AUTOMATIC TELEPHONE OPERATOR USING SPEECH RECOGNITION USING MATLAB

Thi Yi Kywe

***Abstract* — This research paper implement the speech recognition system for automatic telephone operator system by using MATLAB. It is one kind of Speaker-Independent Isolated Word Recognition System. This system is software architecture of the input speech signals and its output is LPT signals. In this system, Mel-frequency Cepstrum Coefficient (MFCC) is used for feature extraction and Vector Quantization (VQ) which uses LBG algorithm is used for feature matching. Final process is the connection to telephone line. By this way the developed system can be applied as the automatic telephone operator. To implement this system MATLAB programming language is chosen.**

***Key Words* — Speech Recognition, MFCC, VQ, MATLAB.**

1) INTRODUCTION

Speech recognition system consists of the process to convert a speech waveform into features that are useful for further processing. There are many algorithms and techniques are used. It depends on features capability to capture time frequency and energy into set of coefficients for cestrum analysis. Generally, human voice conveys much information such as gender, emotion and identity of the speaker. The objective of speech recognition is to determine which speaker is present based on the individual's utterance.

To implement this system, a good quality microphone is required to record the speech signals. This system contains three main modules: feature extraction, feature matching and output signal to parallel port. Feature extraction is the process of extracting a small amount of data from the voice signal that can later be used to represent each speech signal. Feature matching involves the actual procedure to identify the unknown speech signal by comparing extracted features from the speech input of a set of known speech signals and decision making process. [1]

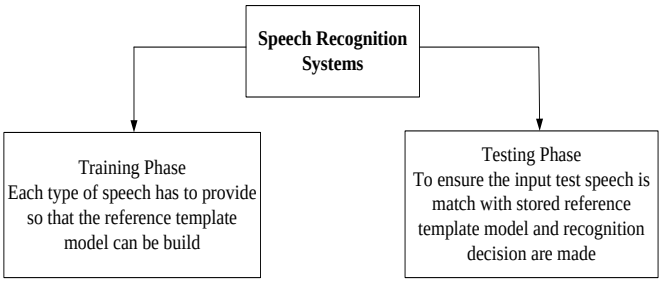

Figure 1.1. Block Diagram of Speech Recognition System

***Thi Ri Kyawe**, Department of Electronic Engineering, Mandalay Technological University, Mandalay.*

A speech analysis is done after taking an input through microphone from a user. The design of the system involves manipulation of the input audio signal. At different levels, different operations are performed on the input signal such as Pre-emphasis, Framing, Windowing, Mel Cepstrum analysis and Recognition (Matching) of the spoken word. [3] The voice algorithms consist of two distinguished phases. The first one is training secessions, and the second one is referred to as operation secession or testing phase as described in Fig. 1.

2) SYSTEM DESIGN

All speech recognition systems contain two main modules: feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each speaker. Feature matching involves the actual procedure to identify the unknown speech signal by comparing extracted features from voice input with the ones from a set of known speech signals.

All speech recognition systems have to serve two distinguishes phases. The first one is referred to the enrollment sessions or training phase while the second one is referred to as the operation sessions or testing phase. In the training phase, the speaker has to provide samples of the speech signal so that the system can build or train a reference model for that speech signal. During the testing phase, the input speech is matched with stored reference model(s) and recognition decision is made. [3]

Automatic Speech Recognition System, only two components: feature extraction and feature matching / recognition are used in this system. This system is implemented using Mel-frequency Cepstral coefficients (MFCC) for feature extraction and Vector Quantization with LBG algorithm for feature matching.

1) Speech Acquistion

The real time speech is to be acquired from computer build in microphone. It is done by wavrecord function. In this function, the parameters such as number of seconds to be recorded, sampling frequency and the recorded data type have to be given. The recorded signal in amplitidue is returned from the function. The following shows the example of code to record the speech file by Matlab software. [3]

2) MFCC Feature Extraction

The implementation steps of feature extraction using MFCC processor for speech recognition system are described as follows:

Frame Blocking: Block speech signal into frames of N samples. [3]

Let N = 256 and M = 100.

Windowing: Window the data with hamming window. The result of windowing is the signal,

$y_1(n) = x_1(n)w(n),\ 0 \le n \le N-1$

Typically the Hamming window is used, which has the form:

$$W(n)=0.54-0.46\cos\left(\frac{2\pi\pi}{N-1}\right),\ 0 \le n \le N-1 \quad (1)$$

Fast Fourier Transform (FFT): Convert each frame of N samples from the time domain into the frequency domain. The FFT is defined as follow:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N}, n = 0,1,2,...,N-1 \quad (2)$$

Mel-frequency Wrapping: Compute the mels for a given frequency *f* in Hz.

$$\text{mel}(f) = 2595*\log10(1+ f/700) \quad (3)$$

Let K=20.

Cepstrum: Convert the log mel spectrum back to time.

mel power spectrum coefficients are $\tilde{S}_k, k = 1,2,..,K,$.

The MFCC's, $\tilde{c}_n$, is calculated as follow:

$$\tilde{C}_n = \sum(\log\tilde{S}_k)\cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{k}\right], \quad n=1,2,\ldots,K \quad (4)$$

Fig.2 shows the flow chart of speech recognition system for phone line numbers Fig.3 shows the flow chart for MFCC processor

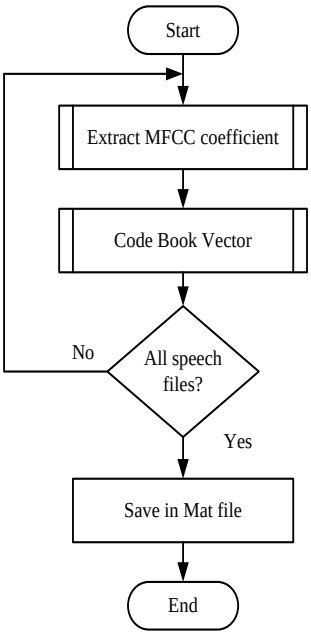

Fig.2. System Flowchart for Feature Extraction

3) LBG Feature Coding

Then the MFCC coefficients can be coded by the LBG algorithm. The implementation steps of feature matching using LBG algorithm for speech recognition system are as follows: [3]

Given T. Fixed ε>0 to be a ``small" number.

Let N=1 and

$$c_1^* = \frac{1}{M}\sum_{m=1}^{M} x_m \quad (5)$$

Calculate $D_{ave}^* = \frac{1}{Mk}\sum_{m=1}^{M}\left\|x_m - c_1^*\right\|$ (6)

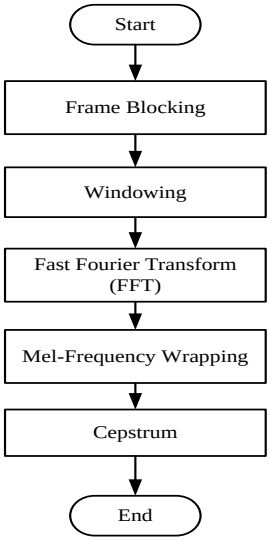

Fig.3. Flowchart of MFCC Processing

Splitting: For i=1, 2, …, N, set

$$c_i^{(0)} = (1+\varepsilon)c_i^*, \quad (7)$$

$$c_{N+i} = (1-\varepsilon)c_i^*, \text{ Set N=2N} \quad (8)$$

Iteration: Let $D_{ave}^{(0)} = D_{ave}^*$. Set the iteration index i=0. For m=1, 2, …, M, find the minimum value of $\left\|x_m - c_n^{(i)}\right\|^2$, overall n=1, 2, …, N. Let n* be the index which achieves the minimum. Set $Q(x_m) = c_n^*$.

Fig. 4 shows the flow chart for LBG algorithm.

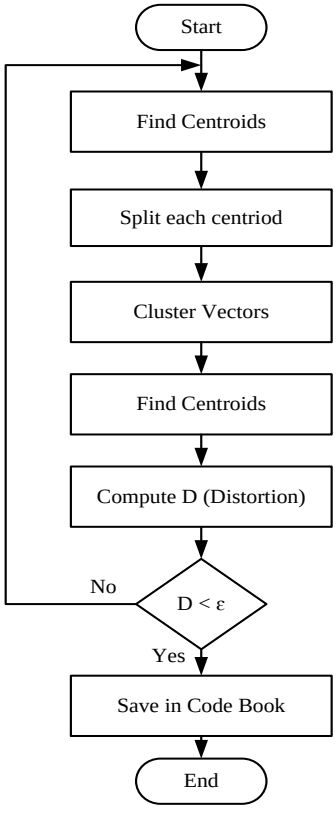

Fig.4. Flowchart for LBG Code Book

3) SIMULATION RESULTS

4) *System Design*

The sound to be recognized is recorded as wave format. The input speech signal is processed for feature extraction. First, the speech signal is blocked into frames of N samples. Then, each frame of N samples is converted from the time domain into frequency domain. Fig.5 shows the power spectrum plot in frequency domain. The power spectrum is then converted into log mel-spectrum and Fig.6 shows logrithmic spectrum plot. In this system, typical values for samples N and M are 256 and 100. But different values for N and M are also tested.

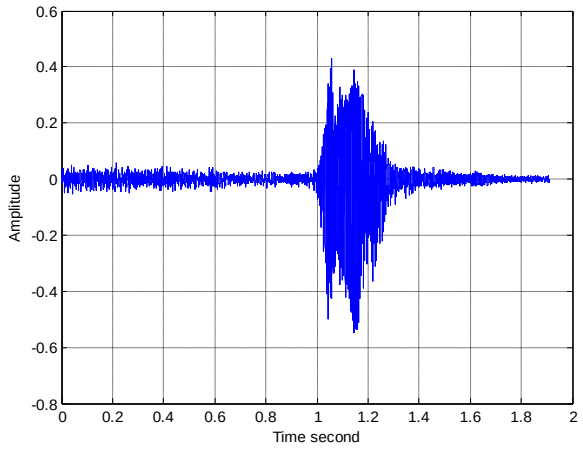

Fig.5. Amplitude Plot of the Input Speech Signal and M.

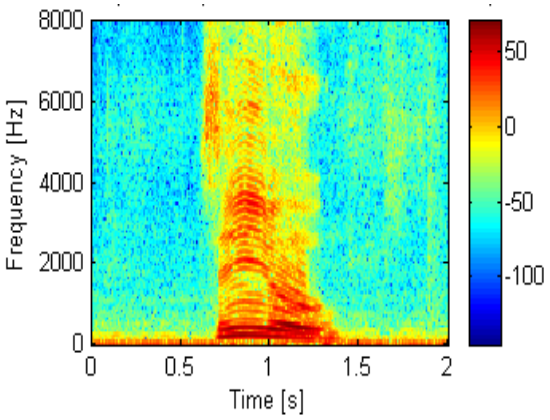

Fig.6. Spectrogram of Speech Signal

The system operation of hardware is tested by combining software and hardware. The parallel port is connected to the telephone line indicator circuit. In testing process, the user can test offline or real time check for telephone operation. The screenshots of software testing are shown in Fig.7.

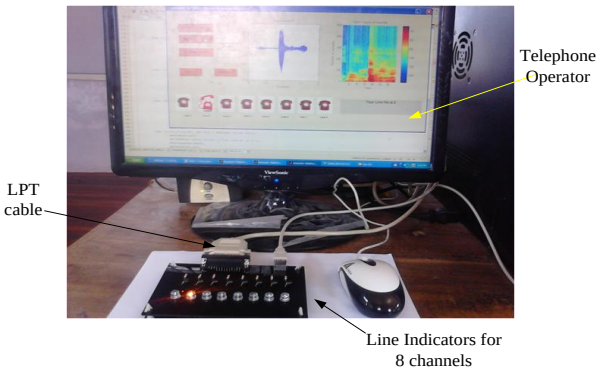

Fig.7. Test and Result of Software and Hardware

In Fig.8, the user can click on the feature extraction button to extract the coefficient features from each speech file. The first figure represents the original speech signal which is plot Amplitude Vs Time in second. And the second figure is the spectrogram of the speech file.

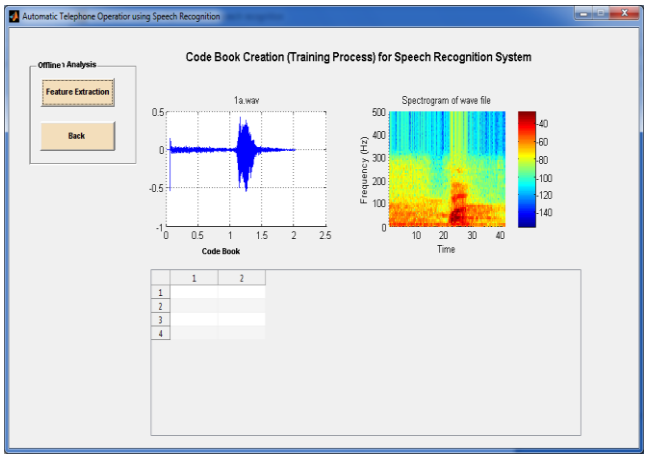

Fig.8. Code Book Creation Program Simulation

Fig.9. shows the feature table of each offline and real time speech trained files.

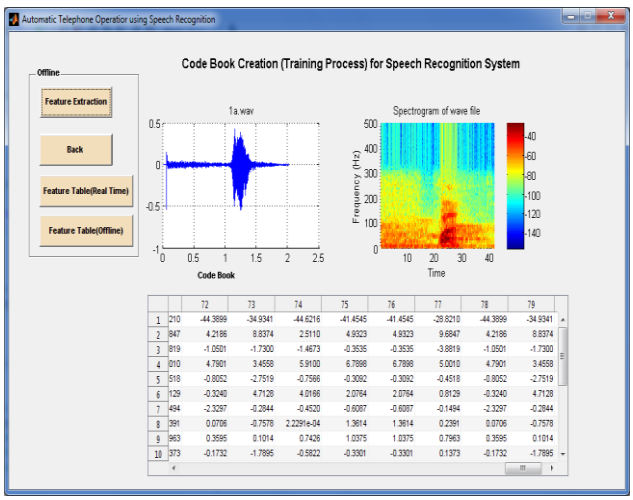

Fig.9. Code book creation with Feature Table

The resultant phone line is displayed as shown in Fig.10. If the tested speech is not recognized or the computed distance value is greater than threshold value, the error message is displayed as shown in Fig.11.

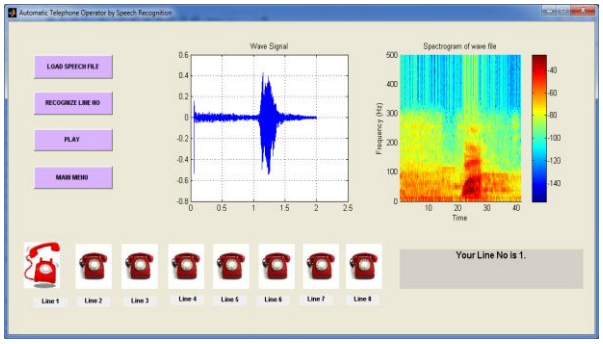

Fig.10. Offline Test

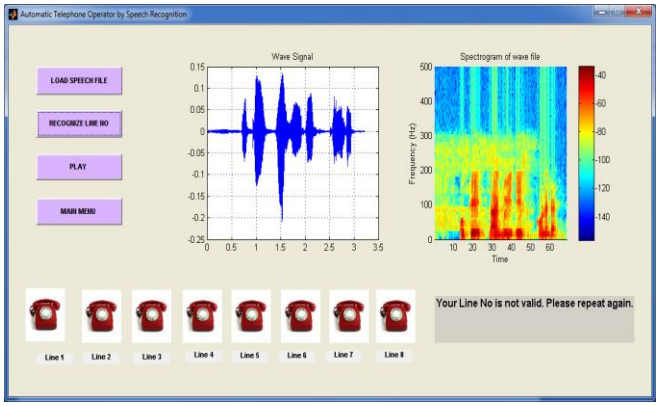

Fig. 11. The Unrecognized Speech and Error Message

The summery of percentage correct for offline test and real-time test are shown in Table 1 and 2. In offline test, the percentage correct is got high result. But in real time test, the system has got low result due to the surrounding noise and other effect such microphone quality and mouse click noise. They can be eliminated by using DSP filters and other voice enhancement programs.

Table 1. Percentage Correct for Offline Test

Telephone Line	No of Test Files	Number of Correct Files	Number of wrong results	% correct
1	10	10	10	100
2	10	10	10	100
3	10	10	10	100
4	10	10	10	100
5	10	10	10	100
6	10	10	10	100
7	10	10	10	100
8	10	10	10	100

Table.2. Percentage Correct for Real Time Test

Telephone Line	No of Test Files	Number of Correct Files	Number of wrong results	% correct
1	10	7	3	70
2	10	4	6	40
3	10	6	4	60
4	10	5	5	50
5	10	4	6	40
6	10	8	2	80
7	10	5	5	50
8	10	5	5	50

5) CONCLUSION

Speech Recognition technique makes it possible to use the speaker's voice to verify their identity and control access to services. In this system, the user has to start the feature extraction step to generate the code book for further processing. The telephone line numbers from one to eight are recorded 10 files for each channel. They are saved in computer storage and extracted the features. The user can also record their speeches by wavrecord command in Matlab. The system has got high performance, but with the use of high quality microphone and some additional features such as pitch, power spectral density, etc.

ACKNOWLEDGMENT

The author would like to express my gratitude to all colleagues at Mandalay Technological University who have contributed to the preparation of this research work.

REFERENCES

[1]Martinez, J., Perez, HEscamilla, E.Suzuki, M.M"Speaker recognition using Mel frequency Cepstral
Coefficients (MFCC) and Vector quantization (VQ) techniques" 22nd International Conference on Electrical Communications and Computers (CONIELECOMP), pp: 248-251, 27 29 Feb 2012
[2]Z. Goh et al., "Kalman Filtering Speech Enhancement Method Based on a Voiced Unvoiced Speech Model",IEEETrans. Speech Audio Processing, vol.7,pp. 510-524, Sep1999
[3]Y. Romanyshyn,V. Hudym, "Wavelet transforms applications for speech signals processing" ,
6th International Conference on The Experience of Designing and Application of CAD Systems in Microelectronics, 2001. CADSM 2001, pp: 297 298, 12 17, Feb 2001
[4] Arvinder Singh, Gagandeep Singh, "Speech Recognition Based System to Control Electrical Appliances" International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 2, pp 81-83, August 2012

Thi Yi Kywe is an academic staff at Electronic Communication Division, Department of Electronic Engineering, Mandalay Technological University. She has completed her master degree in the field of signal processing since 2004 and supervising the research related in the field of signal processing and control engineering.