International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 3, Issue 3, March 2014

VLSI IMPLEMENTATION OF REAL TIME SPEECH RECOGNITION

N.Ulagammal* B.Sarala**
*Department of Electronics and Communication Engineering, Sri Venkateswara College of Engineering
**Department of Electronics and Communication Engineering, Sri Venkateswara College of Engineering

Abstract -**The main aim of this project is to find the authenticationof user'svoice based on FPGA voice processing. Voice samples of authorized users will be trained and stored in VLSI hardware. Whenever the user speaks in front of microphone, the incoming voice compares with the trained voice samples and if they get matched to the matching score of 98%, then the result will be valid. Due to cold or cough or dry throat, the user's voice may differ. So, if the user's voice changes, then the result will be invalid. The application area of this project is in domestic, bank and industries.**

Keywords– **Speech recognition, FPGA, Hidden Markov Model**

I. INTRODUCTION

Speech recognition is a process of automatically recognizing who is speaking on the basis of the individual information included in speech waves. Speechrecognition is the preferred authentication method in commercial transactions or remote personal identification processes carried out by telephone, since it provides security. Speaker recognition represents an accuracy andefficient way of authenticating a person's identity by analysing his or her voice. This identification method is especially suitable in real-life scenarios or when a remote recognition over the phone is required. The speech of an individual was examined and represented by a human expert, who makes a decision on the person's identity by comparing the characteristics in this representation with others. Voice technology offers high levels of confidence and protection.Many organizations like Banks, institutions, industries etc. are currently using this technology for providing greater security to their vast databases.

The design based on FPGA (Field Programmable Gate Array) is a suited way to implement systems that require a high computational capability and the resolution of algorithms in real-time. These devices allow the design of complex digital systems with outstanding performance in terms of execution time. FPGA has advantages of short development cycle, low-cost design and low-risk. In recent years, FPGA has become the key components in high-performance digital signal processing systems in digital communication, network, video and image fields. At a reasonable low-cost, and a reduced time-to-market, these devices allow designing specific hardware architectures devoted to high-speed applications that hardly could be implemented in a different digital device. The good performance of the FPGA, and the flexibility and easiness that available development tools make this device very useful in applications for implementing algorithms with a high computational complexity.

There are many matching speaker verification algorithms including Hidden Markov Models (HMM), Dynamic Time Warping (DTW), and Neural Networks.Modern general-purpose speech recognition systems are based on Hidden Markov Models. Hidden Markov Models are popular because it can be trained automatically and are simple and computationally feasible to use. Dynamic Time Warping is an approach that was historically used for speech recognition. It is now largely displaced by the more successful HMM-based approach. Neural Networks emerged as an attractive acoustic modelling approach and used in many aspects of speech recognition such as phoneme classification, isolated word recognition, and speaker adaptation. In contrast to HMM, neural networks make no assumptions about feature statistical properties and have several qualities making them attractive recognition models for speech recognition.

In this paper, Hidden Markov Model is used among the speech recognition algorithms because it provides high recognition accuracy for word recognition tasks. It is very rich in mathematical structure and forms theoretical basic for use in wide range of applications. Its recognition ability is good for unknown word. The role of HMM recognition is to find out the maximum probability of the HMM.In the existing system, voice process was used just to compare voice and create an id, but full application was not done using voice processing. Itwas like storing the results in system database and it were used for

ISSN: 2278 – 909X

forensic purpose.For a full application, PIC microcontroller is used. A microcontroller is a compact microcomputer designed to govern the operation of embedded systems in motor vehicles, robots, office machines, medical devices, mobile radios, vending machines, home appliances, and various other devices. A typical microcontroller includes a processor, memory and peripherals. PIC16F877A is one of the most advanced microcontrollers from Microchip. This controller is widely used for experimental and modern applications because of its low price, wide range of applications, high quality, and ease of availability. It is ideal for applications such as machine control applications, measurement devices, and study purpose and so on. The PIC16F877A features all the components which modern microcontrollers normally have. A microcontroller depends on the internal memory and has add on features. It has configurations such as oscillator, watchdog timer, power on reset, and brown out detection. These are the advantages of microcontroller over microprocessor.

Flow code algorithm is used for software implementation in PIC microcontroller. Flow code is a very high-level graphical programming language for PIC microcontrollers based on flowcharts. Flow code allows designing and simulating complex PIC-based robotics and control systems simply by drawing a flow chart of the desired program in a matter of minutes, even without any prior programming skills. The great advantage of flow code is that it allows those with little experience to create complex electronic systems in minutes.

II. PROPOSED SYSTEM

Fig. 1 shows the block diagram of proposed speech recognition system. The system takes the input from the microphone and feeds it to the ADC which is sampling at 8 KHz. The ADC output is interfaced to a 256x16 bits memory. The memory stores a 32ms chunk of voice and outputs it to the FFT module which computes the FFT of that voice chunk. The FFT module is interfaced with the FPGA main processor which reads the FFT outputs and it checks whether the voice chunk corresponds to silence or a speech signal. If it detects a speech signal, it stores the next 32chunks of voice (which corresponds to 1.024s (32ms*32) of voice). After it stores the 1 second of voice, it perform various computations like normalization, feature extraction, comparison etc. and checks if the spoken word corresponds to any of the predefined command. If it finds a match, it takes the action corresponding to that particular command.

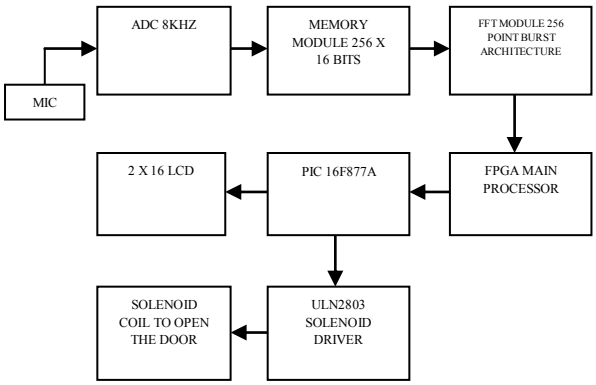


Fig.1 Block diagram of proposed speech recognition system

In case of a user speaks, voice is stored in the FFT of the 1.024 seconds of the voice signal and pass it to the FPGA and store the samples in the SDRAM.The C code then treats this input as an array and performs feature extraction and distance comparison on this input array to check whether it matches with any of the predefined command. If so, it sends out a matching signal which is indicated by the VLSI hardware.

Fig. 2 below shows the voice recognition block diagram. TheVoice Recognition System consists of a sample RAM, Voice Controller and a test RAM. The sample RAM is used to store the digitized voice samples. The test RAM is used to get in the digitized test voice signal to compare with the digitized voice samples. The voice controller has been designed with four states:1.Idle state 2. Config state 3. Check state 4. Compare state.

In the idle state, no operation is performed by the voice controller. In the config state, on giving the config signal, the sample digitized voice data enters through the config-in port and gets stored in the sample RAM through the voice controller. The data enters the sample RAM from the Voice Controller through the sample data-in port and the SRAM-EN pin needs to be high for this purpose. The data can be read from and also write into this sample RAM. While writing or storing the sample digitized voice into the sample RAM, the sample RD-WR pin needs to be high. Thus the digitized voice samples are stored in the sample RAM. The digitized test voice data enters through the test data-in port into the Test RAM. The data then enters the Voice Controller through the Test Data out port when the TRAM-EN pin is high. The Test RD-WR pin has to be low as it reads from the Test RAM. In the check state, the data from the Test RAM and the Sample RAM has to be in the Voice Controller for the checking process and further comparisonprocess to be done. The checking process is done only when the config

ISSN: 2278 – 909X

signal is low. The Voice Controller reads the sample data from the Sample RAM through the Sample Data Out port by making the Sample RD-WR pinlow. Each segment of 8 bit data is taken from both Sample RAM and Test RAM and compared. If they are found to be equal, then the success count is incremented. Else the fail count is incremented. This process continues for each segment of 8 bit data and the success count and fail count are found. If all the 8 bit segments of data are compared, then the view result signal goes high. In the compare state, the success count and fail count are compared. If the success count is greater than the fail count, then the test ok signal goes high indicating that the test word matches with the sample word. Else it means that the comparison process has failed and the word has not matched with any of the stored sample data's. The segment address gives the address of the sample word that has matched with the incoming test word.

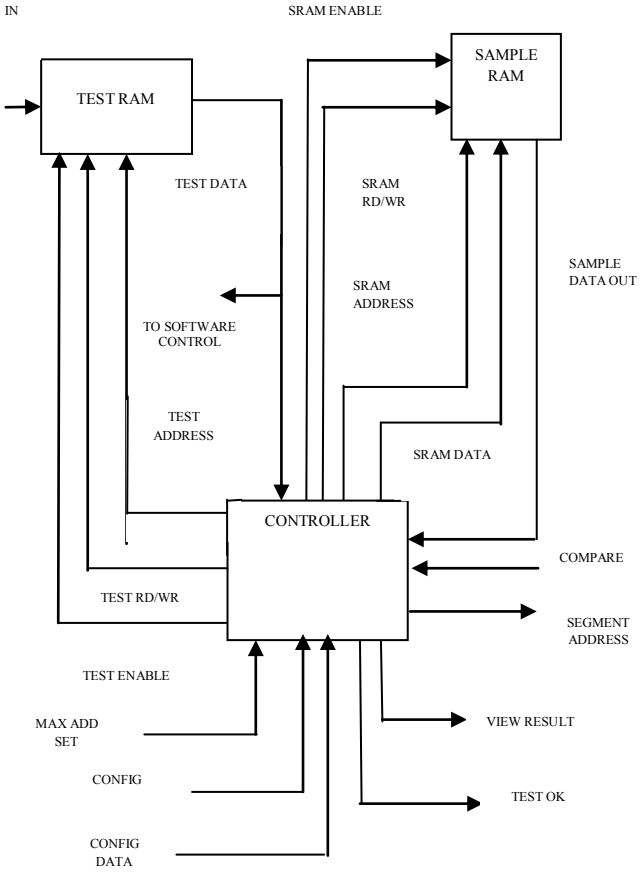


Fig.2 Block diagram of voice recognition

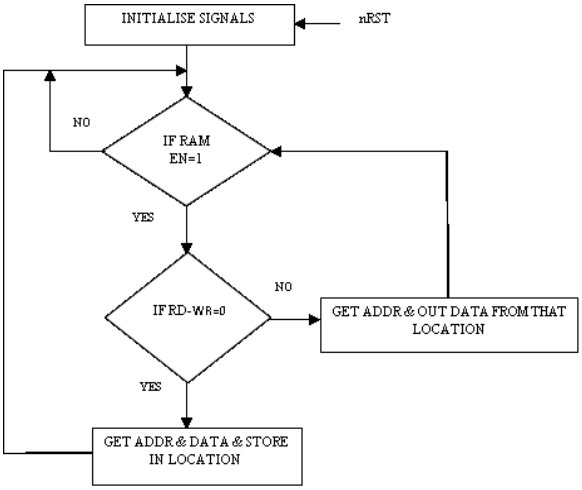


Fig.3 Block diagram of TRAM and SRAM

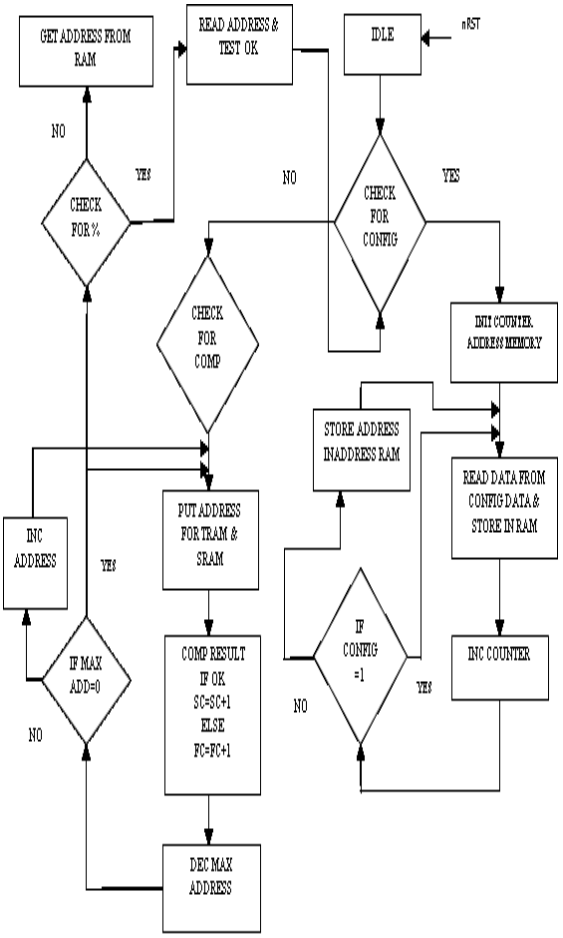


Fig.4 Block diagram of voice controller

III. SOFTWARE IMPLEMENTATION

3.1 FPGA

Aldec Active HDL is used for FPGA simulation. Hidden Markov Model is used in speech recognition. Aldec provides software and

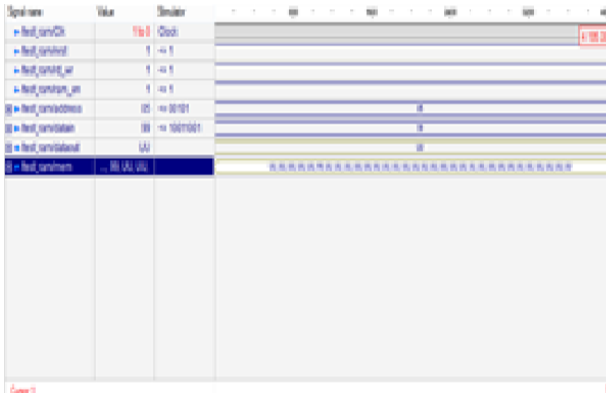
ISSN: 2278 – 909X

hardware used in creation and verification of digital designs targeting FPGA and ASIC technologies.FPGA development environment built around kernel HDL simulator.

FPGA SIMULATION RESULTS

The program coding has been written in VHDL language and simulated in Aldec Active-HDL software.

TRAM



CONTROLLER

SRAM

3.2 PIC16F877A MICROCONTROLLER

Flow code algorithm is used for microcontroller simulation.

PIC MICROCONTROLLER SIMULATIONRESULTS

If the input is voice matched.

If the input is voice mismatched.

IV. CONCLUSION AND FUTURE WORK

This project presents the simulation part of VLSI implementation of real time speech recognition. A FPGA-based Hidden Markov Model speech recognition system was used in this project. The program coding has been written in VHDL language and simulated using Aldec Active HDL software which concludes that the HMM algorithm is implemented successfully. FPGA output is given to PIC microcontroller. The microcontroller part was simulated using flow code algorithm. If input is voice matched, LCD will display the output and solenoid will be active. If input is voice

ISSN: 2278 – 909X

mismatched, LCD will display the output and alarm will be on.

Future work would be the implementation of real time speech recognition using VLSI. The performance of the circuit will be analysed with the hardware part and implementation result will be compared with the simulated result.

REFERENCES

[1] EhKan, P., Allen, T., Quigley, S. F. (2011). FPGA Implémentationfor GMM-based speaker identification, International Journal of Reconfigurable Computing, vol. 2011.

[2] Haitao Zhou & Xiaojun Han (2009). Design and implementation of speech recognition system based on field programmable gate array, CCSE, vol. 3, no. 8.

[3] Kinnunen, T., Karpov, E., & Franti, P. (2006). Real-time speaker identification and verification. IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 1, pp. 277–288.

[4] Kumar Rakesh, Subhangi Dutta & Kumara Shama (2011). Gender recognition using speech processing techniques in Labview. IJAET, vol. 1, pp. 51-63.

[5] Lopez, M., Daugman, J., & Canto, E. (2011). Hardware software Co-design of an iris recognition algorithm. IET Information Security, vol. 5, no. 1, pp. 60–68.

[6] Munteanu, D.-P. & Toma, S.-A. (2010). Automatic speaker verificaiton experiments using HMM, 8th International Conference on Communications (COMM), pp. 107-110.

[7] Pollack, I., Pickett, J.M.,&Sumby, W.(1954). On the identification of speakers by voice. Journal of the Acoustical Society of America, vol. 26, pp. 403-406.

[8] Picone, J.W. (1993). Signal modeling techniques in speech recognition. Proceedings of the IEEE, vol. 81, no. 9, 1215-1247.

[9] Rabiner, L., & Biing-Hwang, J. (1993). Fundamentals of speech recognition. Englewood Cliffs: Prentice-Hall.

[10] Ramos-Lara, R., Canto-Navarro, E., & Puento-rodriguez, L. (2013). Real-time speaker verification system implemented on reconfigurable hardware. J Sign Process Syst, vol. 71, pp. 89-103.

[11] Shingo Yoshizawa & Noboru Hayasaka (2006). Scalable architecture for word HMM-based speech recognition and VLSI implementation in complete system. IEEE Transactions on Circuits and Systems, vol. 53, no. 1.

[12] Shearme, J.N., & Holmes, J.N.(1959). An experiment concerning the recognition of voices. Language and Speech, vol. 2, pp. 123-131.

ISSN: 2278 – 909X