Microcomputer Applications, Vol. 8, No. 1, 1989

A VOICE–INTERACTIVE EVALUATION SYSTEM FOR COMMAND FUNCTIONS IN MILITARY AIRCRAFT

Major Claude Y. Laporte(*), Jean-Louis Houle(**)

Abstract

This paper describes the architecture of a voice–interactive evaluation system designed around the STD bus. The evaluation system includes a microcomputer module, a speech synthesizer module, a speech recognition module, and an application or interface module. The voice–interactive system is used, under the control of a host computer, in a laboratory experiment to simulate the frequency selection of VHF and UHF radios in a military aircraft. The results show that tracking a target while performing radio–frequency selection using a voice–interactive system is more precise than the manual frequency–selection method.

Keywords:

Microprocessor, voice command, speech synthesis, ergonomy, instrumentation.

1. INTRODUCTION

Flying a high–performance aircraft is now more demanding than ever. In the past, high–performance aircrafts were staffed by a pilot and an assistant responsible for the navigation and communication functions. Today, the pilot is left alone to perform all these tasks. In addition, modern aircraft is complicated to fly, and the pilot is threatened by high performance detection systems (radars, satellites) and sophisticated anti-aircraft weapons. Therefore, we propose a way of helping the pilot by allowing him to command, using a voice–interactive system, some of the instruments in his aircraft. The prototype system controls the frequency selection of UHF and VHF radios of an aircraft.

Three qualities are desirable in an evaluation system: it should be flexible, easy to use, and inexpensive. Ideally, it should also be small and easily transportable. The proposed voice–interactive evaluation system is designed around a bus widely used in industry, the STD bus. This bus was selected because it can accommodate most 8-bit microprocessors, its boards have small physical dimensions, and there is a wide diversity of functional boards available.

(*) Assistant Professor, College Militaire Royal de Saint-Jean, Saint-Jean-sur-Richelieu, Canada

(**) Professor, Ecole Polytechnique de Montréal, Montréal, Canada

2. THE STD BUS

Briefly, the STD bus [3] can accommodate most 8–bit microprocessors, and even an 8–bit version of the more powerful 16–bit microprocessors like the Motorola 68008 and the Intel 8088. The STD bus is presently being redefined to handle 16-bit CPU's. CMOS boards are now available and, with the advent of 64HCXXX chips, the temperature range will be extended from 0° to +55° Celsius to −40° to +85° Celsius. The 56-pin bus is divided into five groups: 8 pins for the data bus, 16 pins for the address bus, 22 pins for the control bus, 6 pins for the logic power bus, and 4 pins for the auxiliary power bus. The power bus can accommodate both digital and analog power distributions. All digital signals on the busses are TTL compatible. Physically, the boards are small: 16.5 cm by 11.4 cm (6.5″ by 4.5″). In addition to the wide variety of CPU boards, there are many support modules available: RAM, EPROM, EEPROM, bubble memory, analog–to–digital (A/D) converters and digital–to–analog (D/A) converters, digital input/output ports, serial communication interface, etc. The modules may be purchased separately or as a development system with floppy or hard disks and an operating system.

3. THE VOICE–INTERACTIVE SYSTEM

The voice–interactive system is composed of the following modules: a microcomputer board, speech synthesis boards, a speaker–dependent speech recognition board, and an application board. An additional memory board may be added if necessary. Fig. 1 illustrates the architecture of the voice–interactive evaluation system.

3.1 The Microcomputer Module

The microcomputer module uses an 8-bit MC6809 developed by Motorola. The 6809 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 also holds a 24–pin socket that may be used to accommodate a 2K–byte scratchpad RAM, and another 24–pin socket is available to accommodate a 2K byte EPROM. This EPROM may be used to store the application program. The serial interface may be used to communicate with a host computer, a terminal, or a printer. During the development phase, a host computer

may be used to communicate with the STD bus modules. During field applications, a terminal may be used as a device to monitor the experiment.

3.2 The Speech Synthesizer Modules

In order to evaluate different speech synthesis techniques, four modules were designed and assembled. The first module used the Pulse Code Modulation (PCM) technique. The module is designed such that speech may be digitized and synthesized at three different sampling frequencies: 4, 6 and 8 KHz. The module is built around an analog–to–digital converter (A/D) used to digitize the speech signal and a digital–to–analog converter (D/A) to restitute the speech signal. A pair of low–pass filters (MC145414) are used to filter undesirable noise. 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. It is also equipped with a pair of low–pass filters identical to those used in the PCM module. A third speech module, illustrated in Fig. 2, uses the Linear of Predictive Coefficient (LPC) speech processor TMS5220 manufactured by Texas Instruments. This module can work in two operating modes. In the first mode, the TMS5220 speech processor can synthesize speech by reading the coefficients stored in a 128 Kbit–ROM designed by Texas Instruments, the Voice Synthesis Memory TMS6100. The TMS5220 Speech Processor may directly access up to 16 TMS6100 circuits with no external hardware required. Such an arrangement would provide about 30 minutes of speech. The second operating mode allows the host processor to fetch from its own memory the vocabulary to be synthesized. The second approach is inconvenient in that the host CPU must service the speech processor by performing a read and store series of the LPC coefficients at a rate of approximately 1200 bits per second. Conversely, in the first approach, the host is interrupted only when the speech processor has completed the synthesis of a work or frame. Finally, the last speech module uses the SSI263 phonetic speech synthesizer circuit, manufactured by Silicon Systems, contained in a single 24–pin CMOS integrated circuit.

The system may be operated in two modes. At a low bit rate, the speech quality is comparable to the speech produced by the Votrax SC–01A chip. In this operating mode, the host processor only transmits the codes for the individual phonemes to be synthesized. This explains a low bit rate of approximately 70 bits per second. In the second operating mode, it is possible to dynamically program the contents of the registers of the SSI263. The circuit has a set of five 8–bit registers that allow software control of speech rate, pitch movement rate, amplitude, articulation index, vocal tract filter response, and phoneme selection and duration. In the second operating mode, the bit rate can be approximately 400 bits per second. Another feature of the SSI263, which the SC01 did not have, is an additional set of phonemes that allows the SSI263 to synthesize speech not only in English, but in French and German also.

The speech modules have two features in common. First, each speech synthesizer circuit is interfaced with the host CPU (MC6809) through an identical circuit: a Peripheral Interface Adapter 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, using an LM–386 chip, on watt to an 8–ohm speaker.

3.3 The Speech Recognition Module

The speech recognition module is 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 an EPROM memory circuit containing the processing algorithms. Fig. 3 illustrates the block diagram of the speech recognition module: the circuit has been assembled on two STD bus boards, one board for the analog circuitry and one board for the digital circuitry. As shown in Fig. 3, the analog signal coming from a microphone is first amplified by a programmable amplifier, then a speech equalizer compensates for the roll–off. The signal is then fed to a 16–band audio spectrum analyzer, the ASA–16 developed by Interstate. An 8–bit analog–to–digital converter digitizes the signal coming from the ASA–16 analyzer. All these operations are performed under the control of an 8–bit microprocessor, the MC6803, operating at a frequency of 2 MHz. Once the signal has been stored in RAM, the MC6803 microprocessor performs the operations related to the training or the recognition of words. All the algorithms for performing these operations are stored in a 4–Kbyte EPROM. The firmware is composed of a set of 16 commands: a self–test command checks the RAM circuits and the audio spectrum analyzer. Of course, if this test is successful, it means that most of the digital circuits are operational (CPU, address decoders, EPROM, etc.). The "train" command creates a template of 67 bytes per word of vocabulary, the "recognize" command puts the system in an active mode where each spoken utterance will be given an identification number and a hit score. Finally, a set of utility commands allows the designer to modify parameters such as the input amplifier gain and the reject threshold. The templates generated during a training session may also be saved and retrieved using the upload/download commands.

All communications between the speech recognition module and the STD microprocessor MC6809 are carried out using a parallel interface circuit developed by Motorola: the MC6821 Peripheral Interface Adapter (PIA). The communication protocol is simple: each time a processor wants to talk to another processor, an interruption is generated. For example, when the host processor (MC6809) wants the speech recognition to perform a self–test, it interrupts the MC6803 processor, through the PIA, then sends the appropriate data to initiate a self–test command. Once the MC6803 has completed the self–test sequence, it interrupts the MC6809 and reports the status of the speech circuitry.

When using a speaker–dependent system, one has to remember that the system is trained to recognize the voice

of one user at a time. The templates for a user are stored on RAM on the speech–recognition module. This means that when power is removed, the templates are erased. One solution to this problem would be to store the reference templates of an individual on ROM right after a training session. However, if the evaluation system is to be used by many individuals this approach is not ideal. Another solution would be to use a storage medium that can be nonvolatile and re–programmable. In this way, the storage device may be used to store the templates of an individual during a first training session and could also be used when an individual wants to update selected words of the vocabulary. A suitable storage medium would be an electrically erasable programmable read–only memory (EEROM). A device recently introduced onto the market by Datakey Inc. (Burnsville, Minnesota) provides the ideal storage device for the speech–recognition module. An EEROM chip is encapsulated in thermoplastic material. It has the shape and the size of a key. This storage medium cannot be bent or cracked in normal everyday use. Fig. 4 shows the components of this storage medium: the Datakey storage device, the Keyceptacle used to provide the connection for the key, and the Keytroller used as a communication link between the key and a computer system. Fig. 5 illustrates how the Datakey storage device is interfaced with the voice–evaluation system. A serial interface chip from Motorola, the asynchronous communication interface adapter (ACIA), provides the communication link. Datakey has recently introduced keys that may be interfaced with the computer bus through a Keyceptacle, like a memory chip. The storage capacity of the devices ranges from 1.4 Kbits to 64 Kbits.

3.4 The Application Module

The speech–evaluation system is presently being used in a project to control radio–frequency selection aboard a military aircraft. Ideally, essential controls and displays are intelligently positioned. However, space limitations often make this impossible to achieve. This is particularly true in older aircraft like the CT–133 where additional avionics systems have been added to meet new requirements. In the CT–133 jet aircraft, radios are controlled through remote controllers. These controllers are located in the cockpit while the radio receivers and transmitters are located in the nose of the aircraft. One constraint of the project was that we were not allowed to make any modification, either to the radios or to the remote controllers. We had to tap the communication lines between the controller and the radio in order to get the data.

The frequency selection of the VHF radio set is done through a set of 2 to 5 encoders/decoders, and there is one encoder/decoder per decade. The data lines are normally floating at +20 volts. When a frequency is selected, 2 out of 5 lines are grounded in the remote controller. This data is detected at the input of the radio. We have used National Semiconductor 74C914 level translators to interface the 0/+20 volt data lines from the controllers to the TTL levels of the microcomputer interface chip. Once the data are TTL compatible, an MC6821 parallel interface chip from Motorola is used to read all the data lines coming from the remote controller.

The interface for the UHF radio set is different. In this case, data from the remote controller is transmitted serially to the radio using 9620 differential line drivers. The remote controller continuously transmits pulse trains of 32 bits together with a clock signal. At the input of the radio, the data bits are clocked in a register and decoded. To acquire the data, both data lines and clock lines have to be read. Fig. 6 illustrates the interface circuit for the data lines only. An identical circuit is needed for the clock signal. The signals coming from the controller are fed to F9621 differential line receivers. TTL signals from the F9620 are then read by a parallel interface chip. The program that reads the data is interrupt driven; on every active clock transition the CPU is interrupted and the logic value of the data line is read and stored.

4. DESCRIPTION OF THE TEST BENCH

The test bench was assembled in such a way as to partially reproduce the aircraft environment. It is divided into two parts. First, there is the voice–interactive evaluation system and, second, a host computer. Fig. 7 illustrates the block–diagram of the set–up. On the right is the voice–interactive system with its VHF and UHF remote controllers connected to the application module. Also shown are a headset and microphone hooked to the speech synthesizer and the speech–recognition modules. This system is linked to the host computer through a serial interface port on the microcomputer module. On the left side of Fig. 7 is the host computer, an 8–bit microcomputer manufactured by Southwest Technical Products Corporation. It is composed of a processor module, main and secondary memories, a system monitor, a serial interface port to control a graphics screen, and an interface module to measure pilot stress level and the position of the stick. Finally, a high resolution graphics screen (1024 x 1024 pixels) is used to display a moving target and the position of the tracking joystick. The stress monitoring instrument that was selected is a low–cost device that measures galvanic skin resistance (GSR). Skin resistance increases as a person becomes calm and relaxed, and, conversely skin resistance decreases as a person becomes tense. Fig. 8 shows the instrument, the GSR2 manufactured by Thought Technology of Montreal, and a set of remote electrodes. The instrument can be used without remote electrodes but, in our experiment, since both hands were busy, we had to monitor the relative stress level from the toes of the subject. The GSR2 provides, as an output signal, a tone that varies from 0 to 20,000 Hertz. The frequency of the tone is proportional to the relative stress level. This signal is accessible through a 2.5 mm jack. Fig. 9 shows the block diagram of the interface circuit for the stress instrument. The circuit is also used to measure the X and Y positions of the joystick used in the tracking experiment. The output signal of the GSR2 is fed to a signal conditioning circuit. This circuit delivers a voltage proportional to the relative stress level. This signal is sent to an analog switch and then to an analog–to–digital converter. The block diagram also shows a timer chip (MC6840). This circuit is used to generate an interrupt signal at a rate of 10 Hz. It

is at this frequency that the stress level and the X and Y positions of the joystick are measured. It is also at this rate that the position of the moving target is updated on the graphics screen.

5. DESCRIPTION OF THE EXPERIMENT

Before describing the experiment, we have to describe the layout in the instruments of the aircraft selected for this project. On the left side of the pilot is a lever used to control the speed of his engine. He controls the movements of the airplane with both his feet on the rudders and with his right hand on the stick. In this aircraft the UHF and VHF remote controllers are located on the right side of the cockpit. When the pilot wants to change radio frequency, he has to perform a sequence of movements. First, he has to move his left hand from the throttle to the stick, then he moves his right hand from the stick to one of the remote controllers, he selects the proper radio frequency, and he comes back to normal flying position.

In a first experiment, volunteers were asked to perform a series of ten frequency selections. The host computer, using the voice–interactive system speech synthesizer, instructed the subjects to manually select a frequency. At the same time, the subject was tracking a moving target with a joystick. The host computer updates the position of the target ten times per second, and also measures the position of the joystick and the value of the relative stress level of the subject at the same rate. During a second experiment, the subjects are instructed to perform the frequency selection tasks using the voice–interactive system.

6. RESULTS

Fig. 10 shows the measurements performed during a 45–second time frame in the first experiment. The continuous line illustrates the tracking error and the dotted line illustrates the relative stress level of the subject. It may be noted that the tracking error increases as soon as the subject performs a frequency–selection task. The duration of the task is indicated by a horizontal line above the time axis. The relative stress level also increases when the subject is asked to perform a new frequency selection. As soon as the frequency selection is completed, the tracking error and the relative stress level decrease. Fig. 11 shows the same measurements performed during a voice–interactive session. We may note that the tracking error is lower than in the first experiment and does not increase when the subject is asked to do a frequency selection. We may also note an increase in the relative stress level of the subject during a voice command operation. Table 1 shows the results of the measurements performed on three subjects.

7. CONCLUSION

The voice–interactive system provides the experimenter with a flexible and portable tool to evaluate potential voice applications. It has been used in a project to evaluate the feasibility of voice command for frequency selection. Results show that tracking a target while performing radio frequency selection using a voice–interactive system is more precise than the manual frequency–selection method

REFERENCES

[1] J.H. Harvey, "Single–Board Computers Boost System Throughput" *Computer Design*, Nov. 15, 1989, 45–59.

[2] R. Wiggins and L. Brantingham, "Three–Chip System Synthesizes Human Speech." *Electronics, 51*, (18): 109–116.

[3] STD BUS Technical Manual and Product Catalog, Pro–Log Corp., Feb. 1983.

[4] S. Ciarca, "Build a Third–Generation Phonetic Speech Synthesizer." *Byte, 9*, (3): 28–42.

[5] EDN Board Level Up System Directory, EDN, Apr. 19, 1984, 233–276.

Table 1

Tracking Errors and Stress Levels

	Manual Selection	Voice Command
Tracking errors:		
average	48	23
standard deviation	69	21
Relative stress level:		
average	617	639
standard deviation	67	40
Tracking errors:		
average	43	19
standard deviation	43	9
Relative stress level:		
average	529	510
standard deviation	23	25
Tracking errors:		
average	21	15
standard deviation	16	5
Relative stress level:		
average	755	776
standard deviation	53	52