IJARSE, Vol. No.4, Special Issue (02), February 2015 ISSN-2319-8354(E)

AUTOMATIC SPEECH RECOGNITION FOR NUMERIC DIGITS USING TIME NORMALIZATION AND ENERGY ENVELOPES

N. Sunil[1], K. Sahithya Reddy[2], U.N.D.L.mounika[3]

[1]ECE, Gurunanak Institute of Technology, (India)

[2]ECE, ACE Engineering College, (India)

[3]EIE, Keshav Memorial Institute Of Technology,(India)

ABSTRACT

In this paper 4 KHz band limited signal is sampled at a frequency rate of 8 KHz. The end points of the isolated words are recognized For this, the energy plot of the utterance is obtained and the end points are detected by cutting off the energy region that is less than 10% of the peak energy on either side. Then these data are time normalized and all the digits are set to same number of data. These samples are now segmented into 94 separate segments each 8ms duration. From each segment the zero crossing rate and energy are determined and the energy envelope plot is smoothened by using point average method. The information is then extracted from the above features. Then the energy peak level is classified as high, medium and low energy peak positions represented by number of segments and the zero crossing rate is obtained for each segment. The above information for each digit is identified and stored in the knowledge base.

Keywords: Data Acquisition, Filtering, Finding the Number of Peaks, Locating Starting and Ending Peaks, Plotting the Energy Envelope.

I INTRODUCTION

Speech recognition, often called automatic speech recognition, is the process by which a computer recognizes what a person said. If you are familiar with speech recognition, it's probably from applications based around the telephone. If you've ever called a company and a computer asked you to say the name of the person you want to talk to, the computer recognize the name you said through speech recognition. In speech applications such as dictation software, the application's response to hearing a recognized word may be to write it in a word processor. In an interactive voice response system, the speech application might recognize a person's name and route a caller to that person's phone.

Speech recognition is also different from voice recognition, though many people use the terms interchangeably. In a technical sense, voice recognition is strictly about trying to recognize individual voices, not what the speaker said. It is a form of biometrics, the process of identifying a specific individual, often used for security applications.

II SPEECH RECOGNITION BASICS

Speech recognition is the process by which a computer (or) other type of machine identifies spoken words. Basically, it means talking to a computer, and having it correctly recognizing what you are saying. The following definitions are the basics needed for understanding speech recognition technology.

2.1Utterance

An utterance is the vocalization (speaking) of a word or words that represent a single meaning to the computer. Utterances can be a single word, a few words, a sentence, or even multiple sentences.

2.2 Speaker Dependence

Speaker dependent systems are designed around a specific speaker. They generally are more accurate for the correct speaker, but much less accurate for other speakers. They assume the speaker will speak in a consistent voice and tempo. Speaker independent systems are designed for a variety of speakers. Adaptive systems usually start as speaker independent systems and utilize training techniques to adapt *to the* speaker to increase their recognition accuracy.

2.3 Vocabularies

Vocabularies (or dictionaries) are the list of words or utterances that can recognize by the SR system. Generally; smaller vocabularies are easier for a computer to recognize, while larger vocabularies are more difficult. Unlike normal dictionaries, each entry doesn't have to be a single word. They can be as long as a sentence or two. Smaller vocabularies can have as few as I or 2 recognized utterances (e.g.' wake up"), while very large vocabularies can have a hundred thousand or more.

2.4 Accurate

The ability of a recognizer can be examined by measuring its accuracy or how well it recognizes utterances. This includes not only correctly identifying an utterance but also identifying if the spoken utterance is not in its vocabulary. Good ASR systems have an accuracy of 98%or more! The acceptable accuracy of a system really depends on the application.

2.5 Training

Some speech recognizers have the ability to adapt to a speaker. When the System has this ability; it may allow training to take place. An ASR system is trained by having the speaker repeat standard or common phrases and adjusting *its* comparison algorithms to match that particular speaker. Training a recognizer usually improves its accuracy. Training can also be used by speakers that have difficulty speaking or pronouncing certain words. As long as the speaker can consistently repeat an utterance. ASR systems with training should be able to adapt.

III DIGITAL SIGNAL PROCESSING

Digital signal processing (DSP) is the study of signals in a digital representation and the processing methods of these signals. DSP and analog signal processing are subfields of signal processing. DSP includes subfields like: audio and speech signal processing, sonar and radar signal processing, sensor array processing, spectral estimation, statistical signal processing, digital image processing, signal processing for communications, biomedical signal processing, seismic data processing, Etc.

Digital signals can be easily stored on magnetic media without loss of fidelity and can be processed off-line in a remote laboratory. The DSP allows us to implement sophisticated algorithms when compared to its analog counterpart. A DSP system can be easily reconfigured by changing the program. Reconfiguration of an analog system involves the redesign of system hardware.

Band limited signals can be sampled without information loss if the sampling rate is more than twice the bandwidth. Therefore, the signals having extremely wide bandwidths require fast sampling rate A/D converters and fast digital signal processors. But there practical implementation in the speed of operation of A/D converters and digital signal processors.

IV PROBLEM DEFINITION

The main aim of the present research work is to develop a system that can identify an isolated spoken digit by analyzing the digits. this can be achieved by Knowing the various parameters of the spoken signal and the parameters we consider are energy envelope, locating the starting and the end points, calculating zero crossing rates, calculating number of peaks and peak positions. matlab software will help to realize various parameters in the spoken audio signal.

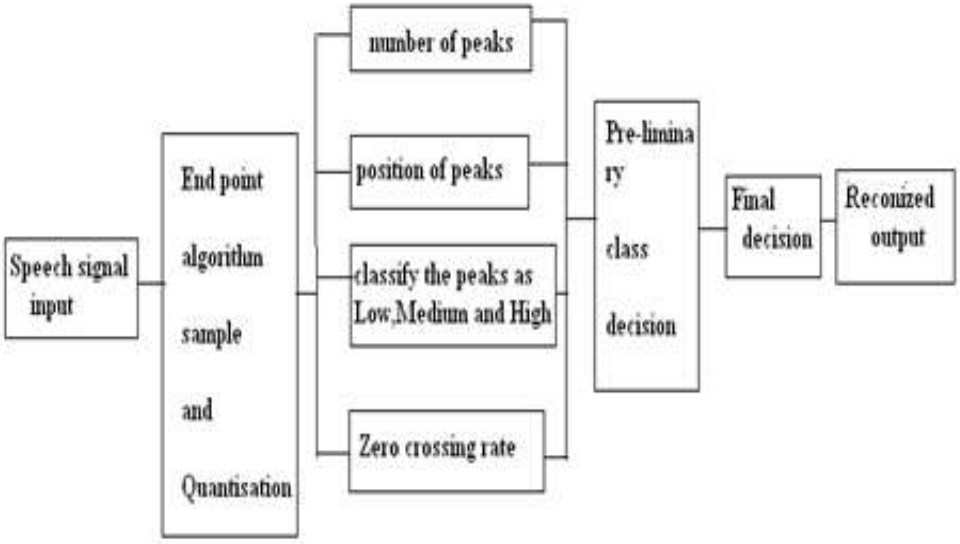
Speech signal recognition has its importance in currently developing fields like voice activated calculator, speaker identification systems, automatic telephone dialing etc.

The various steps involved in recognition are,

- Audio recording and utterance detection
- Pre-filtering (pre-emphasis, normalization, banding, etc.)
- Framing and windowing (chopping the data into a usable format)
- Filtering (further filtering of each window/frame/frequency band)

- Comparison and matching (recognizing the utterance)

V BLOCK DIAGRAM



VI CLASSIFICATION

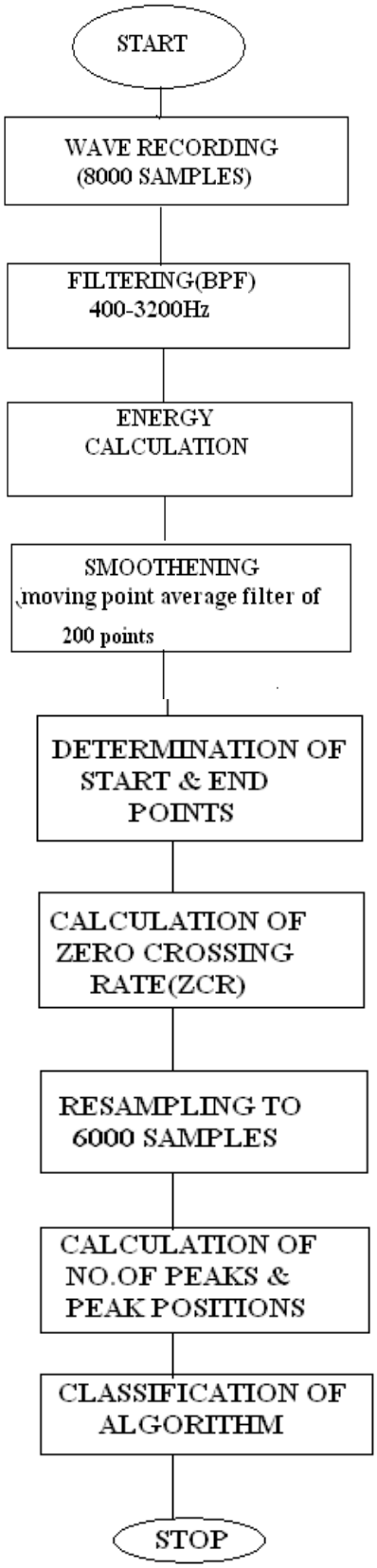
The classification procedure has a tree like structure counting the number of peaks in the energy does the first classification. The digits are classified into three groups. Digits having single peak, two peaks, three peaks. The experiments reveal that digits 1, 2, 4, 6, 9 are having single peak, whereas digits 0, 3, 7 and 8 are having two peaks.

In the group having single peak, next classification is done on the basics of peak position. If the peak position is less than 20 segments, then the digits are either 2 or 6. Information of zero crossing is used to distinguish between the other digits with single peak. For example, the digit 1 has its peak in a segment greater than 40^{th} segment and its zero crossing rate value lies between 14 and 9. digit 9 has a ZCR greater than 14. thus the classification can be done to identify the exact digit among the digits with single peak.

In digits with double peaks, the difference in the peak positions is taken into consideration for classification. If the difference in the peak positions is less than 43 then the digit is either 7 or 3 or 0. so the second position is taken into consideration for classification. If the second peak is less than 60 and ZCR>10.5 then the digit is 7 and the ZCR<10.5 then the digit is 0. otherwise the digit is 3. further classification is done to identify the digit 8.

IJARSE, Vol. No.4, Special Issue (02), February 2015 ISSN-2319-8354(E)

VII FLOW CHART



IJARSE, Vol. No.4, Special Issue (02), February 2015 ISSN-2319-8354(E)

VIII HARDWARE REQUIREMENTS

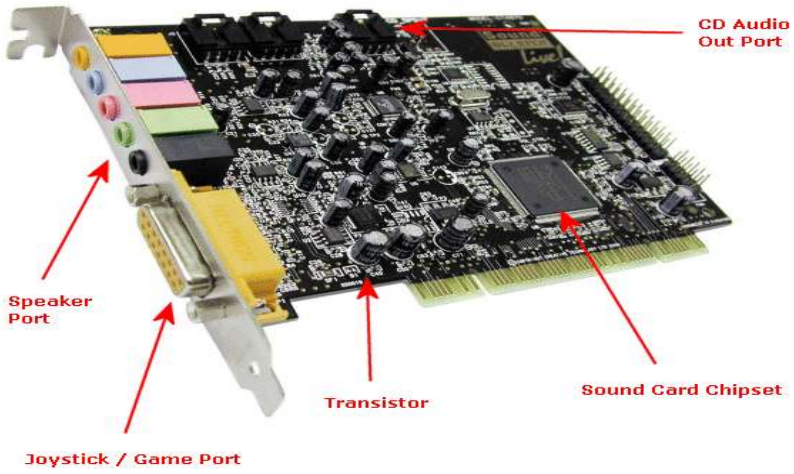
8.1 Microphone

A microphone was used to record the spoken digit. The capacitor or condenser microphone is closest to the ideal, with a Clinical purity of quality. Unlike the magnetic types, the capacitor microphone does not produce its own emf. A DC polarization voltage, somewhere in the 25-150 volts, is fed through a resistor to what is effectively a capacitor plate. A moving plate, which acts as a Diaphragm, forms a second plate of the capacitor and is frequently made of thin plastic Material. As the diaphragm moves, the capacitor changes its value, so varying the current Flow in the circuit. A current operated preamplifier in the body of the microphone then Steps up the signal and provides an output voltage.

Talkmic microphone (49 from iANSYST)- this is a high quality. Pressure gradient, which picks up close sound better than background noise and is attached to the head by means of a loop that fits over the users ear. It is generally very stable and comfortable. Andrea NC61- a good quality microphone which uses active noise canceling to reduce background noise. Less stable to wear than the tlakmic. At the same time as the software has been changing, computer prices have continued to drop and the power of typical desktop machine has grown to such an extent that SR software will run on almost any standard PC manufactured in the past couple of years, provided that it has sufficient memory. Nevertheless, it is important to pay attention to the recommended specification for a particular program. The latest versions of the continuous speech programs have been designed to take advantage of the increased processing power of Pentium IV computers and their equivalents.

8.2 Sound Cards

Sound functions within a PC are generally controlled by a sound card**.** At one time, the quality of sound input signal produced by some cards was not good enough for speech recognition, but this rarely a problem in desktop computers built over the past couple of years. There can still be problems with some laptops where components are closely packed, occasionally leading to electrical interferences.



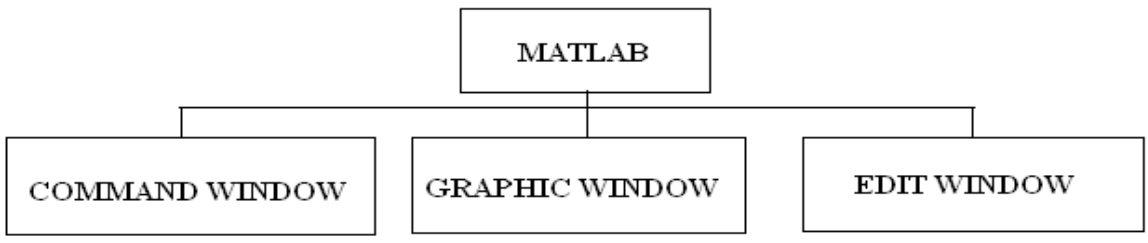
IX SOFTWARE REQUIREMENTS

Matlab is a numerical computing environment and programming language. Created by The Math Works, Matlab allows easy matrix manipulation, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with Programs in other languages. Although it specializes in numerical computing, an optional toolbox interfaces with the Maple symbolic engine, allowing it to be part of a full Computer algebra system. As of 2004, MATLAB was used by more than one million people in industry and academia. Matlab is required to execute the logic designed previously.

There are also several optional TOOLBOXES available from the developers of MATLAB. These toolboxes are collections of functions written for special applications as,

- Symbolic Computation
- Image Processing
- Statistics
- Control System Design
- Neural Networks

9.1 Matlab Windows



- Launch pad
- Work Space
- Command History
- Current Directory

9.2 Types of Files

- M-file: are standard ASCII text files, with a ".m" extension.
- Mat-file: are Binary Data files, with a ".mat" extension.
- Mex-file: are MATLAB callable C-programs, with a ".Mex" extension.

X MATLAB CODING AND RESULTS

Matlab coding is done for data acquisition, filtering of recording data, plotting of energy envelopes, location of starting and ending points, time normalization of data, segmentation, calculation of zero crossing rate, calculating the number of peak positions and detection of digits.

After the coding is done it is simulated in matlab and the following data is obtained in the graphical format for the digit zero and the similar waveforms are obtained for all the remaining digits i.e. from 1 to 9.

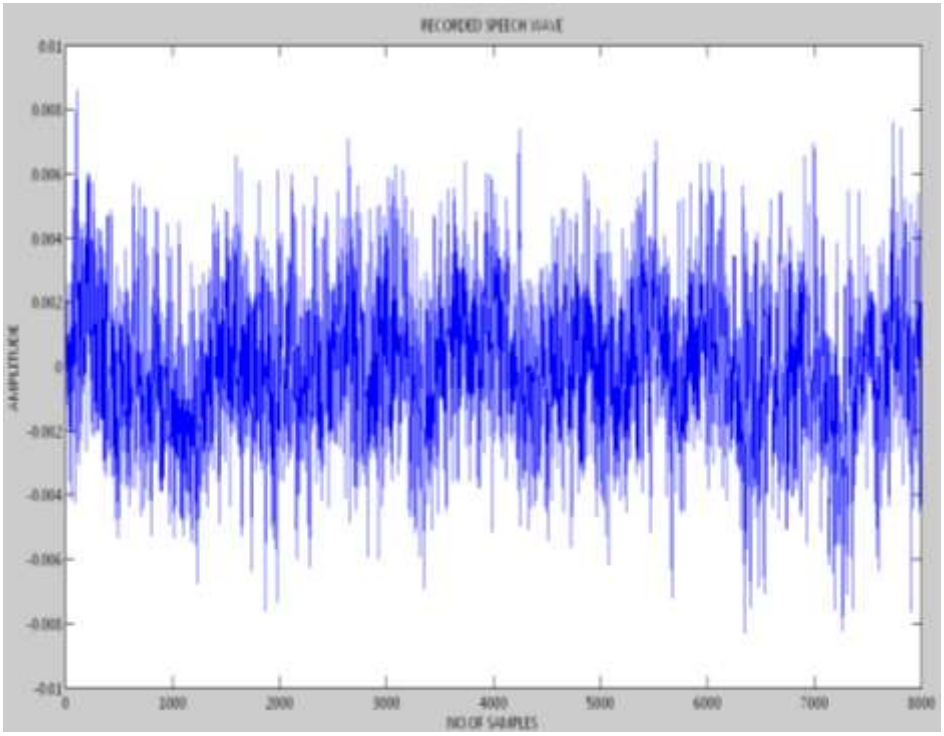


Fig. 10.1 Recorded waveform for digit zero

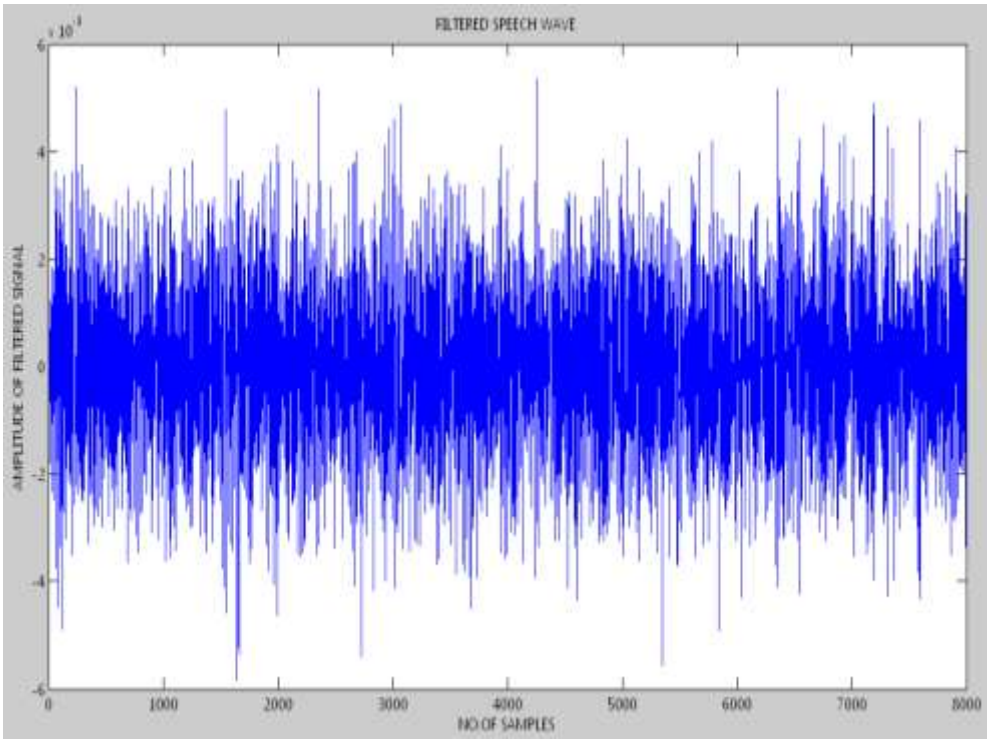


Fig. 10.2. Filtered Waveform for Digit Zero

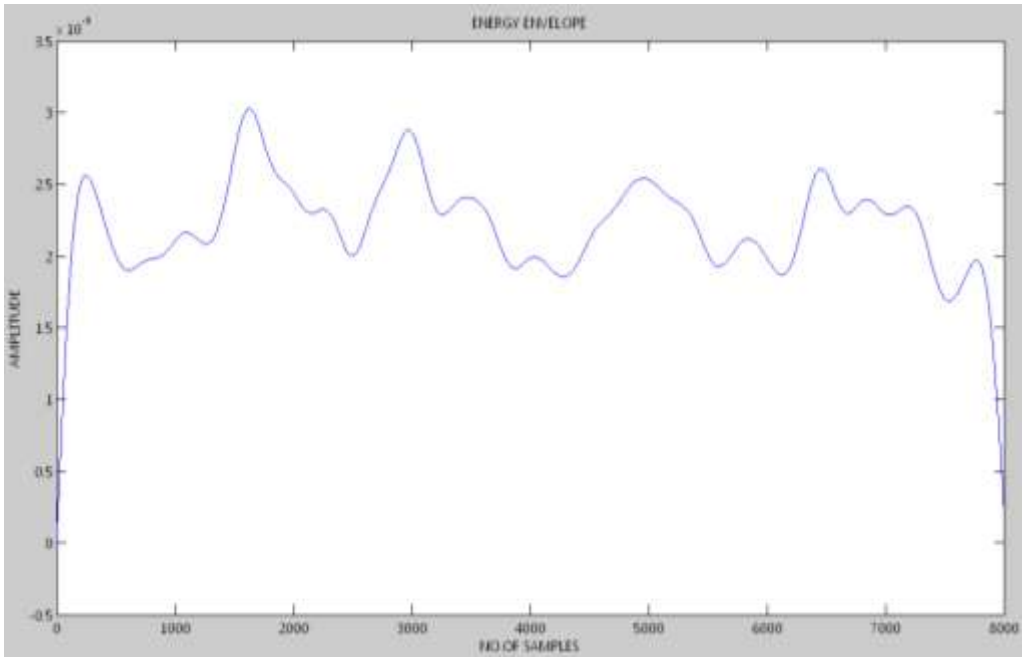


Fig. 10.3. Energy Envelope for the Digit Zero

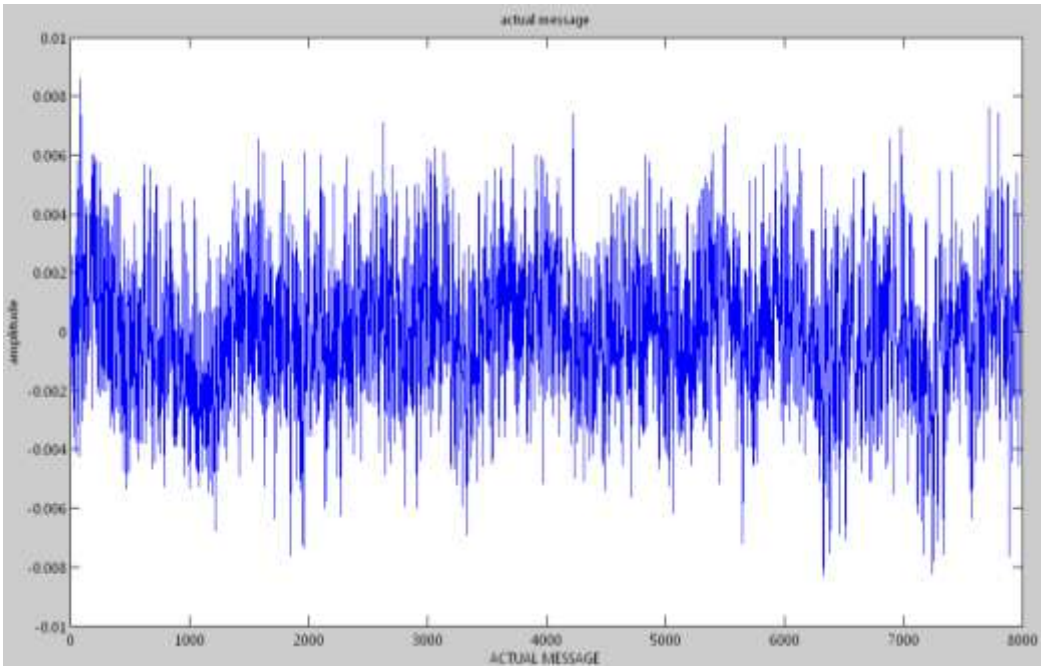


Fig. 10.4. Actual Wave Form for Digit Zero

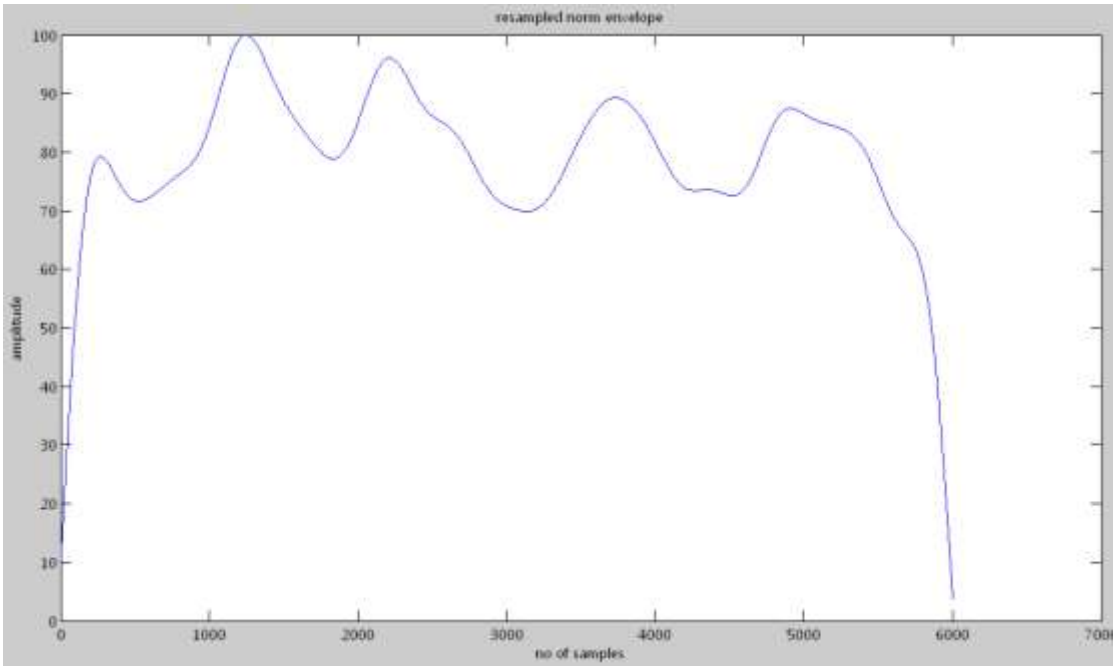


Fig. 10.5. Resampled Normalized Wave Form for Digit Zero

XI CONCLUSIONS

All the digits from '0' to '9' can be recognized using this speech recognition method which involves data acquisition, filtering, finding the number of peaks, locating starting and ending peaks, plotting the energy envelope, segmentation and time normalization. This approach was tested for different voices. A signal analysis package was built, which uses statistical analysis techniques in digital signal processing. The response time for the execution of this is 10 seconds and this can be reduced further by finding the better method for averaging the speech signal. The same approach can be utilized in development of sophisticated system which can recognize more number of isolated words or even continuous speech.

REFERENCES

1. R. Cole, K. Roginski, and M. Fanty.,1992 A telephone speech database of spelled and spoken names. In ICSLP'92, volume 2, pages 891–895.
2. Jurafsky D., Martin J. (2000). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Delhi, India: Pearson Education.
3. Mori R.D, Lam L., and Gilloux M. (1987). Learning and plan refinement in a knowledge based system for automatic speech recognition. IEEE Transaction on Pattern Analysis Machine Intelligence, 9(2):289-305.
4. "Applied Speech Technology". A.Syrdal, R.Bennett, S.Greenspan, 1994
5. "Speech and Audio Signal Processing" Ben Gold, Nelson Morgan, John Wiley, 2002
6. "Automatic Recognition Of Spoken Digits"K.Davis, K.Biddulpl and S.Balashek, J. Acoustic Soc.Am, 1952
7. "Getting Started with MATLAB", Rudra Pratap, Oxford Press.