[redacted]

VOICE INPUT/OUTPUT CAPABILITIES AT PERCEPTION TECHNOLOGY CORPORATION

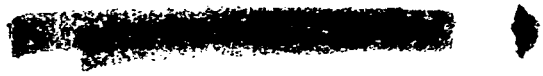
LEON A. FERBER

PERCEPTION TECHNOLOGY CORPORATION
WINCHESTER, MASSACHUSETTS

Perception Technology Corporation was founded in 1969, and began at that time to engage in speech research based upon a Theory of Speech Perception previously advanced by its founder and president, Dr. Huseyin Yilmaz. Since that time, PTC has undertaken and successfully performed a number of research and development programs in speech for various government agencies. As a result of this experience and the backgrounds of PTC personnel, high level capabilities exist in a number of areas related to speech perception.

The Theory of Speech Perception as proposed by Dr. Yilmaz has undergone expansion and refinement over the years, and has been the basis of the speech research effort at PTC. Phenomena predicted by the theory have been verified experimentally, and recognition equipments emulating the human perceptual capability have been constructed. Arising from this work, recognition algorithms and methods have been developed for speaker-independent recognition, recognition of connected speech, and spotting of specific words in unrestricted context. This background has also taken PTC into the voice response field. We have studied both the waveform and spectral natures of speech, and have gained insight into the human facility of speech communication.

The dominant goal of the work at PTC has been the development of effective speech recognition systems. Upon founding of the company, effort was immediately begun on the first PTC recognizer. When completed in 1970 this machine was capable of speaker independent recognition of the digits with a 98% accuracy. Work has continued both under PTC and government sponsorship to expand the utility of this basic system in areas of connected speech, keyword recognition, increased vocabulary, and speaker acceptance. The present capability as recently reported is a recognition accuracy of 99% on a 20 word vocabulary by 50 speakers. An accuracy of 97% has also been realized by a recognizer for connected digits. A more detailed description of capabilities and the performance of the speech recognition systems at PTC is given in the facility section of this paper.



The above discussion is a sample of the capabilities of PTC to carry out programs of research and equipment development. This experience qualifies PTC to undertake related tasks through ability of its personnel to grasp and comprehend high-level concepts such as speech perception, and also through their abilities in implementation of these concepts by computer.

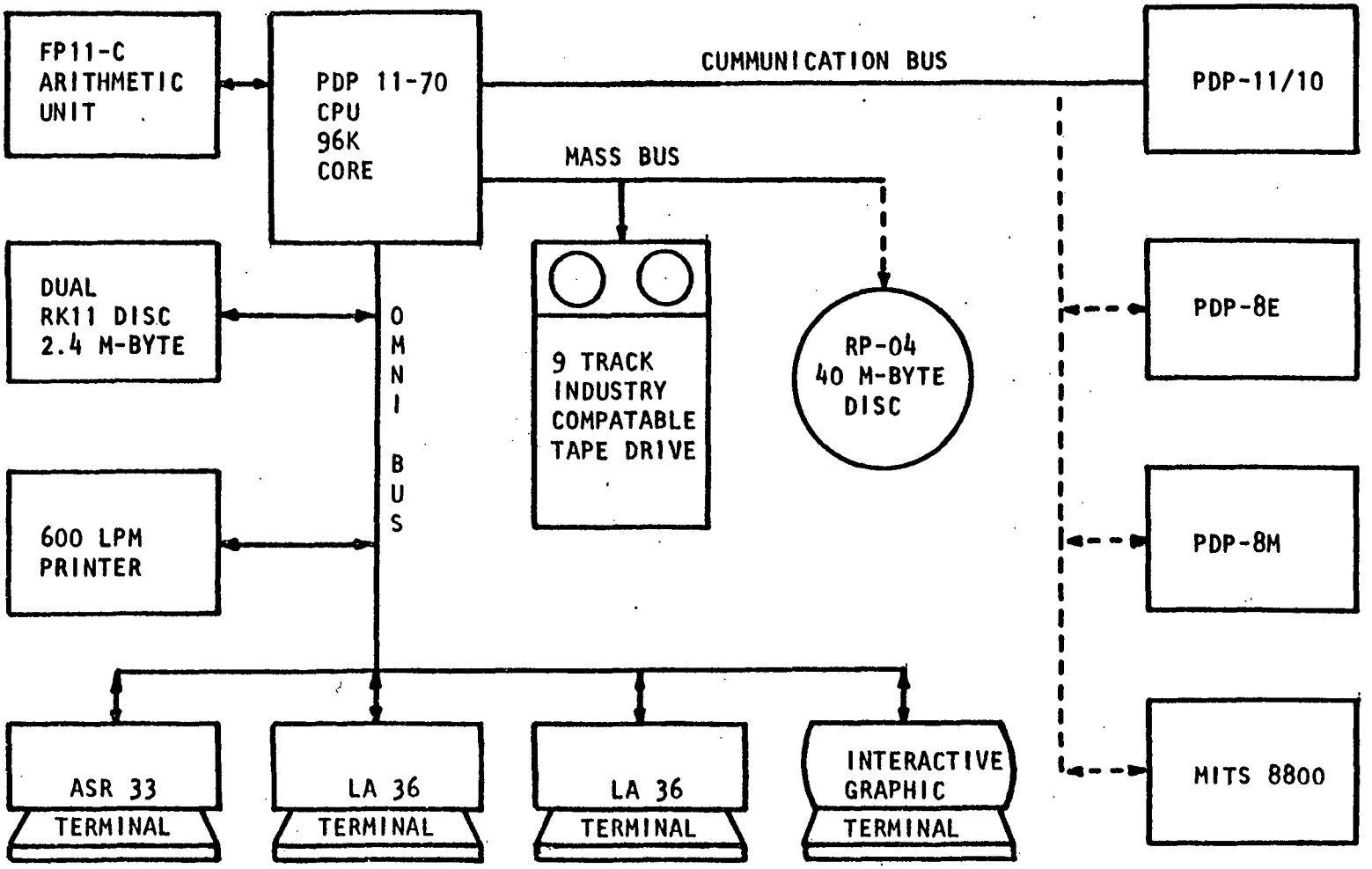
FACILITY DESCRIPTION

Perception Technology Corporation maintains two fully equipped laboratories and a production area. One laboratory is equipped with all the standard and special purpose instruments for R&D in the areas of signal and speech processing, and with instruments and components for breadboarding and testing digital and linear electronic circuits and systems. Another laboratory is equipped for general research in perception and audio perception in particular. It includes equipment to generate speech or to manipulate audio signals to generate a wide range of stimuli required for perception studies in speech. The production area is equipped for assembly of circuit boards and for light manufacturing.

The computer facility configuration shown in Figure 1 is a block diagram showing the major hardware components of the various speech recognition systems. The main system is based on the PDP 11-70 computer operating under RSX-11M. This system is used for software development and for non real-time speech recognition. Most programs are written in FORTRAN IV Plus, evaluated and optimized before conversion to machine language for real-time operation. At the present time this procedure applies only to PDP 11 compatible software. In FY 78 we are planning to have the 11-70 emulate PDP8 and Z80 instructions so that software development for all of PTC's Voice I/O systems can be performed under the main operating system.

There are four additional speech recognition systems, three of which are shown in Figure 2. Two of these are fully operational; the others are under development.

Figure 2a shows the hardware configuration of an on-line data entry system that is planned for FY 78. It is based on software developed for the recognition of digits and control words spoken in connected strings. These programs are now being converted from FORTRAN to assembly language for real-time operation. The system will combine other modes of data entry, such as a digitizing tablet and a CRT, with speech recognition. The system will recognize the English digits spoken in connected strings of random length, and a set of 15 control words.



————————Implemented July 1977 ----------To Be Implemented By March 1978

Figure 1. A Block Diagram of The PTC Computer Facility

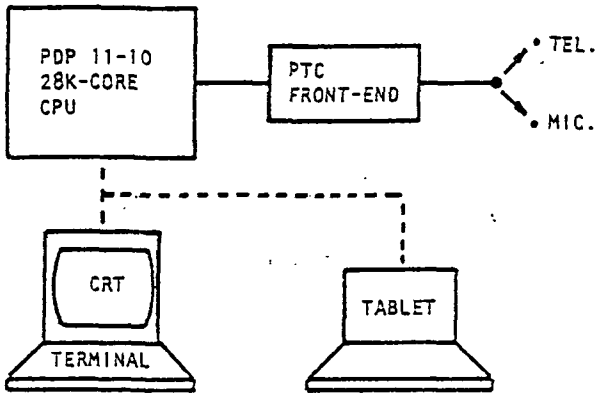


FIG. 2a. Hardware configuration of the real-time, connected speech, recognition system.

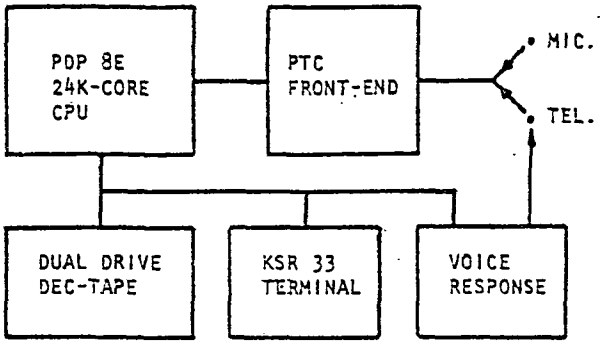


FIG. 2b. Hardware configuration of the Voice I/O on-line system, used for product demonstrations.

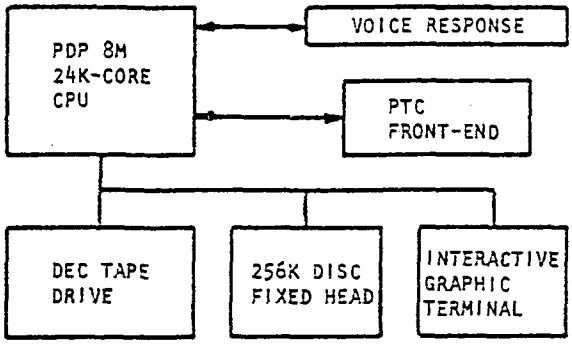


FIG. 2c. Hardware configuration of the Voice I/O non real-time system, used for product software development and connected speech recognition.

Figure 2

The system shown in Figure 2b is a system used for demonstration and evaluation of word recognition. The system is capable of recognizing a syntax-free vocabulary of 30 words spoken in a discrete manner. It is a general-purpose recognizer containing many modes of operation and training schemes. In the speaker independent mode, the vocabulary consists of the 10 digits plus 6 control words. In the trained mode, the vocabulary can be 20-30 words depending on the number of syllables per word. The training has two basic modes of operation, direct training and adaptation. For some applications the two can be combined for increased utility. The direct training consists of repetition of the vocabulary words in sequence or in random fashion using a 32 character alphanumeric display for prompting. In the adaptation mode the system must first be trained for a certain vocabulary, but subsequent speakers use only a few words to get the system adapted to their speech. This system operates with telephone or microphone inputs. The telephone operation is not yet fully interactive; the voice response portion does not yet have a large enough vocabulary for remote prompting and communication.

In FY 78 we are planning to implement the basic recognition portion of the above system on a microprocessor. At the present time we have some of the software operational on an 8080 based development system. The microprocessor based system is expected to be operational by July 1978.

The system shown in Figure 2c is a development system for PDP8 based software. It is also used for testing of "connected speech" recognition and word spotting. The system operates off-line, non real-time and performs recognition on connected digits. Performance tests on this system using constraint-free speech, spoken in random length digit sequences, resulted in an overall recognition accuracy of 97%. This test was done under laboratory conditions using 25 male speakers and results were reported in a technical report No. RADC-TR-76-273.

PTC also maintains a laboratory for general research in perception, and audio perception in particular. The set-up includes equipment to generate speech or to manipulate audio signals for the generation of a wide range of stimuli used in the study of speech perception. This set-up utilizes a PDP8L processor with several software packages. These programs, together with special purpose hardware have been used to implement the following systems:

- An adaptive time compression system for maximizing intelligibility of sped-up speech.
- A digital speech waveform processor with the necessary flexibility for the study and manipulation of signals

in the time domain. This system is also used for synthesizing speech and to generate the data base for the voice response unit.

- A pitch-independent display unit for speech training for the handicapped, based on a color-speech analogy.

SCIENTIFIC & TECHNICAL STAFF

The staff at Perception Technology Corporation consists of seven full-time scientists and engineers with extensive experience in the fields of speech recognition, speech synthesis, speaker authentication and language identification. Other employees include hardware and software engineers with a wealth of experience in system design, circuit design and computer programming. They are augmented by part-time technicians to aid in construction and testing of circuits and systems.

Scientific consultants to and directors of Perception Technology Corporation include: Professor Roman Jakobson of MIT and Harvard University, Professor Harry Levinson of Harvard University and Professor Philip Morse of MIT.

The following pages contain condensed resumes of key company personnel. The information given is pertinent to the fields of research which the company is presently pursuing and does not reflect their overall experience or their achievements in other areas.

HUSEYIN YILMAZ

Dr. Yilmaz recieved B.S. and M.S. degrees in electrical engineering from the Technical University of Istanbul in 1950 and 1951. In 1952, he enrolled as a doctoral candidate at the Massachusetts Institute of Technology and became a research assistant in physics. He received the Ph.D. in theoretical physics in 1954.

From 1954-56, he was a member of the physics department at the Stevens Institute of Technology and in 1956 became a staff member of the National Research Council of Canada. He joined Sylvania Electric Products in 1957, as an engineering specialist pursuing research with emphasis in the fields of atomic physics, theory of relativity, and color perception.

In 1961, Dr. Yilmaz published a mathematical theory of color perception based on adaptive postulates derived from the Darwinian theory of evolution. More recently, he has generalized this theory to embrace other sense perceptions, including the perception of the residue pitch of the human ear, the perception of speech and psychophysics of sensory organization in audio-visual perceptions.

In the spring of 1962, Dr. Yilmaz joined the Research and Development Division of the Arthur D. Little organization of Cambridge, Mass., becoming a member of the Senior Research Staff and a Staff Consultant. During the years of 1962-64, he was also a Research Associate in the Department of Biology at M.I.T.; a guest, for two months in 1964, of the Institute for Perception Research, Eindhoven, Netherlands; and, in 1965-66, a Visiting Professor (full) in Electrical Engineering at M.I.T.

Currently he is concentrating in the fields of speaker-independent recognition of speech, the psychophysical laws, and the problems of audio-visual perception in general. He has also a new statistical approach to quantum field theory which was published in 1969. This work aims at removing field theory divergences by introducing statistical constraints without violating any of the fundamental principles of physics.

As president and principal investigator at Perception Technology Corporation, Dr. Yilmaz follows a highly interdisciplinary approach and tries to join sophisticated ideas and theories with practical engineering applications.

1. "Psychophysics and Pattern Interactions", Models for the Perception of Speech and Visual Form. (Proceedings of a Symposium. Sponsored by the Data Sciences Laboratory, Air Force Cambridge Research Laboratories, Boston, Mass., Nov. 11-14, 1964), Weiant Wathen-Dunn, ed. Cambridge & London: M.I.T. Press, 1967.

2. "On the Pitch of the Residue", Report No. 41, Institute for Perception Research, Eindhoven, Netherlands, 1964.

3. "On Speech Perception", Report No. 42, Ibid.

4. A Program of Research Directed Toward the Efficient and Accurate Recognition of Human Speech. (I). Prepared for the National Aeronautics & Space Administration, Electronics Research Center, Cambridge, Mass. Cambridge: Arthur D. Little, Inc., Dec. 14, 1966, p. 64.

5. "Speech Perception--I", (Vowels), Bull. Math. Biophysics, 29, Dec. 1967.

6. "A Theory of Speech Perception--II", (Consonants), Bull. Math. Biophysics, 30, Sept. 1968.

7. "A Real-Time, Small Vocabulary, Connected-Word Speech Recognition System" (H. Yilmaz, et.al.) Final Report, Contract No. F30602-72-C-0083, 1972.

8. "Perceptual Continous Speech Recognition" (H. Yilmaz, et.al.) Final Report, Contract No. F30602-74-C-0061, March, 1974.

9. "Automatic Speaker Adaptation" (H. Yilmaz, et.al.) Final Report, Contract No. F30602-75-R-0130, July, 1976.

Dr. Yilmaz has given many invited lectures in the U.S. and abroad on speech and color perception. He is the author of numerous internal reports on word spotting and speech recognition published by various government agencies. In addition, he published two books and more than 50 papers and articles in general relativity and psychophysics.

LEON A. FERBER

Mr. Ferber received his B.S. degree in Electrical Engineering from Northeastern University, Boston, Massachusetts in 1969.

Currently, Mr. Ferber is Vice President of Perception Technology Corporation in charge of basic research and product development. He is involved in the design of the company's line of Voice Input/Output products and the implementation of computer based systems for industrial control and material handling. His administrative duties include marketing of Voice Input/Output equipment and contract administration.

Mr. Ferber joined Perception Technology Corporation as an Electrical Engineer to design the digital and analog circuits that went into the construction of the company's first speech recognition system. Subsequently, he was in charge of the design and construction of audio instruments for internal use.

During the years 1967-69, Mr. Ferber worked for Digital Equipment Corporation, Maynard, Mass. His work included design and release to production of circuits for automatic memory test systems, interfacing peripheral equipment to the PDP-8 line of small computers and design of display systems.

1. "A Three Parameter Speech Display", Proceedings of the 1972 International Conference on Speech Communication and Processing, Newton, Mass., April 24-26, 1972.

2. "Speech Perception" Final Report, Real Time, Context Free, Connected Speech Recognizer, Contract No. F30602-74-C-0061, April, 1975.

JAMES SHAO

Dr. Shao received his B.S. degree in Electrical Engineering in 1959 and his M.S. degree in Solid State Physics in 1961, all from the University of Birmingham, England. He received his Ph.D. degree in Physics in 1971 from Massachusetts Institute of Technology.

Presently, Dr. Shao is in charge of development in the area of speech recognition and is the project director on a program to develop a "word spotting" system. His interests are in the areas of speech signal processing and speaker transformation. He also participates actively in the development of computer software necessary for the realization of these processes.

In 1975, Dr. Shao directed the development of a recognizer for unconnected speech. This effort resulted in a product known as PTC VE200.

In 1974, Dr. Shao joined Perception Technology Corporation as a staff scientist to apply symbolic manipulation to the solution of problems in theoretical and applied physics. He participated in the PTC Gravity Research Program and contributed to the study of detection and generation of gravity waves.

From 1972 to the present, Dr. Shao has been a consultant to ERDA at Los Alamos Scientific Laboratory, Los Alamos, New Mexico. He is engaged in the development of software for the Heavy Nucleus Research Program at the Laboratory.

From 1965 to 1968, Dr. Shao was employed by Arthur D. Little, Inc., Cambridge, Mass. he carried out development work on solid state devices and he was in charge of the experiments in their speech research program. During this period, he and Dr. Yilmaz explicitly showed the analogy between color perception and speech perception.

MICHAEL H. BRILL

Dr. Brill joined Perception Technology Corporation in 1977. As a staff scientist he is responsible for the application of speech perception theories in the area of "word spotting", and "connected speech" recognition. His present work includes: Development of feature selection algorithms, application of probability theory and statistics to speech data base generation.

Dr. Brill received his Ph.D. degree in Physics from Syracuse University in 1974; his thesis was "Color Vision: an Evolutionary Approach". He received a M.S. degree in Physics from Syracuse University in 1971 and a B.A. degree in Physics and English from Case Western Reserve University in 1969.

In the period from 1974-77 Dr. Brill was a Post-Doctoral Fellow at M.I.T. working with Professor J. Y. Lettvin on the psychophysics and neurophysiology of the visual system. His work included: computer simulation of information processing in the human visual system, impulse propagation in nerve fibers, and studies of perceptual invariants. He also taught courses and presented lectures on color and vision.

In 1972 Dr. Brill was with the United States Air Force as a 2nd Lieutenant at the IRAP Division, Rome Air Development Center. He monitored contracts on machine recognition of speech and contributed to in-house research on speaker recognition.

HENRY G. KELLETT

Mr. Kellett joined Perception Technology in 1971. He was previously Manager of Acoustic Applications at Peripheral Sciences Inc., of Norristown, Pennsylvania, and has worked as a Senior Research and Development Engineering in Speech Recognition at Philco-Ford and Sperry Rand.

Currently, he is a staff scientist contributing to research and development on government sponsored programs in speech recognition based upon a theory of speech perception and its practical application.

In his present position, he has supervised and contributed to contracts for the National Security Agency and Rome Air Development Center. He has previously been responsible for the design and construction of Speech Recognition equipment at Philco-Ford and Peripheral Sciences Inc.

Mr. Kellett received his education in Electrical Engineering at the University of New Hampshire and the University of Pennsylvania, and holds a B.S. degree in Electrical Engineering.

1. "A New Time Domain Analysis Technique for Speech Recognition", Proceedings of the 1972 International Conference on Speech Communication and Processing, Newton, Mass., April 24-26, 1972.

2. "Experimental, Limited Vocabulary, Speech Recognizer", (Co-author), IEEE Transactions on Audio and Electroacoustics, Vol. AU-15, No. 3, September 1967.

3. "Experimental Speech Recognizer for Limited Word Input", Electronic Communicator, Vol. 2, No. 6, Nov./Dec. 1967.

4. Co-author of numerous technical reports for the National Security Agency, and Rome Air Development Center.

DON DEVITT

Mr. DeVitt joined Perception Technology Corporation in 1977 to take over system development and software operations on the RSX-11 operating system. Presently Mr. DeVitt is working on the conversion of PTC's product software from the PDP-8E to a Z-80 based microprocessor. His objective is to construct a low cost, self-adaptive real time word recognizer.

During 1976, while at Tufts University graduate school, Mr. DeVitt worked with Perception Technology on a voice response system. This system, named the BT-2 Voice Output Terminal, later became a part of PTC's product line.

Mr. DeVitt holds a B.S. degree in Electrical Engineering and a M.S. degree in Computer Science. He received his degrees in 1975 and 1977, respectively, from Tufts University.

Prior to joining PTC, Mr. DeVitt worked for First Data Corporation developing software for interactive graphics and signal processing.

SUMMARY

Recognition methods of connected and continuous speech have been developed by PTC through stratified processing techniques. The smaller, phoneme and syllable, elements are first recognized, then sequences of these are next applied to the large, word and phrase, recognition tasks. This method may be described as a time-warping procedure by which input speech may be recognized even though exact time correspondence does not occur and word boundaries do not correspond with any stored reference data. The methods used in the identification and classification of the phonetic elements are based on a spacial representation corresponding to a perceptual space in which talker and channel transformation are performed. The details of this method are presented in numerous reports that are referenced in the biographical section of this paper. Because of its generality, this method is directly applicable to the implementation of a word identification system. We view all acoustic level speech recognition machines as word spotting systems with appropriate application-oriented constraints. For example, by applying a forced decision threshold and constructing a reference data set for one cooperative speaker, our most general system is reduced to the simplest speech recognizer.

At the present, our main effort is concentrated in the area of word recognition in natural speech. This encompasses two areas of application, keyword spotting and data entry. The keyword spotting

effort is supported by the U.S. Government under contract DAABO-3-75-C-0438. The work in the field of "natural speech" data entry is supported partially by contract No. F30602-77-C-0168 and partially by internal funding. The keyword identification system is targeted as a feasibility study to demonstrate the effectiveness of such a system to perform in a non-cooperative, unknown speaker environment. It is being implemented on a large minicomputer in FORTRAN IV+ and is expected to run in 2-3 times real time. Final evaluation is expected late in FY 78. The data entry system is being implemented on a minicomputer and will operate on-line in real time. A laboratory prototype is expected to be operational in FY 78, and will use speech in combination with other means of data entry. The vocabulary consists of the English digits and command words which may be spoken in connected strings of random length. A similar system operating in an off-line mode was demonstrated at PTC late in FY 76.

BIOGRAPHICAL SKETCH

Leon A. Ferber

Leon A. Ferber is Vice President of Perception Technology Corporation. He is responsible for the development, application and marketing of voice input and output systems.

Mr. Ferber received his B.S. degree in Electrical Engineering from Northeastern University, Boston, Massachusetts in 1969. He joined Perception Technology Corporation in 1969 and designed the company's first word recognition system and numerous speech training equipment for the deaf. Since 1972 he has been project manager of continuing government and internal R&D effort in speech recognition.

During the years 1967-1969 Mr. Ferber worked for Digital Equipment Corporation, designing automatic test systems and graphic displays.

omit to p122

SESSION II

DR. ROBERT BREAUX

NAVAL TRAINING EQUIPMENT CENTER, ORLANDO, FLORIDA

This session presents some of the other applications of speech technology. The first session presented a great deal about artificial intelligence. We heard the terms "man/machine interaction", "command and control systems". These terms, we found, mean different things to different researchers. Yesterday's presentations showed that speech is, in fact, a natural communication channel for the interaction of intelligent entities, a human and a machine. But there was some confusion, I think, yesterday. Those talks could have left the impression that the immediate widespread application of speech understanding must wait for the solution of some significant problem. I will have to agree with that. Before we use speech as an artificial intelligence channel we do have some more work to do. But I also must add that there are commercial firms selling speech products to an ever-growing market. These products are marketed as a way for a company to reduce cost, or to increase productivity among it's people.

Since this market is continually expanding, something must be working in the field of automated speech. So let's shift gears now, and see what these systems are about. Yesterday, we were in low gear, and rightly so. We must have a firm foundation of the potential for automated speech technology. And in low gear yesterday we saw some very powerful potentials. Today, let's shift to drive. We will take a look at how and why commercial off-the-shelf products are being used. But let's also keep in mind that when we shift to drive, we don't want our shiny new technology running away with us, whisking us off to applications for which the technology is not ready. To avoid that, those of you representing government agencies wanting to implement automated speech technology should begin your planning with an analysis of the application. Determine first the extent of an artificial intelligence requirement that you have and this can serve as a measure of how to proceed in your application. One of our efforts at the Naval Training Equipment Center's Human Factors Laboratory, where I am employed as a research psychologist, is an effort for the application of automated performance measurement technology to training.

(This page intentionally left blank)

OMIT

LABORATORY DEMONSTRATION OF COMPUTER SPEECH RECOGNITION IN TRAINING[1]

DR. ROBERT BREAUX[2]

NAVAL TRAINING EQUIPMENT CENTER
ORLANDO, FLORIDA

INTRODUCTION

Background

The Naval Training Equipment Center's Human Factors Laboratory seeks to identify and measure those behaviors which, when improved through training, result in superior performance on the job. Thus, the laboratory seeks to combine new technology developments with current advances in learning/training theory and techniques.

One such technology development is computer speech recognition. The advantage brought to training by this technology is the capability to objectively measure speech behavior. Now, traditional training techniques for jobs which are primarily speech in nature require someone who can listen to what is being said. Otherwise, no measure of the speech behavior is possible. In the U.S. Navy, jobs which are primarily speech in nature include the Ground Controlled Approach (GCA) and Air Intercept (AIC) controllers, as well as the Landing Signal Officer for carrier operations, various Naval Flight Officer positions such as the Radar Intercept Officer, and the Officer of the Deck in ships operations. In addition to the requirement of having an instructor listen to the speech behavior, training in these situations often requires another person to cause changes in the environment which correspond to the trainee's commands. For the GCA and AIC tasks, this takes the form of "pseudo" pilots who "fly" a simulated aircraft target. This 2:1 ratio of support personnel to trainee results in a relatively high training cost.

Previous studies have demonstrated that in analogous situations, it has been possible to achieve savings of manpower and training time while gaining a uniform, high-quality student output by introducing automated adaptive instruction. This advanced technology, if applied to GCA controller training, would bring in its standard benefits such as objective performance measurement and complete individualized instruction.

---

[1]This paper was presented, in part, at the Tenth Naval Training Equipment Center/Industry Conference, 16 November 1977, Orlando, Florida, and published in the proceedings of that conference.

[2]The opinions expressed here are those of the author and do not necessarily reflect the official policy of the United States Navy.

Moreover, for GCA controller students, a more fully automated system could provide greater realism in the performance of "aircraft" under control by accessing directly the computer model of aircraft dynamics rather than relying on the undetermined skills of a variety of pseudo-pilots. Additionally, the rapid processing of an automated system would make possible extrinsic feedback of task performance to the trainee in real-time.

But in order to realize an automated adaptive training system, it is essential that, in addition to values of overall system performance, some relevant aspect of the trainee's activity, in this case his speech behavior, be accessible to the performance measurement subsystem. At this point, our technology review suggested that the state of the art in machine understanding of speech could furnish the means for direct entry of a trainee's advisories. For some whose acquaintance with this possibility is limited to the science fiction of film, television and print media, the response might be "Of course! Why not?" Those more familiar with the problem might say, "Not yet!" The reality is that while computer understanding of continuous unrestricted speech, without pretraining, by any individual who approaches, is still a long way off, there exists today a capability for machine recognition of isolated utterances drawn from a small set of possible phrases. The computer in this case must be pretrained on the language set with speech samples for each individual speaker.

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 setup 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 trainee. Objective scoring is potentially more consistent than subjective ratings. Uniformity can be maintained in the proficiency level of the end product, the trainee. 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 training of the aforementioned Landing Signal Officer, the Air Intercept Officer, the Ground Controlled Approach Controller, and others. Computer speech recognition of human speech 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 automated baggage handling at Chicago's O'Hare airport. 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 Radar (PAR) phase of the 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 indicator containing both azimuth (course) and elevation (glidepath) capabilities. The aircraft target projected on the elevation portion of the indicator 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 cannot catch all the mistakes when there are many.

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 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.

The primary need 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.

Advanced Technology

The major behavioral objectives, then, can more efficiently be achieved through the application of computer speech recognition technology, and thereby the application of advanced training technologies. This is because with objective assessment of what the controller is saying, objective performance meausrement 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, Inc., Cinnaminson, New Jersey.

Three major constraints are imposed by this system. Each user must pretrain the phrases. Recognition does not take place for random, individual words, only predefined 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 50) which are to be recognized. If performance is to be evaluated based upon proper R/T, each phrase must be defined. 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.

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 to the student. The principle being used here is that exposure of a student to certain typical situations will allow him to generalize this experience to real world situations. 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 trainee increases, he is allowed to attempt more complex problems.

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.

RESULTS

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 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 would later 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 phrases were saved for later printout. The no introduction group, on the other hand, was not given practice. Each group then made reference phrases. Reliability data was collected using the procedures described above for R/T practice. A Chi-square value was computed from a 2 x 2 contingency table of frequency of runs in which no recognition errors occurred vs frequency in which one or more errors occured, and whether there had been practice on the phrases vs no practice prior to making the voice reference patterns. 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 pre-practice at the task. Conclusion: Better recognition is achieved when the R/T is voiced consistently and unemotionally.

Training System Evaluation

Twelve recruits were used form the Recruit Training Command, Orlando, who were in their last few weeks and, therefore, were privileged with liberty on the weekend. Each had received assignment to the Navy's Air Traffic Control (ATC) School. Each subject was interviewed for willingness to participate in an "experiment" during liberty hours concerning ATC, and each was informed that for their time they would be paid. Each subject expressed a desire to become an air controlman.

Each subject was issued at the interview those portions of the programmed instruction booklets normally used by the ATC School relating to the Precision Approach Radar (PAR) phase of GCA, and was requested to complete the material prior to arrival at the lab. Each subject was exposed in the lab to approximately three hours of "introduction". During this time the system collected and validated the voice pattern of the subject for each of the PAR phrases. During the between-run intervals, audio recordings were played which explained and reviewed the PAR R/T. Recognition accuracy by the system on the final run of each subject ranged from 81.5% correct to 98.5% with an average of 94.1% correct recognition.

Subjects were then exposed to "free" runs in which they had complete control over the aircraft. It was found that recognition accuracy suffered during the first few runs. The change from a system which fully prompted the subject on the R/T to a full scoring system which required the subject to initiate all R/T, resulted in a noticeable change in the voicings of the R/T. Hesitation, repetition, and corrections were made which, of course, is not within the capability of the speech system to accurately reocgnize. R/T voicing improved with practice, however.

Subsequent School Evaluation

The ATC School was informed of which persons had been exposed to the lab PAR system. Eight of the original 12 subjects completed the 14 week school. Four dropped for "various academic and non academic" reasons, and were therefore dropped from further analysis. During school PAR training which followed exposure to the lab system by about 14 weeks, the subjects' average performance was equal to the school average. A product moment correlation was computed for final score at the school vs complexity level achieved on the lab system. The position correlation $R = .78$, $p<.05$) indicates that better performance on the lab system was related to higher scores at the school. School instructors reported better than average voicings of the R/T by the subjects exposed to the lab PAR system.

The conclusion drawn was that the lab PAR system taught skills similar to those required at the ATC school and, further, that the use of computer speech recognition can be combined with advanced automated training technology to produce an automated training system for the PAR portion of GCA training. Procurement is underway for an experimental prototype system to be installed and evaluated at the ATC school itself.

Where From Here

The technology requirements which follow are based on projections for the next three to five years for proposed applications of automated computer speech recognition in training. The single most important

need is off-the-shelf hardware (e.g., isolated word recognition (IWR) hardware) with software for a limited continuous speech recognition (LCSR) capability. This must have real-time operation with a vocabulary size of 50-100 words. Since training must assume some degree of naivety on the part of the human speaker, training requires a capability to recognize what was said rather than what was meant. Thus, syntax and grammars, which aid processing of the acoustical signal, can in fact be detrimental to training.

Let's consider an example of LCSR and its impact for training. In the GCA approach, a common error is for the trainee to use the word glideslope rather than the correct term glidepath. Now, IWR systems recognize the entire phrase "slightly above glidepath" as one word. So it is seldom that the error is caught when glideslope is used instead of glidepath. With the LCSR capability, however, such errors could be routinely detected. Further, use of syntax as an aid in "understanding" what was meant by the trainee when he erroneously substituted glideslope for glidepath would result in failure to detect that error.

Speaker independence is popular today as a goal for computer speech technology. However, in the training environment the need exists for recognition of speakers from a large cross-section of the population. In fact, there are foreign nationals being trained by some Navy schools. Therefore, emphasis in the training area is for systems which can recognize highly varied speakers, including English speakers whose native language is not English. The IWR system, with its requirement for speaker pretraining, appears to be sufficiently developed to meet this need, particularly if LCSR were included.

Other technology requirements in the training area are reduced hardware costs, less critical microphone placement, and recognition in a noisy environment. Of course, cost is always a factor in any procurement activity. Microphone placement becomes important when the goal of the training system within which the speech hardware operates is a goal of total automated training. The less critical the mike placement, the more inexperienced the user can be. Finally, noise levels cannot always be reduced, as in flight deck operations. With greater noise tolerance, however, greater application could be made for speech recognition. One such example is simulation of flight deck operations for training the Landing Signal Officer.

SUMMARY

A system was described which provided a laboratory evaluation of the feasibility of the use of computer speech recognition in training. Results of the evaluation indicate that training can be enhanced and manpower costs reduced by a careful integration of advanced training technology with off-the-shelf computer speech recognition hardware which is

enhanced with software algorithms designed for a specific vocabulary set. The need was indicated for further research and development via and experimental prototype system to be installed at the Navy's Air Traffic Control School.

REFERENCES

Breaux, R. and Grady, M.W. "The Voice Data Collection Program - A Generalized Research Tool for Studies in Speech Recognition." In Proceedings of the Ninth NAVTRAEQUIPCEN/Industry Conference, Technical Report: NAVTRAEQUIPCEN IH-276, Orlando, Florida, Naval Training Equipment Center, November 1976.

Breaux, R. and Goldstein, I. "Developments of Machine Speech Understanding for Automated Instructional Systems." In Proceedings of the Eighth NAVTRAEQUIPCEN/Industry Conference, Orlando, Florida, Naval Training Equipment Center, November 1975.

Goldstein, I., Norman D.A., et al. "Ears for Automated Instructional Systems; Why Try?" In Proceedings of the Seventh NAVTRAEQUIPCEN IH-240, Orlando, Florida, Naval Training Equipment Center, November 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. Design and implementation of the automated instructor critique software was performed at the Human Factors Laboratory by the author.

BIOGRAPHICAL SKETCH

Dr. Robert Breaux

Dr. Robert Breaux received his Ph.D. in experimental psychology from Texas Technical University in 1974. He is a Research Psychologist in the Human Factors Laboratory at the Naval Training Equipment Center. He has an interest in application of the theoretical advances from the psychological laboratory to the classroom situation. Publications and papers include computer application for statistics, basic learning research, concept learning math models, and learning strategies. He is an instrument rated commercial pilot, and a certified flight instructor.

DISCUSSION

Dr. Robert Breaux

Q: Roland Paine, Systems Control: You mentioned recognizing words, but on this particular training application it seems emotions and the way he controls his voice is very important as well. Have you addressed that issue at all?

A: That's correct. The disc jockey-like voicings are very important to instructor controllers. One of the points that they like about the isolated word recognition systems and the requirement to create voice reference patterns was to require the trainee to speak almost in a monotone, but more importantly, very consistently. Always say the same thing the same way. If a pilot is coming in with icing on his wings and low fuel, he is excited enough. The controller doesn't need to get excited. We need somebody who is calm and cool. We can simulate situations like that to teach the controller how to handle it. There is a potential that with speech technology's requirement to speak very consistently, the instructors feel that there is the potential to improve that portion of training which is concerned with the training of the RT, the Radio Terminology. The students tend to mimic their instructors a great deal, which means they try to go as fast as they can, be very smooth and suave, etc. The instructors really want them to learn the basics right now. You can develop your own technique later. So in that sense, that's one good point about the isolated word recognition systems,

In addition, there is one problem that is very significant in training to me that is different from the problem in the operations area. And that is related, in a way, to syntax and grammar (this is addressed in the paper, by the way). In the training situation, we have a branching factor equal to the vocabulary size because we need to diagnose what the trainee's problem is. He's not an expert in the situation as a pilot would be. We are not talking about having the trainee saying whatever he wants to say and if the system understands him, make the airplane do that. Although that might be a good application in the operational area, it's not in training. We want to teach him to speak the correct phrases. So a speech understanding system that tried to "hear what I mean", may loose a potential to diagnose what the trainee's weakness is at that point in training and, thereby,

loose the potential to determine what sort of situation the trainee may need next. A connected word speech system which could pick out each of the words would be helpful in that sense. Does that answer your question? Any others?

Q: George Doddington, Texas Instruments: Here is the situation. When you are training the controller to do a function where he receives data from a computer and gives data back to the computer, he receives data through a visual display and gives it back to the computer which digests it, recognizes the word and passes it on to the pilot, it seems like an interesting possibility for total automation in this case, where you replace the controller with a computer and the computer then needs to speak to the pilot. What to you think about that idea?

A: Great idea, once you let me describe it this way: Any of you who are pilots realize that you don't trust controllers very much, much less a computer. And even though you might fly a hands off approach on an ACL system, an Automatic Carrier Landing system, you don't fly very far hands off. You're out there ready to grab it. Yes, that's true, and most of the people, a lot of the management-type people who come through our lab, whose job is not concerned necessarily with training or R&D, often make the comment, that gee, what do you need the controller for. And it's certainly a reasonable approach.

Q: George Doddignton, TI: I guess what I am asking is: Is this being considered, are there any programs, have there been any programs, what are the problems? If I were a pilot, I think that I would probably trust the computer more than I would a human, in all seriousness.

A: I won't fly with you. No, I'm kidding. What else can I say? It's a good point.

Q: Wayne Lee, SCRL: If the student is, in fact, going to mimic the instructor and part of his instruction comes from the machine, what quality of speech might be heard. I would't want him to mimic the Votrax we heard.

A: That's a good point. That's been brought up by the instructor controllers themselves.

Q: Wayne Lee: Wouldn't it be very reasonable to just have prerecorded speech that is plugged together and that becomes output?

A: That's a potential that we are considering in the prototype. We'd like to look at a number of ways. As I mentioned the other day, a prototype is a system on which we'll be doing research. I think that was a good point yesterday. We have yet to come out of the lab really. We're going to be in a training situation, but it's going to be a controlled situation and we'd like to look at a number of variables. This again is an R&D effort and when it comes time to procure an operational trainer, if that time comes, then these points should certainly be taken into account, I would think.

Q: Ed Huff, NASA Ames: I don't recall if you mentioned it. What is the language size that you were dealing with and in the course of training, what has been your experience with recognition accuracy? Has that fallen off or improved? And finally, what happens if the recognizer doesn't work properly?

A: First question is vocabulary size, and we are working with a 44 phrase vocabulary. Second question was reocgnition accuracy. Recognition accuracy ranged from about 89 to 97 percent. The third question, in the laboratory version, when I was doing some of the work, I would play an audio tape recorder for part of the time. When the system, the isolated word recognition system, did not understand what was being said, I could replay the audio tape and let the trainee hear what he was saying. In the prototype device we will automate that particular function as well. Essentially, it's a situation which the trainee is trying to learn a number of tasks simultaneously. We hope with advanced training technology that we can reduce these tasks in a small step procedure so that these sort of things don't all hit him at once, and that he won't have trouble voicing his RT. In some situations there are a number of things he must learn all at the same time, not only what to say, but when to say it. He may know exactly what's happening, he's learned that well, and he's just fishing for his RT. He can't think of what to say, and he says "six miles to glide path". You know, little things like this that the system, of course, doesn't recognize. The trainee has the concept; he's fishing for his RT. There are a number of training problems associated with this that are very, very intriguing to me, and that's one of them. Hopefully, we can address some of that in the prototype device. Any other questions?

Q: Dr. Raj Reddy, Carnegie-Mellon University: I have a general comment to make. Those of us who are in artificial intelligence research are constantly faced up to this question of replacing human beings with machines. I think that's a very poor use of words and some of us get carried away with our own enthusiasm. In the long run, I think the way to view this, the use of a computer in general as

an intelligent instrument as we better understand how we can encode more and more of the routine knowledge that an Air Traffic Controller or anyone brings to bear on the problem more of that knowledge can be put into the computer so that the person there can use this facility to do the more important planning and other type of tasks. So the thing we should be talking about is intelligent instruments that would aid all of us whether you are a doctor, an engineer, or whether you're a scientist, in doing your job better, to augment your own intellect. I think that's the way we should think of the use of the computers rather than replacement of a human being by the computer. And I get very sensitive, because those of us who work in the field never think about artificial intelligence as a panacea which will do away with the human beings.

A: That's a good point, and I guess I'm sensitive to it in a way too. And the reason is that we tend to be more intellectual at times than, say another group of people. Keep in mind that not everybody wants to think, not everybody wants to do that kind of a task. There are some people who are very happy about typing away. There are some people who are very happy about various kinds of what we would call non-intellectual tasks. And that's not to degrade it. Not everybody wants to engage themselves in intellectual artificial intelligence. To me it's very difficult, as I said in the opening remarks, to separate speech understanding, communication with an intelligent entity, from the idea of using speech recognition as a tool to reduce cost effectiveness or what. There are a number of areas we could go in with this kind of stuff, and enterprising people, I suspect, hopefully will generate some ideas from this. We have time for a short question.

Q: Roland Paine, Systems Control: You identified this particular program. Would you enumerate some of the others where you are going to be doing more basic and exploratory research with speech technology as affects training in your Center?

A: We would like to explore in some way, artificial intelligence, the kinds of things that have been talked about the past two days, and we're constrained by financial reasons. In general we'd like to see these kinds of systems utilized in training.

Q: Michael Nye, Marketing Consultants: I have a question, but I wanted to make a comment concerning what Raj Reddy said, and that is that I personally believe that one of the limitations or one of the reasons why speech hasn't really, as you can say, taken off in an application environment is that too many times researchers have looked at the conceptual approach without taking a real world appreciation for economics and at such time when economics are presented that there is a cost benefit. Industry and government

applications will come forth very quickly. That's a personal input although I agree with what Raj said. I just wanted to make that comment. My question is when you started in your experimentation of your system, you had some preconceived notion of what you expected, what the limitations and capabilities of this kind of system would be. I'm curious about, based on a few months of practical hands-on experience with technology that is probably limited in scope, what were the things that occurred that you did not expect that caused you to be less enthusiastic about speech understanding systems and what were the positive things that occurred that you didn't expect that made you more enthusiastic about it?

A: Some of the points were made by Mr. Herscher in his paper and I anticipate that he will make them again when he gives his presentation; they concern human factors and the man-machine interaction from a human factors standpoint, logistics of equipment, and this sort of thing. I alluded to one of those earlier about the microphone placement, and things like this. Those are the general kinds of things.