

COMPUTER-AIDED REHABILITATION

Donald V. Torr
Gallaudet College

My topic is computer-aided rehabilitation. What I have attempted to do is devise a means by which I might assist individuals in the development of skills required to decode and encode speech for communication purposes. I will be introducing some ideas which may have application in your field, albeit they may have to be turned $\pm 30^\circ$ to fit your perception of the problem. I have used the term rehabilitation; however, my personal interests, as a result of my experiences at Gallaudet, are perhaps better identified by the term habilitation.

COMPONENTS OF A POSSIBLE TRAINING SUBSYSTEM

My predisposition toward training and the educational process is to attempt to design a system so that the learner may proceed as an individual to the maximum extent possible independent much of the time, but continually aware of the appropriateness his behavior and his need for personal contact with the teacher requires. Figure 1 represents such a system.

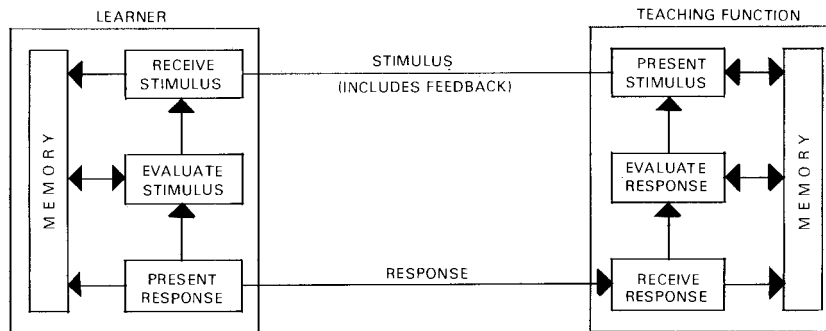


Figure 1: A basic system which will permit learning to take place.

The teaching functions of Figure 1 can be, and generally are, carried out by a teacher or a therapist. For the purpose of this paper I have attempted to mechanize the teaching functions shown in Figure 1 to permit an individual to develop some of the skills necessary to decode and encode speech. This will be the system which I will describe. In practice, both the teacher and the mechanized system would be used to carry out the teaching functions, i.e., they are both subsystems of a larger system, as suggested in Figure 2. In an attempt to simplify things I will emphasize the mechanized subsystem.

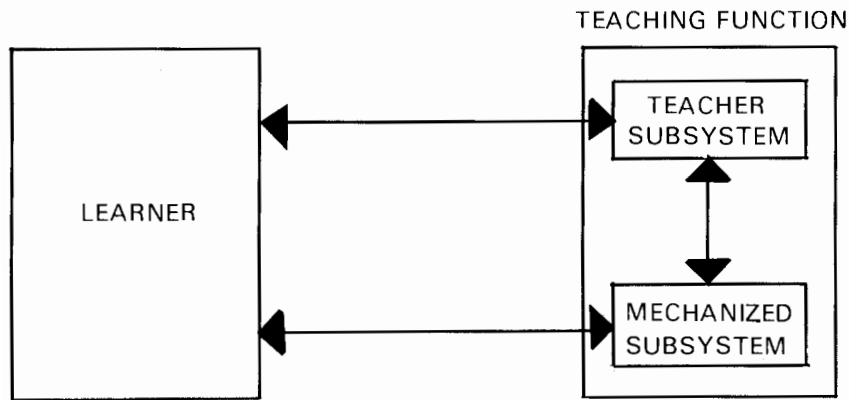


Figure 2: A partially mechanized system.

Figure 3 pictures a system which could be assembled today. I will describe the components briefly.

1. **COMPUTER.** The subsystem I have in mind requires that a variety of devices be controlled and that an ordered sequence of events be followed, based upon unambiguous decision logic. I have allocated these functions to the computer.

I will digress a moment to discuss the computer. At Gallaudet the Computer Center is a component of the Office of Educational Technology. In July 1971 we converted to a time-sharing computer. We installed a Digital Equipment Corporation PDP-10. Our purpose in doing this was to make the computer more readily available to the students and the faculty. This becomes feasible when you employ the concept of time sharing.

Briefly, the idea of time sharing is to use the extremely rapid speed of the computer in a manner which permits it to serve many slow humans at the same time. When I say slow humans, I mean all humans. Our computer will do some 100,000 calculations per second, thus, the machine is typically waiting for the human. Our equipment currently will handle 16 individuals at the same time. Communication between man and machine is by a Teletype or Cathode-ray tube terminal. The equipment can be expanded to handle 127 persons in what appears to be a simultaneous fashion. I hope to expand to this capacity when we can justify it in terms of student and faculty use. I feel we at Gallaudet should try to exploit the computer as we look for ways to improve the education of the deaf.

2. **USER TERMINAL.** We need a means by which we can present information to the human user (the subject, the client, the student). We will provide earphones for sound and speakers or jacks which will permit the user to be trained while using his own hearing aid. We will also wish to display visual information and will use a television screen for this purpose.

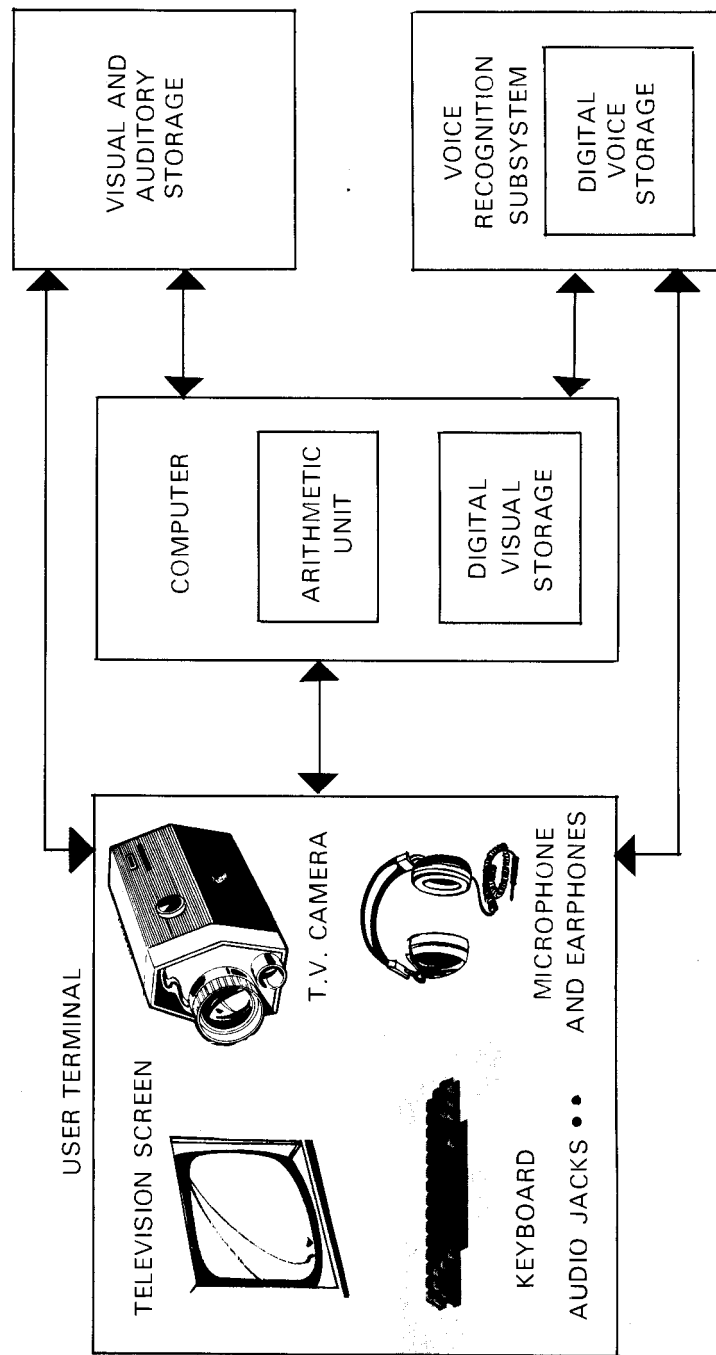


Figure 3: A possible computer-assisted rehabilitation system.

The terminal will also be the means by which the user enters information into the mechanized subsystem. To accommodate this we will provide him with a keyboard for the entry of alphanumeric and other symbolic information. We will also give him microphones for the entry of auditory signals and the television camera for the entry of visual or video signals.

3. ANALOG VISUAL AND VOICE STORAGE. For this system to have utility it will be necessary to provide for the storage of visual and auditory information in an analog form. For this purpose I will use videorecording and playback equipment with a stereophonic sound capability. The video recording equipment will provide storage of visual and auditory stimuli to be used for lipreading or speech reading training as well as auditory discrimination training. It will also be used for speech training.

For the purpose of this paper I investigated, to a modest extent, equipment which could be used to store audio and video information to be retrieved under computer control. I was not impressed by the audio equipment and spent considerable time considering a video disk device. The device permits computer selection of video images. Its storage capacity is limited, its cost is high, and worst of all it was not possible to store the audio information. Since I was looking for a way to store stimulus material for lipreading training and auditory discrimination training, I definitely wanted perfectly synchronized visual and auditory information.

As a result of this finding it became clear that we would have to fabricate this capability if we wanted it. I believe it can be done and am anxious to try it when I can find the necessary resources. It seems to be possible to do it by using a continuous loop of videotape, running past several read and record heads. Information once recorded could be read repeatedly since the tape loop would turn endlessly. The amount of information which could be stored would be a function of the length of the tape loop. By controlling the time at which the tape is read it would be possible to store different messages on the tape and read only the message you wanted as it passed the read head. We could, if you will, "space share" this loop as we are "time sharing" the computer. Thus, students could see different television messages at different terminals. This loop could not store all the information we would need. Three-quarter inch cartridge tape machines are available today which give good color picture quality as well as very good stereo sound reproduction over the range 50-12,000 hertz. These machines can be controlled remotely. I envision them being used here for bulk storage. Each cartridge will hold one hour of video and audio information. Several of these machines would be placed in the system. Each would be loaded with a cartridge carrying one or more classes of stimulus materials, e.g., gross auditory discrimination or finer auditory discrimination material. As the user or student at the terminal required new stimulus material the computer would make that determination, select the television recorder carrying the proper cart-

ridge, and cause that unit to transfer the required information to the tape loop. It would be possible to choose to show the visual information or not. Thus, during training one could show a picture of the source of the sound, e. g., whistle, siren, doorbell, and during testing suppress the visual information.

4. DIGITAL VISUAL AND AUDITORY STORAGE. It will be necessary to display directions, questions, and a variety of alphanumeric information on the television screen in order to instruct and evaluate the user. This information will be stored using storage media associated with the computer *per se*. Files of data on user performance will be maintained in this way.

For speech training it will be necessary to provide temporary storage of digitized voice information obtained from the user. This storage will be provided by the voice recognition subsystem discussed in the next section.

5. VOICE RECOGNITION SUBSYSTEM. Clearly I am attempting to push the technology available today to the limit when I incorporate this subsystem as part of the total design. Nonetheless, the capability I will briefly describe is sold commercially under the name Voice Command System. The system, pictured in Figure 4, will output a digital code in response to a spoken input. I have "talked" to it.

The device divides the audio input into 16 frequency bands between 200 and 5000 hertz. Samples are taken of each of the 16 bands every 1/60 second, multiplexed into a single output, and converted into digital form. This digital output is then "compressed" to a fixed length and stored. The machine is user dependent. It is necessary to train it to a particular user's voice. To do this the person speaks a particular "command" five times. The five compressed codes are then combined into a single digital pattern which includes the tendencies common to the five utterances, as well as variations that occur from utterance to utterance. This final pattern is then stored with other commands (expandable from 24 commands to 128).

The compression process is proprietary to the manufacturer. It is discussed to some extent in a May 10, 1971 issue of *Electronics*.¹ Figures 4.a and 4.b present information contained in that article. On the left, in Figure 4.a, are two spectrograms of one individual's pronunciations of the Spanish word *bueno*. On the right are the compressed equivalents. Every command, regardless of its length is compressed to the same 120 bit length. The vertical lines on the spectrograms identify the compression intervals defined by the machine. The intervals vary in length. The more rapidly the spectral energy changes the shorter the compression interval. The compression algorithm attempts to retain all information related to change and remove information which remains static.

In operation, an incoming command is similarly analyzed, compressed, and matched against all stored commands. The best match (or no match) is then identified and used to establish the digital output of the machine. This output can be used to control other machines.

In operation a pause of at least 250 milliseconds is used to signal the end of a command. The identification is available 20 milliseconds later.

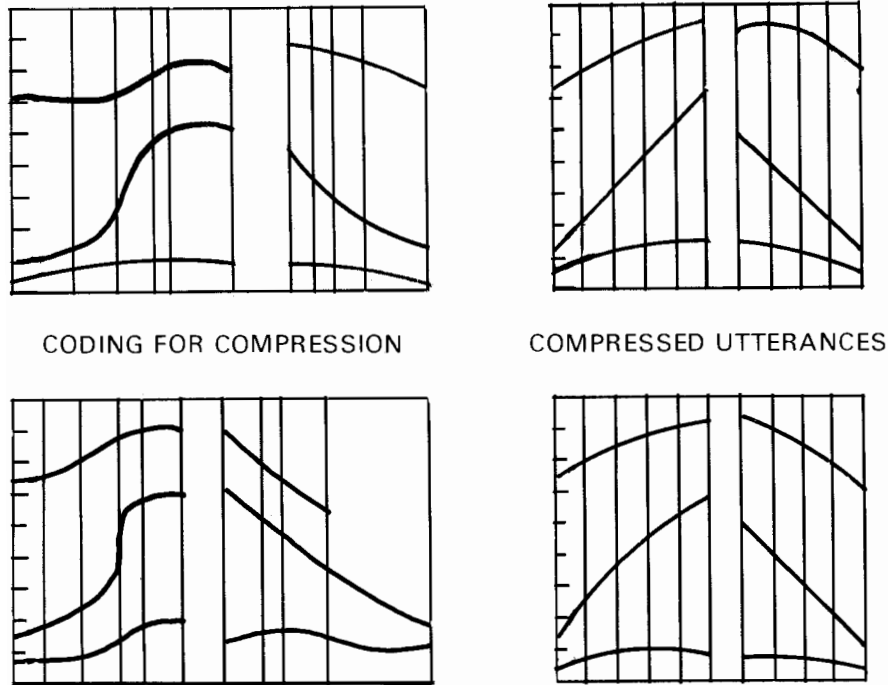


Figure 4a: Compression of two spectrograms for the Spanish word "bueno".

Figure 4.b illustrates the matching process pictorially with one incoming pattern being matched against four stored patterns. This part of the process appears to be simple. The incoming 120 bit word is compared with each stored 120 bit word on a bit by bit basis. When the comparison shows both bits are the same the computer tallies one. The total number of tallies is a measure of the match. The comparison showing the highest total (over some threshold) is selected as the best match. In effect the machine is establishing a digital equivalent of a spectrogram which it then compresses to simplify the identification process and reduce the storage requirement.

There are reservations on the use of this device, but it is used operationally today, e.g., to control conveyor belts. In my estimation it is worthy of study, when used in the manner I will later describe, because it provides a digital equivalent of a spectrogram which can be used in real time. In spite of its limitations it appears to provide a means for giving an individual rapid feedback on the probable correctness or incorrectness of his utterance.

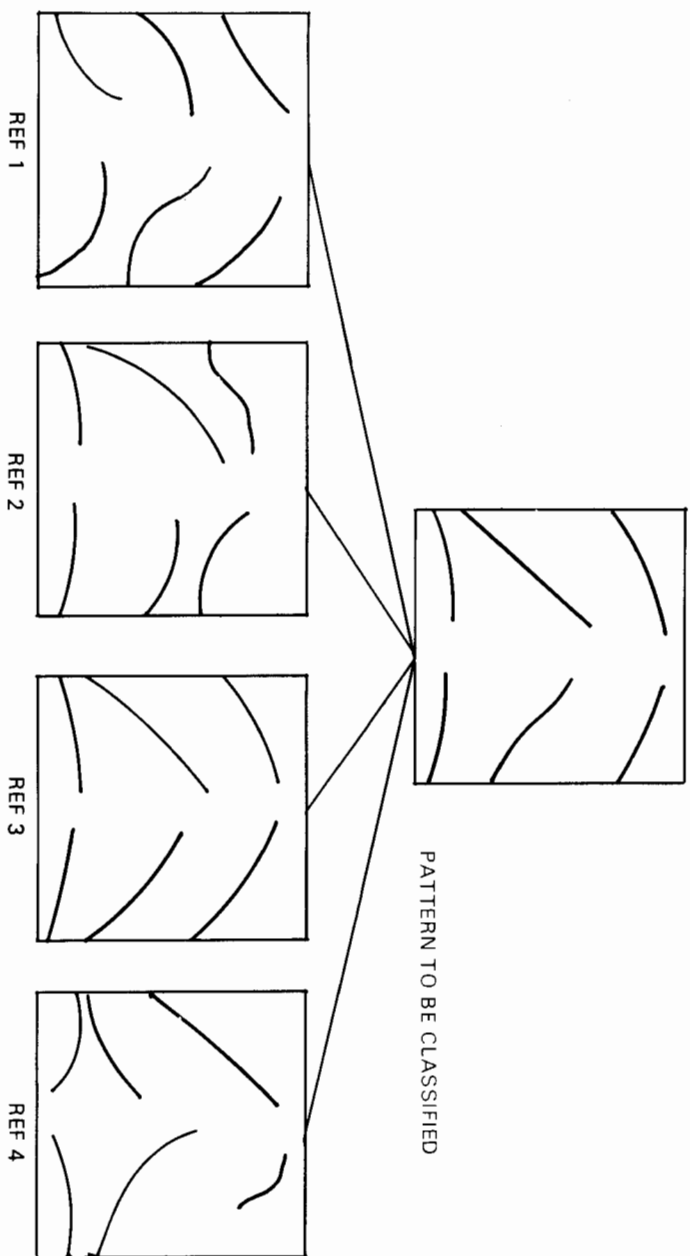


Figure 4b: Matching an incoming word against the stored patterns.

OPERATION OF THE TRAINING SUBSYSTEM

I now would like to present examples of the use of the mechanized subsystem.

1. AUDITORY TRAINING. Let us suppose we have stored a variety of sounds together with correlated visual information in the Visual and Auditory Storage area of Figure 3. Let us further suppose that these sounds are grouped into classes which are graded according to the difficulty of the auditory discrimination to extremely fine discrimination.

A program has been written for the computer which will control the presentation of stimulus material and the storage of user responses. This program would maintain a record of the student's progress through the training materials. It would start a new user on the simplest discrimination task, e.g., sound or no sound, and record his responses as right or wrong. Figure 5 suggests what the user might see

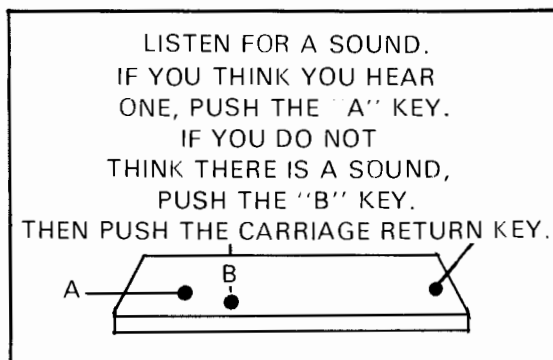


Figure 5: A gross discrimination task.

on the television screen. The system might operate in a test mode or a training mode. In the training mode feedback would be provided as suggested in Figure 6.

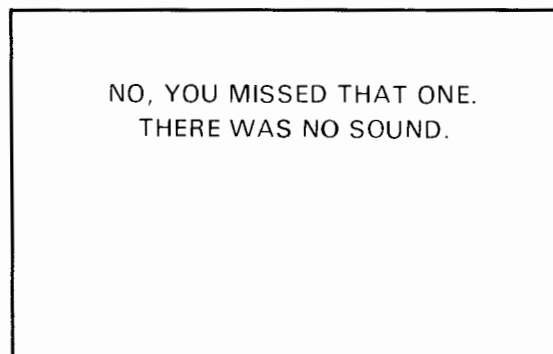


Figure 6: Feedback.

As an aside, it might be appropriate to test the user's ability to detect the presence or absence of all auditory stimulus material. Two uses could be made of this information: a) the computer could subsequently restrict presentations to stimulus materials which the particular user is known to be able to detect, and b) item difficulty data obtained from many individuals might provide a basis for hypothesizing characteristics of the decoding process. In the latter case the data would already be in a computer processable form. One might develop correlation matrixes where each variable would be a stimulus item (e.g., word) and where each pair of data points would be one individual measure of success on each of the paired words. Factor analysis of such a matrix might then reveal unexpected clustering of stimuli. Grouping subjects in different ways might yield different clusters.

Returning to the training process, a threshold, e.g., percent correct on k trials, would have been established so that the computer could determine whether or not to advance to the next level of discrimination difficulty. If the individual did not exceed the threshold over the last " k " trials he would be presented with randomly selected stimuli from the same class. He would thus be trained to discriminate within the class if he possessed the necessary ability. The feedback referred to in association with Figure 6 would give the individual information upon which to establish improved discrimination.

If, or when, an individual exceeded the threshold value, the computer would then move to the next class of stimuli. Figures 7, 8, and 9 suggest stimuli from increasingly more difficult classes. In each

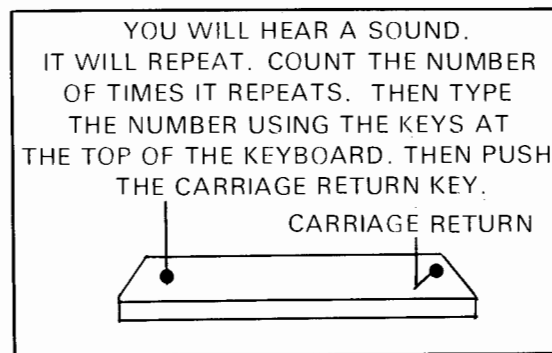


Figure 7: An awareness task.

instance in the training mode the individual would be told if he were right or wrong. If he were wrong he would then be told exactly what he heard and the stimulus or stimuli would be repeated to give him an opportunity to identify perhaps subtle visual and/or auditory clues either consciously or unconsciously. The user could request as many repetitions as he desired. In the case of Figure 9, for example,

the user, who was in error, could be told, "No, it was the second sentence. The first sentence was 'I'm training for the meet.' Here they are again."

YOU WILL HEAR TWO SOUNDS.
DECIDE WHICH SOUND IS A BELL.
THEN TYPE: 1. IF YOU THINK IT WAS
THE FIRST SOUND.
2. IF YOU THINK IT WAS
THE SECOND SOUND.
THEN TYPE A CARRIAGE RETURN. (THE
CARRIAGE RETURN TELLS COMPUTER
YOU ARE FINISHED. USE IT AFTER EACH
MESSAGE TO THE COMPUTER.)

Figure 8: A gross discrimination task.

YOU WILL SEE AND HEAR
A PERSON SAY TWO
DIFFERENT SENTENCES. YOU ARE TO
DETERMINE IF HE SAYS
'IT'S RAINING IN BOSTON.'
THE FIRST TIME OR THE SECOND TIME.

1 = FIRST;
2 = SECOND.




Figure 9: Reception of foreground verbal auditory signals.

For training purposes the ability of the computer to randomly select stimuli will make it possible for a relatively small store of individual stimuli to provide, through permutations, a comparatively large number of practice items.

I believe the above examples should be sufficient to suggest how the system might be used to provide training in auditory discrimination. Clearly one could continue increasing the difficulty of the stimulus material to such a point, for example, that the individual is asked to discriminate between consonant sounds: voiced vs breath, voiced vs voiced, breath vs breath.

2. SPEECH TRAINING. Now, let us consider the possible use of the subsystem for limited speech training. There will be two ways to use the system: a) the teacher and student working at the terminal together and b) the student working alone. In the first case the teacher would work with the student attempting to elicit a particular voice

sample. The student would be continuously videotaped during this period. As he or she listened to the student the teacher would evaluate each voice sample and decide if it were good enough for the student to use as a model. The videotaped sample would be temporarily stored. The teacher would decide whether to keep the sample or not. If the sample were to be retained it would be transferred to more permanent video storage; if not, it would be erased. The teacher would continue working with the student until he had collected five acceptable samples from him. Then a new utterance would be selected.

When the student elected to study on his own, the computer would retrieve the 5 required stored samples of each utterance and "train" the Voice Command System. The computer would then present the student with the printed word or consonant sound on the television screen with instruction to say it. If the student were correct he would be told so in printed form. In the event the student did not properly enunciate the stimulus word the system would only indicate the lack of a match (or conceivably show a match on the wrong word). Modification of the equipment would permit better feedback to the student. It would be possible to tap into the electronics of the Voice Command System to obtain the compressed pattern of the student's attempt at the word or sound, as well as the stored pattern of the given sound. This compressed spectrographic information could then be presented to the student to be used to guide his next attempt.

The limitations of the Voice Command System are such that the voice samples would have to be one second or less in duration. It would probably not be desirable to link words. If attempts were made to string words into sentences or phrases, a 250 millisecond pause would be required between words, thus introducing an unnatural rhythm. Care would have to be taken in the selection of the different word samples to ease the discrimination task for the device.

CONCLUDING REMARKS

I have attempted to conceive of a computer application which might have utility for auditory training and perhaps for limited speech training. In my opinion a prototype of the system I have described could be in operation by this time next year, at which time its usefulness could be determined. I am investigating ways to make this happen.

BIBLIOGRAPHY

1. Glenn, James W. and Hitchcock, M. H. "With a speech pattern classifier, computer listens to its master's voice." *Electronics*, May 10, 1971.