This paper describes the design and development of a voice operated multilingual information display and retrieval system. The words or commands spoken by a speaker can be displayed on the multilingual terminal for verification and data entry applications. Previously stored information can also be accessed through a spoken word command and displayed on a CRT terminal.

The system consists of an isolated word recognition unit and a CRT terminal. The design aspect and the functioning of the system are described in detail. The code generated by the recognition unit corresponding to the recognized word is sent by the interface card and displayed on the terminal. The terminal uses a software called GIST and has the capability of displaying some Indian and Roman scripts. The information can also be displayed in different font and graphic forms.

1. INTRODUCTION

Speech is the most natural and efficient means for humans to communicate with each other as well as with machines. Among the human-machine interaction systems, speech recognition is an attempt to have machines respond to spoken commands. Attempts are being made to have automatic speech recognition systems for various complexities. However, the success so far is limited to hand small vocabulary and to great extent externalizing systems.

This paper describes the development of a voice operated multilingual information display/retrieval system and the flexibility of its practical applications. Both the word recognition unit and the interface card as well as the interface card and display system are Z80A microprocessor based hardware modules.

Interesting applications are to be anticipated with automatic speech recognition systems in the areas of industrial controls, transport, communication and military technologies. As one of the interesting applications, the automatic spoken word recognition system can be successfully used for the display and retrieval of stored information on a CRT by voice commands. The CRT used for this purpose is equipped with a GIST card which accepts serially the ASCII codes of the information to be displayed. To do this a controlling unit has been designed and developed at CEERI.

The unit is capable of applying the serial ASCII codes of information desired to be displayed on the CRT.

With GIST card incorporated in the above system for the CRT is to display the information is any of the several Indian scripts and Roman script for English. Later it can be multiplexed or to one of the selected scripts on the screen. Since India is a multilingual country, the objective of developing the voice operated information display system is to have interaction with computers using multilingual scripts.

2. DESCRIPTION OF THE SYSTEM

The system consists of two major parts, one is a word recognition unit and the other is an interface unit for information display. These two units are suitably interfaced as shown in the figure 1.

2.1 Spoken Word Recognizer

The recognition processor is a single board microcomputer based on a Z-80 A CPU. It consists of 64 Kbytes of RAM/PROM, Parallel interface adapter (PIA), Counter Timer Circuit (CTC) and two Asynchronous Adapters (ACIA). One ACIA interface to the video terminal controller unit.

The system is speaker dependent but it can be easily trained for a new speaker. The speaker has to speak through a close talking microphone. This speech signal is fed to signal conditioning circuitry which consists of pre-amplifier, equalizer, automatic volume control etc. The pre-processed signal is then passed through an audio spectrum analyzer chip ASA-16. It has 16 contiguous band pass filters covering the frequency range of 200Hz to 7KHz. Each bandpass filter is followed by a low pass filter with cut off frequency of 25 Hz. This arrangement gives the slowly varying signal proportional to sound pressure level in that channel. The ASA-16 also contains a channel multiplexer which provides spectral data for each of the 16 frequencies. The multiplexed signal is allowed to pass through the GIST analog-to-digital converter and this digitized data is stored in the memory.

The underlying principle behind the isolated word recognizer is pattern matching. In this case, patterns of all the words corresponding to a vocabulary are obtained and their template in the system memory. Next, the pattern of the test word is generated and coas
pared with the templates recorded earlier. In case of a correct match there is a display of the correctly recognized word on the front panel of the machine. Corresponding to the matched word the system generates a specific code. The code is accepted by the interface unit through I/O device (2880 PIO). This code is further processed for another comparison with pre-stored codes. Having compared with the correct code this interface unit sends the ASCII data of the stored information to CRT terminal through serial communication.

2.2 Interface Unit

This unit is also based on ZS80A CPU which is driven by 4.0 MHz clock. It uses a special purpose integrated circuit designed to ease the problems associated with serial communication i.e.: Universal Synchronous/Asynchronous Receiver and Transmitter (USART). The unit provides a three line serial communication port. One line transmits character to the terminal, second line receives characters entered at the terminal keyboard and the third line simply provides a common system ground connection. Information is put on these wires one bit at a time and in this way it takes eight separate transmissions to send a whole byte. The baud rate i.e. speed at which the data is transmitted is chosen to be 9600 in this case. However, this rate depends upon the specific application and the equipment involved. Since there is no handshaking in this I/O process, it is the microprocessor's responsibility to send characters faster than the speed at which the terminal can receive. Similar method is used to accept the incoming characters at the frequent intervals to avoid missing data being sent by the terminal.

To communicate with the terminal, the transfers to and from interface unit are conveyed by using the American standard EIA-RS232 C, because the data are no longer activated by current intensity but by voltage level i.e. +12V and -12V. Also such levels are not TTL compatible and cannot be provided by the microprocessor. Thus a physical interface is interposed using a driver and a receiver circuit. The functional diagram of the complete system is shown in Fig.2.

3. Testing/Features of the System

The system is tested with 10 words vocabulary spoken by both male and female speakers. Though it can accommodate 40 words vocabulary, the words are spoken once into a microphone to generate the reference pattern. Corresponding to 10 words vocabulary 10 pages of stored information can be retrieved. Transcript provides emulation for the DEC VTs52 and ANSI compatible terminals for display of English as well as Indian scripts. Another mode displays high resolution, print quality characters for Indian scripts which can be used for text processing requirements.

ACKNOWLEDGEMENT

The authors are grateful to Dr. W.S. Khokhle, Director CEERI for kind support and encouragement for the above work. The authors wish to acknowledge the financial support of Dept. of Electronics, Govt. of India.

REFERENCES


