GUIDES FOR

VOICE TECHNOLOGY IN NAVY TRAINING SYSTEMS

Dr. Robert Breaux
Naval Training Equipment Center, Orlando, FL

Dr. Michael E. McCauley
Canyon Research Group, Westlake Village, CA

Dr. Paul E. Van Hemel
Ergonomics Associates, Inc., Orlando, FL

ABSTRACT

Computer speech recognition provides the enabling technology for the use of automated performance measurement and instructor support features to allow "instructorless" training for those tasks which are primarily speech in nature. The design guides take the approach that automated, "instructorless" training is achievable through the use of computer software models of the instructor and the task. Human factors design guidelines are provided for the integration of speech technology with the software models. The Navy has built two prototype training systems using voice technology to capture student behavior, evaluated one of them, and is in the process of evaluating another. Further, the training implications for airborne applications of voice technology have been developed. The emergence of voice technology as one solution to the manpower shortage has provided justification for these efforts. This paper describes how voice technology can make the transition from R&D to application.

INTRODUCTION

Background

For several years, the Human Factors Laboratory at the Naval Training Equipment Center has been developing techniques for the application of voice technology to training systems. There has emerged from that work an identification of three areas of expertise which will be required in future efforts to apply voice technology to Navy training systems.

The first area is Systems Engineering. Obviously, knowledge is required in hardware capability and in programming requirements of voice technology subsystems. The second is Human Factors. Obviously, too, knowledge is required in man-machine interface design requirements. The third area is relatively new, the Voice Technology Specialist. Knowledge is required in the integration of Systems concepts with Human Factors design concepts. What makes the Voice Technology Specialist separate is the requirement to recognize the unique characteristics of Voice Technology. Talking to a machine with today's technology is not like talking to another person. Special skills are required to properly integrate a voice subsystem, and this paper will address what those skills are.

Each of the three areas of expertise has a critical role to play in the design and implementation of the training system using voice technology. This paper seeks to define those roles in terms of the tasks which must be performed. It is proposed that the first two roles will merge to form the third as more and more voice subsystems are applied to training systems. The method of this merger is described in the conclusion of this paper.

Growing Popularity

Science fiction has identified numerous applications for voice technology, but the technology will have to be able to converse as smoothly as the conversation between two people for those applications to be feasible. Industry has shown cost payoffs with today's technology with applications in the data entry and the secure access areas. So, we can expect to see a slow but growing development from current technology capabilities to "natural" communication with machines. The length of time that it takes for the growth will depend upon the speed at which voice technology is popularized by reduced cost and upon the speed at which need is created for more sophisticated applications of man-machine dialogue. The Navy is shaping this development by its applications of voice technology.

This paper will briefly describe two Navy training prototypes. They have taught us much about the unique characteristics of voice subsystems and have allowed development of the voice technology design principles discussed here.

The two prototypes are for controller training. The Navy has built (1) and evaluated (2) a system for training a portion of the task of the Air Traffic Controller, and has built (report in preparation) and is evaluating (work in progress) a second system for a portion of the task of the tactical controller, the Air Intercept Controller. In each case, voice technology is used within a package of instructional features. Voice technology serves to capture the vocal behavior of the trainee *so that other automated subsystems can* evaluate the performance of the trainee. Other subsystems then select the next task so as to teach the trainee the required next logical skill. This package of instructional features is both the reason why voice technology can be made to work, as well as why the application principles are so complicated. That is, a non-trivial dialogue between a person and a machine requires an intelligent person and an intelligent machine.

Future Payoffs

The Navy's cost payoff from voice technology is currently in the area of Manpower, Personnel and Training. Voice technology as part of an instructional features package can serve to substitute for personnel actions and in some cases for personnel themselves. Thus, fewer people can perform more tasks in less time because technology is performing the functions once done by people. In the 1990's, voice technology combined with artificial intelligence (AI) is expected to pay off additionally in terms of reducing the complexity of operating sophisticated training devices so that training tasks such as tactics or gaming are instructionally manageable and performance is measureable. This will be important in the 1990's, even if the manpower shortage is reduced.

Organization of Remainder of Paper

In the three sections which follow, the discussion will be concerned with tasks required of the Systems Engineer, the Human Factors Engineer and the Voice Technology Specialist. The Systems Engineer will be required to design systems in terms of two dynamic, interactive models. The task model generates events to be presented to the trainee. The instructor model delivers instruction. The voice technology subsystem must then be designed to pass sufficient data to the models to allow the student to interact with them. Of course, a model can be composed of submodels, depending upon requirements.

The Human Factors Engineer will be *required to design the feedback* subsystem. For example, people can use a simple frown during conversation to indicate a word wasn't understood. However, a training system using voice recognition must provide a functionally equivalent source of feedback that doesn't distract the trainee from the task.

The Voice Technology Specialist must integrate these two systems. Voice technology data capabilities such as indications of potentially confusing phrases or hesitant speech must be matched with an appropriate instructionally relevant encouragement *from the system so that the trainee* receives functionally the same type instruction a teacher would deliver in a similiar situation.

THE SYSTEMS ENGINEER

The Role of Models

Supporting software models will be required when a speech recognition device is included in a complex Navy training system. Models are not needed when voice is merely a data entry device; a keyboard replacement. The real power in a training system with voice, however is achieved through the relationship between voice recognition and two major models: the model of the task and the model of the instructor. As an example, consider the Ground Controlled Approach Controller Training System (GCA-CTS), designed to provide automated instruction of Precision Approach Radar (1, 2). The task model is the pilot/aircraft model which simulates the responses to the student controller's verbal transmissions, such as "Turn right heading 160." A closed-loop system is achieved by the student observing the simulated radar scope and making appropriate transmissions which are "understood" by the voice recognition system, resulting in the appropriate changes in the radar display. In this case, the integration of voice recognition with the pilot/aircraft model has enabled a real-time interactive simulation. An added benefit is that no human "pseudo pilot" is required to manually simulate the pilot responses. Therefore, the combination of voice recognition and pilot/aircraft modeling has achieved a more controlled response of the simulated pilot while eliminating the need for one support person for each trainee. In a similar development, a pilot/aircraft model has been designed by Hooks (3) in support of automated training for the Landing Signal Officer (LSO).

Another type of model which is closely related to the use of voice technology for training is the instructor model. Several studies have dealt with the issue of how to move toward *"instructorless" training through the use* of software models of the instructor and the task (4, 5). Instructor model functions for an LSO training system will soon be completed (6). Advances in the field of artificial intelligence (AI) promise to enhance the development of "intelligent" instructor models (7), and the NAVTRAEQUIPCEN currently is initiating work on this issue.

Instructor models will never totally replace a human instructor for long periods of training, but instructorless (or nearly instructorless) training is feasible now and can provide the capability both to relieve instructor manpower shortages and to promote effective, objective, consistent training. Instructor models in voice-interactive training systems must be capable of several functions, including the following:

- Provide Instruction
- Measure and Evaluate Performance
- Provide Performance Feedback to the Student
- Decide on the Appropriate Individualized Instruction (remediation, task difficulty, etc.)
- Keep Records of Students' Progress
- Communicate Relevant Information to the Human Instructor.

These functions are not limited to training systems with voice technology, but the voice capability carries with it a special set of considerations for the design of the instructor model. For example, automated instruction can take full advantage of voice interaction by using speech generation as well as video display to demonstrate proper and improper procedures. Speech recognition and speech generation can provide the basis for a natural language interface between the student and the simulated instructor. The student can use voice to query the system, and ask for review or additional information. Voice interaction with the automated instructor is a natural communication medium. The focus here is to make the student an active part of the instructional process, rather than a passive recipient of information.

The performance measurement and evaluation function in a voice interactive training system for verbal tasks will be critically dependent on the accuracy of voice recognition. Accurate performance measurement for air traffic controller training, for example, can be severely degraded by speech recognition errors. Careful design of the supporting software is required to assist in the discrimination between a student's performance error and a speech recognition error. The importance of accurate discrimination between these two types of errors is obvious when one realizes that incorrectly attributing a speech recognition error to a student error can be carried through to faulty performance feedback, syllabus decisions, and record keeping. Task oriented software based on AI principles appears to be a promising approach to this problem.

The instructor model also may be assigned the responsibility of managing the support requirements of the voice recognition subsystem, such as collecting voice reference patterns, providing instruction on "how to talk to the system," monitoring confusion matrices to prompt voice retraining, and supporting voice retraining when requested by the student. These functions must be accomplished through the coordination of the instructor model and the voice subsystem. The design goals are to: minimize the time required for the student to learn to use the voice system; avoid long and tedious voice training (data collection) sessions; and assist the student in maintaining high recognition accuracy over time. This topic is discussed later under Reference Pattern Formation.

A student model is another candidate for inclusion in an automated training system. The function of the student model is to generate inferences about the changing state of knowledge/skill of the student. These inferences are used to select appropriate individualized instruction. Adaptive training in the form of either multiple syllabus branches or variable task difficulty can be supported by a student model, as discussed in several NAVTRAEQUIPCEN reports (4, 5, 8). The design of the student model in a voice-interactive system is not particularly unique, other than its dependence on the voice recognition system (and the performance measurement system) to provide the input data regarding the students' current level of (verbal) performance.

Voice Recognition Issues Affecting Training System Design

A number of voice technology issues will be encountered by the systems engineer during implementation of voice recognition technology. Although there may not be simple answers to these issues, Table 1 is presented in the belief that identifying some of the potential problems can be beneficial.

Clearly, the list of issues in Table 1 is not exhaustive. It is merely a sample. In general, a major category of issues to be confronted is speaker variability (fatigue, sore throat, lip-smacking, non-meaningful sounds such as "ah," amplitude variation with situational stress, etc.). In short, all the variabilities that make speech more interesting (to the human listener) than a predictable monotone are what make automated speech recognition a challenge. In addition to speaker variabilities, there will be environmental conditions to be confronted, depending on the training application. Ambient noise, particularly of the impulsive type, may be a factor. Consistent microphone placement is important. Motion and vibration in an

Table 1

Issues and Suggestions for Implementing Voice Recognition

ISSUES	POTENTIAL SOLUTIONS/SUGGESTIONS
Vocabulary Definition	Do it early, ideally during a full ISD process. Consider stylization requirements and natural pauses. Avoid confusible items. Consider word/phrase lengths and time required for speaking.
Voice System Selection	Consider objectives/requirements of training system, e.g., "real-time" vocabulary size, isolated vs. connected speech, and sampling requirements. (See 9)
Speech Sampling (Voice Data Collection)	*More research is needed on opposing viewpoints:* (1) Train (sample) in random order. Seek consistency and sampling in context of task. (2) Sample repetitively on the same word/phrase seeking variability in reference patterns. Double-map difficult or alternative words (e.g., "nine" and "niner").
New Users	A "voice recognition test" mode should be available. Increasing competence (and recognition accuracy) is to be expected over time for a new user.
Recognition Feedback	Essential. Alpha numeric, audio or situational feedback information should be provided immediately, ideally without interfering with the speaker's primary task.
Maintenance of Recognition Accuracy	Automatic or manual procedures are needed to facilitate accuracy maintenance. Voice test function is needed. Access to "retraining" is needed. A confusion index is recommended.
"Understanding" Software	*Supporting software (AI?) is recommended to enhance speech* recognition through the use of syntactic, semantic, and task information.
Training System Integration	All software "downstream" of speech recognition, such as performance measurement and adaptive syllabus control, must be designed to minimize the impact of speech recognition errors.

operational training environment could introduce additional sources of variance.

Navy Training Systems

The two prototype Navy training systems that have included voice technology have been relatively sophisticated, featuring real-time interactive simulation, automated performance measurement, and automated instruction. However, simpler voice applications in training also may be quite worthwhile, such as a voice-interactive CAI system. One can envision a simple CAI system with a very limited vocabulary, perhaps consisting of eight words such as "ALPHA, BRAVO, CHARLIE, DELTA, NEXT, YES, NO, and REVIEW." Advances in the technology of integrated circuits reportedly will enable a single-chip recognition device with an eight word vocabulary to be marketed within the next year for less than $100.00 (10). These advances in hardware technology may reduce costs to the point of popularizing voice technology in a wide variety of entertainment and consumer products. This presents both a challenge and a benefit to the designer of training systems oriented to complex, technological jobs within the Navy. The benefit will come by the very "demystification" of automated speech recognition. Just as today's recruits find hand calculators commonplace, nearly every Navy recruit in 1990 will have operated some sort of voice system. This will tend to facilitate user acceptance and prevent the equivalent of "mike fright" for new users. The challenge for the training system designer will be to select the appropriate voice recognition device from the projected large number available, and adroitly to integrate it into the training system.

Design Guideline Development

In addition to the implementation of models, the systems engineer must integrate data flow between the models and the voice technology subsystem. Some of the issues to be considered in that include strengths and weaknesses of various features of speech recognition systems, such as:

- Isolated Word Recognition (IWR) or Connected Speech Recognition (CSR)
- Speech Stylization Requirements
- Speaker Dependence/Independence
- Vocabulary Size

- Voice Data Collection
  - *Number of Samples*
  - Procedures for
- Voice Test and Retraining
- Confusion Matrices
- Recognition Accuracy
  - *Misrecognition*
  - Non-recognition
- Recognition Speed
- "Understanding" Software.

The promised advantages to be gained from voice technology, however, can only be gained by careful design of the entire training system to be compatible with the capabilities and limitations of the trainee and the voice subsystem. The design guidelines currently under development (11) are intended to consolidate the lessons learned from previous prototype system developments, to project the trends of voice technology, and serve as a sourcebook for specifying design options and choosing among them, based on the objectives of the particular training system being developed.

There are numerous physical, environmental, and human factors issues in the training applications of voice technology. An "up front" Instructional Systems Development (ISD) process is necessary, and, at the other end, a generous allocation of time for system test and "debugging" is required to achieve an effective voice-interactive training system. The systems engineer must therefore work closely with ISD personnel in order to consider all relevant issues.

HUMAN FACTORS CONSIDERATIONS

The Task Analysis

The development of a training device, especially a major device for hands-on training, must be based on an analysis of the tasks or skills to be trained. *This training task analysis* provides the basic data for determination of training objectives and resulting training system performance requirements. In a training task analysis, the emphasis *should be on relating system functions to* trainee perceptions and responses, determining special skill requirements, and making certain that each system output critical to task performance is *linked to an operator response.*

One of the recommendations from a recent study of the training implications of airborne applications of voice *technology (12)* was that task analyses and other front-end analyses for training systems employing voice technology should be performed by professional personnel who thoroughly understand the human factors of voice-interactive technology. Furthermore, personal hands-on experience with voice-interactive systems was recommended as a *means* of assuring familiarity with the human factors of *voice technology. Without such personal* experience, there is a danger that some important factors in learning to use automated speech recognition and synthesis will not be reflected in the *development of instructional objectives.*

In the design of voice technology for systems to train speech-based tasks that do not themselves use voice technology --that is, in using voice technology as a training medium-- particular care must be taken that the constraints imposed by voice technology do not cause the training tasks to *misrepresent the speech-based tasks being* trained. Personal voice technology experience can help instructional designers to use training media based on voice technology in ways that foster rather than hinder accomplishment of training objectives.

Reference Pattern Formation

Most of the presently available automated *speech recognition systems are* "speaker-dependent", that is, they require that each operator "train" the system by providing examples of that speaker's pronunciation of the words to *be understood. This voice reference* pattern formation process is not required for speaker-independent systems, but such systems can generally recognize only within a vocabulary limited to a very few items.

The pattern registration process in speaker-dependent systems requires a speaker to pronounce several times any word or phrase to be recognized. A composite of the several pronunciations is stored and the recognizer system can *then compare future utterances* with stored composites as the basis of its recognition magic. The formation of reference patterns is extremely important to recognition accuracy, since word or *phrase recognition occurs when an* utterance is judged by the computer to match one reference pattern better than any other.

It is to be expected, then, that if a word or phrase is spoken in a particular way during reference pattern formation, and then spoken differently later, it may not be recognized correctly. The subtlety of differences which can interfere with recognition becomes clear only after one has attempted to use an automated speech recognition device using today's technology. Differences which are not at all apparent to the speaker may result in non-recognition or misrecognition by the system, leading to considerable frustration on the part of the user.

Differences between the context in which reference pattern formation occurs *and in which recognition is attempted can* *result from changes in physical* conditions, such as noise, vibration, G-forces, and other factors. The substantial research on the effects of physical context has been summarized by Coler (13) and others. Although not as well studied, differences resulting from changes in psychological context have often been noted informally by many researchers (14, 15).

Among the most widespread observations are 1) that words trained individually may not be recognized when later embedded in longer utterances, and 2) that a speaker is often misrecognized when speaking in a stressful situation if the voice reference patterns have been entered in a non-stressful setting. Such misrecognition may be self-perpetuating, since it induces additional stress, which leads to further misrecognition. If speech recognition is to work well in a variety of psychological contexts, it is probably necessary to perform voice reference pattern formation under conditions that effectively simulate the range of operational situations to be encountered. It may be possible to obtain speech samples which are sufficiently typical of the trainee's normal voicing by collecting them during the practice of correct terminology during training, as was done for parts of the GCA-CTS vocabulary (1).

It is desirable that trainees spend as little time as possible in speaking for the sole purpose of registering voice reference patterns, an activity with little training value to the trainee. However, if voice reference pattern registration is well integrated into the training program, it can occupy considerable time, and that time will also be beneficial to the trainee. It is essential that this be done to avoid wasteful use of trainee time, and also to avoid trainee boredom or loss of interest. Fortunately, such integration of reference pattern registration into substantive training exercises also serves to increase the likelihood of proper psychological context for reference pattern registration.

Recognition and Re-training

All currently available Automated Speech Recognition systems have performance limitations which render them less efficient and less adaptable than a human listener. Although some limitations are not easily surmounted, others stem from conditions which can be controlled to minimize their detrimental effects on recognition.

We have already discussed what is perhaps the most important of these controllable factors, which is the context in which voice reference pattern formation occurs. Another factor known to affect recognition accuracy is variability among individual users in their ability to "talk to a machine." Some users are consistently well understood by Automated Speech Recognition devices, while others have persistent problems, probably in part because their speech is more variable (16).

If we acknowledge that currently available Automated Speech Recognition systems, and systems likely to be fielded in the near future, do not always recognize speech with high accuracy, it becomes necessary to assess the effects of non-recognition or misrecognition on the performance of man-machine systems. In particular, we are interested in the effects on the user and on interactions with the speech recognition system.

In the design of training systems using voice technology, the design of system feedback on recognition accuracy is critical, requiring exacting human factors analysis for each specific application. This is necessary because recognition failure induces frustration and stress in the user, increasing the probability that the next utterance will be misrecognized, producing further frustration and stress in a vicious cycle. To break this cycle, we must provide the user with easily interpretable feedback concerning recognition accuracy and with natural ways of responding to misrecognitions. A well-designed system will notify the operator that it has not understood, or will display what was understood to have been said, thus giving the speaker a chance to detect and cope with a misrecognition problem. The speaker can repeat the utterance, or retrain the system if necessary. A system which merely fails to respond appropriately to an utterance will leave the speaker not knowing what has gone wrong, nor what can be done to set it right. This extremely frustrating experience can have a strong negative influence on the user's acceptance of the training system and attitude toward training. Thus the effectiveness of training systems using voice technology may be expected to vary directly not only with recognition accuracy but also with the usefulness of the feedback on recognition accuracy.

The voice reference pattern updating or re-training capability of training systems using voice technology must be designed with reference to human factors. Because an effective training program for a speech-based task will lead to constant improvements and changes in a trainee's speech behaviors, frequent updating of reference patterns may be expected. This process may be performed openly, in such a way that the trainee knows that it is taking place (and giving the trainee a measure of control over the process), or it may be integrated into the training so as to be transparent or unnoticed by the trainee. If the updating is transparent, there must also be a provision whereby the trainee can deliberately test and update the voice reference patterns if poor automated speech recognition performance occurs at any time during training. A convenient test and update capability will aid in preventing the frustration which arises from incorrect recognition, and will foster trainee perceptions of control over the training process. It will help prevent an *adversary relationship from developing* between the trainee and the training system.

Human-machine Interface for Voice Technology--Alice's Doorway

In the report on training implications of airborne applications of voice technology (12), an analogy was made between the user of voice technology and Alice in Wonderland. In one adventure, Alice is trapped behind a small door through which she can see a marvelous garden with bright flowers and cool fountains, just out of reach beyond a narrow passageway. Reaching the full benefits of voice technology may also be seen as requiring a user to traverse a narrow passageway, a channel restricted by the human factors peculiar to current speaker-dependent voice-interactive systems, as suggested by Figure 1. Human engineering of the man-machine interface can provide the "magic" to allow the user access to the garden of benefits.

As Figure 1 shows, the human factors that must be reflected in the design of the man-machine interface for voice technology are not limited to speech control factors resulting from the characteristics of present-day automated speech recognition systems. They also include general factors resulting from use of auditory and voice channels for information exchange characteristic of artificially intelligent systems.

In the category of general factors, one of the more difficult problems to deal with is likely to be resistance by experienced training designers and users to changes induced by the introduction of voice technology in training. Heavy involvement in the design of the instructor interface by personnel who will use the system for instruction, coupled with thorough training in its use, can help combat this resistance.

Other difficult problems stem from teaching users the reality of dealing with limited intelligence machines, and from tendencies to view the machine as an adversary. These problems can be especially troublesome because voice technology systems usually do not provide feedback and verification of the sorts people expect in interaction with another intelligent entity. Simply stated, users may attribute too much intelligence to a machine just because it talks and listens, and they can become angry when it fails to live up to their expectations, especially if the basis for its misbehavior is not readily apparent. The human-machine interface must be designed to help the user suppress inappropriate responses that might be directed toward another human and to strengthen appropriate responses that take into account the limitations of the machine.

In the category of speech control factors, the human- machine interface must be designed to aid the user in dealing with any constraints on user speech patterns peculiar to that automated speech recognition system. For example, it has been pointed out that recognition accuracy can be critically dependent upon characteristics of speech utterances that are not usually attended to by a speaker. Feedback on user inputs may have to be structured to help speakers attend to such subtle characteristics.

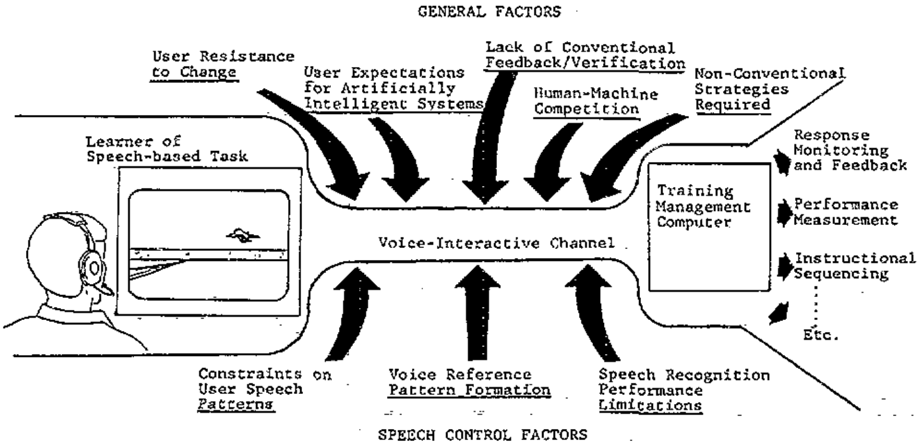

Figure 1. Human factors constraints resulting from use of the voice-interactive channel in training systems.

Voice reference pattern formation/updating and speech recognition performance limitations also have an impact on the design of the human-machine interface. Even for a system in which voice reference pattern formation and updating are sequenced within training program content so that little or no trainee time is spent solely for those activities, recognition failures may be perceived by the trainee as correctible by better reference pattern formation. To maintain user acceptability, the system must permit a user the option at frequent intervals or at will of voice recognition test and reference pattern update.

If the human-machine interface for voice technology in training systems is engineered with the user's capabilities and limitations as a guide, the restrictiveness of the voice-interactive channel can be minimized. Human factors engineering and a well-designed training program can provide the "magic" to allow the user access to the garden of benefits made possible by creative application of voice technology.

VOICE TECHNOLOGY SPECIALIST

The emergence of a Voice Technology Specialist is required for success in complex applications and high payoff in voice technology. To encourage that trend, a short course is being developed at the Naval Training Equipment Center, Human Factors Laboratory, to provide hands-on experience with voice technology. It is expected to be ready for evaluation early in 1982 by a select group of Naval Training Equipment Center instructional systems design personnel, then to be offered Government-wide.

Skills of the Voice Technology Specialist

Models. Training tasks to which voice technology can be applied are characterized by requirements for the trainee to process information, manage data, and make decisions. The model of the task which must generate the data required for proper training is characterized by requirements for generation of variability. For example, a pilot/aircraft model for a controller trainee must be capable of flying various approach profiles which exhibit typical situations. The voice technology specialist must ensure that the model is properly integrated with the voice subsystem so that the decisions of the trainee as spoken to the system are accurately reflected in the response of the model. For example, voice recognition systems often confuse "port" with "four" (they may sound different to you and me, but electronic ears hear things differently). Thus, the integration task of the voice technology specialist requires identification of potential recognition confusions unique to the particular hardware and vocabulary being used. Then, logical substitutions can be made by the system itself when a trainee's decision doesn't make sense.

Similarly, the instructor model must evaluate decisions of the trainee and provide appropriate feedback, remedial training, and critique. The voice technology specialist must ensure that the model is properly integrated with both the task model and speech subsystem so that the trainee doesn't become confused or frustrated. For example, confusion resulted in the GCA-CTS prototype system during a particular two-part maneuver. The first part required a number of decisions in a relatively short time. If any were not recognized precisely, the trainee was not told why and also could not proceed. Since the task was timed, there occurred a build-up of stress which reduced further the trainee's chance of being understood. Thus, the integration task of the voice technology specialist requires identification of potentially stressful situations. Then, the voice recognition subsystem can be designed to relax recognition accuracy requirements until the trainee becomes proficient at the task.

Voice Technology. Certain areas within voice technology are constantly improving and changing. The voice technology specialist must keep abreast of developments in speaker dependent/independent systems, isolated/connected word recognition, vocabulary size vs. system cost trade-offs, accuracy of recognition, and software access to voice recognition parameters such as a confusion matrix, threshold values, and timing of incoming speech.

Human Factors. One area of human factors has been concerned with training panel layout design. In the voice technology area, however, the other design criteria, discussed earlier in this paper, become important. In particular, there is a learning-to-talk-to-the-machine phenomenon in which most people can achieve 99.9% recognition accuracy with a few hours practice. What appears to happen is that a person develops consistency in speaking a particular phrase and develops uniqueness in how that one is said versus how another acoustically similar phrase is said. Exaggeration of consonants is the simplest technique.

The voice technology specialist must analyze the application for critical voice technology features, then optimize the man-machine interface design for user friendliness. Stress is particularly important to consider in training system design. The higher the stress, the lower the likelihood of recognition. In training, stress often comes about from poor, erroneous, or no feedback, or from too fast-paced responding. The voice technology specialist must design a system that combines the "intelligence" of the task and instructor models with the recognition parameters of the voice subsystem to produce a dynamic, interactive, voice-based training system.

CONCLUSION

This paper has concerned itself with technology transfer. An emerging technology, computer voice technology, has been described in terms of the skills required to apply the technology to real-world training situations. Transfer to engineering is being conducted via development of design guides for system models and subsystems. Transfer to instructional systems development is being conducted by development of Human Factors principles of feedback and exploitation of available voice quality data from the system. A Voice Technology Specialty appears to be emerging from a combination of the System Engineering skills and Human Factors skills. This emergence is required for continued application of voice technology.

REFERENCES

1. Hicklin, M., Barber, G., Bollenbacher, J., Grady, M., Harry, D., Meyn, C., & Slemon, G. Ground Controlled Approach Controller Training System Final Technical Report. Technical Report NAVTRAEQUIPCEN 77-C-0162-6. Orlando, FL: Naval Training Equipment Center, 1980.

2. McCauley, M. E. & Semple, C. A. Precision Approach Radar Training System (PARTS) training effectiveness evaluation. Technical Report NAVTRAEQUIPCEN 79-C-0042-1. Orlando, FL: Naval Training Equipment Center, 1980.

3. Hooks, J. T. Pilot behavior models for LSO training systems. Technical Report NAVTRAEQUIPCEN 80-C-0063-1. Orlando, FL: Naval Training Equipment Center, in press.

4. Chatfield, D. C., Marshall, P. H., & Gidcumb, C. F. Instructor model characteristics for automated speech technology (IMCAST). Technical Report NAVTRAEQUIPCEN 79-C-0085-1. Orlando, FL: Naval Training Equipment Center, 1979.

5. Chatfield, D. C., Klein, G. L., & Coons, D. The role of artificial intelligence in voice based training systems. Technical Report NAVTRAEQUIPCEN 80-C-0061-1. Orlando, Fl: Naval Training Equipment Center, in press.

6. McCauley, M. E. & Cotton, J. C. Automated instructor models for LSO training systems. Technical Report NAVTRAEQUIPCEN 80-C-0073-1. Orlando, FL: Naval Training Equipment Center, in press.

7. Barr, A. & Davidson, J. Representation of knowledge. In A. Barr & E. A. Feigenbaum (Eds.), Handbook of artificial intelligence. Stanford, CA: Stanford University, 1980.

8. Chatfield, D. C. & Gidcumb, C. F. Optimization techniques for automated adaptive training systems. Technical Report NAVTRAEQUIPCEN 77-M-0575. Orlando, FL: Naval Training Equipment Center, 1977.

9. Lea, W. A. (Ed.) Trends in speech recognition. Englewood Cliffs, NJ: Prentice-Hall, 1980.

10. The coming wave of electronic ears. Business Week, April 6, 1981, 40B, 40F.

11. Cotton, J. C. & McCauley, M. E. Voice technology design guides for Navy training systems. Technical Report NAVTRAEQUIPCEN 80-C-0057-1. Orlando, FL: Naval Training Equipment Center, in press.

12. Van Hemel, P. E., Van Hemel, S. B., King, W. J., & Breaux, R. Training implications of airborne applications of automated speech recognition technology, Technical Report NAVTRAEQUIPCEN 80-D-0009-0155-1. Orlando, FL: Naval Training Equipment Center, 1980.

13. Coler, C. R. Automated speech recognition and man-computer interaction research at NASA Ames Research Center. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, Texas, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

14. Breaux, R., Curran, M., & Huff, E. (Eds.) Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

15. Harris, S. (Ed.) Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

16. Doddington, G.R. Speech systems research at Texas Instruments. In R. Breaux, M. Curran, and E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffett Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.