D44

N79-17519

SPEECH AS A PILOT INPUT MEDIUM

R.P. Plummer
University of Utah *
Salt Lake City, Utah 84112

C.R. Coler
NASA-Ames Research Center
Moffet Field, CA 94035

ABSTRACT

An automatic speech recognition system is currently being employed to investigate the use of speech as an input medium from pilots to computers on board aircraft. Such a system would allow pilots to provide inputs without the eye-hand coordinations required with keyboards, switches, etc. One stimulus for this work is literature, briefly reviewed in this paper, demonstrating the effectiveness of speech as compared with other means of communication.

The speech recognition system under development is a trainable pattern classifier based on a maximum-likelihood technique. An adjustable uncertainty threshold allows the rejection of borderline cases for which the probability of misclassification is high. The syntax of the "command language" spoken may be used as an aid to recognition, and the system adapts to changes in pronunciation if feedback from the user is available. Words must be separated by .25 second gaps.

The system runs in real time on a mini-computer (PDP 11/10) and has been tested on 120,000 speech samples from 10- and 100-word vocabularies. The results of these tests were 99.9% correct recognition for a vocabulary consisting of the ten digits, and 99.6% correct recognition for a 100-word vocabulary of flight commands (using command language syntax), with a 5% rejection rate in each case. With no rejection, the recognition accuracies for the same vocabularies were 99.5% and 98.6% respectively.

Plans for the system include fixed-base flight simulations, a motion simulator study, and in-flight tests.

INTRODUCTION

The increasing use of computers on board aircraft requires that increasing attention be paid to the design of the pilot-computer interface. The airborne use of computers usually takes place concurrently with other tasks, with time constraints on the interaction, and with a need for high accuracy of inputs and intelligibility of outputs.

The research reported here is concerned with the selection of an input medium for airborne computers, and, specifically, with the use of an automatic speech recognition system that allows inputs to be given verbally. The attractiveness of spoken inputs in the cockpit environment stems mainly from the fact that a large percentage of the workload is visual and manual. It is felt that the use of another communication channel (speech) for providing computer inputs will be less disruptive of (and less disrupted by) other tasks than the use of a manual input system.

BACKGROUND

Although considerable literature exists on the development of speech recognition systems [7, 9], less work has been done on the effectiveness of using such a system as a communications medium. This section reviews briefly work relating at least indirectly to the use of speech in cockpits.

Braunstein and Anderson [1] performed an early study comparing the speed and accuracy of speaking and keypunching digits. Their subjects, who had no prior keypunching experience, were able to read digits aloud twice as fast as they could keypunch, even with several hours practice. Accuracy of speaking was determined by human judges and found to be slightly better than that of keypunching.

A recent study by Williams [10] measured the keystroking ability of commercial and airline pilots. On a five-minute typing test, the subjects averaged 95.35% correct keystrokes. This provides a useful figure for comparison with the accuracy of speech recognition systems.

A general discussion of the use of speech for man-computer interaction has been given by Turr [8]. He cites the following attractive features of speech:

(1) the independence of speech from the visual channel and manual activities,
(2) the omnidirectional nature of speech,
(3) the ability of a speaker to communicate simultaneously with computers and humans, and
(4) the simplicity of converting speech to electronic form.

Turn also discusses the difficulties in implementing speech recognition systems. These lie mainly in the area of continuous speech recognition; he points out that isolated word recognition is already a reality. (The system discussed below uses isolated words.)

Also relevant to the question of speech as a communications medium is the work of Chapanis, et al, on interactive communication [2, 3, 4, 5]. Most important from the standpoint of

* Presently on leave to NASA-Ames Research Center.
This work was supported under NASA Grant NGR 45-003-108.

man-computer interaction are the following results [3]:

(1) Problems are solved significantly faster in communication modes that have a voice channel than in those that do not.
(2) Oral communication is highly redundant and most communication can be carried on effectively with a small, carefully selected set of words.

In summary, the work cited suggests that a speech recognition system would provide a natural, accurate, and rapid means of communicating with computers, especially in environments where the visual and manual workload of concurrent tasks is high.

AN ISOLATED WORD RECOGNITION SYSTEM

As a tool for experimentation in the cockpit, an automatic speech recognition system has been constructed. The system recognizes isolated words, that is, words separated by pauses of at least .25 seconds. The resulting "staccato" style of speech is not felt to be a problem for the anticipated command-oriented applications.

An utterance is digitally encoded by the use of 16 bandpass filters, sampled at 60 Hz. A time-warping algorithm divides the utterance into 8 subintervals (of possibly unequal duration), such that the amount of spectral change is the same within each subinterval. The data within each subinterval is then reduced to a 15-bit representation, producing a 120-bit encoding of the uttrerance.

Recognition is achieved by applying a maximum-likelihood pattern classification technique [6] to these 120-bit patterns. The system is trained to a particular speaker's voice by providing it with a set of samples of each word in the vocabulary to be spoken (the number of "training" samples of each word is generally between 5 and 25). These samples are used to estimate the probabilities of the occurrence of a 0 or 1 at each of the 120 bit positions for each vocabulary word. Given an unknown utterance to be classified, the probabilities are used to compute a similarity score for each vocabulary word, and the unknown is classified as being an example of the word with the highest score.

Three additional features augment this basic recognition scheme. In many applications it is preferable to reject (fail to classify) an utterance rather than misclassify it. The system rejects those utterances whose classification is "uncertain", where uncertainty is measured by computing the ratio of the second highest score to the highest score. A word is rejected if its uncertainty exceeds a preselected threshold (if no rejection is desired, the threshold is made >1).

A second feature of the system concerns the fact that even trained speakers vary their pronunciation of words slightly over time. Thus, the characterizations of the vocabulary words obtained from the training samples become less accurate as the speaker subsequently uses the system for recognition. In applications where feedback from the user is available, the system uses the words spoken to continually update its probabilities, thus compensationg for pronunciation shifts. Feedback may simply inform the system whether its classification was correct, in which case updating is done after each correctly classified utterance. If the system is also told the correct classification for each missed utterance, then updating can always be done.

Finally, in applications where the user is speaking a "command language" of known structure, the syntax of the language may be used to determine the subset of vocabulary words that are possible at each point in a command. For example, after "landing gear", the only meaningful words might be "up", "down", and "status". Recognition could be done only against this small subset rather than the entire vocabulary of, say, 100 words. This technique provides a considerable hedge against any degradation of performance with increasing vocabulary size.

The system is implemented on a PDP 11/10 minicomputer. About 13K 16-bit words of data storage are required for a 100-word vocabulary, and recognition time is slightly less than .5 seconds.

RECOGNITION RESULTS

The system has been tested extensively, on both 10- and 100-word vocabularies. Speech samples for the vocabulary consisting of the ten digits were obtained from 20 subjects, with each subject providing 25 training samples and 100 recognition samples of each word. As shown in Table I, recognition with this vocabulary is near perfect, especially when rejection is allowed. Syntax was not involved in this experiment, but feedback for updating was provided.

TEN DIGIT VOCABULARY

% Correct	% Rejected
99.5	0.0
99.9	5.0

Table I

A 100-word vocabulary of flight commands was tested with a group of 10 subjects, with each subject providing 25 training samples and 100 reconition samples of each word. A command language using this vocabulary was constructed; its syntax grouped the commands into 15 subsets ranging in size from 3 to 10 words (average size = 8.7 words). Table II shows the recognition results with and without rejection and with and without the use of syntax. Feedback for updating was again provided.

100 FLIGHT COMMAND VOCABULARY

	% Correct	% Rejected
without syntax	93.2	0.0
	95.7	5.0
with syntax	98.6	0.0
	99.6	5.0

Table II

FUTURE PLANS

The results to date demonstrate that an effective speech recognition system has been constructed. A set of experiments will begin shortly that will compare the system with a keyboard input device from the standpoint of accuracy and speed in laboratory conditions, and in conditions of noise and turbulence similar to those encountered in aircraft. The speech recognition system will then be used for providing inputs to a 4-D area navigation system in a full mission flight simulation, and, ultimately, in actual flight tests. These experiments should provide conclusive evidence on the viability of automatic speech recognition in the cockpit environment.

REFERENCES

1. Braunstein, M. and Anderson, N.S. A Comparison of the Speed and Accuracy of Reading Aloud and Keypunching Digits. IRE Transactions on Human Factors in Electronics, HFE-2, 1, 56-57, 1961.
2. Chapanis, A. Interactive Human Communication. Scientific American, 232(3), 36-42, 1975.
3. Chapanis, A. Interactive Human Communication: Some Lessons Learned from Laboratory Experiments. Paper presented at the NATO Advanced Study Institute on Man-Computer Interaction, Mati, Greece, Sept., 1976.
4. Chapanis, A., Ochsman, R.B., Parrish, R.N., and Weeks, G.D. Studies in Interactive Communication: I. The Effects of Four Communication Modes on the Behavior of Teams During Cooperative Problem -Solving. Perceptual and Motor Skills, 38, 343-374 1974.
5. Kelley, M.J. Studies in Interactive Communication: Limited Vocabulary Natural Language Dialogue. Doctoral Dissertation, The Johns Hopkins University, 1975.
6. Nilsson, N.J. Learning Machines. New York: McGraw-Hill, 1965.
7. Reddy, D.R. Speech Recognition by Machine: A Review. Proceedings of the IEEE, 64, 501-531, 1976.
8. Turn, R. The Use of Speech for Man-Computer Communication. Rand Corporation Report R-1386-ARPA, 1974.
9. White, G.W. Speech Recognition: A Tutorial Overview. Computer, May, 1976.
10. Williams, D.H. The Keystroking Ability of Commercial Pilots. NASA Technical Memorandum TMX-73,160, 1976.