International Journal of Advance Research in Science and Engineering
Vol. No. 5, Special Issue No. 01 , May 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

IMPLEMENTATION OF SPEAKER RECOGNITION SYSTEM ON FPGA USING LOGICORE IP

Sachin S Kodaganur[1], Sadashiva V Chakrasali[2]

[1]ECE, MSRIT, (India)

[2]Assistant Professor, ECE, MSRIT,(India)

ABSTRACT

In any biometric systems, the response time is proven to be critical. In the past decades, speaker recognition software systems have been running on a general microprocessor having a sequential operation which tend to be slow. As a result, FPGA implemented systems gained preference because of the dedicated hardware. This paper presents a Speaker Recognition system implemented on Spartan 6 using Xilinx LogiCORE. MFCC is used to extract the features and Minimum Euclidian Distance Classifier has been used at the decision logic. It has proven that this implementation is simpler compared to generic methods. The device utilization is optimum.

I INTRODUCTION

Nowadays, biometric identification and authentication are preferred compared to the traditional methods like pin and passwords. The method of identification based on biometric characteristics is safer, unique to each individual and have more reliability.

One such, which uses the biometrics is speaker recognition system. Identifying a person from characteristics of his/her voice (voice biometrics) is Speaker recognition. Ituses the voice biometrics and acoustic features of speech which are different for different individuals. The acoustic patterns show anatomy like the size and shape of the vocal tract as well as the learned behavioural patterns like voice pitch and speaking style [1].

As how any other biometric system works, speaker recognition system also has 2 parts: Registration and Verification. Registration is the training phase, where in the voice sequences from different speakers are collected and the required features are extracted and stored. In verification, also called testing phase, the real time voice sequence is taken, the same features are extracted and compared with the stored feature data. The best match is selected.

There is a significant advantage for speaker verification as it uses a low-cost sensor device to collect voice samples. Few parameters are extracted from the voice signals for further processing. Few of the approaches for extracting these parameters are Reflection Coefficients (RCs), Linear Predictive Coding (LPC), Linear Prediction Cepstral Coefficients (LPCC) and Mel-Frequency Cepstral Coefficients (MFCC). Among these different approaches, the MFCC has an improvement on the recognition accuracy and more robust in the presence of background noise and hence used widely [2]. After extracting the parameters, they are compared

with the previously stored features using a simple classification algorithm like Minimum Euclidian Distance and the user's identity is verified.

II SPEAKER RECOGNITION SYSTEM

2.1 Introduction

Any speaker recognition system consists of a training phase and a testing phase. Fig [1] shows a generic block diagram of a speaker recognition system. The input for both training and testing phase is the human speech which is a slowly time varying signal whose frequency varies between 500Hz – 2 kHz. The input speech is digitalized by sampling. The MFCC block extracts the 13 MFCCs which is stored or compared based on training or testing phase.

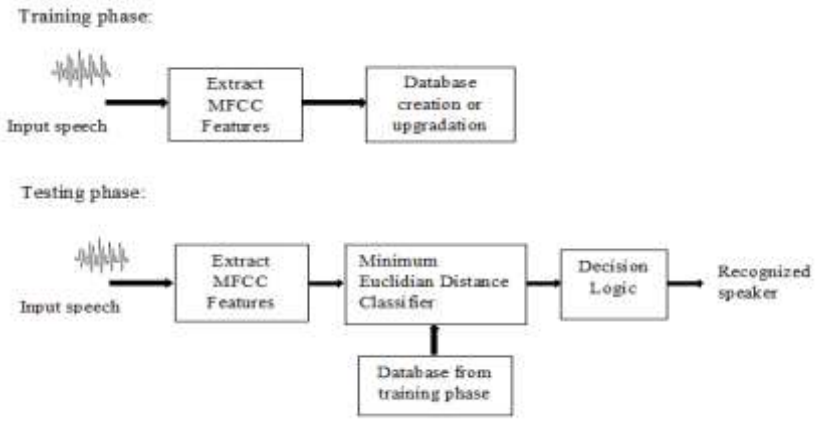


Fig 1: Block Diagram of Speaker Recognition System

2.2 Mel Frequency Cepstral Coefficient (MFCC)

One of the efficient feature extraction technique used in most of the speaker recognition system is Mel-Frequency Cepstral Coefficients commonly termed as MFCC. Fig [2] shows the MFCC block. The digitalized speech signal is divided into frames having N = 256 samples each with an overlap of 50%. So, in each new frame there are 128 new samples [5].

The frames are normalized by removing frame's mean from each speech sample.

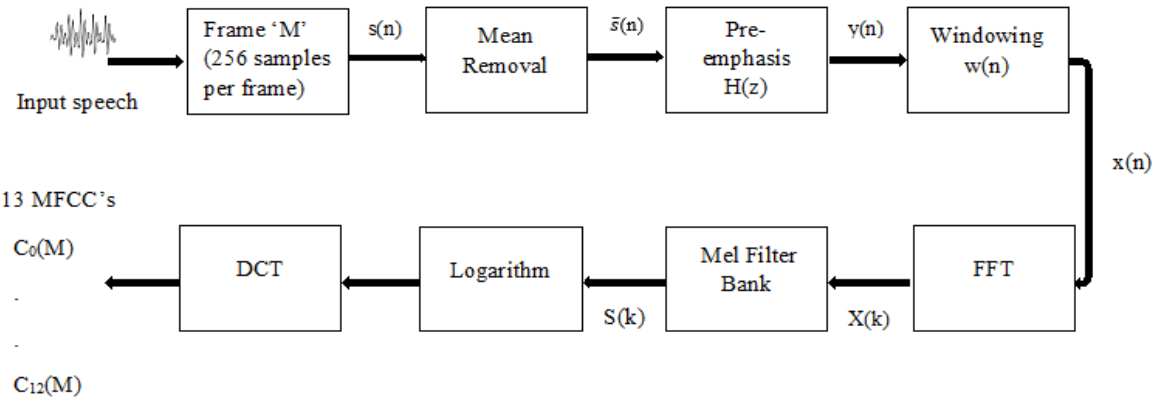
$$\bar{s}(n) = s(n) - \bar{S} \text{with}$$

$$\bar{S} = A \sum_{n=0}^{N-1} s(n)$$

for $0 \leq n \leq N-1$ and $A = 1/N$ (1)

The high frequency components of the normalized frame having lower amplitudes is compensated by applying a pre-emphasis filter.

International Journal of Advance Research in Science and Engineering
Vol. No. 5, Special Issue No. 01 , May 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354



$$H(z) = 1 - Bz^{-1}$$

Fig 2: MFCC Feature Extraction Block of MthFrame

The output of the pre-emphasis filter is given by

$$y(n) = \bar{s}(n) - B.\bar{s}(n-1)$$

with $0 \leq n \leq N-1$ and B=0.97 (2)

The output signal of the pre-emphasis filter is smoothened using a Hamming window w(n) which is defined as

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \\ 0 \qquad\qquad\qquad ,otherwise \end{cases}$$

(3)

The result of windowing is signal x(n)

$$x(n) = y(n).w(n) \qquad (4)$$

The time domain signal x(n) is converted to a frequency domain X(k) with the help of Discrete Fourier Transform (DFT).

$$X(k) = \sum_{n=0}^{L-1} x(n).e^{-j\frac{2\pi kn}{L}}, 0 \leq k \leq L-1$$

(L = 256) (5)

Mel filter banks are used to convert the frequency scale in Hertz to Mel scale signal. 13 band pass mel filter banks are used in the proposed work. The magnitude of mel scale signal is passed through a logarithmic block. This signal is given to a Discrete Cosine Transform (DCT) block whose output is 13 cepstral coefficients of the particular frame.

$$C_n(M) = \sqrt{\frac{2}{K}} \sum_{k=1}^{K} (\ln S(k)) \cos\left[n(k-0.5)\frac{\pi}{K}\right]$$

where K = 13 and $1 \leq n \leq 12$

(6)

2.3 Minimum Euclidian Distance Classifier (MEDC)

It is one of the simplest method used in the decision logic of a Speaker Recognition System [4].

It works as follows,

The average of feature vectors of all the analysis frames is computed for both training and testing data.

$$\bar{C}_i^{ts} = \frac{1}{M} \sum_{m=1}^{M} C_i^{ts}[mL, n]$$

$$\bar{C}_i^{tr} = \frac{1}{M} \sum_{m=1}^{M} C_i^{tr}[mL, n]$$

(7)

Further, the distance between them is calculated using Mean square difference.

$$D = \frac{1}{R-1} \sum_{n=1}^{R-1} \left(\bar{C}_i^{ts}[n] - \bar{C}_i^{tr}[n]\right)^2$$

(8)

This D value is compared with the threshold T and decision is made.

If, $D > T$, speaker is recognised.

III IMPLEMENTATION

The above mentioned speaker recognition system is implemented using LogiCORE IP on Spartan6. Simulation is done in Xilinx ISE Project Navigator software. The optimized parameterizable cores for Xilinx FPGAs are generated and delivered by LogiCORE IP. It is a faster way to create DSP functions, storage elements, math functions and other such modules.

The input speech signal is sampled at 8 kHz using Data Acquisition System (DAS) of the FPGA and are stored in the internal memory. Fig 3 shows the mean removal, pre-emphasis and windowing blocks designed using LogiCORE IP modules accumulator, multiplier, subtractor and memory. The FFT, mel filter bank and DCT modules are shown in fig 4. Memory 1 contains the Hamming window coefficients of length 256. Similarly, memory 2 stores the 13 band pass Mel filter bank coefficients and memory 3 contains DCT coefficients. The square root and logarithm modules are built using Radix-2 algorithm. The block diagram for MEDC is shown in Fig 5.

IV RESULTS

The simulation result for the system is shown in Fig 6. The timing diagram output of all the blocks in MFCC is shown. Fig 7 shows the device utilization summary. It can be seen that by using LogiCORE IP, the utilization of slice registers in 25% and slice LUTs is 56%. And similar other utilization is shown.

V CONCLUSION

In this paper, Speaker Recognition system is implemented on FPGA using LogiCORE IP. The feature extraction technique used is MFCC which is more efficient compared to other methods. The MEDC classifier is simple and thus GMM can be used to get more accurate results.

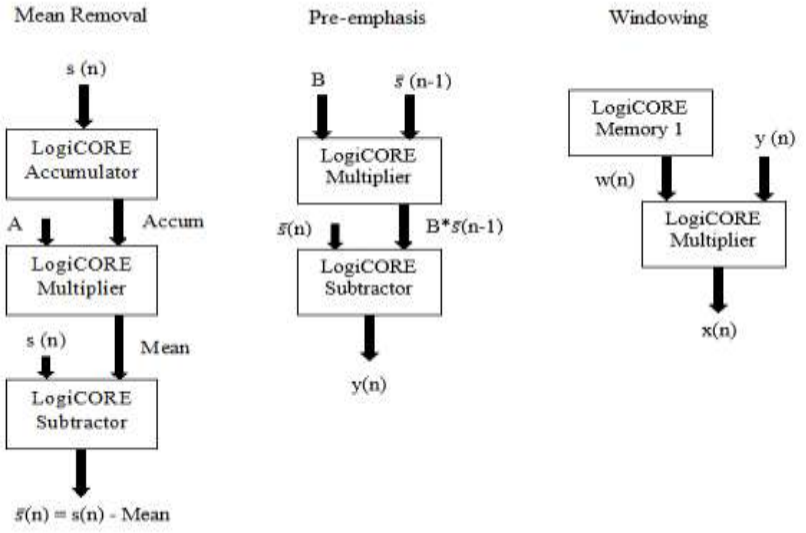


Fig 3: Mean removal, pre-emphasis and windowing blocks

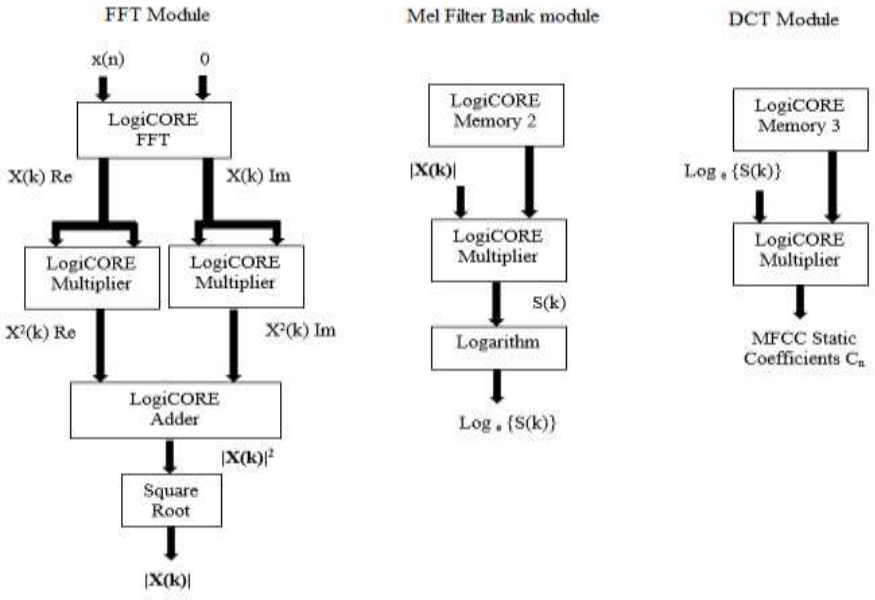


Fig 4: FFT module, Mel filter bank module and DCT module

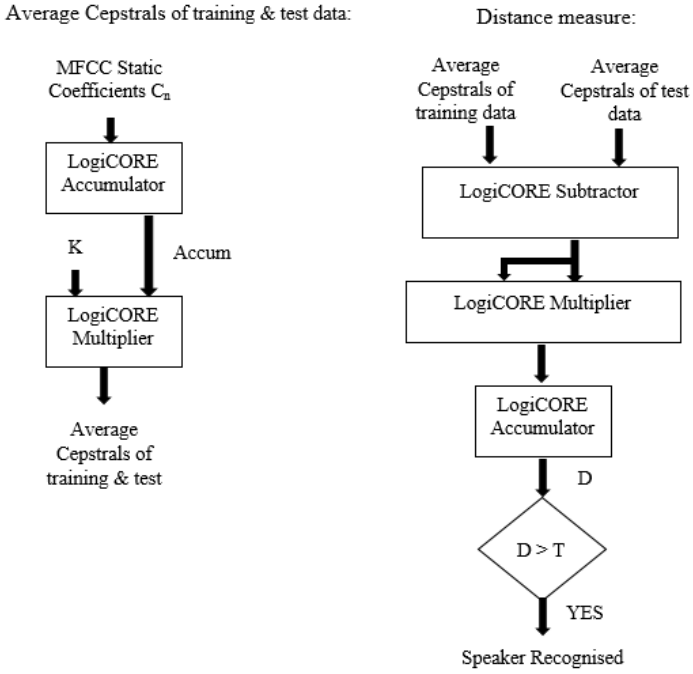


Fig 5: MEDC Block

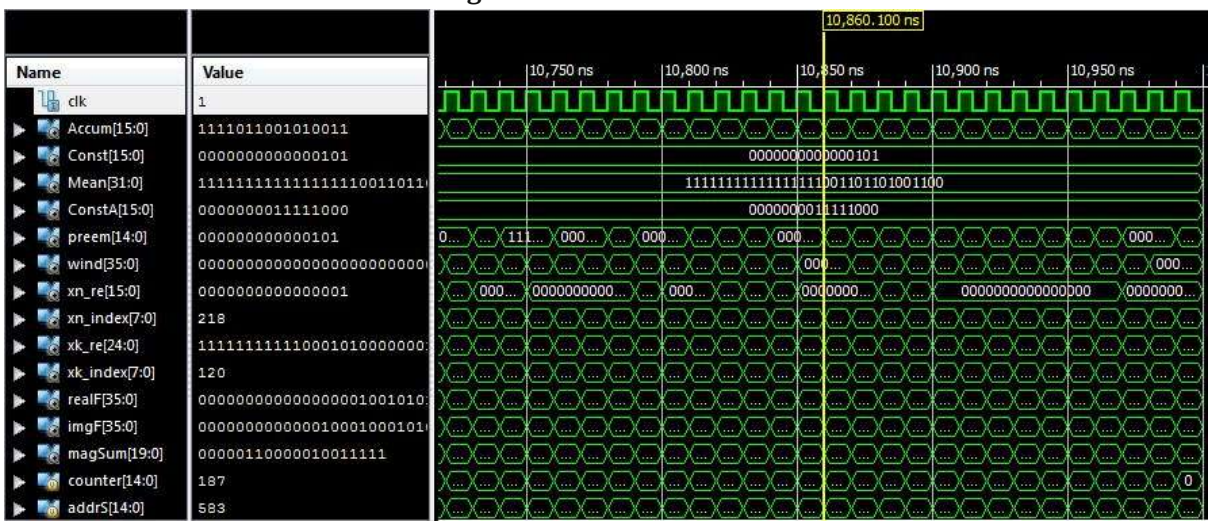


Fig 6: Timing Diagrams of Blocks in MFCC

Device Utilization Summary (estimated values)			[-]
Logic Utilization	**Used**	**Available**	**Utilization**
Number of Slice Registers	7758	30064	25%
Number of Slice LUTs	8563	15032	56%
Number of fully used LUT-FF pairs	6010	10311	58%
Number of bonded IOBs	2	226	0%
Number of Block RAM/FIFO	18	52	34%
Number of BUFG/BUFGCTRLs	1	16	6%
Number of DSP48A1s	8	38	21%

Fig 7: Device Utilization Summary

REFERENCES

1. Kinnunen, Tomi, and Haizhou Li. "An overview of text-independent speaker recognition: From features to supervectors." *Speech communication* 52.1 (2010): 12-40.
2. Rabiner, Lawrence, and Biing-Hwang Juang. "Fundamentals of speech recognition." (1993).
3. Kan, Phak Len Eh, Tim Allen, and Steven F. Quigley. "A GMM-based speaker identification system on FPGA." *Reconfigurable Computing: Architectures, Tools and Applications*. Springer Berlin Heidelberg, 2010. 358-363.
4. Quatieri, Thomas F. *Discrete-time speech signal processing: principles and practice*. Pearson Education India, 2002.
5. Ramos-Lara, Rafael, et al. "Real-time speaker verification system implemented on reconfigurable hardware." *Journal of Signal Processing Systems* 71.2 (2013): 89-103.