DEVELOPMENTS OF MACHINE SPEECH UNDERSTANDING FOR AUTOMATED INSTRUCTIONAL SYSTEMS

ROBERT BREAUX and IRA GOLDSTEIN
Naval Training Equipment Center

INTRODUCTION

This paper describes further progress in the development of automated adaptive instructional systems. The technology of machine speech understanding is the most recent addition to our repertoire of training technologies, and it gives us automated training capability in areas thus far unamenable to advanced training techniques. The concept and preliminary functional design of one such system were presented in the Proceedings of the Seventh NAVTRAEQUIPCEN/Industry Conference (Goldstein, Norman, et al., 1974). For convenience, this paper will review the earlier work and then go on to provide implementation details and research results.

Automated Adaptive Instruction

Automated adaptive training has a number of advantages over the more traditional approaches to training. Automation of training relieves the instructor of busywork chores such as equipment set-up and bookkeeping. He is thus free to use his time counseling students in his role as training manager. In adding the adaptive component, efficiency is increased with more training per unit time. Individualized instruction, with its self-paced nature, maintains the motivation of the student. Objective scoring is potentially more consistent than subjective ratings. Uniformity can be maintained in the proficiency level of the end product, the student. But, tasks requiring verbal commands have thus far been unamenable to automated adaptive training techniques. Traditionally, performance measurement of verbal commands has required subjective ratings. This has effectively eliminated the potential development of individualized, automated, self-paced curricula for the training of the Landing Signal Officer, the Air Intercept Controller, the Ground Controlled Approach Controller, and others. Computer speech recognition of human voice offers an alternative to subjective performance measurement by providing a basis of objectively evaluating verbal commands. The current state-of-the-art has allowed such applications as baggage handling at Chicago O'Hare. A more sophisticated recognition system is required for training, however. To that end, the Naval Air Systems Command and the Advanced Research Projects Agency have supported the Naval Training Equipment Center Human Factors Laboratory in efforts to establish design guidelines for training systems which combine automated adaptive training technologies with computer speech recognition technology. The particular application chosen is the precision approach phase of the Ground Controlled Radar Approach (GCA).

TRAINING REQUIREMENTS

The GCA Application

The task of the GCA Controller is to issue advisories to aircraft on the basis of information from a radar scope containing both azimuth (course) and elevation (glidepath) capabilities. The aircraft target projected on the elevation portion of the scope is mentally divided into sections by the Controller. This is because the radio terminology (R/T) for glidepath is defined in terms of these sections. Thus, at any one point in time, one and only one advisory is correct. Conversely, each advisory means one thing and only one thing. This tightly defined R/T is perfect for application of objective performance measurement. The drawback, of course, is that performance is verbal and has thus far required subjective ratings. In addition, the time required for human judgment results in inefficient performance measurement. The instructor simply cannot catch all the mistakes.

Needs and Objectives

The major behavioral objective of current GCA training is to develop the skill to observe the trend of a target and correctly anticipate the corrections needed to provide a safe approach. The standard R/T is designed to provide a medium to carry out this objective, and GCA training exposes the student to as many approaches as possible so that the trainee may develop a high level of fluency with his R/T. Safety is paramount, and is stressed heavily. The most difficult portion of the approach, course corrections, is seen by

trainees and instructors alike as being second in importance to safety for a good approach. This is because glidepath R/T is only a post hoc advisory service, while the controller is in positive control of heading. At his discretion, for example, a controller may take away a pilot's gyros; he may initiate a change to a no-gyro approach if the pilot appears to be having difficulty executing course changes.

The primary need, then, to fulfill its objective is for GCA training to teach the skill of extrapolation. A controller must recognize as quickly as possible what the pilot's skill is. He must recognize what the wind is doing to the aircraft heading. Then he must integrate this with the type aircraft to determine what advisories to issue.

Current Technology

Two approaches in training these primarily conceptual skills (the significance of which is discussed in the Training Modes Section) have been the use of live aircraft and the use of simulators. Live aircraft are expensive in terms of fuel. Simulators are expensive in terms of manpower. The most popular radar simulator uses an operational radar scope with an artificial target generator to produce the simulated aircraft. The naive student is the most expensive to train in this situation. He requires an instructor to prompt him on his R/T and another person to operate the target generator. Thus, there is an expensive 2:1 ratio of support personnel to student.

The type of "pilot" in the simulated approach remains relatively constant. There is little that can be done with existing devices in the way of exposing the trainee to pilots of varying skill level. Indeed, what often happens is that trainees spend half their time as "pilots" and half their time as controllers. Naturally there is a tendency to be as good a pilot as possible, thus ensuring himself of the same when he is the controller. What happens, then, is that as controller skill improves so does "pilot" skill. What would be desired instead is a broader range of pilot types presented in a systematic way.

Existing technology is just beginning to employ capability to vary systematically simulated wind conditions. Device 15G14 was the first to employ this technique. Obviously, there are few approaches made in calm wind. Thus, the addition of wind adds fidelity to training. More importantly, it develops the primary skill of course correction computation. However, little guidance is available as to how wind should be varied. What is desirable is that wind components be capable of varying systematically with controller skill. But, the conceptual relationship between wind speed, wind direction, aircraft type, and controller skill is not currently a part of training for the air controlman.

Advanced Technology

The major behavioral objectives, then, can more efficiently be achieved through the application of computer voice recognition technology, and thereby the application of advanced training technologies. This is because, with objective assessment of what the controller is saying, objective performance measurement is possible, and thus we have the capability of individualized instruction. The use of simulated environmental conditions allows the development of a syllabus of graduated conceptual complexity. The integration of these components results in an automated, self-paced, individualized, adaptive training system. The job of the instructor now becomes one of training manager. His experience and skill may be exploited to its fullest. The training system can provide support in introducing the student to the R/T. The instructor can scan the progress of each student and provide counseling to those who need it. Simple error feedback is provided by the training system. Only the instructor can provide human-to-human counseling for specific needs, and the training system provides more time for this valuable counseling.

TRAINING SYSTEM OVERVIEW

A training system for the GCA controller was determined to require four subsystems, speech understanding, pilot-aircraft model, performance measurement, and a syllabus. The speech understanding subsystem was developed around the VIP-100, purchased by the Naval Training Equipment Center from Threshold Technology, Inc., Cinnaminson, New Jersey. The incoming speech signal is sampled every two milliseconds by special circuitry which determines the presence or absence of each of 32 features designed to characterize acoustic energy. The successive binary samples must be stored in a buffer until the end of the speech is recognized, usually by the feature called Long Pause. Each utterance must be no longer than two seconds. These data are time normalized with software to reduce the buffer to a standard size and group the primary data.

It must be pointed out that this is not a science fiction or "Star Trek" type recognition system. One cannot simply select a person from the population at random and expect recognition to be at an acceptably high level without employing special procedures.

Three major constraints are imposed by this system. Each user must pre-train the phrases. Recognition does not take place for random, individual words, only pre-defined phrases. Each phrase is repeated a number of times and a Reference Array is formed representing the "average" way this speaker voices this particular phrase. Thus, the second constraint is that there must be a small number of phrases (about 100) which are to be recognized. If performance is to be evaluated based upon proper R/T, each phrase must be defined. The existence of an R/T implies a finite number of phrases. The third constraint, due to performance measurement requirements, is that there be no ambiguous phrases -- right or wrong depending strictly on who the instructor is. Technically, the GCA application appears to be conformable to these constraints. The result of pre-training with its implication for consistent, unemotional vocal delivery, remains to be evaluated.

To achieve high fidelity, simulation makes use of various math models: The model of the controller is at the focal point of all other models, and serves to provide criteria to the performance measurement system. A model of the aircraft and pilot allows for variation in the complexity of situations presented the student. The principle being used here is that the exposure of a student to certain typical situations will allow him to generalize this experience to real-world situations. Of course, these situations must be presented systematically if efficient learning is to be achieved. For example, two concepts which must be learned by the GCA controller are recognizing positions on glidepath and computing course corrections. Since the latter is more difficult, situations requiring it must be introduced only after the former is well established. The pilot model allows for systematic presentation of various skill levels of pilots. In addition, the equations used in modeling the pilot and aircraft responses also allow for introduction of various wind components. The adaptive variables, pilot skill, aircraft characteristics, and wind components, are combined systematically to produce a syllabus graduated in problem complexity. As the skill of the student increases, he is allowed to attempt more complex problems. Which specific problem is to be presented is a fact determined by his history of performance, not simply his current score.

Since the score is determined by the performance measurement system, the heart of scoring is the model controller. As it often happens, what constitutes "the" model controller is a matter of some discussion among GCA instructors. Thus for automated training applications, one must determine the concepts which are definable, such as how to compute a turn, and leave other concepts to be developed by the instructor-student apprentice relationship. Such things as when to issue "approaching glidepath" is defined broadly and is therefore a judgment call in large part. Nevertheless, a great deal of effort has gone into development of a model controller. With field evaluations, it is anticipated that further refinements will be possible. The scoring system reflects on a one-to-one basis what the model controller is evaluating. Thus, implementing improvements will be relatively straightforward.

RESULTS: RECOGNITION RELIABILITY

Confusions

A major problem was discovered in the speech understanding software (SUS) in that phrases containing "above" and "below" were often confused (e.g., the phrase "above glidepath" was recognized equally often as "below glidepath" and as "above glidepath"). A number of phrases exist in the R/T containing "above" and "below," so this was not a trivial problem.

Say Again Rule for the Human

The initial thought was to develop some "say again" rule for the human. For example, when humans are in conversation where a phrase is not understood, the rule used is to speak each word more slowly or more distinctly when repeating the phrase. This rule of speech allows discrimination of the specific words within the phrase. Thus, it was assumed that some rule of clarification was needed for human-machine conversation to allow the machine to discriminate words in a repeated phrase. The recognition algorithm accumulates a score by comparing bit settings corresponding to acoustic energy features in the input (spoken) phrase with the Reference (pre-trained) phrase. Visual inspection of the scores for the features of the phrase "slightly above glidepath" compared to "slight below glidepath" indicated similarity of features for both the beginning and end of the two phrases

(scores of 32), but the middle of the phrases were dissimilar (scores of 16).

For clarification of what the scores mean, it must be noted that the software *standardizes all input into 32 units. Each* time unit may have any number of 32 feature bits set depending upon the particular phrase. That is, regardless of the time length of the input phrase, the features are fitted or normalized into a 32 by 32 array (features by time units) for storage purposes. The effect of this procedure is that features of words within a long phrase are compressed relative to those of words within a short phrase. A reference pattern of the relatively expanded features for the single word "above" is quite different from that portion in the reference for "well above glidepath" that contains the relatively compressed features of "above." Logicon, Inc., of San Diego, California, has added an "attribute" code to the phrases which saves the actual time length prior to normalization. Thus when input occurs, all references within that time range or all references which took about that long to say during pre-training are compared to the input by computing an index of similarity (I). The algorithm adds or subtracts points depending upon a same-different comparison of features. The reference with the highest I is selected as representing the input phrase if its I is sufficiently higher than the next highest I (i.e., for recognition, there must be $I_{max} - I_{max-1} \geq C$, where I_{max} is the largest index of similarity, I_{max-1} is the next largest, and C is some minimum difference criterion).

It was concluded that the reason for poor discrimination between similar phrases was that the uniqueness between the two phrases, which resulted from the features for "above" and "below" is very slight. The features are compressed into so few time units that $I_{max} - I_{max-1} \leq C$. For example, the compressed features of "above" in the phrase "well above glidepath" that has been normalized into 32 time units occurred in about six time units. Since the phrases "well above glidepath" and "well below glidepath" are unique only in the words "above" and "below" or in about six time units, the difference between their indices is based on about 1/5 of the phrase. The difference between "well above glidepath" and "well below glidepath" is understandably often smaller than criterion.

Say Again Rule for the Machine

It was determined that a rule for discriminating similar phrases should be developed for the machine rather than for the human. The algorithm was proposed to be a two-step process. A separate index of similarity was computed between the two reference features which had produced I_{max} and I_{max-1}. *In this way, time units in which the unique* words occurred could be discerned -- the index should be low (16) at those points. Then, a new similarity index was computed, but only using that word (those time units) in the phrase which was different. This technique was observed as successful (100 percent recognition accuracy) for two humans in one sampling of the following phrase pairs: well above glidepath vs well below glidepath, above glidepath vs below glidepath, slightly *above glidepath vs slightly below glidepath,* turn left heading vs turn right heading, well left of course vs well right of course, and for slightly left of course vs slightly right of course.

It was concluded that this two-step algorithm would improve recognition accuracy and be generally applicable across people and vocabulary. The important point is that the hardware has been ignored. Emphasis has *been on the software stage of recognition.*

The Problem of Novelty

In an attempt to verify the recognition algorithms, naive adult males were employed as subjects. It was soon discovered that probability of correct recognition was as low as 50 percent in the beginning and that phrases had to be retrained to increase recognition reliability. It was hypothesized that the novelty of "talking to a machine" was a significant factor in the low recognition reliability. If this initial novelty could be reduced, it was thought reliability would also increase. Four adult males and four adult females were used to compare an introduction method vs a no-introduction method. The introduction group was given R/T practice, saying the GCA phrases as they later would in an actual prompted run. The model controller was utilized to anticipate for the subject an optimum response every four seconds. This prompt was presented graphically on the display, as the aircraft made the approach. The subject spoke the phrase, then both the prompt and the understood phrase were saved for later printout. The no-introduction group was not given practice. Each group then made Reference phrases. Reliability data was collected using the procedure described above for R/T practice. A Chi-square value was computed from a 2 x 2 contingency table of frequency runs in which no recognition errors occurred vs frequency in which one or more errors occurred, and whether there had been practice on the phrases vs no

practice prior to making the voice tape. It was found that X^2 (1) = 3.12, p $<$.10, indicating a relationship. A correlation was computed for the groups vs the number of different phrases which were not recognized on a run, with R=-.33, p $<$ .10, indicating a tendency for fewer errors with practice at the task. It is proposed that the following procedure will aid in reducing errors by the SUS: Ten runs to introduce the student to the vocabulary, create a reference pattern for each phrase by uttering repetitions of each phrase and each digit, five more runs checking the recognition accuracy, then a re-make of those phrases for which recognition accuracy is low. Conclusion: Better recognition is achieved when the R/T is voiced consistently and unemotionally.

RESULTS: TRAINING MODES

Basic Modes

The laboratory model system developed for research purposes provides three training modes. The first mode fulfills the prerequisite of voice recognition that the computer have available a reference pattern of each phrase consistent with the way the student will say each phrase. So, after the student has accustomed himself to the GCA task and has developed a consistent way of saying each phrase, reference arrays are created.

Since the task will eventually require the trainee to recall the phrases from memory, the reference arrays are created the same way. The phrase is presented for memorization on a CRT for 5 seconds, then erased. Next, the prompt "SAY IT" appears, expecting the trainee to voice the phrase from memory as he would in a scoring run. In addition, for most phrases a target appears on the CRT which corresponds to the phrase. Thus, the trainee gets some exposure to the correspondence between the GCA vocabulary and target position. This mode allows research on the question of how many repetitions of each phrase are necessary to obtain a reliable reference pattern. As discussed previously, two repetitions appear to be sufficient for the GCA application. However, future application such as the Air Intercept Controller and others could require a different number of repetitions.

A second mode is required to introduce the naive trainee to the complex GCA R/T priority scheme. This mode relieves the instructor of hours of sitting behind the trainee, telling him everything to say. Even the more studious trainees need initial practice getting accustomed to the precision of GCA R/T. For example, something must be said at least every five seconds. Until the trainee can generate phrases himself quickly enough to maintain the rapid pace of the approach, he needs constant prompting.

An added feature of the prompting mode is the collection of reliability data. Since it can be assumed quite readily that the trainee will, in fact, say what he is told, the prompt, a comparison is possible between what is understood and what is prompted. If a particular phrase is consistently being mis-recognized, retraining is possible for that phrase. The retraining procedure must entail an increase in the number of repetitions over the initial number. Both the spoken phrase and the erroneously understood phrase must be retrained. Obviously, it is important that recognition errors not be confounded with trainee judgement errors. Thus, this reliability check serves as important a function as the prompting of the phrases.

It may be possible to adjust the software when increase in the number of repetitions does not improve reliability. That is, various parameter values such as minimum similarity criterion and difference criteria may be adjusted readily to increase the sensitivity of recognition. Research is needed to determine the optimum set of parameter values for a given vocabulary. However, it is also possible that speaker characteristics rather than vocabulary characteristics will be useful in determining the values of recognition parameters.

The third mode, the scoring mode, is, of course, what training is all about. Since scoring is automated, the instructor can devote more time to students who need it. Once the trainee can consistently anticipate the prompts in mode two and is confident in his ability, he may test his skill against an objective performance measurement system. The self-paced nature of adaptive training adds to the motivational incentive of objective scoring to maintain a high level of interest in the serious student. Better students can progress on their own, weaker students can seek help. Feedback is provided the instructor that is useful in counseling the trainee on his weak points as well as reinforcing his strong points.

The key to teaching concepts effectively is in the order of presenting the stimuli, or order of the problems. The type problem presented is determined by a syllabus. The syllabus is intended to introduce each concept of the task systematically, fading in more slowly those

concepts which are more difficult or complex. For example, glidepath advisories have a one-to-one correspondence with target position. Thus, the rules for glidepath R/T are simple concepts (associations), and can be best learned through memorization. Course changes, on the other hand, must be computed from trend of the target. The concept is more complex, a conjunctive rule based on wind and course deviation trends. So the initial problems of the syllabus concentrate on glidepath R/T by simulating an approach of a slow aircraft in a no-wind condition. Successful completion of these problems develops basic R/T skills and allows the trainee to advance to problems introducing complex conceptual rules such as course changes, then problems of variable wind speed, wind direction, requiring even closer concentration on the course corrections. The task of a GCA controller is to anticipate the necessary corrections to maintain a safe approach. The intent of the syllabus is to teach these skills of anticipation in a systematic way, thus increasing training efficiency.

Adaptive Training

Traditionally, adaptive training varies problem difficulty without regard to informing the subject of the rationale for the change. In fact, difficulty is varied within a single "trial" on some types of aircraft stick manipulation tasks. This traditional mode of adaptive training applied to, essentially, motor tasks is possible for the conceptual skills in controller training as well. At the end of an approach, a score is calculated to determine the next problem type. Thus, the trainee could conceivably work problem after problem with all feedback grouped at the end or at the beginning of a run.

A second type of adaptive training, however, can utilize feedback as errors occur, thus cuing the student on his weaknesses at the point in time when they show up. In this way, the serious student may determine for himself that certain basic concepts are misunderstood and it is time to stop and get help, or to attribute errors to lack of practice, say, and to proceed on. More important, however, is that poor habits are eliminated quickly. The student cannot advance in the syllabus at minimum passing score because he is continually making some minor mistake, without being aware of that mistake. Instead, the errors are pointed out as they occur and are thereby extinguished early in training. The third training mode employs as an option both of these types of adaptive training.

OTHER APPLICATIONS FOR CONTROLLER TRAINING

General Research

The job of a controller, in general, is to issue verbal commands based upon some combination of visual and auditory cues. In addition, his repertoire is a tightly defined set of words and phrases. This description applies to the GCA Controller as well as the Air Intercept Controller, the Landing Signal Officer, and the Officer of the Deck. Once an objective behavioral assessment system is developed for these jobs, it is possible to implement other advanced training technologies. Computer Speech Recognition is one possibility for providing this objective assessment.

The GCA Controller Training System developed for laboratory use has employed a modular design. It is anticipated that application to other controller jobs will need only slight modifications. In particular, control models specific to the job type will have to be programmed. They, of course, will determine the most appropriate control message at each point in time. The existing performance measurement subsystem can be used to accumulate a count of message discrepancies between trainee and model. Since the speech understanding subsystem is application-independent, no changes will be needed there. Thus, studies on the recognizability of R/T from other controller type jobs can be made with the current research system.

Voice Response

Controllers must also respond to requests. The GCA application described has not addressed this problem. However, a computer speech synthesizer has recently been acquired. This capability opens the way for increased fidelity of controller training. Typical pilot requests, for example, could be simulated as part of the training syllabus. To provide more complex problems for the trainee to handle, voice synthesizers under the systematic control of a syllabus could present unusual requests or initiate simulated emergencies.

SUMMARY

A complex research system has been described which is to be used for R&D for controller training. Work has been described which sought to enhance the state of the art in computer voice recognition capability. The problems of real-time voice recognition were discussed, and some ideas for their solution advanced. A description of a

modularized laboratory version of a GCA controller training system was given. A departure from the traditional form of adaptive training was taken. Emphasis was given to informative feedback during the course of an approach, as well as summary information at the end, to facilitate teaching the concepts. Student and instructor need feedback, but of a different sort and at different times. Capabilities were described for research on the vocabularies of other controller-type tasks. The modularity of this system requires replacement of only the simulated task display, controller model, and performance measurement system for conversion to the appropriate training system. Procurement of a voice synthesizer was noted and advantages to training of such a capability were indicated.

REFERENCE

Goldstein, I., Norman, D.A., et al. "Ears For Automated Instruction Systems: Why Try?" in Proceedings of the Seventh NAVTRAEQUIPCEN/ Industry Conference, NAVTRAEQUIPCEN IH-240, Orlando, Florida, November 19-21, 1974.

ACKNOWLEDGEMENT

The design and implementation of the vast bulk of the software for the Speech Understanding Subsystem and the Performance Measurement Subsystem were done by M. W. Grady, M. J. Barkovic and R. M. Barnhart of Logicon, Inc., San Diego, California. J. P. Charles and L. D. Egan were, successively, Project Manager for Logicon under NAVTRAEQUIPCEN Contract N61339-74-C-0048.

ABOUT THE AUTHORS

DR. ROBERT BREAUX is a Research Psychologist in the Human Factors Laboratory at the Naval Training Equipment Center. He has interest in application of theoretical advances from the psychological laboratory to the classroom situation. Papers include computer application for statistics, basic learning research, concept learning math models, and learning strategies. He is an instrument rated commercial pilot.

MR. IRA GOLDSTEIN is a Research Psychologist in the Human Factors Laboratory at the Naval Training Equipment Center. His professional activity over the past dozen years has focused on the role of computers in behavioral science research and their employment to improve training. He has contributed a number of technical papers on computer-controlled measurement of human performance in perceptual-motor and decision-making tasks. Before joining the Center in 1969, Mr. Goldstein was employed in private industry for 4 years. Previously, he had been with the United States Air Force's Decision Sciences Laboratory from 1958 to 1965.