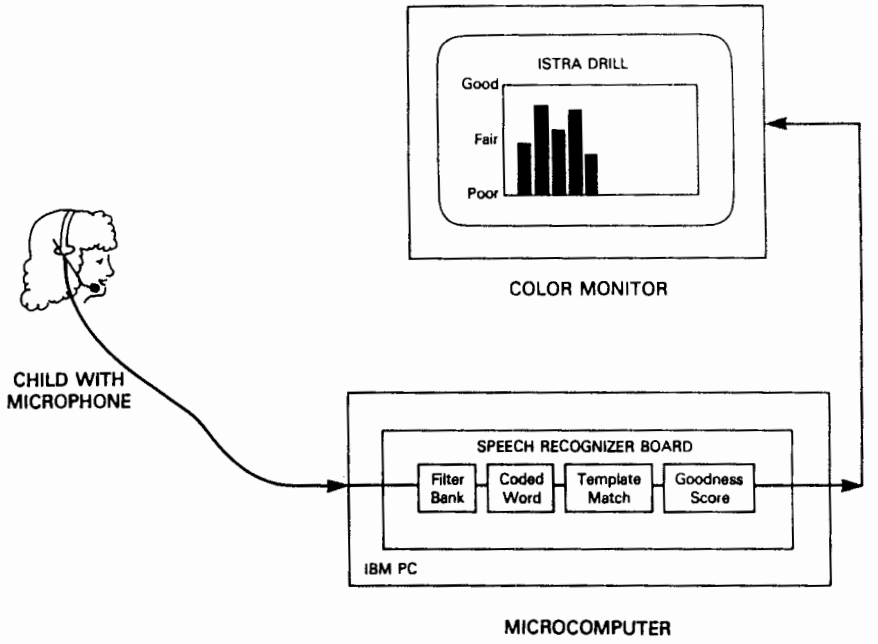

Figure 1. This is a schematic representation of the components used in the ISTRA speech production trainer. The child speaks into a microphone. The speech is sampled by the speech recognition board and matched to a stored template of the best productions of that word previously produced. The matching score provides a measure of speech quality that is displayed as visual feedback on the computer monitor.

Other possibilities for translating *acoustic signals* into visual displays, include single parameters, such as fundamental frequency (de Bot, 1983), or vowel targets (Povel & Arends, 1991). Visual information can also represent *articulatory gestures* that are not easily observed from the face, such as tongue placement provided by the palatometer (Fletcher, 1989). At the present time speech technology has sufficiently matured so that a number of variables can be simultaneously captured by a computer, analyzed, and displayed in glorious detail on a computer monitor. Indeed the number of different systems developed in the laboratory is quite large (for reviews see: Bernstein, Goldstein, & Mahshie, 1988; Lippmann, 1982). The important issues for research appear, therefore, not to revolve around the technology needed to capture information about speech but rather, around how to use this information effectively in speech training, and how to evaluate whether or not speech improves with the use of CBST systems.

EFFECTIVE FEEDBACK

Varieties of Feedback

The effective implementation of visual feedback in CBST systems requires both an understanding of the specific nature of *feedback* and a well-designed human interface of the composite display of that feedback. Two papers, one by Bernstein (1989) and one by Watson and Kewley-Port (1989), have attempted to describe the nature of feedback in some detail. The focus of Watson and Kewley-Port's categorical analysis of feedback considered: (a) the *physical measurement* on which feedback is based (acoustic, articulatory movement, or electrophysiological); (b) the *model* of correctly produced speech that is stored in the computer (e.g., teacher generated, computed); and (c) the *level of detail* represented in the computer displays (single parameters vs. multidimensional displays). This categorical analysis helps describe differences between kinds of feedback as illustrated by the following examples. These systems were chosen from only those that have been used in clinical trials by persons with a hearing impairment.

Physiological measurements of speech articulation can provide feedback about fundamental frequency, tongue placement, nasal airflow. Several physiologically-based systems have been developed specifically for individuals with a profound hearing loss (Palatometer by Kay: see Fletcher, 1989; CISTA by Matsushita: see Youdelman & Levitt, 1991). A virtue of these systems is that one (or more) of several one-dimensional parameters can be displayed, allowing the client to focus in detail on particular aspects of articulatory gestures. On the other hand, it is not clear what visual features of the time-varying signals convey *meaningful* information concerning differences in articulation. Since the same acoustic sound can be articulated in more than one way, supplying a model of correct production is a significant problem for physiological feedback. Thus, it is mandatory that a specially trained teacher be continuously present to interpret the physiological displays to clients.

Training aids which are based on acoustic input from a microphone are easier to use in school environments than are physiological systems. Some examples of these systems demonstrate issues concerning the *model* of speech provided. A system may have no model if the feedback is very simple, such as representing loudness by a balloon that inflates (e.g., SpeechViewer by IBM: see Ryalls, 1989). In these cases, the feedback may be said to have face validity in the sense that change in the acoustic measurement and in the feedback are highly correlated. Nonetheless, the *validity* of this simple feedback should be established. That is, the balloon could inflate following a linear scale of pressure measurement, or a logarithmic (dB) function or another mathematical transform. Establishing which of these representations provides the "best" measure of loudness in terms of perception or production is a research question, one that should be asked but is often ignored. The need for research is more acute as the feedback becomes more complex. For example, consider the very complex detail presented in whole spectrograms. Even though the use of spectrographic dis-

plays as models was shown to improve production in a group of deaf teenagers for a limited set of fricatives (Maki, 1983), such displays are inherently very complex and the validity (or even appropriateness) of this type of model ought to be systematically investigated.

Two groups of researchers have been very concerned with validating feedback in training systems. The first is Povel and his colleagues in the Netherlands who have developed an acoustic trainer for vowels (Visual Speech Apparatus, see Povel & Arends, 1991; Arends et al., 1991). The display consists of "ellipses" locating the correct targets for the production of individual vowels. While these targets are related to traditional measurements of the first and second formant frequencies, the two dimensions are actually derived from a more complex, principle component analysis of vowels. These targets have been experimentally validated separately for speech from men, women, and children using perceptual measures (Povel & Wansink, 1986).

Another group has investigated the validity of evaluative feedback concerning the *speech quality* of utterances used in ISTRA (see Watson, Reed, Kewley-Port, & Maki, 1989). In this case feedback is based on a "goodness" metric derived from the output of commercial speech recognizers. Watson et al. (1989) and subsequently Anderson and Kewley-Port (1993) have developed procedures to establish whether goodness metrics can substitute for human judgements of speech quality (e.g., "That's good," "Oh what a good one," etc.). These studies demonstrated that valid goodness metrics can be obtained from some recognizers (see below).

Relation Between Human and Computer-Based Feedback

During the course of investigating how computer feedback can substitute or enhance human feedback in some of the specific cases noted above, several more general issues have arisen. First, there is a need to identify and describe different categories of judgements of disordered speech made by teachers during speech training. Three logically distinct categories of judgements of disordered speech can be described. The first is *identification,* in which the word produced is judged as either correct or incorrect, or as consisting of a particular set of phonemes or features. The second is the evaluation of the *speech quality* of utterances. Finally, the speech teacher may provide an *interpretation* of the production, such as describing where the tongue was placed on the palate. For the first two categories, computer-based feedback can, in principle, substitute completely for the human judgement. For interpretive judgements, computer-based feedback can provide a qualitatively different kind of feedback than can the human observer alone, but either the teacher or the student would still be needed to do the interpretation.

In traditional speech training methods human teachers usually judge disordered speech in one of these three categories as the basis for their feedback. Before a computer can replace feedback given by humans, more knowledge concerning how these judgements are made is needed. Research on human performance in

identification and speech-quality tasks has been reported from several different points of view but it appears that little is known about human performance in interpretative tasks. In the preceding chapter, Metz and Schiavetti (1994) have reviewed studies of humans assessing the "intelligibility" of speech produced by individuals with a hearing impairment. Assessment tests may involve identification tasks in terms of closed- or open-set responses of read materials, or speech quality rating tasks, such as the National Technical Institute for the Deaf (NTID) Speech Intelligibility Rating Scale (Subtelny, 1977). However, the purpose of these tests, namely measuring intelligibility in terms of communicated meaning, is not the same as assessing phonetic detail in individual utterances in speech drill. Thus specific studies of the latter behavior are also needed. In the case of replacing human feedback with computer-based feedback for judgements of speech, there is an even clearer need for specific research on human performance.

Recently Anderson and Kewley-Port (1994) have proposed methods to compare human performance with the performance of speech recognizers in judging speech. Two types of identification tasks were compared, the identification of phonemes in minimal-pair utterances and the identification of disordered speech as either correctly or incorrectly produced. In addition, correlational methods, derived from Watson, Reed, Kewley-Port, & Maki (1989) were developed to compare human and computer ratings of speech quality in isolated words produced by normal-hearing, misarticulating children. First, the reliability of humans to rate the quality of disordered speech (from very poor to normal) is calculated. Then a measure of the "true quality" of the speech utterances is derived from averages of ratings from only the reliable listeners. Finally, the correlation between the goodness scores from the recognizer and the true-quality ratings indicates whether the goodness scores can substitute for human judgements of speech quality in speech drill.

The first and very important step in making these comparisons is the development of an appropriate database of speech samples. Anderson and Kewley-Port (1994) noted that for speech disorders in normal-hearing, misarticulating children, results of research on disordered phonology limits the large number of possible phonetic substitution errors to around 25 common errors (see Elbert, Rockman, & Saltzman, 1980). This number can be further reduced and ordered by referring to the *implicational hierarchy* of phonological errors described by Dinnsen, Chin, Elbert, and Powell (1990) and Gierut (1992). Thus, Anderson and Kewley-Port (1994) recorded a database of speech (called ISPEED) focused on a modest number of phonological errors. Talkers included normal-hearing adults and normal-hearing, misarticulating children. Evaluation of that database indicated that speech content is sufficient to establish the performance of speech recognizers for most speech errors that might occur in the speech of normal-hearing persons.

In traditional speech drill the feedback provided to the student is determined by the judgements of *trained* teachers. Thus, in our investigation graduate students in speech and hearing were selected as the human judges (Anderson &

Kewley-Port, 1994). Their performance was then compared to analogous computer-based judgements of speech in the ISPEED database obtained from three commercial speech recognizers. The study yielded two different sets of conclusions. Surprisingly, the trained human judges were not particularly reliable at either the correct/incorrect scoring or the speech-quality-rating tasks when listening to large numbers of repetitions of the same word (reliability coefficients from 0.4 to 0.8 depending on the specific words and disorder). Measures of interrater correlation even from reliable raters also varied substantially depending on the particular utterances and disorders (from 0.6 to 0.9). Even in the correct/incorrect scoring task, the trained listeners were not able to agree if there were many, some, or no correct productions in judging a /w/ for /r/ substitution error (also, see Shriberg, 1972). These results clearly indicate that more research is needed to understand how trained teachers judge disordered speech. This knowledge can then be used to establish baseline performance for analogous judgements using speech technology.

The investigation of the three speech recognizers evaluated by Anderson and Kewley-Port (1994) revealed that the performance on the identification and correlational tasks were in contraposition. Two recognizers compared favorably to human performance in rating speech quality, but they were not very accurate in the identification tasks. Performance for the other recognizer was just the reverse. Clearly, future research on the development of algorithms that can provide valid feedback for both identification and speech quality tasks is needed. One hopeful line of research is that of Deng and his colleagues who are developing Hidden Markov Model (HMM) algorithms based on articulatory feature models (Deng, 1991; Deng & Erler, 1993; Deng & Sun, 1993). Deng has been collaborating with us in revising these algorithms specifically to perform well on the ISPEED database. In the future, collaboration between researchers who study disordered speech and those who develop DSP algorithms will be needed in order to make significant progress in providing valid feedback for speech training.

Unfortunately, the total number of datasets in ISPEED for which criteria of human performance have been established is quite small. Recording speech from normal-hearing, misarticulating children that ranges from incorrectly produced to correctly produced utterances must take place over months. It is also very time consuming to obtain human-listener judgements. It seems clear that if computer-based judgements of speech are going to be validated, a much larger database will be needed to investigate human judgements of disordered speech. Moreover, speech from talkers who have a severe-to-profound hearing impairment may present a different set of problems that need to be evaluated separately. The token-to-token variability along both segmental and prosodic dimensions in the speech of individuals with a hearing impairment is quite high in contrast to the rather consistent errors found in the speech of normal-hearing, misarticulating children. We have begun to collect speech samples from children with hearing impairment at the Central Institute for the Deaf, and will repeat

evaluations of the identification and speech-quality tasks previously undertaken for normal-hearing, misarticulating children.

Feedback: Summary and Conclusions

In summary, the variety of possible types of feedback for speech training is very large and even the small number of CBST systems that are commercially available present a bewildering set of feedback capabilities. Further research on the production and perception of disordered speech is needed to establish that the information conveyed in any particular form of feedback is valid for speech training. (We might provide a cautionary note that even valid feedback can be displayed in confusing, distracting, or inappropriate ways rendering it unusable for the student.) Also, each type of feedback including that provided by human teachers alone can provide only a limited amount of information in relation to the overall goal of teaching intelligible natural speech. Thus, in the future we must strive to develop within a single CBST system a variety of feedback options (in different speech drill formats) and thereby allow the speech teacher to select the most appropriate ones for the student. Of course such a system would be useful to the speech teacher only to the extent that the variety of drills are integrated into a structured, speech-training curriculum.

CLINICAL EVALUATION OF SPEECH TRAINING AIDS

Ultimately, the success of a computer-based speech trainer will be measured by clinical evaluations, that is demonstrations that speech improves as a result of training on that system. We need to be alert to the danger that the enthusiasm of clinicians and children for the colorful, video-game formats of the speech drills will be accepted as a measure of clinical effectiveness. Studies of treatment efficacy are very expensive and time consuming and, as recent conferences and position papers have noted (Kewley-Port, 1990; Olswang, 1990), there has been a paucity of research on the evaluation of traditional treatment methods. This leads to the perplexing situation that there are few models of clinical trials that might be adapted to the evaluation of CBST efficacy. Moreover, there is little data in the published literature on the efficacy of traditional methods that can then be compared to evaluations of computer-based methods.

In fact, computer-based trainers require several different kinds of evaluation before they can be used efficiently in treatment. Kewley-Port (1990) has described four levels of evaluation associated with CBST systems. In brief these are: (a) tests of the *acceptability* of the system by speech teachers and students, (b) *beta-test* evaluations at sites remote from the developers, (c) tests of *clinical effectiveness* conducted in collaboration with the developer, and (d) *independent verification* of clinical effectiveness. Acceptability evaluations are conducted with a few target users of a CBST system, by means of observation, discussion, and possibly questionnaires. Information gathered during this phase assists in the development of a system. In the next step, so called beta-testing, systems

are installed in sites where they will be used routinely. The purpose is to fine-tune the device using feedback from the speech teachers. Sometimes this step involves questionnaires which can elicit testimonial data on user satisfaction. Such testimonials are then all too often included in promotional materials. We must all be cautious not to mistake these claims for genuine evaluations of clinical effectiveness.

Studies of the clinical effectiveness can use one of several methods to demonstrate that speech improves as a result of training on a particular CBST system. Developers of CBST systems in academic settings have used single-subject designs (Kewley-Port et al., 1991), and pre- and post-test measures of groups using the systems (Arends et al., 1991; Fletcher, 1989; Kewley-Port, 1990). Occasionally designs that incorporate experimental and control groups have been used, such as those in progress in our laboratory. Of course such studies might be somewhat contaminated by the direct role taken by the developer. Nonetheless, it would seem that some positive results of clinical effectiveness should precede wide distribution or sales of CBST systems and the developers themselves are in the best position to conduct the initial investigations. In fact, the validity of the results of such studies can be established through the use of *independent listening juries* (i.e., evaluations of speech by individuals who are not involved in the training).

Truly independent clinical evaluations of CBST systems are, of course, the most desirable. Practically speaking, a trainer would have already achieved some acceptance before time and money would be committed to such studies. A few have been reported, for example studies with the VisiPitch (by Kay Elemetrics; de Bot, 1983), for fundamental frequency training using a tactile aid (McGarr, Youdelman, & Head, 1989), and with the CISTA palatographic displays (Youdelman & Levitt, 1991).

CONCLUSIONS

This report has discussed the dilemma of conducting clinical trials of computer-based training. On the one hand it is obvious that such trials must be conducted in order to establish the validity of this approach to rehabilitation. On the other hand, the long and costly process of conducting those evaluations deters even the undertaking of clinical trials. In a similar vein, Gagné (1994) in this volume has advocated systematic, evaluative research throughout the development of computer-based speech *perception* trainers, although this research is also difficult to carry out. It has been argued in this chapter that research is needed at all levels in the development of speech production trainers as well, starting with the DSP algorithms that analyze speech and ending with independently conducted, controlled studies of clinical effectiveness.

Perhaps we should step back and consider the justification for the development and evaluation of computer-based trainers. For individuals with a significant hearing impairment who choose to communicate orally intense speech drill is the only way to achieve high quality and intelligible speech. This drill, tradition-

ally provided by human teachers, requires more time and more money in proportion to the severity of the hearing impairment. Speech technology already makes an enormous variety of feedback (and therefore drill) possibilities available. The cost of the required research to produce a valid and effective trainer is justified to the extent that CBST systems can enhance speech training. That is, either the speed of the training process, improvement in speech quality, student motivation, or improved efficiency of human resources may justify the cost of CBST systems. It is the opinion of most professionals who work with CBST systems the justification of those costs exists and the paths to successful development are now known. Keeping in mind that speech technology generates many kinds of information that can be displayed as various forms of feedback, research efforts should concentrate on establishing effective feedback and on how to integrate different kinds of feedback into a useful training curriculum. This research effort should remain very close to the clinical process and to various levels of clinical evaluation in order to chart a path towards effective commercial speech training systems.

ACKNOWLEDGEMENTS

This work reflects the effort of a strong team of collaborators who have worked on the ISTRA project over the last 8 years. The author gratefully acknowledges the contributions of Prof. Charles S. Watson; Prof. Daniel P. Maki; Patricia Cromer, CCC-SLP, CED; and Anne Summers, CCC-SLP. Portions of this manuscript were presented at the 1992 Biennial International Convention of the Alexander Graham Bell Association, San Diego, CA, June, 1992. This work is supported in part by an NIHDCD SBIR grant, DC-00893 to Communication Disorders Technology, Inc.

REFERENCES

Anderson, S., & Kewley-Port, D. (1994). *Evaluation of speech recognizers for speech training applications.* Manuscript submitted for publication.

Arends, N., Povel, D., Van Os, E., Michielsen, S., Claassen, J., & Feiter, I. (1991). An evaluation of the visual apparatus. *Speech Communication, 10,* 405-414.

Bernstein, L.E. (1989). Computer-based speech training for profoundly hearing-impaired children: Some design considerations. *The Volta Review, 91,* 19-28.

Bernstein, L.E., Goldstein, M.L., Jr., & Mahshie, J.J. (1988). Speech training aids for hearing-impaired individuals, I: Overview and aims. *Journal of Rehabilitation Research and Development, 25,* 53-62.

Blamey, P.J., & Alcántara, J.I. (1994). Research in auditory training. In J.-P. Gagné & N. Tye-Murray (Eds.), *Research in audiological rehabilitation: Current trends and future directions* [Monograph]. *Journal of the Academy of Rehabilitative Audiology, 27,* 161-191.

Blamey, P.J., Dowell, R.C., Clark, G.M., & Seligman, P.M. (1987). Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant. *Journal of the Acoustical Society of America, 82,* 38-47.

Boothroyd, A. (1972). Sensory aids research project at the Clarke School of the Deaf. In G. Fant (Ed.), *Speech communication ability and profound deafness* (pp. 367-377). Washington, DC: A.G. Bell Association for the Deaf.

Boothroyd, A. (1985). A wearable tactile intonation display for the deaf. *IEEE Transactions of Acoustics, Speech, and Signal Processing, ASSP-33,* 111-117.

Boothroyd, A., & Hnath-Chisholm, T. (1988). Spatial, tactile presentation of voice fundamental frequency as a supplement to lipreading: Results of extended training with a single subject. *Journal of Rehabilitation Research and Development, 25,* 51-56.

Brooks, P.L., Frost, B.J., Mason, J.L., & Gibson, D.M. (1986). Continuing evaluation of the Queen's University tactile vocoder I: Identification of open-set words. *Journal of Rehabilitation Research and Development, 23,* 119-128.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induces practice behavior. *Language and Speech, 26,* 331-350.

Deng, L. (1991). The semi-relaxed algorithm for estimating parameters of hidden Markov models. *Computer Speech and Language, 5,* 231-236.

Deng, L., & Erler, K. (1993). *Microstructural speech units and their HMM representation for discrete utterance speech recognition.* Manuscript submitted for publication.

Deng, L., & Sun, D. (1993). *A statistical framework for automatic speech recognition using the atomic units constructed from overlapping articulatory features.* Manuscript submitted for publication.

Dinnsen, D.A., Chin, S.B., Elbert, M., & Powell, T.W. (1990). Some constraints on functionally disordered phonologies: Phonetic inventories and phonotactics. *Journal of Speech and Hearing Research, 33,* 28-37.

Elbert, M., Rockman, B., & Saltzman, D. (1980). *Contrasts: The use of minimal pairs in articulation training.* Austin, TX: Exceptional Resources, Inc.

Engebretson, A.M., Morley, R.E., & Popelka, G.R. (1987). Development of an ear-level digital hearing aid and computer-assisted fitting procedure. *Journal of Rehabilitation Research and Development, 24,* 555-564.

Faulkner, A., Ball, V., Rosen, S., Moore, C.J., & Fourcin, A. (1992). Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and audiological abilities. *Journal of the Acoustic Society of America, 91,* 2136-2155.

Finley, C.C., Wilson, B.S., Zerbi, M., Hering, D., Van den Honert, C., & Lawson, D.T. (1991). *Speech processors for audiological prostheses* (NIH Contract N01-DC-9-2401). Research Triangle Park, NC: Neuroscience Program Office, Research Triangle Institute.

Fletcher, S.G. (1989). Visual articulatory training through dynamic orometry. *The Volta Review, 91,* 47-64.

Gagné, J.-P. (1994). Visual and audiovisual speech perception training: Basic and applied research needs. In J.-P. Gagné & N. Tye-Murray (Eds.), *Research in audiological rehabilitation: Current trends and future directions* [Monograph]. *Journal of the Academy of Rehabilitative Audiology, 27,* 133-159.

Gault, R.H. (1926). Touch as a substitute for hearing in the interpretation and control of speech. *Archives of Otolaryngology, 3,* 122-135.

Gierut, J.A. (1992). The conditions and course of clinically induced phonological change. *Journal of Speech and Hearing Research, 35,* 1049-1063.

Kewley-Port, D. (1990). Cross-disciplinary advances in speech science. *ASHA Reports #20: Proceedings of The Future of Science and Services Seminar,* 69-85.

Kewley-Port, D., Watson, C.S., Elbert, M., Maki, D., & Reed, D. (1991). The Indiana Speech Training Aid (ISTRA) II: Training curriculum and selected case studies. *Clinical Linguistics and Phonetics, 5,* 13-38.

Kewley-Port, D., Watson, C.S., Maki, D., & Reed, D. (1987). Speaker-dependent speech recognition as the basis for a speech training aid. *Proceedings of the 1987 IEEE International Conference on Acoustics, Speech, and Signal Processing,* 372-375.

Leek, M.R., Dorman, M.F., & Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 81,* 148-154.

Levitt, H. (1988). Recurrent issues underlying the development of tactile sensory aids. *Ear and Hearing, 9,* 301-305.

Lippmann, R.P. (1982). A review of research on speech training aids for the deaf. In N.J. Lass

(Ed.), *Speech and language: Advances in basic research and practice, vol. 7* (pp. 105-133). New York: Academic Press.

Lippmann, R.P., & Watson, C.S. (1979). New computer-based speech training aid for the deaf. *Journal of the Acoustical Society of America, 49,* 467-477.

Maki, J. (1983). Application of the speech spectrographic display in developing articulatory skills in hearing impaired adults. In I. Hochberg, H. Levitt, & M.J. Osberger (Eds.), *Speech of the hearing impaired: Research, training, and personnel preparation.* Baltimore: University Park Press.

McDermott, H.J., McKay, C.M., & Vandali, A.E. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *Journal of the Acoustical Society of America, 91,* 3367-3371.

McGarr, N.S., Youdelman, K., & Head, H. (1989). Remediation of phonation problems in hearing-impaired children: Speech training and sensory aids. *The Volta Review, 91,* 7-18.

Metz, D.E., & Schiavetti, N. (1994). Current and future directions in research on speech intelligibility assessment of persons who are deaf. In J.-P. Gagné & N. Tye-Murray (Eds.), *Research in audiological rehabilitation: Current trends and future directions* [Monograph]. *Journal of the Academy of Rehabilitative Audiology, 27,* 237-249.

Morley, R.E., Engebretson, A.M., & Trotta, J.R. (1986). A multiprocessor digital processing system for real-time audio applications. *IEEE Transactions of Acoustics, Speech, and Signal Processing, ASSP-34,* 225-231.

Nickerson, R.S., Kalikow, D.N., & Stevens, K.N. (1976). Computer-aided speech training for the deaf. *Journal of Speech and Hearing Disorders, 41,* 120-132.

Olswang, L.B. (1990). Treatment efficacy research: A path to quality assurance. *ASHA, January,* 45-47.

Povel, D.J., & Arends, N. (1991). The Visual Speech Apparatus: Theoretical and practical aspects. *Speech Communication, 10,* 59-80.

Povel, D.J., & Wansink, M. (1986). A computer-controlled vowel corrector for the hearing impaired. *Journal of Speech and Hearing Research, 29,* 99-105.

Ryalls, J. (1989). Comparison of two computerized speech training systems: SpeechViewer and ISTRA. *Journal of Speech-Language Pathology and Audiology, 13,* 53-56.

Shriberg, L.D. (1972). Articulation judgments: Some perceptual considerations. *Journal of Speech and Hearing Research, 15,* 876-882.

Shriberg, L.D., Hinke, R., & Trost-Steffen, K. (1987). A procedure to select and train persons for narrow phonetic transcription by consensus. *Clinical Linguistics and Phonetics, 1,* 171-189.

Subtelny, J. (1977). Assessment of speech with implications for training. In F. Bess (Ed.), *Childhood deafness* (pp. 183-194). New York: Grune and Stratton.

Van Tassel, D.J. (1993). Hearing loss, speech, and hearing aids. *Journal of Speech and Hearing Research, 36,* 228-244.

Watson, C.S., & Kewley-Port, D. (1989). Computer-based speech training (CBST): Current status and prospects for the future [Monograph on Sensory Aids for Hearing-Impaired Persons, N. McGarr (Ed.)]. *The Volta Review, 91,* 29-46.

Watson, C.S., Reed, D., Kewley-Port, D., & Maki, D. (1989). The Indiana Speech Training Aid (ISTRA) I: Comparisons between human and computer-based evaluation of speech quality. *Journal of Speech and Hearing Research, 32,* 245-251.

Weisenberger, J.M., Broadstone, S.M., & Saunders, F.A. (1989). Evaluation of two multichannel tactile aids for the hearing-impaired. *Journal of the Acoustical Society of America, 86,* 1764-1775.

Weisenberger, J.M., Craig, J.C., & Abbott, G.D. (1991). Evaluation of a principal-components tactile aid for the hearing-impaired. *Journal of the Acoustical Society of America, 90,* 1944-1957.

Yeung, E., Boothroyd, A., & Remond, C. (1988). A wearable multichannel tactile display of voice fundamental frequency. *Ear and Hearing, 9,* 342-350.

Youdelman, K., & Levitt, H. (1991, July). *Speech training of deaf children using a palatographic display.* Presented at International Symposium on Speech and Hearing Sciences, Osaka, Japan.