Computer-Assisted Speech Training for the Hearing Impaired

M.J. Osberger, R.P. Lippmann*, M.P. Moeller, and J.M. Kroese
The Boys Town Institute for Communication Disorders in Children

**Massachusetts Institute of Technology Lincoln Laboratories*

A computer aid is being developed for use in a speech-training program for hearing-impaired children. The aid consists of a small computer and analog analysis equipment which includes pitch and nasality detectors and a bank of filters. It evaluates a child's utterances using a template comparison similar to that used in speaker-dependent word-recognition systems. The aid is being used in a speech-training program to supplement teacher training by providing the time-consuming speech drills that appear to be necessary to establish automatic, correct speech patterns. Preliminary data collected with a small group of hearing-impaired children indicate that the aid can be used to effectively provide the drill component of the speech-training program. The results also show that more speech drills can be performed per unit time with the aid than with the teacher alone.

It has been well documented that the speech of severely and profoundly hearing-impaired children is largely unintelligible (Brannon, 1966; Markides, 1970; Smith, 1975). Improving the speech skills of deaf children has been a concern of educators and clinicians for many years because the adequacy of such skills can have considerable influence on the social, educational, and career opportunities available to these individuals. We, along with others (e.g., Calvert & Silverman, 1975; Ling, 1976; Monsen, 1978), do not consider poor speech intelligibility to be a necessary consequence of severe or pro-

M.J. Osberger, Ph.D., is a Research Associate and Coordinator of Clinical Research in Aural Rehabilitation at Boys Town Institute for Communication Disorders in Children, Omaha, Nebraska. R.P. Lippmann, Ph.D., formerly a Research Associate and Head of the Communication Engineering Laboratory at Boys Town Institute, is now a member of the Technical Staff at MIT Lincoln Laboratories, Lexington, Massachusetts. M.P. Moeller, M.S., is Coordinator of the Clinical Program in Aural Rehabilitation at Boys Town Institute, and J.M. Kroese, M.A., is a Research Assistant in the Aural Rehabilitation Laboratory at Boys Town Institute.

found hearing loss. This notion is supported by the good speech intelligibility of some profoundly deaf children and adults.

The poor speech intelligibility of most profoundly hearing-impaired children is not surprising in view of the quality and quantity of the speech training which they receive. A recognized problem is that speech pathologists and teachers of the hearing-impaired often are not adequately trained to provide speech services to hearing-impaired children. Increased inservice training and improved programming at the preservice level of training have been suggested as ways of alleviating this problem (Hochberg & Schmidt, in press). Also, it is only within recent years that well-defined, systematic speech-training procedures have been developed for use with the hearing impaired. These speech-training strategies are gradually replacing eclectic approaches which were often successfully applied only by a master teacher.

An unresolved problem is the lack of carryover of speech skills into spontaneous speech outside the therapeutic environment. Although this is one of the greatest problems faced in teaching the hearing impaired, there has been little research concerning this issue. Clinical experience suggests that frequent, repetitive speech drills are necessary to develop, stabilize, maintain, and promote carryover of speech skills beyond the training setting (Ling, 1976). The effectiveness of this approach, however, remains to be empirically demonstrated.

The limited quantity of speech training remains a critical problem. It is not unusual to find that one speech teacher or clinician is responsible for providing services to as many as 40 or 50 children in a school. Often, a child is seen for individual speech therapy only once or twice a week for a brief session, or the child receives instruction with a large group of other children. In addition, many teachers of the hearing impaired do not provide developmental or maintenance speech training in the classroom (Hochberg & Schmidt, in press). The quantity of speech services is further reduced when the children reach the intermediate school years where academic subject matter may receive greater emphasis than speech.

The inadequate quantity of speech services appears to be at least partly a result of limited budgets which restrict the number of speech teachers in a school. One approach to this problem is to use computer-assisted instruction to supplement teacher training. A computer aid could provide, without supervision, repetitive speech drills and appropriate feedback using a scheme similar to that employed by speaker-dependent word-recognition systems. We are in the process of developing such a computer aid. If effective, and produced at low cost, such aids could be used to relieve a teacher of the time-consuming drill that appears to be necessary to establish automatic, correct speech patterns. In the clinical trials currently underway, we also intend to develop and evaluate systematic speech-training procedures designed to promote carryover of speech skills learned during training sessions to those used in conversational speech.

The primary purpose of this paper is to give a progress report on our early experience with the computer speech-training aid, including a description of the aid and the ways it has been used. In addition, we will present a brief overview of the speech-training program and preliminary data on the effectiveness of the aid in providing speech drills.

METHODS

The Speech-Training Program

An overview of the speech-training program is presented, and it is important to note that we have collected our data using only a small part of the program. Further investigations will include a more extensive evaluation of the speech-training procedures.

The speech-training procedures are based on the program described by Ling (1976) with modifications suggested by Osberger, Johnstone, Swarts, and Levitt (1978) and Osberger (in press). Emphasis is placed on speech drills to achieve automaticity of speech production. In the training tasks, new speech patterns are added to the drill sequence whenever the child achieves reliable, intelligible production of the current sequences. This procedure is designed to both add new skills to the speech repertoire and to promote concurrently the transition of already well-developed patterns into meaningful use. Throughout training, the acquisition of speech skills is fostered by teaching-ordered target behaviors and subskills. A primary objective of the drill is to promote automaticity so that conscious attention to articulation does not interfere with higher level activities required for fluent speech.

Four types of drill techniques are used to develop automaticity of speech production. The first three drills have been developed by Osberger et al. (1978), and the fourth technique has been developed for this project to promote carryover. The types of drill techniques are:

1. Imitation–child imitates auditory-oral model
2. Production on Demand–spontaneous production elicited by visual cueing (graphic display or written word)
3. Discrimination--identification of auditory signal using a simple pointing response (to facilitate development of an auditory-kinesthetic feedback loop)
4. Self-Evaluation–the child is prompted to make a judgment regarding the intelligibility of self-utterances; both the accuracy of the production and the accuracy of the judgment are evaluated

The first three drill techniques are applied in parallel, with the fourth procedure being introduced after a child achieves a specified level of performance. We consider the fourth drill, self-evaluation, to be an important step in the carryover process because a child must learn to evaluate and modify her or his productions via her or his own internal feedback system if the desired production patterns are to become firmly established.

Training is planned to proceed in systematic steps from syllables to words to short phrases to sentences to spontaneous conversation. In addition to the automaticity training to promote carryover, the program will include extra-clinic training by parents and teachers in situations outside the speech-training sessions. The overall structure of the training program is such that linguistic complexity and functional communication demands gradually increase, and reinforcement gradually decreases.

Although some work will involve the suprasegmental characteristics of speech, the early phases of the speech-training program will focus on segmental skills, starting with vowels and progressing to consonant production. Our rationale is that recent research (Boothroyd, Archambault, Adams, & Storm, 1975; Houde, 1973; Osberger & Levitt, 1979) has demonstrated that intelligibility does not necessarily increase when suprasegmental characteristics of speech improve. In contrast, high correlations have been reported between intelligibility and number of segmental errors (Smith, 1975). We do not mean to imply that good suprasegmental production is not a desirable goal in a speech-training program for the hearing impaired, but our main concern at this time is the improvement of intelligibility. We plan to add suprasegmental training to the segmental training scheme in later stages when work is initiated with phrases and sentences.

A pre-training evaluation indicated that the six children who participated in the study were able to produce many of the vowels accurately in nonsense syllables. Intelligible productions of the target vowels in words, however, were inconsistent. Therefore, training focused on the carryover of skills to meaningful speech (words) and the subsequent stabilization of these skills. To this end, speech drills using a production-on-demand task were performed. The children's production of drill words (VC, CV, or CVC monosyllables) was elicited with written words. The consonants which appeared in the drill words with the target vowels were those judged by the teacher to be established in a child's speech. The role of the teacher and computer in the speech-training program is described in a later section.

The Computer Speech-Training Aid

Previous Speech-Training Aids. In order to appreciate fully the characteristics of the computer aid used in our research, it is necessary to give a brief description of previous speech-training aids. The majority of speech-training aids that have been used in the past have provided very simple displays of a limited number of acoustic characteristics of speech. Use of such displays has not resulted in dramatic improvements in intelligibility, probably because only limited aspects of the child's deviant speech patterns were corrected. In contrast, aids which provided more complex displays, such as a real-time speech spectrum, contain so much detail that it is often difficult for the child to locate and reproduce the target features. In addition, previous aids, whether using a simple or complex display, required the presence of a skilled teacher who had to guide and monitor training, interpret the display, and provide feedback to the child. Even if a child could work independently on some of these devices s/he still had to evaluate her/his productions by complex external criteria. If a child's self-evaluations were incorrect, poor production patterns were reinforced. With no exceptions, all speech-training aids developed in the past have been relatively "unintelligent" displays that relied on the teacher or student to guide the speech-training procedures or make decisions concerning the correctness of utterances based on the display. In contrast, the present computer aid automatically performs a complex acoustic analysis of the child's productions, evaluates predetermined aspects of the child's speech, guides and monitors training, and gives immediate feedback regarding the correctness of production.

Computer-Aid Hardware. The speech-training aid, first described by Lippmann and Watson (1979) consists of a small computer and analog analysis equipment which includes pitch and nasality detectors and a bank of filters. There are currently two models of the aid, a laboratory-based system which uses a PDP-11/14 computer and a portable aid which uses an LSI-11/23 microcomputer. The portable aid is mounted on a cart with all ancillary devices and has been designed so that it can be moved from a laboratory setting into an educational setting with hearing-impaired children. Because the majority of our data have been collected with the portable aid, we will restrict our description of the hardware components to this model.

A block diagram of the portable aid appears in Figure 1. An utterance is analyzed using pitch and nasality detectors and a filter bank. The filter bank consists of 16 contiguous filters which span the region 100 to 8000 Hz and have bandwidths approximately equal to psychoacoustic "critical bands" (1/3 octave filters except for the lowest two which are approximately one octave wide). The filters are followed by level detectors which include RMS averaging and logarithmic conversion and have a dynamic range of approximately 60 dB. This speech-analysis system is similar to that used by Klatt (1979); Pols (1977); and Searle, Jacobson, and Kimberly (1980) in speech-recognition systems. Not shown in the figure is a speech-whitening filter and associated level detector, the output of which is used to determine the presence or absence of speech energy. There is also an 800 Hz low-pass filter and level detector. The output of this filter is used, along with other signals, to determine which parts of an utterance are voiced. The LSI-11/23 has 32K words of memory, two RK05 disk drives, and a hard-copy printer.

The video terminal is used to communicate with the child. At no time is the terminal used to present a speech display that the child attempts to match with her or his own production. Rather, the student receives instructions, speech prompts, and feedback messages on the screen. The screen can also be used to generate printed text or various types of graphic displays. A loudspeaker is mounted in the vicinity of the student so that prompts and instructions can be delivered aurally to the child, if desired. Not shown is a second terminal which can be used by the teacher to monitor and control the training session.

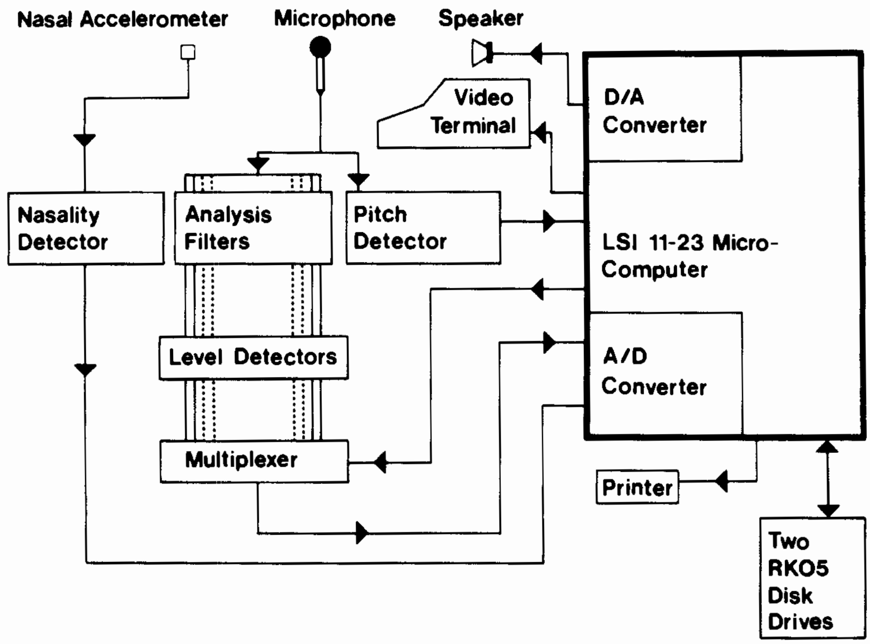

Figure 1. Block diagram of the portable computer speech-training aid.

Computer-Aid Software. As mentioned previously, the computer aid automatically evaluates a child's productions and provides rapid feedback regarding the correctness of the production. This is accomplished by a template comparison, similar to that used in speaker-dependent word-recognition systems. Prior to a training session, tape recordings are made of a child producing multiple tokens of an utterance. Templates are then made of those recorded words in which the target sound in the utterance was judged by the clinician to have been produced most intelligibly by the child. Qualitative production aspects are ignored by the clinician in selecting words for the computer; intelligibility of production is the only selection criterion. When a computer template is made, the aid performs a real-time acoustic (nasality, duration, level, pitch, and spectral) analysis; and the computer stores this information. When a child returns for a training session, her or his goal is to produce all tokens of a target sound as intelligibly as the sound in the reference (template) utterance. Thus, a child is trying to match her/his own previous best production. During the training session, each word produced by the child is acoustically analyzed in real-time and then compared to the reference utterance. Of course, the aid is not judging intelligibility per se but is making physical comparisons between the acoustic characteristics of an utterance previously judged to be intelligible by the teacher and a child's attempt to replicate the utterance.

The acoustic comparison between the reference and live samples involves sampling and storing the outputs of the filter bank and detectors every 10 msec. while speech is present. During the time the aid is actively looking for speech, samples are stored in a ring buffer. After an utterance begins, samples are stored in a linear array until the utterance ends. Techniques similar to those described in Martin (1976) are used to mark the beginning and end of utterances and reject extraneous noise. Using such techniques, it is unnecessary for a child to operate any controls. Only spoken utterances are required. Once samples are obtained, these samples are amplitude- and time-normalized.

The comparison between the live and reference samples yields a difference score. The larger the number, the greater the difference between the two utterances. The difference scores are computed in a manner that allows spectral, temporal, pitch, and nasality features to be differentially weighted. If the difference score is greater than a preselected criterion value, the child receives negative feedback. Otherwise, feedback is positive.

In the present study, vowel production was analyzed using only spectral information. Duration, pitch, nasality, and level were not evaluated. Spectral comparisons were performed only on that part of the word which was judged by the aid to contain the vowel. The child's production of the consonants in the training words was not included in the acoustic analysis.

Figure 2 depicts the events which occur during a training session. In this example, the child is prompted to say the word "cat." An acoustic analysis is performed on the live utterance and a comparison is made between the acoustic characteristics of the utterance and the reference utterance. The area between the dashed lines comprises that segment of the word (i.e., the vowel) that is acoustically analyzed. A difference score is computed and the appropriate feedback is then given to the child.

The aid has three control modes. In the teacher-controlled mode, feedback and training are under direct control of the teacher who sends commands and messages with an auxiliary video terminal. In the computer-teacher-controlled session, the computer controls training and decides on feedback after each utterance. The teacher can observe the training and override any decision made by the computer. Finally, in the computer-controlled session, the computer controls all aspects of training without the necessity of teacher

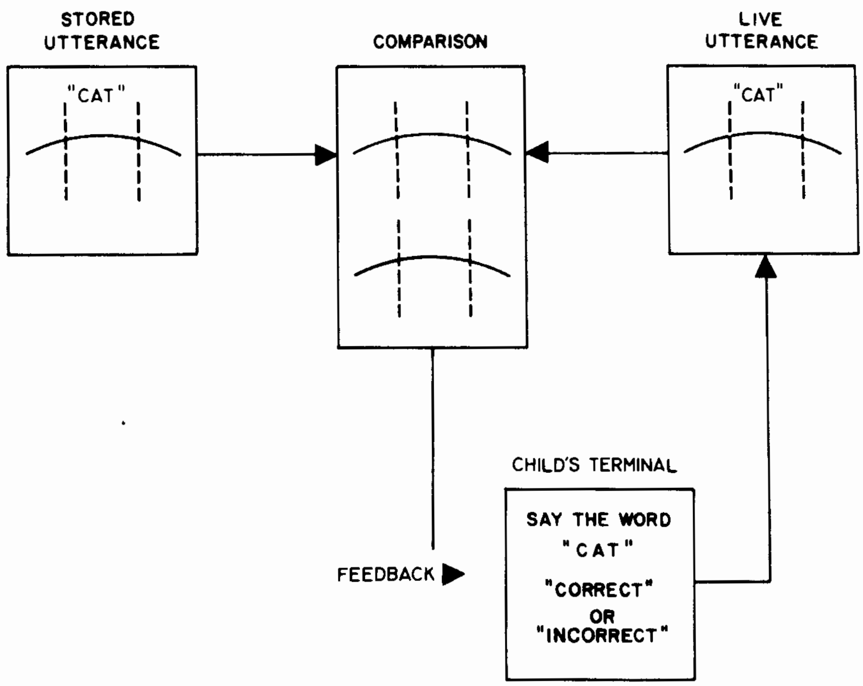

Figure 2. Diagram of the events which occur during a training session with the computer aid (the acoustic representation of a target speech sound and output from one low-frequency channel of the filter bank; time vs. level is displayed by the curve).

supervision.

The aid is designed to be used with programmed instruction and operant-conditioning techniques. It will guide a child through specific training tasks, provide feedback after each utterance, keep detailed records on performance and the acoustic characteristics of the utterance, and use special features designed to motivate subjects.

Subjects

Six children, 9 to 12 years of age, with prelingual severe and profound sensorineural hearing losses participated in the study. They were all educated in Total Communication programs; four of the children attended a state residential school for the deaf, and two of the children were educated in a resource classroom for the hearing impaired in a public school system with partial mainstreaming. Relevant information regarding age, sex, nonverbal IQ, and degree of hearing loss appears in Table 1.

The children were seen for a twenty-minute speech-training session four

Table 1
Summary of Relevant Information on the Children Participating in the Study

Computer-Trained Subjects	Age	Sex	Nonverbal I.Q. Range–WISC-R	PTA Better Ear (dB HL)
JW	10-2	F	High Average	97
LD	12-7	M	Very Superior	92
TT	12-5	F	High Average	93
CB	12-4	M	Average	105
TN	11-0	M	Average	103*
RM	11-5	M	High Average	85*

*Two-frequency average

times a week. It should be mentioned that two control groups were also included in the study. One control group of children received systematic speech training from a highly experienced speech pathologist who was a member of the project staff; the other group received only speech training normally provided to all children by the school-related agency.

The Role of the Teacher and the Computer in the Training Program

Instruction and drill were divided between the clinician and the computer. The computer conducted the drill, but it did so under the supervision of the teacher. The teacher-computer control mode was used; and the teacher, independent of the computer, rated the intelligibility of the children's productions. The final decisions on feedback were based on the teacher's judgments. This scheme was used in order that the accuracy of the aid in "rating" intelligibility could be checked.

The clinician's responsibilities consisted of the following: (a) evaluation of needs and selection of speech targets; (b) establishing new skills, using a variety of teaching techniques if necessary; (c) stabilizing emergent skills to a level where the techniques available from computer-assisted instruction were sufficient, and (d) preparation for the drill, including selection of utterances to be used in template comparisons, determining reinforcement schedules, and preparing text for instructions and feedback messages. The role of the aid in this evaluation study was to present speech prompts to the child, determine when to advance to the next stimulus, provide feedback, and keep extensive performance records.

RESULTS

The data presented should be viewed as preliminary in nature because the first evaluation period was only six weeks long. This is an extremely short period of time when one considers that years are usually devoted to teaching profoundly hearing-impaired children to speak intelligibly.

First, we were interested in determining if more drill work was accomplished per unit-time by the computer than by the teacher. To determine this, tally was made of the number of drill words produced in each session by the children trained by the computer and the children trained by the project speech pathologist. An analysis of these data showed that the children trained by the aid produced many more utterances per session than the children trained by the teacher. Table 2 shows that, on the average, the children trained by the computer produced 127 words per drill session, whereas the children trained by the teacher produced only 74 words per session. Thus, almost twice as many more words were drilled by the aid than by the teacher. The rapid pace at which training progressed in the sessions with the aid also permitted a greater number of different words to be drilled than was possible in the sessions conducted by the teacher. As shown in Table 2, an average of 55 different words were produced per session when the aid was used; slightly less than half of this number were produced by the children working with the teacher.

Table 2

Average Number of Words and Number of Different Words Produced Per Training Session by the Two Groups of Children

	Total Number of Words		**Number of Different Words**	
	Mean	**Range**	**Mean**	**Range**
Computer-Trained	127	104-158	55	43-68
Project-Teacher-Trained	74	58-92	25	18-37

The children's performances in the computer-based training sessions were monitored to assess learning rates. The performance measure was the number of target vowels produced intelligibly in a variety of monosyllabic words. It should be recalled that the final decisions regarding the intelligibility of a child's productions was made by the supervising teacher and not by the computer. Some examples of the children's performances with the computer-aid are plotted in Figures 3 and 4. In Figure 3, the average percentage of intelligible productions of the vowel /ɪ/ (as in hid) in monosyllabic words is plotted for the group of children. During each session, the target vowel was produced in 10 to 20 different words, with each word produced two consecutive times. During the initial drill session, the children produced the target vowel intelligibly 73% of the time. This level of performance ensured that the children experienced a modest degree of success in producing the desired speech targets. It may appear to be a high level of performance, but our training goal was to elicit intelligible productions 90 to 100% of the time and to stabilize this level of performance. The data show a gradual increase in percent of intelligible productions of the target vowel across training sessions. At the end of the last session, the children, on the average, were producing the target vowel intelligibly in words 93% of the time. This represents a 20% improvement in performance between the first and last sessions.

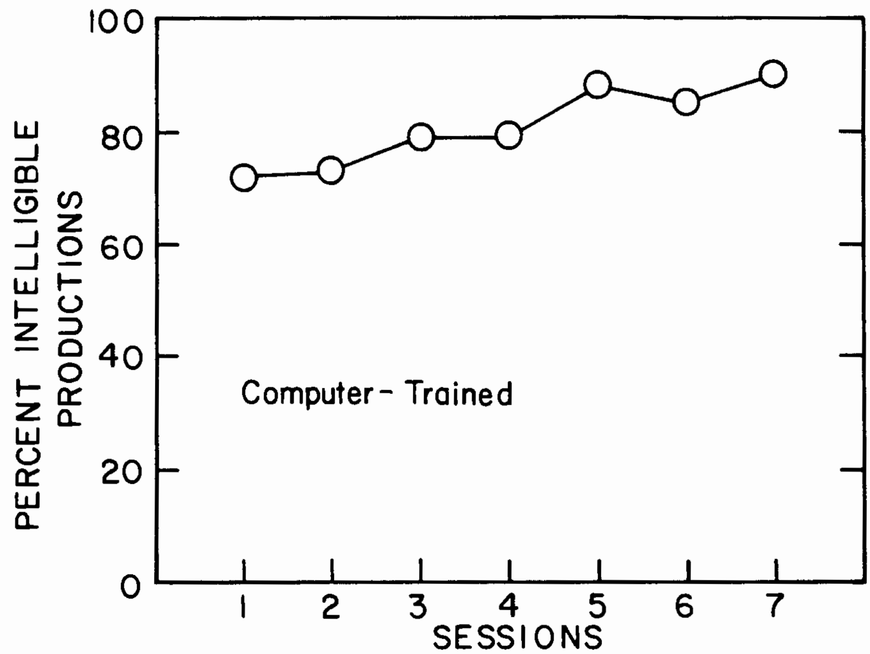

Figure 3. Percent intelligible productions of the vowel /ɪ/ (as in hid) elicited from the children drilled by the computer (score averaged across trials in each training session and across children).

Evidence of learning becomes more apparent when the data for individual children are plotted. In Figure 4, data obtained on two children are plotted. In order to illustrate best the effects of learning, performance on individual trials within a session is plotted. A trial consisted of 8 to 10 words, each produced two times in succession. As the data show, the percentage of intelligible productions by Child 1 for the vowel /i/ (as in heed) gradually increases after Trial 3, and for trials 9 to 12, 90 to 100% of the target-vowel productions are intelligible. Child 2 shows a slightly different learning pattern for the vowel / æ/ (as in had) in that there is a sharp increase in intelligible vowel production during Trial 5, with a subsequent stabilization of this performance level.

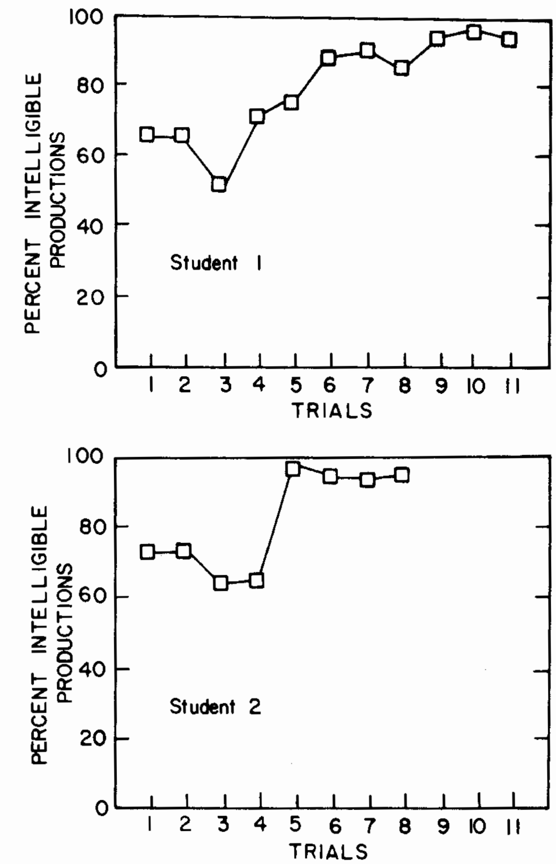

Figure 4. Percent intelligible productions of the vowel /i/ (as in heed) for Child 1 (upper plot), and the vowel / æ/ (as in had) for Child 2 (lower plot) on individual trials during training sessions.

DISCUSSION

Our preliminary data indicate that the drill component of a speech-training program can be provided by a computer. For the limited speech drills under study, the children showed evidence of learning, and there were measurable improvements in intelligibility. From these preliminary data, it appeared that the computer-aid procedure was at least as effective as the teacher in performing the speech drills. However, additional studies are required to determine the range of types of tasks for which computer-aided drills are successful.

Our data show that more drill can be completed per unit-time by a computer than by the teacher. A major reason for the large difference in number of words produced by the two groups of children is that a teacher often engages in conversation, instruction, and other communicative interchange with the child, reducing the time available for drill. This, in turn, reduces the amount of practice that a child receives in producing the target speech pattern. When the computer is used, the speech prompt, the child's response, and the feedback all occur within seconds.

The children enjoyed using the computer; and even after six weeks, they remained highly motivated to perform the speech drills with the aid. Clinical experience suggests that in traditional speech therapy sessions, lack of motivation may not be restricted to the children but may be a teacher problem as well. Often, teachers resist drill work because they contend that the children become bored with the task. More often than not, it is the teacher who becomes bored with this type of activity and understandably so. A computer, however, never becomes bored or disinterested with its assigned job.

The drill was provided by the computer in the present study under the direct supervision of the teacher. This was required because the aid was not always accurate in judging the "intelligibility" of the children's speech. One reason for this appeared to be that the children produced multiple tokens of a particular word with large acoustic variations, yet the utterances were still perceived as intelligible. As a result, the computer rejected intelligible productions. We intend to refine further the performance of the computer by carefully examining the acoustic characteristics of intelligible and unintelligible productions and then modifying and improving our analysis procedures to be sensitive to the acoustic features which are important for the speech skills being taught.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Ray Kent and Dr. Charles Watson for their assistance during the course of the study and preparation of the manuscript; Mr. Mike Collins, Mr. Tom Creutz, and Mr. Richard Wenner for their programming support; and Mr. Doug Adams and Mr. Alex Grossman for their technical assistance. We are very grateful to the students, teachers, and administrative personnel of the Omaha Public School System and the Nebraska School for the Deaf for their cooperation during the study. This work was supported partially by the Biomedical Research Development Grant from NIH and Grant No. NS16247 from NIH/NINCDS.

REFERENCES

Boothroyd, A., Archambault, P., Adams, R., & Storm, R. Use of computer based aids for deaf persons. *The Volta Review,* 1975, *77,* 178-193.

Brannon, J.B. The speech production and spoken language of the deaf. *Language and Speech,* 1966, *9,* 127-135.

Calvert, D., & Silverman, S.R. *Speech and deafness.* Washington, D.C.: A.G. Bell Association for the Deaf, 1975.

Hochberg, I., & Schmidt, J. A model inservice-preservice training program to improve the speech of hearing-impaired children. In I. Hochberg, H. Levitt, & M.J. Osberger (Eds.), *Speech of the hearing impaired: Research, training, and personnel preparation.* Washington, D.C.: A.G. Bell Association for the Deaf, in press.

Houde, R.A. *Instantaneous visual feedback in speech training for the deaf.* Paper presented at the Annual Convention of the American Speech and Hearing Association, 1973.

Klatt, D.H. Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics,* 1979, *7,* 279-312.

Ling, D. *Speech and the hearing-impaired child: Theory and practice.* Washington, D.C.: A.G. Bell Association for the Deaf, 1976.

Lippmann, R.P., & Watson, C.S. New computer-based speech-training aid for the deaf. *Journal of the Acoustical Society of America,* 1979, *66,* S13(A).

Markides, A. The speech of deaf and partially hearing children with special reference to factors affecting intelligibility. *British Journal of Disorders of Communication,* 1970, *5,* 126-140.

Martin, T.B. Practical applications of voice input to machines. *Transactions of Institute of Electrical and Electronic Engineers: Acoustics, Speech, and Signal Processing Society,* 1976, *64,* 487-500.

Monsen, R.B. Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research,* 1978, *21,* 197-219.

Osberger, M.J. The development and evaluation of some speech training procedures for hearing-impaired children. In I. Hochberg, H. Levitt, & M.J. Osberger (Eds.), *Speech of the hearing impaired: Research, training, and personnel preparation.* Washington, D.C.: A.G. Bell Association for the Deaf, in press.

Osberger, M.J., Johnstone, A., Swarts, E., & Levitt, H. The evaluation of a model speech-training program for deaf children. *Journal of Communication Disorders,* 1978, *11,* 293-313.

Osberger, M.J., & Levitt, H. The effect of timing errors on the intelligibility of deaf children's speech. *Journal of the Acoustical Society of America,* 1979, *66,* 1316-1324.

Pols, L.W. *Spectral analysis and identification of Dutch vowels.* Amsterdam: Academische, 1977.

Searle, C.L., Jacobson, J.L., & Kimberly, B.P. Speech as patterns in the 3-space of time and frequency. In R.A. Cole (Ed.), *Perception and production in fluent speech.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Smith, C.R. Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research,* 1975, *18,* 795-811.