

# Implementation of Feature Extraction Algorithm of Speech Signal in FPGA

<sup>1</sup>Anushree Supatkar , <sup>2</sup> Dr. S.N.Mali

<sup>1</sup>Student, SITS, Narhe, Pune

<sup>2</sup>Principal, SITS, Narhe, Pune

**Abstract** - A speech signals are used in biometric recognition technologies and communicating with machine. The features should describe each segment in such a characteristic way that other similar segments can be grouped together by comparing their features. There are enormous interesting and exceptional ways to describe the speech signal in terms of parameter. The speech signal contains many information. To obtain statistically relevant data it is important to obtain parameter or extract feature from speech signal. The Mel Cepstral coefficients are considered as features of the speech signal. The simulation of feature extraction is carried out in the Xilinx ISE design suite 13.2. Further the algorithm of feature extraction is implemented on FPGA spartan hardware. The FPGA has high computational capacity, accuracy and speed so the results are obtained in reduced time.

**Key Words:-** Field programmable gate array, Feature extraction, Mel-cepstral coefficient, Xilinx ise design suite 13.2

## 1.INTRODUCTION

The recognition system includes two phases consist of feature extraction and classification. recognition system for feature extraction process, initially there is one input signal, the input signal may be any signal. First the signal is pre-processed, pre-processing includes removal of noise, amplification and so on. Then by implementing the procedure on signal in Xilinx design suite and the FPGA hardware, the features of speech signal is extracted. The features are the mel cepstral coefficients of the segments of the signal. These features are then stored and are used for further process of classification.

Speech signal are very important signals, they are used in many applications. In the developing technologies it is very important to communicate with machine. Speech signals are used in recognition technologies. For the speech signal the feature extraction is the process to extract the information related to language or speech. The feature extraction of speech signal is based on the short term of the amplitude of the spectrum of the speech signal. During feature extraction the voice recording is cut into windows of equal length, these cut-out samples are called frames which are often 10 to 30 ms long. The short

section of speech signal are separated from the spectrum and then are processed. Fig. 1 presents the basic block diagram.

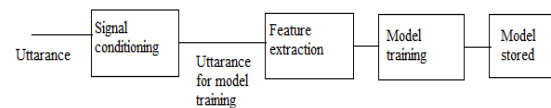


Fig 1. Basic block diagram

Basically identification or authentication using speaker recognition consists of four steps: 1. Voice recording 2. Feature extraction 3. Pattern matching 4. Decision (accept / reject)

There are various algorithm that are used to detect the required feature of speech signal.

- Mel-frequency cepstral coefficient (MFCC)
- Linear-scale filter bank cepstral coefficient (LFCC)
- linear predictive cepstral coefficient (LPCC)
- linear predictive coefficients (LPC)

Mel-scale frequency cepstral coefficient (MFCC) are mostly used algorithm for feature extraction and speech recognition.

### 1.1 Feature Extraction

Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

### 1.2 Feature Extraction of Speech Signal

Speech is the natural and easiest way to communicate between humans. It is desired to interact with computer hardware in the same convenient way as with humans. The main issues, which prevent to design a reliable and universal method for speech recognition, are the environmental impacts, poor pronunciation, speech variability, limited vocabulary size, and a similar phonetic transcription of the words. These issues influence features extraction in a speech. Therefore, an evaluation of algorithms is important especially for the real-time recognizers.

Feature extraction is the process of taking out linguistic information from an uttered speech signal for utilizing in recognition. Short sections of these speech signal are isolated and given for processing. This processing is repeated for the entire duration of the waveform. The result of this operation is a new sequence of features. Mel-scale frequency cepstral coefficients (MFCC) are the most frequently used for speech recognition. This is because MFCCs considers observation sensitivity of human ear at different frequencies, and hence, is appropriate for speech recognition. Feature Extraction plays a major part in the speech recognition algorithm.

### 1.3 Mel Frequency Cepstral Coefficient (MFCC)

The first step in any automatic speech recognition system is to extract features i.e. identify the components of the audio signal that are good for identifying the linguistic content and discarding all the other stuff which carries information like background noise, emotion etc. The main point to understand about speech is that the sounds generated by a human are filtered by the shape of the vocal tract including tongue, teeth etc. This shape determines what sound comes out. If we can determine the shape accurately, this should give us an accurate representation of the phoneme being produced. The shape of the vocal tract manifests itself in the envelope of the short time power spectrum, and the job of MFCCs is to accurately represent this envelope. This page will provide a short tutorial on MFCCs. Mel Frequency Cepstral Coefficients (MFCCs) are a feature widely used in automatic speech and speaker recognition. They were introduced by Davis and Mermelste in in the 1980's, and have been state-of-the-art ever since. Prior to the introduction of MFCCs, Linear Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficients (LPCCs) and were the main feature type for automatic speech recognition (ASR).

### 2. Proposed System

The vector floating point unit computation can be performed with vectors stored in external memory and scaler provided. Likewise result can be stored in external memory. The bus

interface connects the VFPU to the external processor through system bus. It manages writing register. [12]The memory FIFO reads the vectors from external RAM and fills the memory in the first in first out pattern. Multiplexing is the generic term used hereto describe the operation of sending digital signals over a common transmission line to the FPU. The pipeline FPU operates its function of mathematical operation of addition, subtraction and so on also the exponential and logarithm function. These are used to obtained the features of the signal based on the mel frequency cepstral coefficient. To ease the operation the floating point is converted into fixed point during process and finally the output is again obtained in floating point.

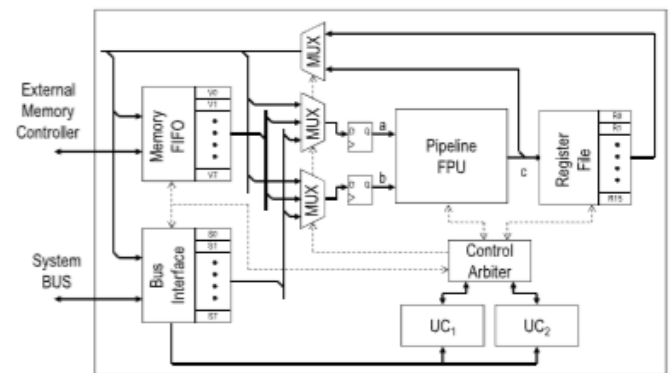


Fig 2. Block diagram of proposed system [1]

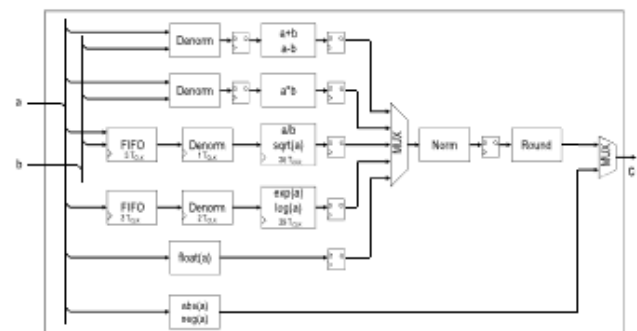


Fig 2. Internal Schematic of FPU [1]

The VFPU executes computations on vectors of arbitrary size using operands of single precision (32-bit). Using this format, the total compatibility between the data shared by the microprocessor and the VFPU is ensured. Although a design based on half-precision (16-bit) would consume less hardware resources, in such case a block for data conversion should be included to guarantee the compatibility between both formats, which may introduce for some stages a penalty on the execution time. Furthermore, using half-precision the computations performed in some stages may lead to produce

under ow (over ow) errors, which should be conveniently managed to avoid their potential effect on the recognition process. Computations can be performed with vectors stored in external memory, scalar numbers provided by the microprocessor, or any combination of them. Likewise, the result of any computation can be placed on an external memory, or read by the microprocessor. The internal architecture of the VFPU is designed to optimize vector computations based on the execution of a set of basic floating-point operations. In this way, these computations can be performed without unnecessary accesses to external memory, which are used to store temporary results.

### 3. Results

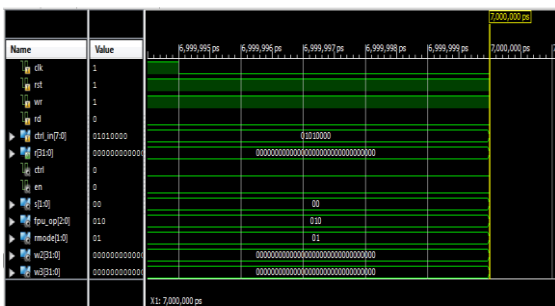


Fig: Final output, matching of signal

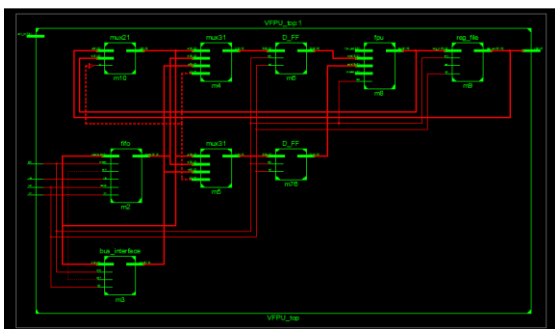


Fig: Final output, matching of signal.

The clock frequency used to obtain the experimental results is 40 MHz. Program and data are located in a 2-MB SRAM external memory. This memory is connected to both the microprocessor and the VFPU, which have direct access to read and write data. The performance of the VFPU was compared with two systems of similar features: the FPU provided by Xilinx and the ARM Cortex A8 microprocessor. Experimental results show as each frame is processed by the VFPU in 3.64105 clock cycles, which represents an acceleration factor of 11.2 and 15.41 when compared with systems based on an ARM-NEON microprocessor and the FPU of Xilinx, respectively.

Logic Utilization	Used	Available	Utilization
Number of slice register	272	34576	0
Number of Slice LUT	740	27288	2
Number of fully used LUT ff pairs	146	866	16
Number of bonded IOBs	27	218	12
Number of block RAM/FIFO	5	116	4
Number of BUFG	3	16	18

Table:-Table of device utilization

### 4. CONCLUSIONS

For the feature extraction of the speech signal, the MFCC system is studied, hence the feature extraction system based MFCC can be implemented. A generic architecture of VFPU that solves all the vector floating-point computations involved in the algorithm. Additionally, the architecture provides a high flexibility. The implementation of the system on the FPGA platform improves the accuracy as computation capability is high. The faster system can be implemented VFPU based on MFCC, as MFCC is faster than all other methods. The extracted features can be further used in speech recognition system by SVM classifiers.

### REFERENCES

- [1] Rafael Ramos-Lara, Mariano López-García, Enrique Cantó-Navarro, Luís Puente-Rodriguez, “SVM Speaker Verification System Based On A Low-Cost FPGA”, ©2009 IEEE.
- [2] Micha Staworko, Mariusz Rawski, “FPGA Implementation of Feature Extraction Algorithm for Speaker Verification”, MIXDES 2010, 17th International Conference Mixed Design of Integrated Circuits and Systems , June 24-26, 2010, Wroclaw, Poland .
- [3] Shing-Tai Pan, Member, IEEE, and Xu-Yu Li, “An FPGA-Based Embedded Robust Speech Recognition System Designed by Combining Empirical Mode Decomposition and a Genetic Algorithm”, IEEE Transactions on Instrumentation and Measurement, Vol. 61, No. 9, September 2012.
- [4] Genevieve I. Sapijaszko Department of EECS, Wasfy B. Mikhael, “An Overview of Recent Window Based Feature Extraction Algorithms for Speaker Recognition”, ©2012 IEEE
- [5] Saambhavi.V.B., S.S.S.P.Rao and P.Rajalakshmi, “Design of Feature Extraction Circuit for Speech Recognition Applications”.
- [6] Xiaohui Hu<sup>1</sup>, Haolan Zhang<sup>2</sup>, Lvjun Zhan<sup>1</sup>, Yun Xue<sup>1</sup>, Weixing Zhou<sup>1</sup>, Gansen Zhao, “Isolated Word Speech

Recognition System Based On FPGA”, Journal of Computers, Vol. 8, No. 12, December 2013.

[7] Karthikeyan Natarajan, Arun. S, Murugaraj. K, Mala John, “An Application Specific Matrix Processor for Signal subspace based speech enhancement in noise robust speech recognition applications”, © 2007 IEEE.

[8] Veton Z. Këpuska, Mohamed M. Eljhani, Brian H. Hight, “Wake-Up-Word Feature Extraction on FPGA”, World Journal of Engineering and Technology, 2014, 2, 1-12.

[9] Taabish Gulzar, Anand Singh, Dinesh Kumar Rajoriya and Najma Farooq, “A Systematic Analysis of Automatic Speech Recognition: An Overview”, International Journal of Current Engineering and Technology.

[10] Hitesh Gupta, Deepinder Singh Wadhwa, “Speech Feature Extraction and Recognition Using Genetic Algorithm”, International Journal of Emerging Technology and Advanced Engineering.

[11] Naufal Alee, Phaklen Ehkan, R.Badlishah Ahmad, Naseer Sabri, “Speaker Recognition System: Vulnerable and Challenges”, Naufal Alee et.al / International Journal of Engineering and Technology (IJET).