International Journal of Mini and Microcomputers, Vol. 10, No. 2, 1988

A DISTRIBUTED COMPUTER SYSTEM FOR THE DEVELOPMENT OF VOICE INTERACTIVE APPLICATIONS

Claude Y. Laporte* and Jean-Louis Houle**

ABSTRACT

This paper describes the architecture of a distributed computer system designed for the development of voice interactive applications. The system is composed of two main elements: a microcomputer and a front-end processor. The front-end processor is a speech transaction processor used as a man-machine interface. It is designed to receive verbal commands and deliver aural information and feedback to the operator of the system.

Key Words

Microprocessor, voice command, distributed processing, development system

INTRODUCTION

Speech technology is presently mature enough to start designing innovative applications. But since it involves man-machine dialogues, it needs fine tuning before a system is left to an operator. In this paper, we describe a development system that has been designed to properly evaluate, develop, and validate such man-machine interfaces. We describe the architecture of a front-end processor and its associated firmware. We then describe its applications to radio frequency selection and robot controlled machining tasks.

1. THE FRONT-END SPEECH PROCESSOR

The speech processor designed around the widely used STD bus is composed of the following modules: a microcomputer board, a speech synthesis board, and a speaker-dependent speech recognition board. Since the hardware of these modules has been presented in previous publications [2, 3], we will only briefly describe the hardware of each module and will present, in more detail, the software that has since been developed.

1.1. The Speech Synthesizer Modules

Four speech synthesizer modules were designed and assembled. The first module uses the pulse code modulation (PCM) technique. It is designed in such a way that speech can be digitized and synthesized at three different sampling frequencies: 4, 6, and 8 kHz. The second module uses a continuously variable slope delta (CVSD) modulation circuit—the HC-55564 manufactured by Harris semiconductor. The module can encode speech at different bit rates since it is equipped with an on-board clock sampling circuit. A third speech module uses the linear predictive coefficient (LPC) speech processor TMS5220 manufactured by Texas Instrument. The bit rate is approximatively 1200 bits per second. Finally, the speech module uses the SSI263 phonetic speech synthesizer circuit, manufactured by Silicon Systems, contained in a single 24-pin CMOS integrated circuit. It can synthesize speech in French and German as well as in English. The bit rate is approximatively 100 bits per second. The speech modules have two common features. First, each speech synthesizer circuit is interfaced with the host CPU MC6809 through an identical circuit: a MC6821 peripheral interface adapter (PIA) chip. Second, all speech modules have two outputs: a low level output that can drive an amplifier and a high level output that can deliver 1 W to an 8-ohm speaker, through an LM-386 chip.

1.2. The Speech Recognition Module

The speech recognition module is a speaker-dependent device built around the voice recognition chip set, VRC100-2, manufactured by Interstate Electronics Corporation. The chip set consists of a 16-channel audio spectrum analyzer, the ASA-16, and a 4K-bytes EPROM memory circuit containing the processing algorithms. The system has been designed to recognize up to 200 words. Each word must be neither shorter than 80 milliseconds nor longer than 1.25 second. A user must pause approximately 160 milliseconds between words to allow the system to recognize word boundaries. The response time is a function of the vocabulary size, since each word pronounced must be compared to every template stored in a RAM. As an example, the response time, with a vocabulary of 50 words, is 100 milliseconds. Figure 1 shows the block diagram of the speech recognition module: the circuit has been assembled on two STD bus boards. An 8-bit MC6803 microprocessor operating at a frequency of 2 MHz performs the operations related to the training of the system or the recognition of words.

1.2.1. Description of the development setup

We describe in this section the method followed to render the modules operational. Once the boards were wire-wrapped, we verified all the connections and installed the chips. As expected, the system did not work completely the first time. Since the programs stored on the EPROM are proprietary, we did not have access to the listing of the different routines. It was therefore not possible to find the problems with the boards configured as a speech recognition system. We then improvised a debugging tool. Since the instruction set of the MC6803 processor is an enhanced instruction set of the MC6800 microprocessor, we used a MC6800 microcomputer as a development system. We developed programs on the MC6800 microcomputer and downloaded them, through the PIA of the speech recognition system, to the MC6803 microprocessor. Figure 2 illustrates the setup. The MC6800 microcomputer was fitted with a parallel interface board which uses a PIA as a parallel interface chip. On one end, a 40-pin clip is mounted on top of the PIA, on the other end of a 40-pin cable is attached a 40-pin dual-in-line connector. This connector is

*Physics Department, College Militaire Royal de Saint-Jean, Québec, Canada, J0J 1R0.
**Electrical Engineering Department, Ecole Polytechnique de Montréal, Montréal, Québec, Canada H3C 3A7.

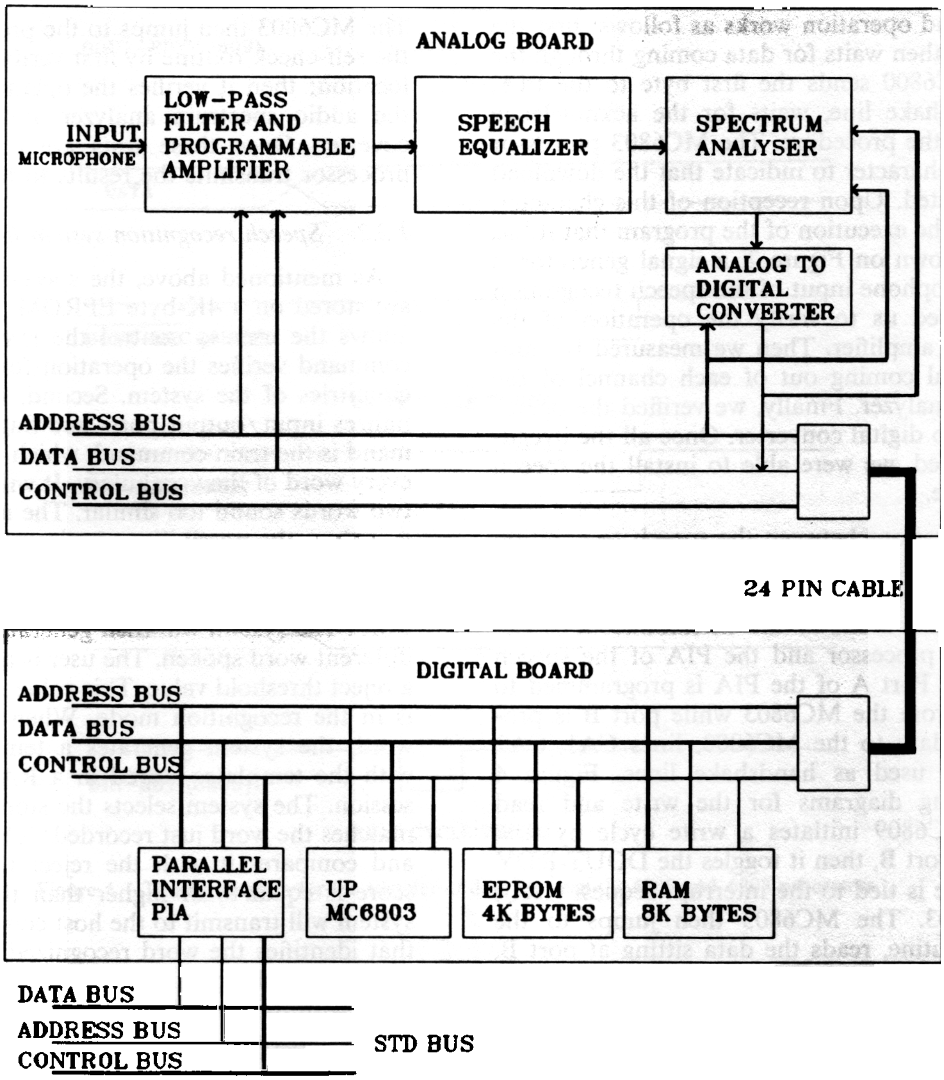

Figure Speech recognition block diagram

inserted into the socket of the PIA of the speech recognition system. The first program was written to exercise the entire address range of the MC6803. This allowed us to verify the proper operation of the address bus, the read/write, reset, and clock lines. We were also able to check all the address decoding logic circuits. Then we wrote a routine to test the RAM chips and to verify the operation of the PIA used to communicate with the microcomputer module of the STD bus. Satisfied that all the digital circuits were operational, we wrote a loader, for the MC6803 processor, to download routines from the MC6800 microcomputer to test the operation of the analog cir-

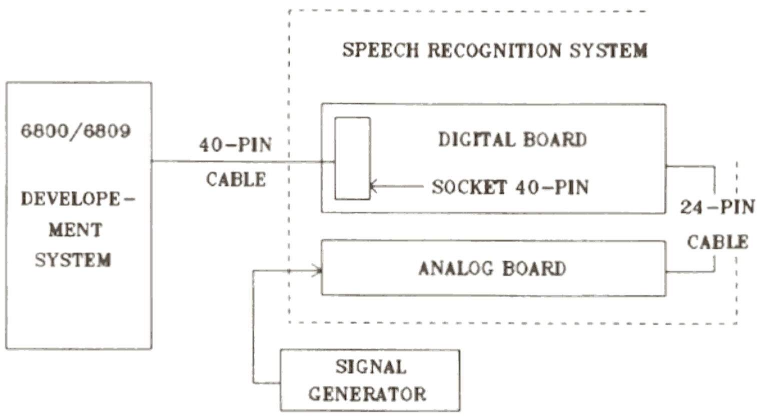

Figure 2. Interconnections between the 6800 microcomputer and the speech recognition system.

cuitry. The download operation works as follows: first the MC6803 is reset; it then waits for data coming through the PIA. Next, the MC6800 sends the first byte to the PIA, activates the handshake line, waits for the acknowledge signal, and repeats the procedure. The MC6803 processor looks for a unique character to indicate that the download operation is completed. Upon reception of this character, the MC6803 starts the execution of the program that it has just received. As shown on Figure 2, a signal generator is plugged to the microphone input of the speech recognition module. This allowed us to verify the operation of the programmable gain amplifier. Then we measured the amplitude of the signal coming out of each channel of the ASA-16 spectrum analyzer. Finally, we verified the operation of the analog to digital converter. Once all the irregularities were corrected, we were able to install the speech recognition firmware.

All the communications between the speech recognition module and the microcomputer module are done using a parallel interface circuit MC6821 peripheral interface adapter (PIA). Figure 3 shows the interconnections between the MC6809 processor and the PIA of the speech recognition module. Port A of the PIA is programmed to read data coming from the MC6803 while port B is programmed to write data to the MC6803, lines CA1/CA2 and CB1/CB2 are used as handshake lines. Figure 4 illustrates the timing diagrams for the write and read operations. The MC6809 initiates a write cycle by first storing its data on port B, then it toggles the DOUT-RDY line (CB2). This line is tied to the interrupt request (IRQ) pin of the MC6803. The MC6803 then jumps to the interrupt service routine, reads the data sitting at port B, and toggles the DOUT-REC line (CB1) of the PIA to signify that the data transfer is completed. A similar sequence of operations is done when the MC6803 processor wants to transmit data to the MC6809 processor. As an example, when the MC6809 processor wants to check the working condition of the speech recognition modules, it transmits the command code 11 to the MC6803 processor. The MC6803 then jumps to the proper program and starts the self-check routine by first verifying every RAM storage location; then it verifies the operation of each channel of the audio spectrum analyzer and the analog to digital coverter. Once these tests are completed, the MC6803 processor transmits the results to the MC6809 processor.

1.2.2. *Speech recognition system firmware*

As mentioned above, the speech recognition algorithms are stored on a 4K-byte EPROM. A set of 16 commands allows the user to control the system. First, the self-test command verifies the operation for the analog and digital circuitries of the system. Second, the reset command initializes input/output ports and flags. Third, a major command is the train command, which creates the templates of every word of the vocabulary. It will also inform the user if two words sound too similar. The user can use this feature to reduce the possibility of misrecognition. When the user enters the training mode, the system asks how many words he wants to record and how many times he will repeat each word. The system will then generate the template for every different word spoken. The user is also requested to specify a reject threshold value. This value is used when the system is in the recognition mode. When the user pronounces a word, the system generates a template and compares it with the templates stored in a RAM during the training session. The system selects the stored template which best matches the word just recorded. It computes a match score and compares it with the reject threshold. If the match score is equal to or higher than the reject threshold, the system will transmit to the host computer an index number that identifies the word recognized. Otherwise, the system will inform the host computer that it did not recognize a word with the precision requested. If the reject threshold is set too low, the system may make mistakes (misrecognition). Inversely, if the threshold is set too high, the system may not be able to identify a word (rejection). In the latter case, the user is requested to repeat the word. The system also reports the index number of the word which achieved

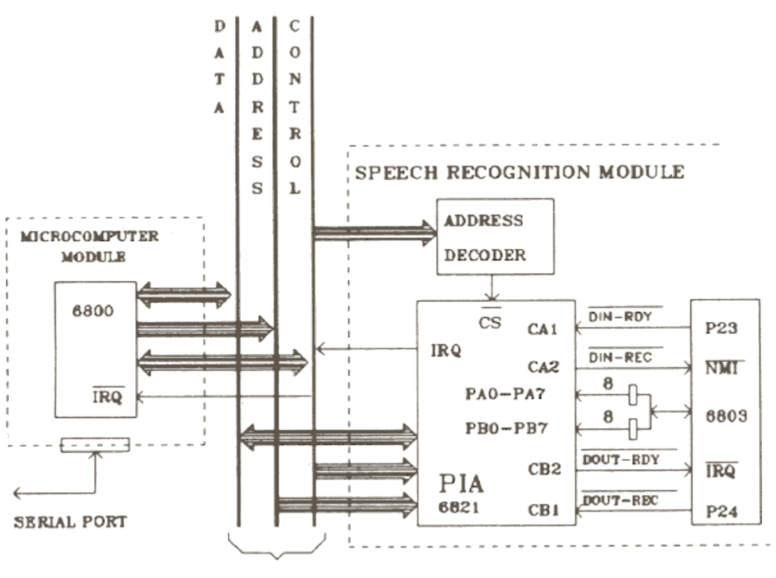

Figure 3. Interconnections between the 6800 processor and the 6821 PIA of the speech recognition module

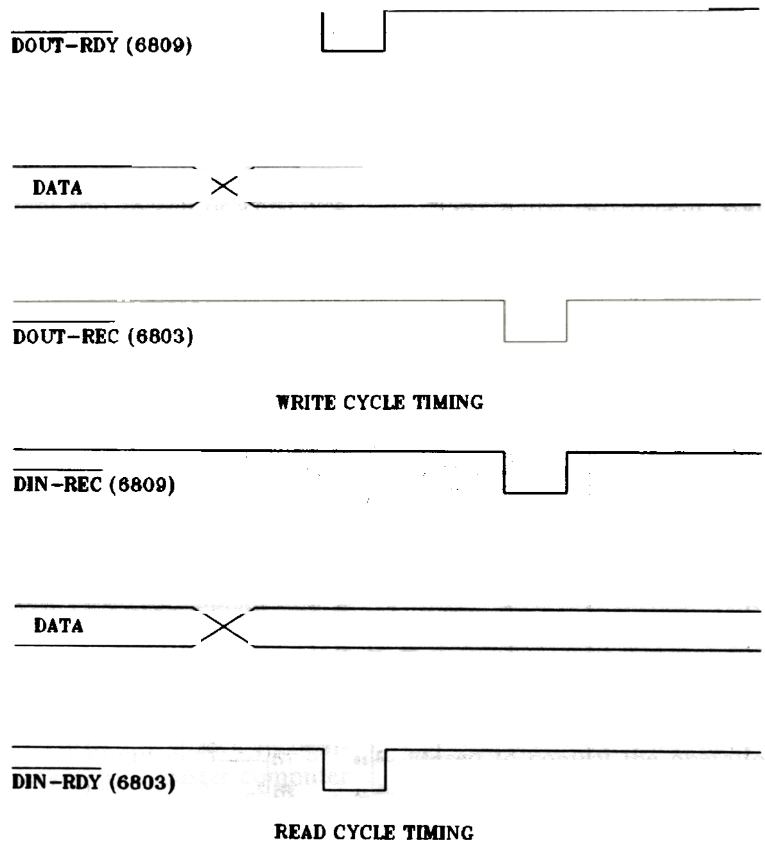

Figure 4. Timing diagram for the transfer of data between the 6809 and 6803 processors.

the next best matching score, as well as giving its matching score. This way, a user is in a position to identify words that may be misrecognized. The user may then modify his vocabulary update command, mostly used when the system has problems identifying a particular word. The user may then update the template. As an example, the user could have been nervous when he trained the system for the first time. As he gets more comfortable with this new interface, his voice changes. Using the update command, he will adjust the template to the way he is now pronouncing. The set gain command allows the user to compensate for the positioning of the microphone and for variation in the loudness of the voice from one speaker to another. Since the templates are stored in RAM, the user may want to save them before turning off the system. The upload command transmits to a host computer the templates of the vocabulary. Inversely, the download command transmits to the speech recognition system the template requested. Using these two commands many subjects can use the speech recognition system. The other commands allow the user to read and write parameters like the reject threshold or the value of flags.

Before we started developing any application program for the speech recognition system, we decided to write a monitor that allowed us to familiarize ourselves with this new man-machine interface. The monitor program was written on an MC6809 development system and downloaded on an EPROM. This chip was inserted on the microcomputer module (see Section 2.3). The monitor allows the subjects to use any of the 16 commands, through a CRT terminal and create, update, or modify a vocabulary. The monitor was very useful; it allowed us to discover problems that would have been hard to detect once an application program was written. The problem was that the rejection rate was too high, because new users were tense when they first trained the system—people are not used to talking to a machine. As the subjects become familiar with the system, the error rate decreased substantially.

For our purposes, we have used only two RAM circuits, out of a maximum of 8, to store the reference templates. This means that we can accommodate up to 50 words of vocabulary. This was thought to be quite acceptable for our applications.

1.3. The Microcomputer Module

The microcomputer module uses an 8-bit MC6809 CPU developed by Motorola. The MC6809 was selected because of its powerful instruction set and also because of previous in-house experience with Motorola's CPU and interface circuits. The board holds a 24-pin socket for a 2k-bytes scratchpad RAM and another 24-pin socket for a 2k-bytes EPROM. A 6850 asynchronous communications interface adapter (ACIA) has been used to communicate between the MC6809 and the outside world. A 555 chip, wired as an astable multivibrator, has been used as a baud rate generator. Another 555, wired as a monostable multivibrator, has been used as a reset circuit for the processor. Figure 5 shows the circuit diagram of the microcomputer board.

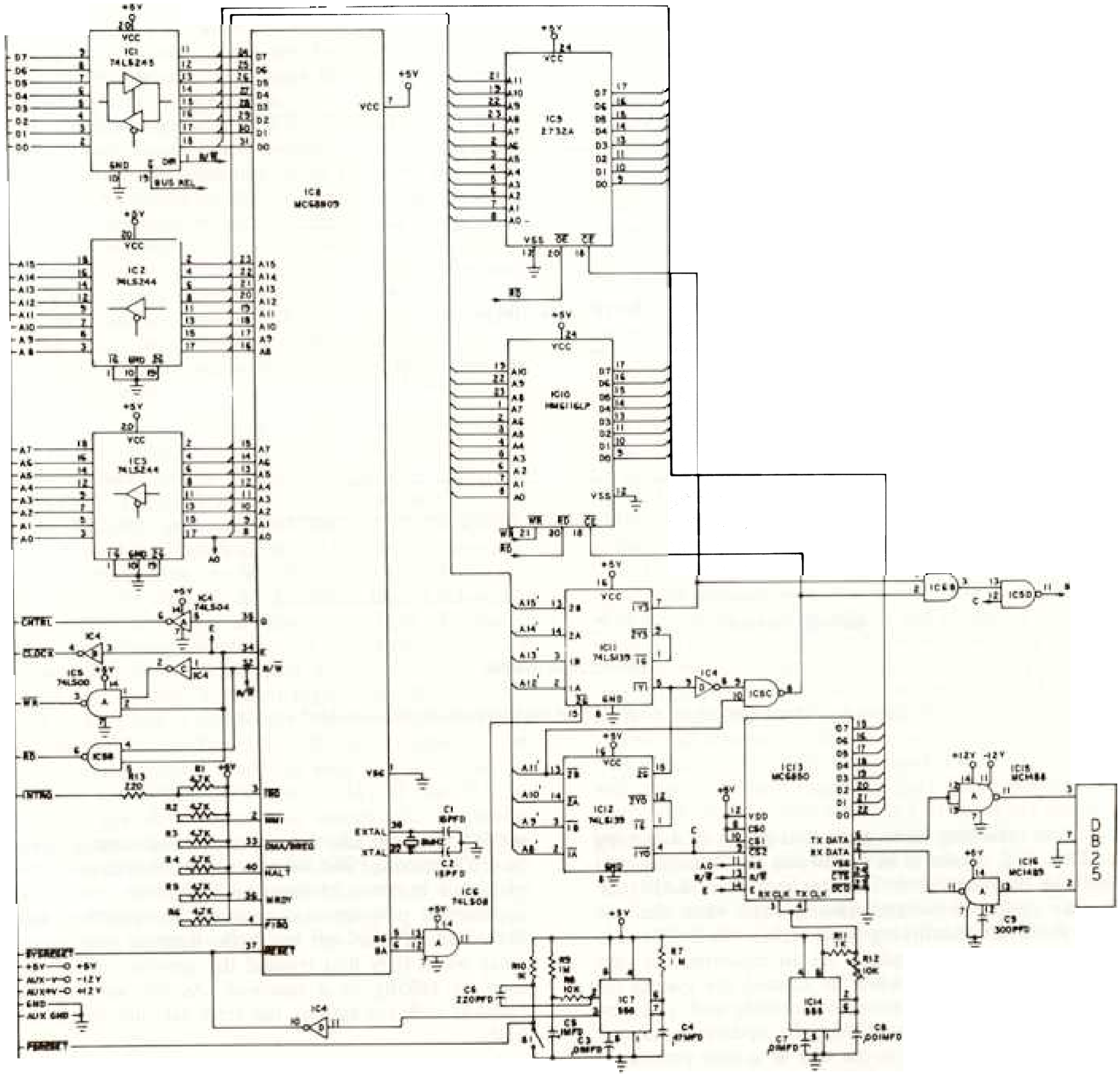

Figure 5. Circuit diagram of the microcomputer board.

The microcomputer module can be used together with the speech recognition system as a training aid in order to familiarize users with this new man-machine interface. We have developed a monitor program that allows a user to enter in the training and recognition modes of operation as described in Section 2.2. When the modules (speech recognition, speech synthesis, microcomputer) are configured as a front-end speech processor, the microcomputer board is fitted with another EPROM monitor. This monitor allows the master computer to send commands to the speech processor. The speech processor will then execute the command and return a date to the master processor, if requested. Since the user has a choice of four different speech synthesizer modules, he must, upon system initialization, inform the master processor. During the setup phase, the master computer will download, to the MC6809 module, the software routines for the configuration selected.

2. DESCRIPTION OF THE MASTER COMPUTER

We have used, as a master computer, an 8-bit microcomputer manufactured by Southwest Technical Products. Its processor is an MC6809 chip developed by Motorola. The Southwest PC was configured with 56K bytes of RAM although up to 768K bytes are addressable. The main bus is a 50-pin bus while the input/output bus has 30 pins. We used floppy disks as secondary storage but hard disks are also available. We also used off-the-shelf parallel and serial interface modules. Unfortunately the 30-pin input/output bus cannot accommodate direct memory access operations. The operating system, Flex, written by Technical Consultant System, is easy to use and has all the utilities to develop application programs. The user may also select other operating systems like Uniflex or OS-9 from Microware Inc.

A library of utilities was written in Assembler and in Basic. These programs allow the master computer to control all the functions of the front-end speech processor. A user can easily and rapidly develop an application and test it.

3. APPLICATION OF THE SYSTEM TO RADIO FREQUENCY SELECTION

The speech evaluation system is used in a project to control radio frequency selection of an aircraft. Subjects were asked to track a moving target on a CRT screen using a joystick, while they were directed to make radio frequency selections. A test bench was assembled to partially reproduce the aircraft environment. It was divided into two parts: the front-end speech processor system and a master computer. Figure 6 illustrates the block-diagram of the setup. On the right is the speech processor system with VHF and UHF radio controllers connected to an interface module. Also shown are a headset and microphone connected to the speech synthesizer and the speech recognition modules. This system is linked to a master computer through a serial interface port on the microcomputer module. On the left side of Figure 6 is the master computer: It is the 8-bit microcomputer described in Section 3. A parallel interface module is used to measure the pilot stress level and the position of the stick. Finally, a high resolution graphic screen (1024 by 1024 pixels) is used to display a moving target and the position of the tracking joystick.

Once the system is powered, a menu is displayed to the operator. He may select one of the two familiarization programs. The first one asks the user to track a moving target, using a joystick, while he is verbally directed, through a speech synthesizer module, to manually change frequencies of radio sets. If the operator selects the second program, he is first asked to train the speech recognition system and then perform a series of voice activated frequency selections.

As a first experiment, volunteers were asked to perform a series of ten manual frequency selections. The master computer, through the speech synthesizer module, instructed the subject to manually select a frequency. At the same time, the subject was tracking a moving target with a joystick. The master computer updated the position of the target, measured the position of the joystick, and the value of the relative stress level of the subject. During a second experiment, the subjects were instructed to perform a series of ten frequency selections using the speech processor system. Results show that tracking a target while performing radio frequency selections, using a voice interactive system, is more precise than the manual frequency selection method, while stress level is not significantly changed.

4. APPLICATION IN A MACHINING ENVIRONMENT

In this application, the development system is being used to control the operations of scaled-down machining process. Figure 7 shows a block-diagram of the setup. Figure 8 is a photograph that shows the equipment used. Briefly, parts are brought on a conveyor belt, manufactured by Rhino Robots Inc. [8], to a vision system, manufactured by Micromint Inc. [9]. The task of this system is to identify the object. A computer-controlled robot arm, manufactured by Microbot Inc. [7], is then instructed to pick up the part and place it on the pneumatic clamp of a computer-controlled milling machine, manufactured by Dyna Electronics Inc. [6]. The milling machine will then perform a series of operations for this particular part.

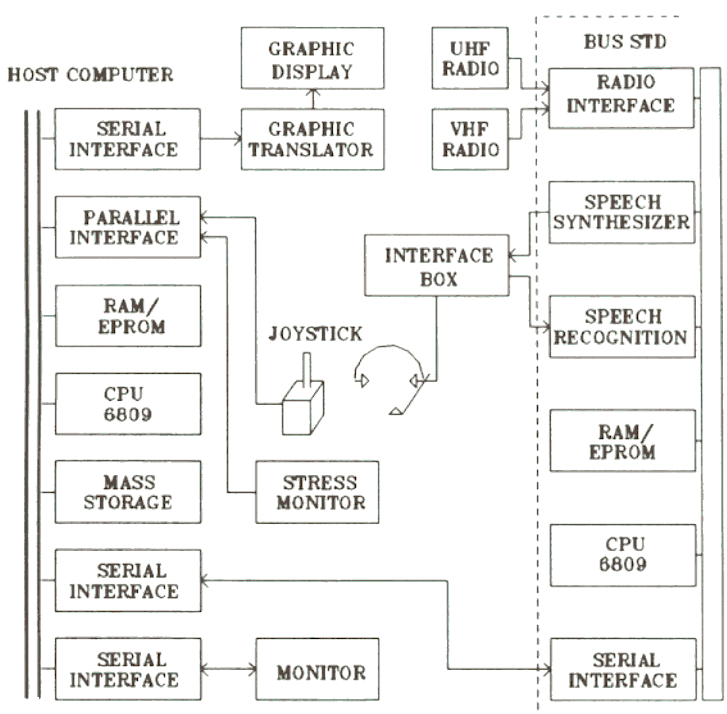

Figure 6. Block-diagram of the test bench.

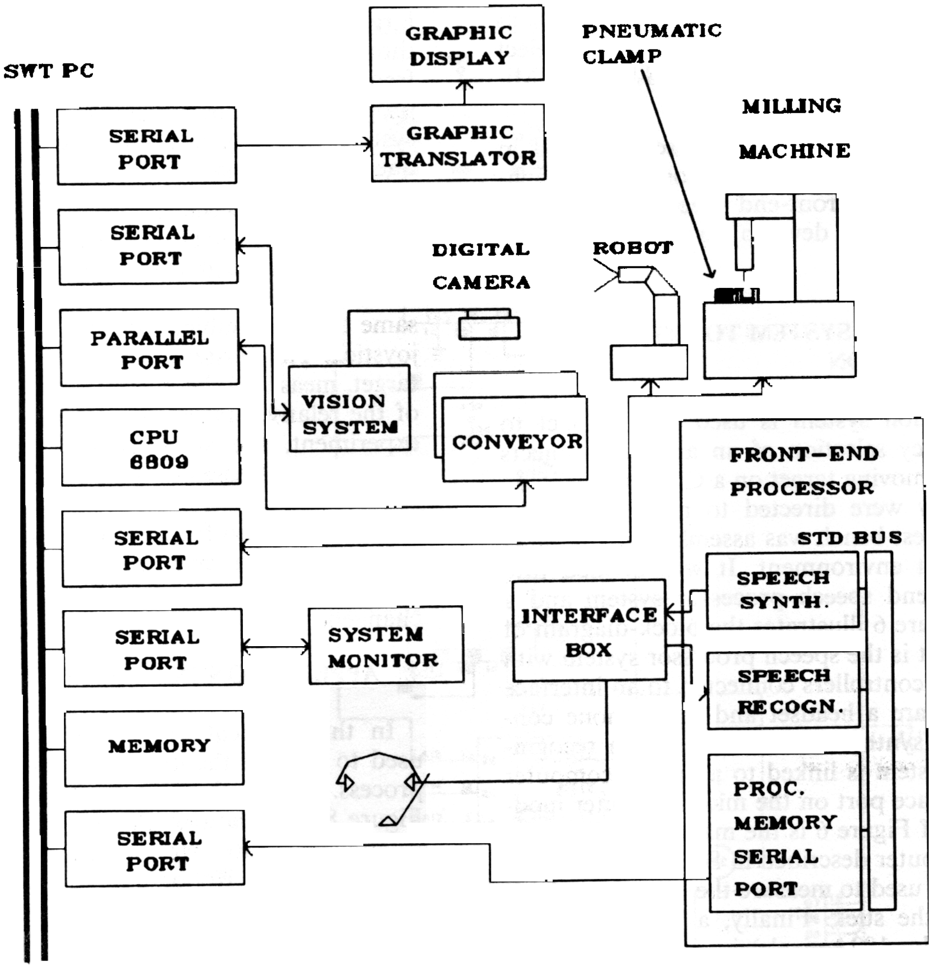

Figure 7. Block diagram of the machining application.

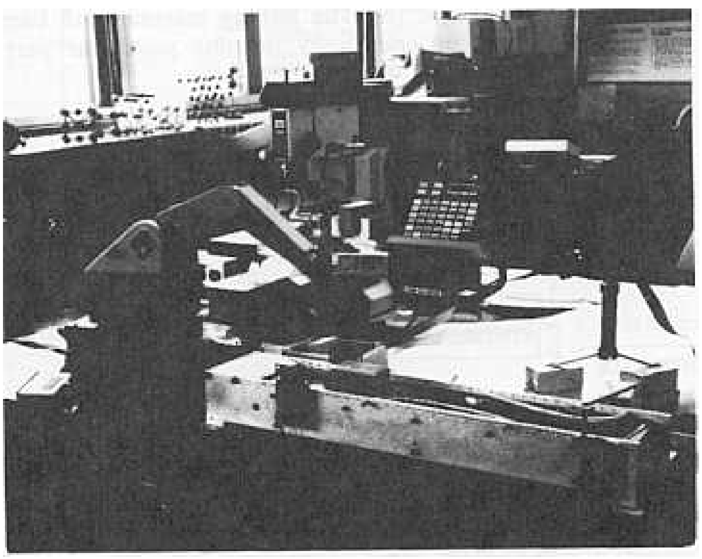

Figure 8. Vision System and other equipment

Once completed, the robot arm will pick up the part and store it in a bin. The operator has full control of the operations and uses the speech system to perform his task.

5. CONCLUSIONS

This system was used in a feasibility study to control radio frequency selection in a military aircraft. Results showed that tracking a target while performing radio frequency selection using a voice interactive system was more precise than the manual frequency selection method. The voice interactive system is presently being used in a project to develop an industrial application.

REFERENCES

[1] Harvey, J.H., "Single-board Computers Boost System Throughput," *Computer Design*, Nov. 15, 1985, pp. 45–59.

[2] Laporte, C.Y., A Description of the Architecture of a Voice I/O Evaluation System and its Application in a Military Jet Aircraft, Proc. Speech Tech'85, Media Dimension, New York, pp. 313-316.

[3] Laporte, C.Y., and Houle, J.L., "Design of a Voice Interactive System for Command Functions in Military Aircrafts," Proc. of IASTED Int. Symposium: Applied Informatics, Innsbruck, Austria, Feb. 18–20, 1986, pp. 23–28.

[4] Wiggins, R., and Brantingham, L., "Three-Chip System Synthesizes Human Speech," *Electronics*, Vol. 51, No. 18, pp. 109–116.

[5] Ciarca S., "Build a Third-Generation Phonetic Speech Synthesizer," *Byte*, Vol. 9 No. 3, 1984, pp. 28–42.

[6] Dyna Electronics, Inc., 2346 Walsh Ave., Santa Clara, California 95051.

[7] Microbot Inc., 453-H Ravendale Drive, Mountain View, California 94-43.

[8] Rhino Robots Inc., 3402 North Mattis Ave., P.O. Box 4010, Champaign, Illinois 61820.

[9] Micromint Inc., 561 Willow Ave., Cedarhurst, New York.