Automatic Spectral Analysis for Speech to Text Conversion with Embedded Execution

P. Yasotha[1] N. Rajasekaran[2] M. Nathiya[3]
[1]P.G. Student [2,3]Assistant Professor
[1,2,3]Department of Electronics Engineering (Embedded systems technologies)
[1,2,3]Kongunadu College of Engineering, Trichy, Tamil Nadu, India

Abstract— Speech signal important role in the Digital signal processing. In this paper speech sample observed with MFCC for the improvement of speech feature representation of the HMM based training approach. MFCC is used to extract the voice features from the voice sample and also HMM is used to the recognize the speaker based on the extracted features. I order to recognize the speaker first train the extracted feature by using HMM parameters and then compared to the original voice signal. If it is matched the voice signals to produce the text by using MATLAB Software. The Simulation results show an improvement of speech recognition with respect to computational time, learning precision for a speech recognition system. Here MATLAB output is interface to the Microcontroller by using UART and display through the LCD module.

Key words: Speech Recognition, Mel Frequency Ceptral Coefficients, Hidden Markow Model, PIC Microcontroller, URAT Interface

I. INTRODUCTION

Speech recognition is a topic that is very useful in many applications and environments in our day today life. Generally speech recognizer is a machine which understands human and their spoken word in some way and can act thereafter. For example, in a car environment to voice control non critical operations, such as a dialing a phone number. To control the traffic by using voice command

The most well-liked spectral based parameter used in recognition approach is the Mel Frequency Cepstral Coefficients called MFCC. In established methods, Automatic Speech Recognition (ASR) is based on statistical models (usually HMMs) of speech units. These models are trained on a large amount of data recorded by many speakers. For a large vocabulary system, the speech units will be at the level of individual speech sounds, phones.

Using sequential cepstral derivatives improves the speech frame feature vectors. After preprocessing the input speech samples to extract feature vectors, the system builds the codebook. The code book is the reference code space that we can use to compare input element vectors. The weighted cepstrum matrices for different users and digits are compared with the codebook. The nearest equivalent codebook vector indices are sent to the Baum-Welch algorithm for training an HMM model. Finally the recognized speech is converted to the corresponding text.

Hidden Markov Model (HMM) is a natural and highly robust statistical methodology for automatic speech recognition. It was tested and proved noticeably in a wide range of applications.

II. RELATED WORK

Isolated speech recognition by using the Mel-Scale Frequency Cepstral Coefficients (MFCC) and Dynamic Time Warping (DTW). Several features are extracted from speech

DTW technique is used for feature matching and also to solve the time alignment problem. Testing the speech samples via maximum likelihood classifier and support vector machine. Those are selected to train and test on two-dimensional MFCC dataset and then compared to each other for performances on correct classification.

GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm from a well-trained speech signal. GMM are not suitable for the environmental conditions

III. FEATURE EXTRACTION USING MFCC

Speech is usually segmented in frames of 20 to 30 ms, and the window analysis is shifted by 10 ms. Each frame is converted to 12 MFCCs plus a normalized energy parameter. Assuming a sample rate of 8 kHz, for each 10 ms the feature extraction module delivers 39 numbers to the modeling stage. This operation with overlap among frames is corresponding to taking 80 speech samples without overlap and on behalf of them by 39 numbers. In fact, assuming each speech sample is represented by one byte and each characteristic is represented by four bytes. If a sample rate of 16 kHz is assumed, the 39 parameters would represent 160 samples.

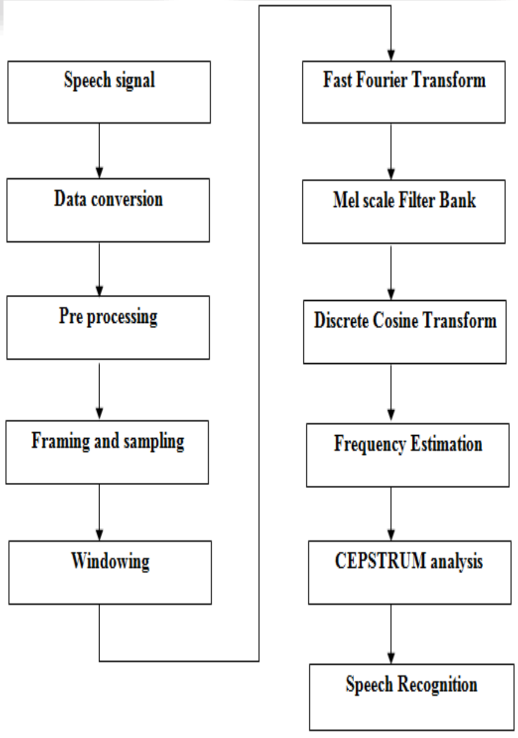

Fig.1 Block diagram of speech recognition

In signal processing sampling is the reduction of a continuous signal to a discrete signal. A common example is the conversion of a sound wave to a chain of samples. A sample refers to a value or set of values at a point in time and/or space. A sampler is a subsystem or operation that extracts samples from a continuous signal as shown in the fig.1.

The Hamming window is defined as

$$Y(n) = X(n) \times W(n)$$

Where,

$W(n) = 0.54 - 0.46\cos(2\pi n/M-1) \quad 0 \le n \le M-1$

After that Fast Fourier transform is used for output samples as

$$N_{output} = (N_{input}/2) + 1$$

Where,

Noutput= No of output samples
Ninput = No of input samples
Noutput = (256/2) + 1
Noutput =129

$$Y(w) = FFT[h(t)*X(t)] = H(w)*X(w)$$

Where,

X (w), H (w) and Y (w) are the Fourier Transform of X (t), H (t) and Y (t) respectively.

MFCC is used to extract the unique features of human voice. It is used to calculate the coefficients that represent the frequency Cepstral these coefficients are based on the linear cosine transform of the log power spectrum on the nonlinear Mel scale of frequency. HMM recognition is extracted feature vectors of MFCC are trained into HMM.

DCT is used to separate the image into parts. It is similar to the Discrete Fourier Transform, and also it transforms a signal or image from the spatial domain and frequency domain. The Cepstrum is a common transform used to gain information from a person's speech signal. It can be used to separate the excitation signal which contains the words and the pitch and the transfer function which contains the voice quality.

Finally, the speech will be recognized and display text output in the MATLAB Software.

IV. SOFTWARE

MATLAB IDE is an integrated development environment that provides development engineers with the flexibility to develop and debug firmware for various Microchip devices.

MATLAB is a high-level technological computing language and interactive environment for algorithm improvement, data visualization, data analysis, and numeric computation. Using the MATLAB product, you can solve technical computing problems faster than with traditional programming languages, such as C, C++, and FORTRAN.

We use MATLAB in a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modeling and analysis, and computational biology. The features are:

– High-level language for technical computing.
– Development environment for managing code, files, and data.
– Interactive tools for iterative investigation, design, and problem solving.
– Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, and numerical integration.
– 2-D and 3-D graphics functions for visualizing data.
– Tools for building custom graphical user interfaces.

V. EXPERIMENTAL RESULTS

As shown in the below figure to setup the experimental results for speech to text. The speech signal in the form of time domain as shown in the fig.2.Speech signal is nothing but a voice signal. The input signal is converted into the number of frames as shown in the fig.3.

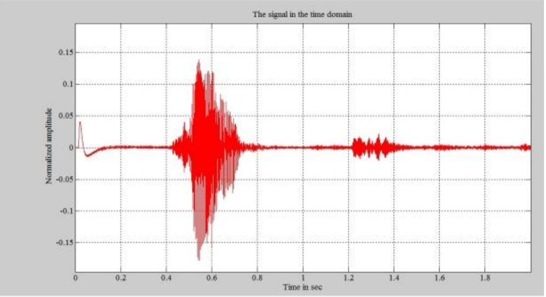

Fig. 2: input signal

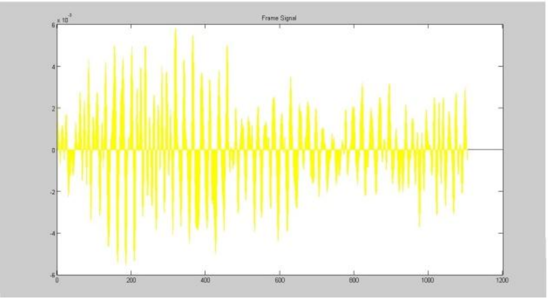

Fig. 3: Frame signal

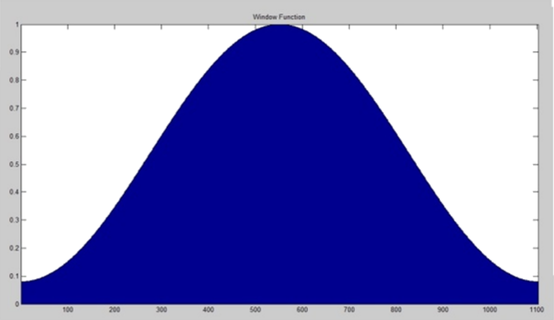

Fig. 4: Hamming Window

Hamming window is used to minimize the maximum side lobe. Amplitude versus samples as shown in the fig.4.The Ceptrum breakdown is an independent variable of a cepstral diagram is called the quefrency. The quefrency is a assess of time, while not in the logic of a signal in the time domain as shown in the fig.5.

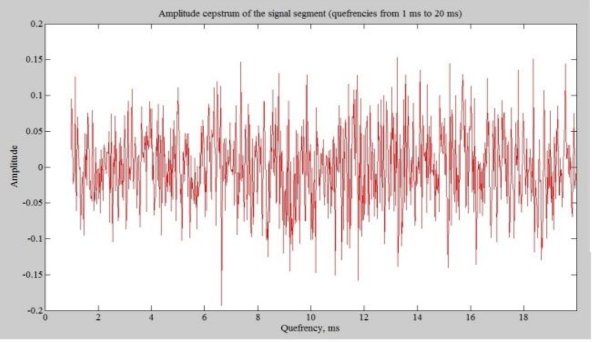

Fig. 5: Amplitude cepstrum of the signal segment

Finally, the output is the text with respect to the speech signal. Several speech signals is used to converted the speech into the text.

VI. CONCLUSION

It is concluded that the proposed research uses the technique of MFCC to extract unique and reliable human voice feature pitch in the form of Mel frequency and trained and recognized using HMM log likelihood methodology. MFCC algorithm calculates cepstral coefficients of Mel frequency scale. In this paper we depicted the real time software implementation of speech recognition using MATLAB coding. And also we have generated codes with the help of MATLAB Programming which requires .wav format speech signals. The experimental results were analyzed with the help of MATLAB and it is proved that the results are efficient.

REFERENCES

[1] Shivanker Dev Dhingra , Geeta Nijhawan , Poonam Pandit, "Isolated Speech recognition using MFCC and DTW", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, August 2013.
[2] Manan Vyas,"A Gaussian Mixture Model based Speech Recognition System using MATLAB", Signal & Image Processing, An International Journal (Sipij) Vol.4, No.4,August 2013.
[3] Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk, "Speech Recognition using MFCC" ,International Conference on Computer Graphics, Simulation and Modeling, July 2012.
[4] Ibrahim Patel1,Dr. Y. Srinivas Rao, "Speech Recognition using HMM with MFCC- An Analysis using Frequency Specral Decomposion Technique ",Signal & Image Processing , An International Journal(Sipij) Vol.1, No.2,December 2012.
[5] Lindasalwa Muda, Mumtaj Begam ,I. Elamvazuthi ,"Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques" ,Journal of Computing, Volume 2, Issue 3,March 2010.
[6] Suresh Gyan Vihar, Jaipur, Rajasthan,"Speech Recognition using MATLAB", International Journal of Advances in Computer Science and Cloud Computing, Issn: 2321-4058 Volume- 1, Issue- 2,November 2013.
[7] Ahmed Besbes, Panagiotis Koutsourakis ,Loic Simon ,Chaohui Wang, "Cepstral Analysis ",Signal Processing, December 12, 2008.
[8] Professor Deepa Kundur,"Cepstral Analysis", http://www.comm.utoronto.ca/~dkundur/course/discrete-time-systems.