USER ACCEPTANCE IN AN AUTOMATED SPEECH TECHNOLOGY BASED TRAINING SYSTEM

MARY HICKLIN AND GAIL SLEMON
Logicon, Incorporated

INTRODUCTION

Computers are no longer deaf and dumb! The automated speech technologies (AST) have made it possible for computers to converse with their users. Now computers can actually understand the spoken word and can speak as easily as they can print. In theory, simulation of the verbal behavior of persons in the environment is now possible. In practice, the current state of the art imposes some rather severe limitations on such simulations. This paper describes some practical problems encountered in the development of a speech technology based training system for use in an environment where user acceptance is of paramount importance. The design solutions to these problems suggest that the state of the art has advanced to the point where speech technologies can be integrated successfully into today's operational simulators, albeit with care in demanding environments.

Background. Logicon's Advanced Systems Department has long been concerned with the automation of training. Since 1969, we have been studying the application of the automated speech technologies to the training environment. In 1974 we delivered a laboratory version of the Ground Controlled Approach Controller Training System (GCA-CTS) to the Naval Training Equipment Center (NTEC) in Orlando, Florida to demonstrate the feasibility of a speech recognition based training system.

Since that time, we have incorporated the speech technologies in other systems including the Automated Flight Training System designed for the Air Force and an Automated Command Response Verification System, a maritime safety system studied for the Department of Transportation. We have also continued to refine the original GCA-CTS in accordance with the results of experiments conducted by the Scientific Officer, Dr. Robert Breaux. We are presently building an experimental prototype version of the GCA-CTS for evaluation at the Naval Air Technical Training Center in Memphis, Tennessee. The remainder of this discussion will focus on this GCA-CTS work, which poses the most challenging user acceptance problems encountered to date.

The GCA-CTS. The GCA-CTS will employ commercially available hardware including two Data General Eclipse S/130 minicomputers, a Threshold Technology Threshold 500 speech preprocessor, a Federal Screw Works VS 6.4 Speech Synthesizer, and a Megatek MG552 Graphic Display Processor. The GCA-CTS will be designed to train students to interpret precision approach radar information and to give the well-defined verbal advisories which will enable the pilot to make a safe approach even in conditions of low visibility and without NAVAID receiving equipment in the aircraft.

The problem of training GCA controllers was originally chosen for study for several reasons. First of all, NTEC was interested in applying the concepts of automated adaptive training to air controller training. Until the advent of automated speech recognition however, machine-measurement of verbal task performance was not feasible. Automated speech recognition now makes it possible for the computer to determine the accuracy of what was spoken and grade student performance accordingly. With detailed performance assessment, it can automatically structure the training course to conform to the needs of the individual student. Secondly, present training for one student requires the attention of both an instructor and a pseudo pilot who controls a simulated radar target. This expensive resource allocation made the task a likely candidate for automation. Finally, since the verbal advisories are standardized and precise, the requirements seemed to be within the scope of the isolated word recognition art. Commercially available speech recognition systems are capable of recognizing a limited vocabulary of words or phrases, spoken in isolation, using reference patterns collected from the individual talker. A GCA controller training system therefore provided an ideal test bed for the concept of applying the AST to the training environment.

The system which is currently being designed will serve as an experimental task trainer for use by U.S. Navy enlisted men and women. It will provide relatively standard simulation of the radar environment and aircraft dynamics. No pseudo pilot will be needed because it will employ a pilot simulation whose ears will be the speech recognition capability and whose voice will be the speech synthesizer. Furthermore, the routine duties of the instructor will be simulated so that the GCA-CTS can provide instruction, present problems based upon the trainee's level of skill, supply performance assessments, and even administer and score a final examination. Simulation of these routine instructional duties will free the instructor to fill the more demanding role of training manager and will provide him with the data to do this more effectively.

USER ACCEPTANCE CONSIDERATIONS

The GCA-CTS is an ambitious project both in terms of the scope of the required simulation and because of its reliance on state of the art technologies. The successful laboratory demonstrations have focused upon the technical risk areas and have revealed potential risks in the area of user acceptance. If the system proves difficult to use, irritating to listen to, or fails consistently to recognize the spoken advisories, the trainee's frustration will probably impede learning.

The following paragraphs describe many potential AST related user acceptance problems and present the diverse solutions which will be employed in the GCA-CTS. These include:

- a. teaching the student to use the system properly
- b. design of a training course which complements AST requirements
- c. modification of state of the art speech recognition algorithms to accommodate the GCA vocabulary
- d. addition of speech understanding logic to augment speech recognition
- e. provision of effective feedback
- f. use of discretion in the application of speech generation.

The very diversity which makes the topics difficult to weave into a logical progression illustrates the point that a concern for user acceptance must pervade all aspects of system design.

Training GCA-CTS Users. Proper use of the training system will be the first topic addressed in the training program. The rules of microphone placement and speech level production are critical to good speech recognition, but are easy to learn and employ. There is also a less obvious component of proper system use which has been labeled "learning to talk to the system." Introspection, if it may be admitted, suggests that the components of this art include confidence, naturalness of speech and consistency. The skill is easily acquired, yet time for acclimatization is very important to good speech recognition. During this period the trainee will be encouraged to experiment with the system to learn that it really can recognize what he says, and to discover the limitations inherent in automatic speech recognition. With this background, he will be ready to use the system effectively.

A Training Course Which Complements AST. There are many possible approaches to teaching GCA controller skills. The present course requires that the trainee master the entire vocabulary and try to put it to use the first time he plays the role of final controller. At first the instructor stands by and prompts him, and the pseudo pilot (a trainee also) can easily understand the advisories despite stylization problems and even word substitution errors. During the 5 days of training, errors decrease dramatically and the trainee emerges as a qualified controller.

This training philosophy is not suitable for an automated training system however. The automated system has certain strengths and certain limitations when compared to the human instructor. The automated system cannot match the verbalization error tolerance of the human listener for example, but it can be much more attentive than an instructor who is responsible for several students. It has the flexibility to simplify the environment and even stop the approach if necessary to illustrate its points. It can also contrive practice approaches which require use of only that material learned to date.

To take advantage of the tremendous power of the automated training system and to minimize the impact of its shortcomings, a training course has been designed by our training technologists to maximize the training effectiveness of the GCA-CTS. This training will be in accordance with the principles of errorless learning so the trainee should never learn to make mistakes. Furthermore, this training strategy will actually take advantage of what is often considered to be a drawback in state of the art speaker dependent speech recognition, namely the requirement to configure the system for the individual's speech patterns. Briefly, the training strategy involves dividing the GCA controller task into its component parts so that one topic can be presented at a time. The simulated environment will be manipulated to provide illustrative examples while the instructor simulation provides prompts. The trainee's speech patterns will be collected at this time and used to create the reference patterns for speech recognition. Thus as the trainee learns the procedures and phraseology, the system will be learning to recognize his voice. After this phase, the system will present tasks for the trainee to perform without prompts. These will be scored. Their purpose will be to allow the student to practice the new material and integrate it with the old.

This training strategy is expected to impact user acceptance in several ways. First, use of the system should prove rewarding.

Initial speech recognition problems will be minimized by the small vocabulary and thorough training so the goals set forth should be readily attainable. This initial success is expected to bolster the trainee's confidence in the system, thus, providing an important ingredient for future recognition success.

Secondly, the formidable task of configuring the system to recognize the large GCA-CTS vocabulary has been integrated with task training in a way that will be transparent to the user. User acceptance will not be hindered by an enmasse onslaught of phrase repetition requests. In addition, the strategy employed in speech pattern collection should ensure representative reference patterns which further promote speech recognition accuracy.

Finally, the automated adaptive system will devote all of its resources to teaching the individual user. It will ensure that each topic is mastered in a systematic way at the trainee's own pace. Any needed remediation will be provided automatically. Because of this highly individualized instruction, the trainee will be expected to attain proficiency in all of the GCA control tasks. The proficiency level attainable through thorough, systematic, task-oriented learning can be its own reward and should enhance user acceptance.

Augmentation of Speech Recognition Algorithms. In addition to these efforts to ensure good speech recognition by training the user carefully, the traditional speech recognition algorithms must be enhanced to accommodate the challenging vocabulary. The long vocabulary list of approximately 100 phrases contains many similar items and consists of both very long and very short utterances. Obviously this vocabulary cannot be modified in the interests of good speech recognition because it is defined by the FAA and used by all GCA controllers; instead, speech recognition will have to be modified.

In order to recognize the relatively large vocabulary in real-time and with high accuracy, the reference patterns will be partitioned by the phase of the approach, the previous transmission, the length of the phrase, and its intended destination. Using all these clues, the branching factor can be reduced to about 50 items in the worst case. This is similar to the branching factor tested successfully in the laboratory GCA-CTS.

To distinguish vocabulary items which vary only slightly, a scheme devised by Dr. Breaux and tested in the laboratory GCA-CTS will be employed. The technique involves comparing similar reference patterns to find those time slots which are significantly different. These rows are then correlated with the corresponding rows of the input feature pattern. The procedure effectively causes the pattern recognition algorithm to weight the distinctive portion of the utterance more heavily than the similar portions.

The GCA-CTS vocabulary, with its combination of long and short utterances, is a problem for the speech recognition algorithm, which depends on a time normalization scheme to extract relevant information from input utterances. The input is compressed into a pattern of standard length in which only those features which are reliably present are set. This compressed array is compared to similarly constructed reference patterns. The optimum length of these compressed arrays is related to the length of the input utterance. This becomes problematical when phrases as diverse as "eight" and "going further above glidepath" must be recognized. This will be solved by employing two sizes of reference patterns based on the length of the utterance. This may even prove helpful for vocabulary items of medium duration. On an experimental basis, both reference patterns will be used for these items and the results will impact the speech recognition module's measure of confidence in its decision.

The Requirement for Speech Understanding. To ensure user acceptance, the system must not make the mistakes which strict reliance upon speech recognition in this demanding environment could produce. There are sources of information in the simulation environment as well as a priori information which can be used to ensure the system understands what was spoken.

First of all, the speech understanding module will atempt to distinguish between controller errors and recognition errors, and will compensate for the latter. An example will help clarify this point. To the machine ear, the advisories "well above glidepath" and "well below glidepath" sound very much alike. Confusing these two is a gross error which the trainee is not likely to make. If such an error occurs, the system will assume the trainee has given the correct advisory. Trainee errors, on the other hand, are more likely to involve a confusion between contiguous aircraft positions such as between "above glidepath" and "slightly above glidepath." These phrases are readily distinguishable to the speech recognition algorithm, so the system will be able to score the incorrect advisory.

The speech understanding module will also compensate for typical trainee stylization errors whenever possible. GCA-CTS will employ isolated phrase recognition and therefore there must be slight pauses between the digits given in turn commands. Because of their previous training, students are expected to forget this important requirement. Audio tapes reveal that a frequent error is a failure to pause before the first digit of a heading. Trainees will say for example, "turn right heading one (pause) five (pause) five." This utterance will probably trigger a low confidence recognition of "turn right heading" followed by correct recognition of two digits. Since the first digit is always the same, speech understanding will assume this error occurred if two reasonable digits follow a heading command and will compensate for it. In parallel development work, we are investigating an extremely promising limited continuous speech recognition algorithm called LISTEN (Logicon's Initial System for the Timely Extraction of Numbers) which would obviate the need for stylization of heading commands. The exciting results of the first phase of this work were presented at the December, 1977 Conference on Voice Technology for Interactive Real-Time Command/Control Systems Application at NASA, Ames Laboratory, *Sunnyvale, California. The technique is being* tested now in a laboratory training system, and we expect it will soon be ready for incorporation in the GCA-CTS and other operational trainers.

Finally, speech understanding will monitor the speech level to detect fluctuations significant enough to affect recognition accuracy. It will inform the trainee through the simulated pilot when transmissions become weak.

Effective Feedback. Despite the precautions taken to ensure errorless learning, consistently error-free performance in the complex GCA *controller environment will probably* remain a lofty ideal which can only be imperfectly realized. Therefore, effective feedback will be provided whereby the user can understand and learn from his mistakes. In this unique training environment, these mistakes will include stylization errors which cause recognition failures. Dr. Breaux added a replay capability to the laboratory GCA-CTS and it was found to be an effective feedback technique for procedural errors. In that system the speech synthesizer repeats the trainee's advisories and gives rule explanations when an error is encountered. He noted, however, that "message not understood" reports proved especially frustrating to the students. Many times a student was convinced that he had uttered the correct advisory but he had no way to argue with the computer or to understand why the recognition failure occurred. For example, he would remember that he had said "slightly above glidepath" but would not realize he had paused in mid-phrase, making the advisory unintelligible to speech recognition.

To add to replay's usefulness as a feedback technique and to reduce the frustrations associated with recognition failures, a speech input digitizer will be designed to record the trainee's advisories. The replay will then consist of an actual recording of the trainee's speech, synchronized with the visual display. From this, the student should be able to understand the cause of any recognition failures which occur. In addition to enhancing the student's acceptance of training system decisions, the replay will provide the instructor with an excellent tool. After the run, the student and instructor can review the run together. This will provide an excellent forum for discussion of such things as subtle points of style.

Use of Synthesized Speech. Synthesized speech is the ideal channel for some system outputs such as pilot responses and instructor comments. There are other situations in which it may be necessary to use it with discretion. For example, many persons adapt to its slightly artificial sound easily, but some express annoyance at having to listen to it for long periods of time and, therefore, its usefulness for presentation of instructional materials is limited by user acceptance considerations. The GCA-CTS design supports multimedia presentations of such materials, and synthesized speech, employed sparingly, is expected to enhance these presentations.

Use of the speech synthesizer could influence user acceptance in an indirect way as well. Verbal prompting is employed as an effective teaching technique in present controller training. The laboratory system also provides prompts using the speech synthesizer, but there is some debate as to whether or not synthesizer prompting is effective for good reference pattern collection. Some observe an unwilling mimicry of the speech synthesizer's inflection and timing. Others do not notice this effect. One thing is certain, reference patterns created from unnatural verbalizations do not provide the best possible recognition accuracy. The efficacy of the use of synthesized prompts in the collection of reference patterns will be carefully considered during GCA-CTS development.

SUMMARY

There are many challenging problems facing the system designer in the application of the automated speech technologies to simulation. We have attempted to show that effective strategies can be designed to employ these technologies in a training environment. These strategies include techniques for augmenting shortcomings both in the state of the art and in user performance to produce a system which can respond intelligently, be acceptable to the user, and can provide excellent training. These concepts will soon be ready for testing, and we look forward to a successful field evaluation.

ABOUT THE AUTHORS

MS. MARY HICKLIN is Project Leader for the Ground Controlled Approach Controller Training System (GCA-CTS), experimental prototype, at Logicon, Inc. She has worked with all versions of the GCA-CTS for the last four years and was responsible for software design and implementation of the Automated Command Response Verification System, and for much of the implementation of LISTEN, Logicon's Initial System for the Timely Extraction of Numbers, a real-time continuous speech recognition system developed for the Naval Training Equipment Center. Her experience with performance measurement for automated adaptive trainers has included the development of the GCA-CTS performance measurement routines as well as implementation of the real-time performance measurement routines used in the air-to-air intercept and ground attack radar modes of the Automated Adaptive Flight Training System (AFTS). She holds a degree in biology from San Diego State University where her graduate work in systems ecology included the development of computer simulations of ecosystems.

MS. GAIL SLEMON, a member of the Technical Staff of Logicon, Inc., is currently working on design of the Ground Controlled Approach Controller Training System (GCA-CTS) experimental prototype. Past experience includes software system design and implementation of several different prototypes for the Micrographics Automated Intelligent Terminal (MAIT) which incorporates microprocessor controlled microfiche readers with hard copy and audio outputs along with keyboard and touch panel inputs for maintenance/training and other data base sequence defined applications. Prior to this, she was involved with System Engineering Level Evaluation and Correction (SELECT). Ms. Slemon received a B.S. degree in electrical engineering from Cornell University concentrating in computer science.