VOICE CONTROL OF COMPLEX WORKSTATIONS

Jeffrey L. Scruggs

Texas Instruments, Inc.
Speech And Image Understanding Lab
Computer Science Center
P.O. Box 655474, MS 238
Dallas, Texas 75265

ABSTRACT

This paper describes the use of a speaker-dependent connected word recognition system to control an Air Traffic Control (ATC) demonstration workstation, and the work which went into developing that speech system. The workstation with speech recognition was demonstrated live at an Air Traffic Controller's Association convention in 1987. The paper discusses the purpose of the demonstration workstation, and highlights the development of the speech interface. This includes: a brief description of the speech hardware and software, and overview of the speech driven workstation functions, a description of the speech vocabulary/grammar, and details the enrollment and training procedures used in preparing the controllers for the demonstrations. Although no quantitative results are available, the paper discusses the potential benefits of using voice as an interface to this type of workstation, and highlights limitations of the current speech technology and where more work is required.

INTRODUCTION

For many years, speech has been recognized as one of the preferred man-machine interfaces. Within the last decade, with the advent of low-cost speech processing hardware and software, we have begun to see commercial applications which utilize speech as an interface between man and machines. There have been many successful systems providing voice response applications; for example, voice mail. Applications using speech recognition have generally been limited to areas where the vocabulary required to interact with the system has been small, and where the words can be spoken in isolation. Many of these successes have come in the areas of factory quality inspections and inventory control.

In the last 3 or 4 years, the speech industry has begun to address the problems associated with using speech to interface with more complex tasks. Some of these tasks have included: voice control of AFTI F-16 cockpit systems, dictation using a voice actuated typewriter (VAT), voice control of stock market order entry stations, and medical transcription terminals. All of these systems require higher levels of speech recognition performance than the earlier applications. These new applications require larger vocabularies, more connected speech capabilities, and easier training mechanisms. In addition, these new applications continue to require system recognition accuracies of greater than 95%. The efforts to meet the needs of these more complex tasks have met with varying degrees of technical success, but none of these systems have yet achieved widespread commercial success to date.

This paper describes an effort to use a commercially available speech recognition product from Texas Instruments as an interface to a complex workstation, an Air Traffic Controllers Workstation. This was a joint effort between the Ground Systems Group of Hughes Aircraft and the Computer Science Center Speech And Image Understanding Lab of Texas Instruments, Inc. The goal was to develop a speech interface for Hughes-designed demonstration ATC workstation. The workstation was displayed at the 1987 Air Traffic Controllers Association Convention in Los Angeles, and at 1987 Radio Technical Commission for Aeronautics in Washington DC. The speech system was developed for demonstration purposes only.

OVERVIEW OF THE SYSTEM

Figure 1 shows the main hardware components of the demonstration workstation. The controller's console was composed of a 20-inch square color display on which the operator could view the air traffic in his own and adjoining sectors. The console incorporated several interface technologies including a keyboard and track ball (used in current workstations), as well as a touch panel and speech recognition interface. Any of these devices could be used interchangeably by the controller to manage the console. A Sun workstation was used as the console controller during the demonstration, while a PC containing the TI speech recognition system was the other major piece of hardware. The PC was connected to the Sun via a RS-422 serial communication link. In order to provide realistic data during the demonstrations, the Sun workstation hd been preloaded with a scenario from the Los Angeles International Airport area. The microphone cable was patched directly from the console to the speech hardware.

The purpose of the demonstration was to show how the use of various interface technologies made it easier for the controller to manage his or her workstation. There was no simulation of the link between the aircraft and the ground. Instead, the controller just managed the workstation in front of him. During the demos, the airspace scenario was free-running on the console, while the controller demonstrated the features of the man-machine interface while using the various input devices.

DESCRIPTION OF THE SPEECH SYSTEM

The speech system used was the Texas Instruments LR2000 recognition system. The hardware is a single board option for IBM PC's and compatibles based on the TMS32010 digital signal processing chip. The board is a flexible speech peripheral capable of performing a wide variety of speech processing tasks including: record / playback, text-to-speech, speech recognition (both isolated and connected speech), and speaker verification. In addition, an application software development kit is available to allow users to write custom applications utilizing any of these speech capabilities.

Figure 2 shows a block diagram of the speech recognition process. The lowest level is a speaker-dependent word hypothesizer which has two inputs: the real-time speech input, and previously stored vocabulary word templates. As the user speaks, his input speech is compared against the templates; when a match is found the hypothesizer outputs a result to the second level of the system. This second level is the sentence recognizer. This subsystem compares compares the output of the word hypothesizer with a previously defined grammar structure, and outputs recognized sentences to an application program on the PC. The grammar structure is a finite-state grammar which describes all the valid sentences in the application domain. The vocabulary and grammar used in the ATC demonstration are described in the next section.

The advantages of this two-level decision structure are two-fold: first, the robustness of the recognition is improved since more global knowledge of the application environment is available at the recognizer level (in the form of the application grammar), and two, by using improved training procedures the users can speak to the system using connected speech.

DESCRIPTION OF THE VOCABULARY AND GRAMMAR

The task of determining where to use speech as an interface was a cooperative effort including Hughes human factors experts, the controllers who would be demonstrating the systems, and the Texas Instruments speech application developer. There were three principal areas where speech was considered: voice recognition driven by the radio uplink to aircraft, voice recognition to control the workstation console, and speech synthesis to notify controllers of conditions requiring attention. Since the scenario to be demonstrated did not include a simulation of the radio uplink, that area was rejected. The other two areas were both considered very promising for using speech, and were both within the capabilities of TI's speech product. However, due to time constraints in preparing for the convention, only the console control voice recognition was actually implemented.

The voice commands fell into two major categories: console display control and aircraft situation acknowledgement. The console display control functions were concerned with how the data was displayed on the main console color display. These included displaying data from other control sectors, changing the display range in miles, and highlighting critical flight data elements for specific aircraft. The aircraft situation acknowledgement commands included: acknowledging aircraft alerts, acknowledging flight plan postings, marking handoff of aircraft between sectors, and assigning altitudes, beacon codes, and preferential routes. A vocabulary of 94 words was defined to provide these functions. Table 1 provides a list of the vocabulary used in the demonstration.

As previously mentioned, the LR2000 recognition system requires both a vocabulary list and a application grammar. Definition of a grammar for the ATC workstation application did not prove very difficult since the controllers are already trained to use a standard "language" when controlling their airspace. Figure 3 shows a portion of the system grammar describing the acknowledgement of alerts and the highlighting of flight data elements (FDE). The symbol **<flid>** indicates that the controller could at that point in the grammar say any of the flight identifiers which were available in the scenario and had been programmed into the grammar. The symbol **<1 - 8>** indicates that the controller could say any digit from one to eight. Due to the limited nature of the demonstration, the possible flight identifiers were restricted to those occurring during the scenario. This restriction was also required due to the vocabulary size limitations of the TI speech system. This size limitation is related to the processing power and memory space available on the speech hardware, and is not a physical limitation of the recognition algorithm.

ENROLLMENT AND TRAINING STRATEGY

Because the speech recognizer was speaker dependent, the system had to be trained to recognize each individual speaker. For the three day convention, 8 controllers were chosen to demonstrate the ATC workstation. Each controller was required to enroll the complete 94 word vocabulary.

The enrollment strategy for the LR2000 system is a two-step process where each word is said once as part of a sentence and once in isolation. These two "templates" are then used to create a single recognition template for that word. The resulting template incorporates "coarticulation" effects which normally prevent speech recognition systems from being used with connected or conversational speech. The initial template creation took approximately 45 minutes for the 94 word vocabulary. Subsequently, each controller maintained their templates by periodically repeating a set of sentences which included all 94 words of the vocabulary. This update process required about ten minutes per repetition.

In order to ensure good performance at the convention, the updates were performed at different times of the day over several weeks. In this way, each template included the daily variations which all of us have, along with any long-term variations which might appear. During this process, several of the controllers had colds or allergies; the effects of which were included in the update process. The updates were supervised by one of the controllers who was trained to monitor the template creation and update function. By the time of the convention, the 8 controllers had updated each word at least 7 times. In addition, at the convention, each of the controllers was again updated to accomodate the new acoustic environment in the convention hall.

RESULTS AND CONCLUSIONS

The goal of the speech demonstration was met. The system demonstrated that today's speech recognition systems, in particular TI's LR2000 connected word recognition, are capable of being used in a complex workstation environment. The system was demonstrated with speech by the 8 controllers for three days approximately 6 hours each day. During that time, the controllers were able to demonstrate the operation of the ATC console using all the various interface technologies, including voice. The overall impression of both the viewing audience and the controllers using the system was that a speech recognition system providing connected speech recognition can be a useful part of an improved man-machine interface for advanced ATC work-station.

While this demonstration was a success, there are many areas that must be considered by a system designer before a speech interface is actually implemented. Most of today's speech systems are speaker-dependent, and thus require user enrollment and training. Many of the systems available do not have connected speech capabilities, they require a pause to be inserted between each word. TI's LR2000 connected word recognition system is an exception. Most systems also have limitations on the number of words which can be recognized at one time. In addition to these concerns, the system designer also must look at the recognition accuracy which is required and how recognition error recovery will be handled.

This is not to say that speech systems do not have a role to play. Recognition provides an excellent interface for tasks where an operator's hands and eyes are busy. In addition, in systems where the operator is required to manage a large variety and amount of data, the addition of speech as an alternative input device may provide an improved man-machine interface.

Training systems utilizing recognition could provide high quality, lower cost operator training where the recognizer would be used to determine the correctness of communication between an operator and other people. An example might be to use recognition to mimic the role of the pilot in aircraft under the control of an air traffic controller. Other speech technologies can also be used to provide more effective workstations. Speech output can provide audible warning or help messages. Speaker verification can be used to ensure that only authorized personnel log into a workstation.

In conclusion, today's speech systems can provide for a more effective man-machine interface in the complex workstations required to manage complex tasks.

ACKNOWLEDGEMENTS

The author would like to acknowledge the efforts of the many individuals at Hughes Aircraft Co. who contributed their time and energies to making the speech portion of the demonstration a success. In particular, he would like to recognize Augie Beining, for his part in having the speech interface included in the first place. In addition, the author thanks the controllers who worked to learn the system and did such a fine job in demonstrating it. In particular, he would like to recognize Larry Suppan who was responsible for the enrollment and training of the rest of the controllers.

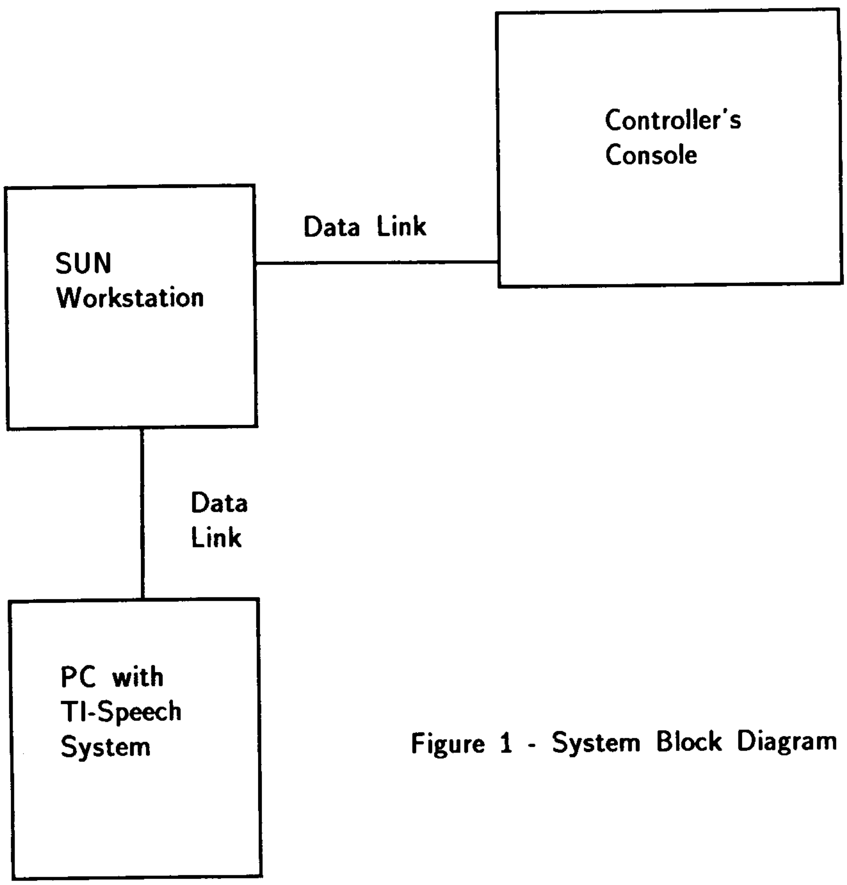


Figure 1 - System Block Diagram

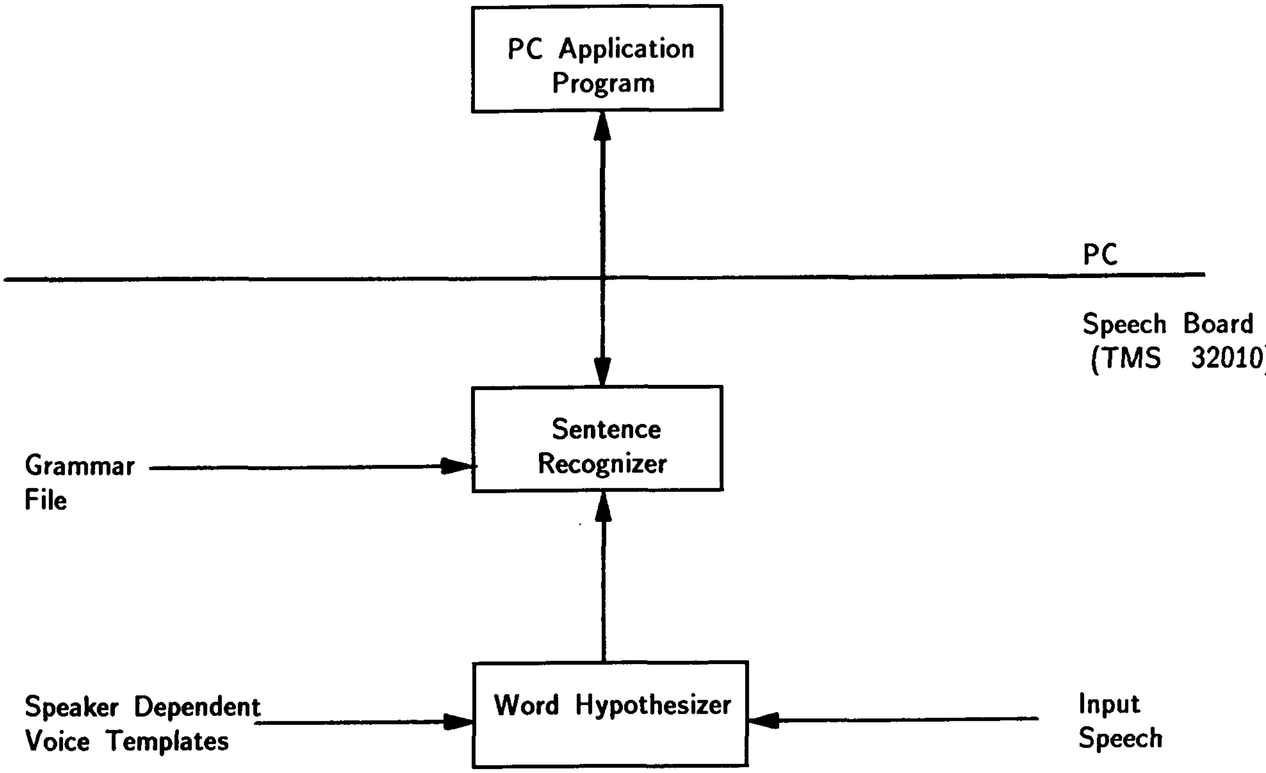
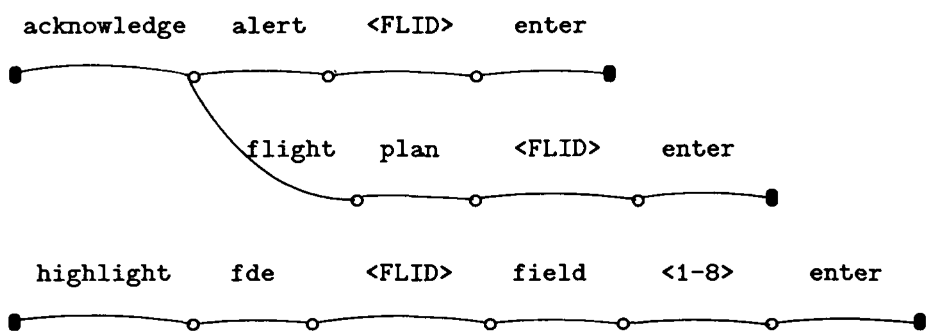


Figure 2 - LR 2000 Block Diagram

ZERO	OH	ONE	TWO
THREE	FOUR	FIVE	SIX
SEVEN	EIGHT	NINE	NINER
TEN	"EELEVEN"	"UHLEVEN"	THIRTEEN
FIFTEEN	SEVENTEEN	EIGHTEEN	TWENTY
THIRTY	FORTY	FIFTY	ENTER
"ENNER"	THOUSAND	AIR FORCE	AMERICAN
DELTA	UNITED	PSA	NOVEMBER
ROMEO	SNOW	NAVY	MIKE
WHISKEY	DROP	TRACK	FLIGHT
PLAN	HISTORIES	SITUATION	INSET
RANGE	MARKS	START	VELOCITY
VECTOR	MILE	HIGHLIGHT	FDE
ACKNOWLEDGE	MOVE	ALERT	QUICK
LOOK	DATA	BLOCK	HANDOFF
ACCEPT	FIELD	TYPE	INITIATE
SECTOR	CANCEL	POINT	OUT
ASSIGN	ALTITUDE	REPORTED	INTERIM
BEACON	CODE	PREFERENTIAL	ROUTE
CHANGE	FREQUENCY	LEVEL	CESSNA
SPEED	FIX	VENTURA	TIME
ESTIMATED	DIRECT	SANTA	BARBARA
PALMDALE	CORRECTION	NORTH	SOUTH
EAST	WEST		

TABLE 1 - VOCABULARY LIST



(continued for remainder of grammar)

Example Sentences: "Acknowledge Alert American Ten Enter"

"Highlight FDE November 9871 Delta Field 3 Enter"

FIGURE 3 - PORTION OF APPLICATION GRAMMAR