Implementation Speech Recognition for Robot Control Using MFCC and ANFIS

Trima Mustofa
Electrical Engineering Post Graduated Program, Sultan Agung Islamic University
Jl. Raya Kaligawe KM 4 Semarang 50112, Indonesia
e-mail: soetadiwirya@gmail.com

Abstract

Speech recognition is one of machine intelligence field that grow rapidly, it is recognized by almost all technology devices that equipped by voice command. The process of converting voice patterns into text through complex transcription process. There is an enormous complexity involved in analyzing speech input, such as variations in pronunciation, accent, speaker physiology, emphasis and acoustic environmental characteristics in generating hundreds of different phoneme classifications for each sound. A sound pattern recognition system is required to process sound signals quickly and realtime in recognizing voice input with accurate results. In this case MFCC is an appropriate method to apply sound extraction because it presents the signal well. While ANFIS is needed for sound characteristic learning because it has advantages possessed by fuzzy inference system and artificial neural network system. In this research tested through robot control.

Keywords*: ANFIS, MFCC, Speech Recognition, Robot Control*

1. Introduction

Speech recognition is one of machine intelligence field that grow rapidly, it is recognized by almost all technology devices that equipped by voice command. It has attracted researchers to make speech recognition as an important discipline and created technological impact on society, and it is expected to be further developed in the field of machine interaction with humans.

Recognizing sound patterns is a difficult issue. The ultimate goal of all speech recognition research is to create an intelligent system, that listens to what the user says and then performs the instructions ordered. One simple example is the use of voice commands to replace keyboard functions on a computer or smartphone. The process of changing the sound pattern into the text looks simple, but in reality, it should pass a complicated transcription process.

One of the major problems in recognizing and understanding computer speech and language is to overcome a lot of searching space. Currently, the most widely used of identifier system is based on Hidden Markov Modeling (HMM). In this system, typically, each word is represented as a phoneme sequence, and each phoneme is associated with a phoneme sequence, then each phoneme is associated with a state sequence. If the number of vocabulary increases, the search space size too. As the size of the search space increases, the sound pattern recognition performance decreases.

Another problem in speech recognition is the enormous complexity involved in analyzing speech input. Variations in pronunciation, accents, speaker physiology, emphasis and acoustic environmental characteristics typically produce hundreds of different phoneme classifications for each sound. It leads to many phoneme classifications that produce many word probabilities at each point generated. All of these words can then be combined to generate hundreds of possible sentences for each speech. Search space generated becomes very large. Therefore, it takes a sound pattern recognition system that can process voice signals quickly to filter out all the wrong probabilities and recognize real-time sound input with accurate results in this case MFCC is the right method because this method can work well in presenting signal. While ANFIS is needed for sound characteristic learning because it has advantages possessed by fuzzy inference system and artificial neural network system.

In developing speech recognition technology that previously only aimed at how to change the voice into the text. Now the focus of research is needed to make speech recognition technology into integrated sound pattern recognition systems, that is, the output of voice pattern recognition can be used to complete a task or a control. One of the implementation is on voice command on robot control. It underlies the research to implement the concept of voice recognition into a voice recognition system with MFCC and ANFIS with the output of the voice recognition is being used as a command for navigation of control robots.
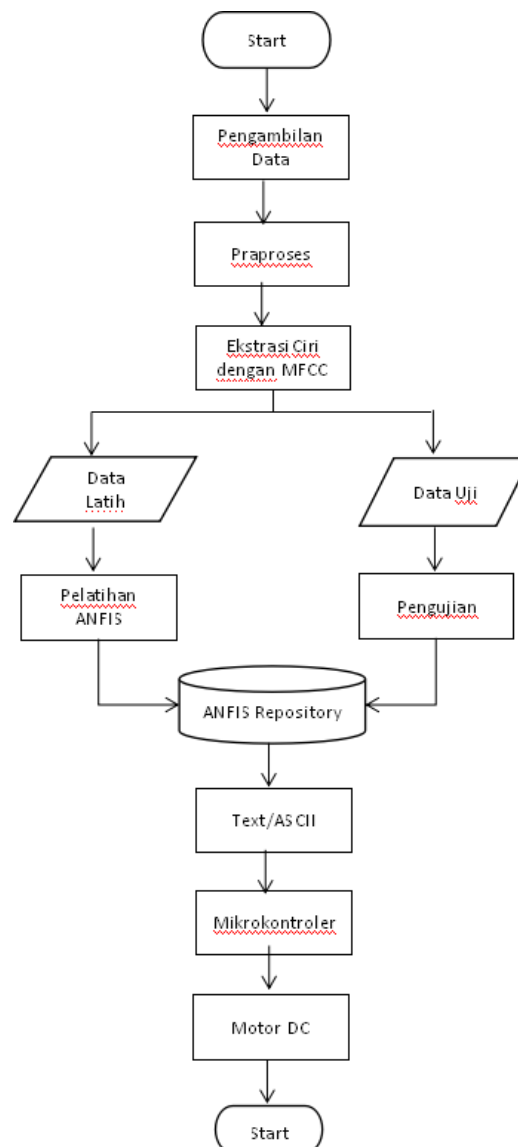
2. Research Method

Figure 1. Research Flowchart

In this study there are two methods used for the recognition of sound patterns, they are MFCC and ANFIS. The MFCC method is used for sound character extraction while ANFIS is used to recognize previously extracted sound patterns. As for the pattern recognition results are tested on the mobile robot.
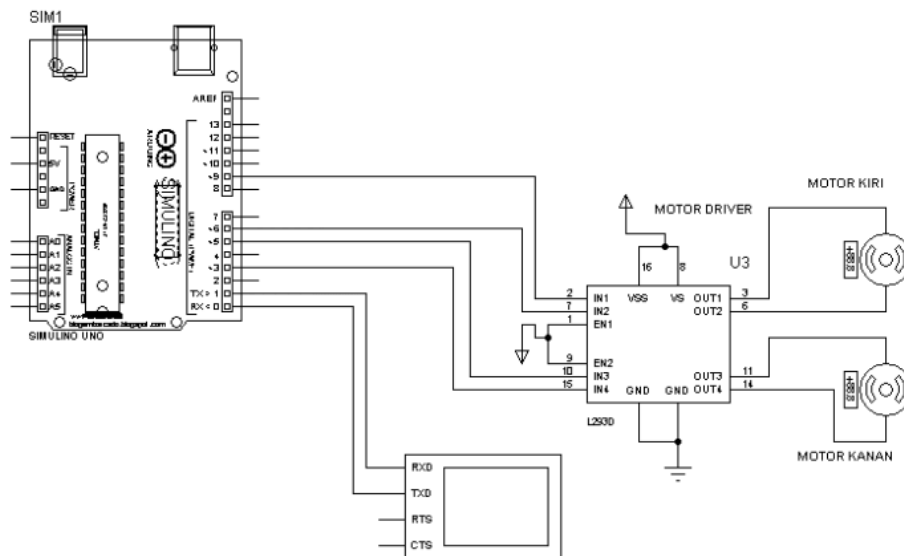
Figure 2. Block Diagram of Robot Control

Figure 2 is a series of control robots that means for testing speech recognition made. In the circuit that consists of arduino as controller for robot movement, Motor Driver is a part that serves to drive a DC motor where the change in direction of the DC motor depends on the value of the input voltage of the driver itself and the DC motor serves to move the wheel. Meanwhile, to move the robot is using voice commands.

Voice recognition process begins with sound feature extraction performed by ANFIS method. The simple workings of MFCC are based on the different frequencies that can be heard by the human ear, thus able to express the sound signals as how humans present them. The process of MFCC are as follows:
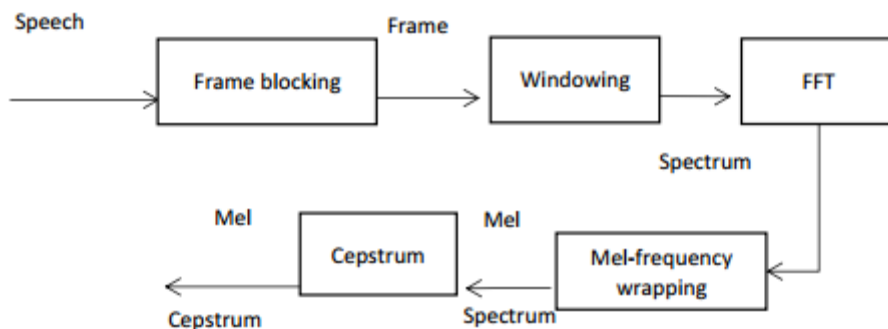

Figure 3. Algoritma MFCC

2.1. Preemphasis

At this stage, it is used to flatten the signal spectrum and increase the signal's authenticity in subsequent signal processing.

$$y(n) = x(n) - ax(n-1) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

2.2. Frame Blocking

In this process, the sound signal is segmented into multiple overlap frames. This is done in order to prevent lost signals (deletion).

2.3. Windowing

Analog signals that have been converted into digital signals are read frame by frame and in each frame are done windowing with certain window function. The windowing process aims to minimize unsustainability signal at the beginning and end of each frame. If we define the window as w (n), 0 ≤ n ≤ N - 1, where N is the number of samples in each frame, then the result of windowing is a signal:

$$y1(n) = x1(n)w(n), 0 \leq n \leq -1 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(2)$$

2.4. Fast Fourier Transform (FFT)

FFT is a fast algorithm of Discrete Fourier Transform (DFT) which is useful for converting each frame with N samples from time domain into frequency domain, as defined as follows:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\partial \, jkn/N} \dots\dots\dots\dots\dots\dots\dots\dots\dots(3)$$

n= 0, 1, 2, … , n− 1 dan n= √−1.

The result of this stage is usually called as a spectrum or periodogram.

2.5. Mel-Frequency Wrapping

The perception of the human auditory system to the frequency of the sound signal can not be measured on a linear scale. For each tone with the actual frequency, f, measured in Hz, a subjective pitch is measured on a scale called "mel". The mel-frequency scale is a low frequency that is linear under 1000 Hz and a high frequency that is logarithmic over 1000 Hz. The following equation shows the relation of the mel scale to the frequency in Hz:

$$F_{mel} = \begin{cases} 2595 * \log 10\left(1 + \dfrac{F_{Hz}}{700}\right) & F_{Hz} > 1000 \\ F_{Hz}, & \\ & F_{Hz} < 100 \end{cases} \dots\dots\dots\dots\dots\dots\dots(4)$$

2.6. Cepstrum

At this stage, it will convert the mel-spectrum into time domain by using Discrete Cosine Transform (DCT). The result is called mel-frequency cepstrum coefficient (MFCC).
Here are the equations used in cosine transformations:

$$C_j = \sum_{i=1}^{M} X_j \cos\left(j(i-1)/2\frac{\pi}{M}\right) \dots\dots\dots\dots\dots\dots\dots\dots(5)$$

The speech recognition process begins with sound input and the voice signal in the process of using the MFCC method as has been described in the previous chapter. From the extraction process, it will get the data train. Voice extraction process in this research takes some sample of voice that will be used as data train. This recording process is using Matlab 2013 a tools, from the recording process will produce coefficients for each recording done. There are 6

characteristics of each sound produced then marked with the target. As for the words of the recording is the word "Maju", "Mundur", "Kanan", "Kiri" and "Berhenti". Each respondent performed 10 times recording for all words. Here are the results of the extraction process and its characteristics.

```
Command Window
>> program_rekam_suara
Warning: WAVRECORD will be removed in a future release. Use AUDIORECORDER instead.
> In wavrecord at 42
  In program_rekam_suara at 2
Warning: WAVPLAY will be removed in a future release. Use AUDIOPLAYER instead.
> In wavplay at 41
  In program_rekam_suara at 11

koef =

    3.1729    1.9254    2.0427    1.1537    0.5545    -0.0041
```

Figure 4. Result of a sound recording process that produces a coefficient for the word Maju.
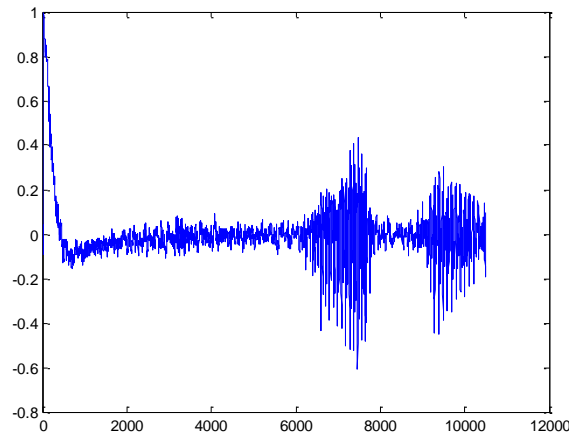

Figure 5. Signal Output Speech

Training data obtained by sound pattern extracting process is the learning process of voice data by using ANFIS. Here is the plot:
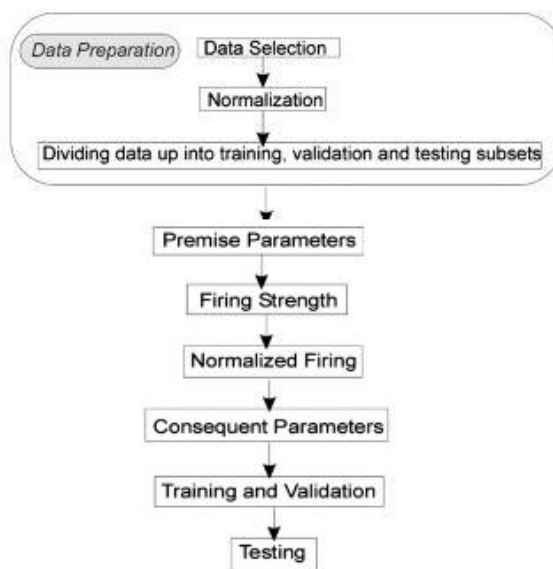

Figure 6. Speech Recognition Method with ANFIS

The next step for voice pattern recognition is to prepare voice data that will be used as a feature. At this stage a sound recording process is done to get a sound sample. Furthermore, the process of normalization If the process of normalization has been done, then the sound data is ready used for train data. The training data is then processed by using the ANFIS method that begins with fuzzyfication process up to the pattern recognition test.
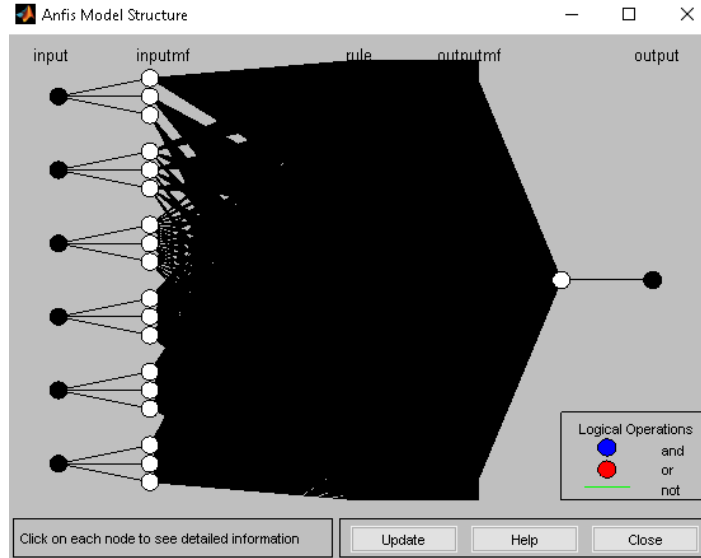

Figure 7. Structur of ANFIS

Figure 7 is ANFIS Structure in this research, where in ANFIS structure has 5 layer, thay are input layer, inputmf, rule, outputmf, and output.

In order to know the performance of ANFIS that has been done with training the train data, the next step is to plot the train data and ANFIS with the same input value.

3. Results and Analysis

After designing both hardware and software, then the next process will be the testing process of some parts that are designed. Starting from the sound recorder process for voice data is continued with a speech recognition process created by using GUI in MATLAB up to trial process on hardware devices.

3.1. Speech Recording Proccess

The first step to voice pattern recognition is to prepare the voice data that will be used as training data. The results of the sound recording process are as follows:

Table 1. Adult Male Speech Extraction Results Using MFCC

characteristic 1	characteristic2	characteristic 3	characteristic4	characteristic5	characteristic 6	Target
2.5261	0.6112	0.5964	0.4197	0.1889	0.0296	1
2.7744	1.3999	1.3613	0.9507	0.42	0.0607	1
5.9635	2.1767	3.0628	-1.821	1.9641	-0.4829	2
3.3914	0.1815	1.0173	0.0006	1.6087	-0.2434	2
2.9803	1.2161	1.7639	-0.5574	0.2981	-0.2163	3
2.1963	0.6764	0.6884	0.5286	0.2854	0.0762	3
2.175	-0.5392	1.0219	-1.0913	2.3186	-0.1043	4
3.7334	-0.8777	1.139	-1.2096	2.7306	-0.1517	4
-1.5991	0.10649	1.0455	0.3473	-0.9423	-0.6354	5
-1.5202	-0.19757	0.9519	0.60638	-0.6871	-0.6904	5

3.2. Learning Speech Recognition

After obtaining the data of voice training, the next step is to conduct the process of learning sound by ANFIS. The following is the results of the voice data training.
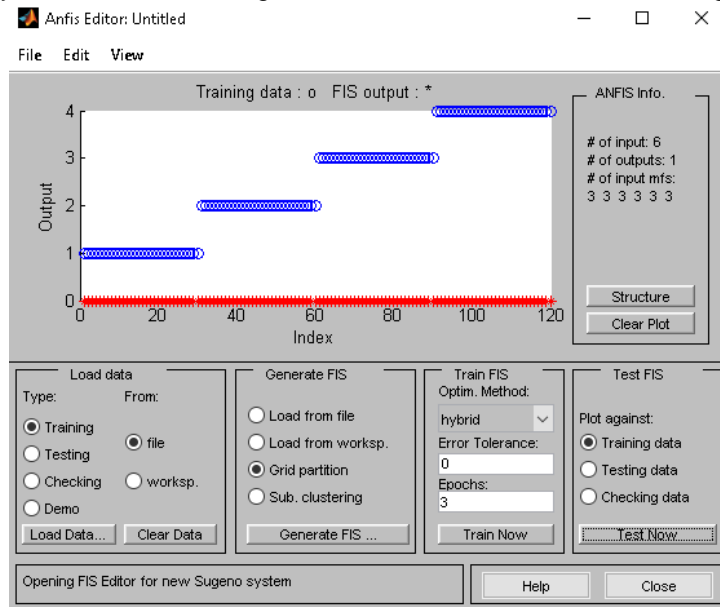

Figure 8. Output of ANFIS traning

In Figure 8 we can see ploting of training data and ANFIS, and the result of ANFIS has been able to map input and output well. The next step is to test the trained sound data by using ANFIS. Here is an example of testing the word "maju".

```
>> mfcc_anfis
        out =

            0.8

    MAJU
```

Figure 9. Results of learning output by ANFIS

After the test data that has been tested the value is considered in quite accurate, the next step is to make the GUI for testing the train data to make it easier in the process of analyzing test data.

3.3. Testing Speech Recogntion On Robot

The result of the ANFIS sound recognition pattern then there are experiments on the control robot of this research done on simulation in Matlab 2013a, the step is by connecting serial port on arduino with display for robot movement test as in Figure 9.
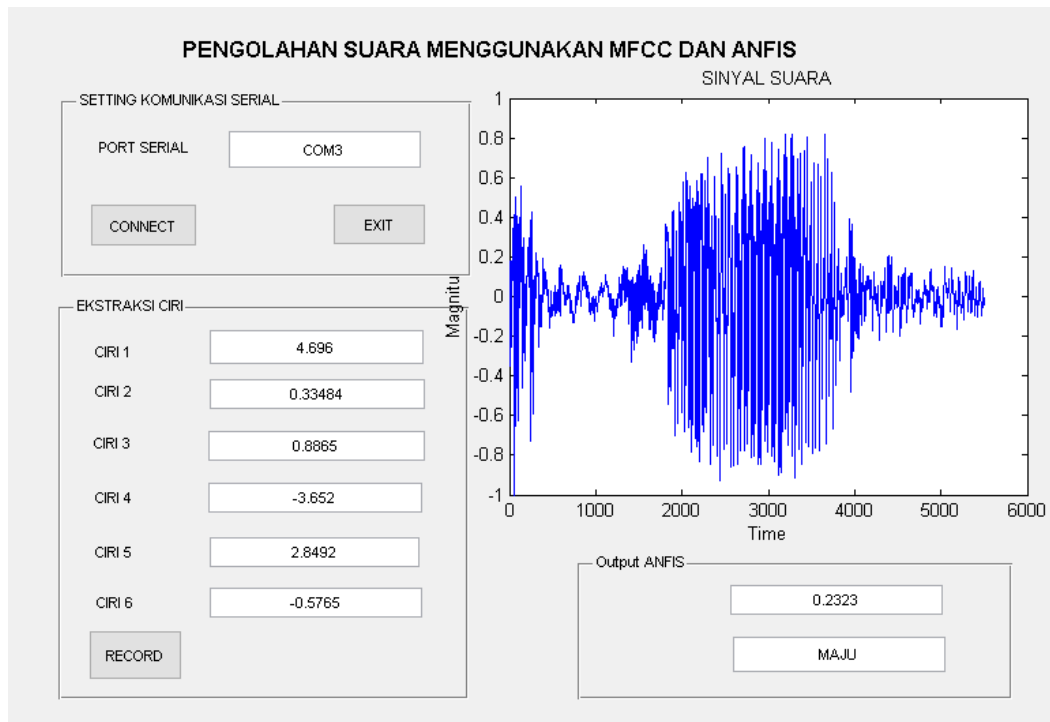
Figure 10. Speech Recognition Test Results

If that is connected, then it can be tested for speech recognition on the robot control For the test process by pressing the record button on the GUI then say the word "MAJU", "MUNDUR", "KANAN", "KIRI", and "BERHENTI" The results of the test of speech recognition that has been tested on the random sample are as follows:

Table 2. Speech Recognition Test Results

No	The Spoken Word	Target	Output	Result
1	KANAN	2.25 - 3.25	2.9803	Valid
2	KANAN	2.25 - 3.25	2.3689	Valid
3	KANAN	2.25 - 3.25	1.2890	Valid
4	KIRI	3.26 - 4.25	2.9876	Invalid
5	KIRI	3.26 - 4.25	3.7334	Valid
6	KIRI	3.26 - 4.25	4.1141	Valid
7	MAJU	0 - 1.25	0.8234	Valid
8	MAJU	0 - 1.25	0.1889	Valid
9	MAJU	0 - 1.25	0.5964	Valid
10	MUNDUR	1.26 - 2.25	1.9641	Invalid
11	MUNDUR	1.26 - 2.25	2.1767	Valid
12	MUNDUR	1.26 - 2.25	2.5787	Valid
13	BERHENTI	4.26 - 5.00	4.5671	Valid
14	BERHENTI	4.26 - 5.00	3.5678	Invalid
15	BERHENTI	4.26 - 5.00	3.9876	Invalid

4. Conclusion

After going through the process of testing speech recognition that is facilitated by robot control, this study can be concluded that Speech recognition using MFCC and ANFIS method can be applied into robot control navigation system. The results showed that the applied method was able to give good performance.

Reference

[1] Lee, K.F. and Hon, H.W., 1989. Speaker-independent phone recognition using hidden Markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *37*(11), pp.1641-1648.

[2] Sanjaya WS, Mada. 2016. Panduan Membuat Robot Cerdas Menggunakan Arduino dan MATLAB. Jogjakarta : Andi Ofset.

[3] Srivastava, N., 2014. Speech Recognition using Artificial Neural Network. International Journal of Engineering Science and Innovative Technology (IJESIT), 3(3).

[4] Antanas Lipeika, Joana Lipeikiene, Laimutis Telksnys, "Development of Isolated Word Speech Recognition System", INFORMATICA, Vol. 13, No. 1, pp. 37–46, 2002.

[5] Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal, "Bangla Speech Recognition System using LPC and ANN", Seventh International Conference on Advances in Pattern Recognition, IEEE, February 2009.

[6] Vimal Krishnan V R, Athulya Jayakumar, Babu Anto.P, "Speech Recognition of Isolated Malayalam Words Using Wavelet Features and Artificial Neural Network", 4th IEEE International Symposium on Electronic Design, Test & Applications, 2008

[7] Ahmad A. M. Abushariah, Teddy S. Gunawan, Othman O. Khalifa, Mohammad A. M. Abushariah, "English Digits Speech Recognition System Based on Hidden Markov Models", International Conference on Computer and Communication Engineering, May 2010.

[8] R. B. Shinde , Dr. V. P. Pawar, "Vowel Classification based on LPC and ANN", International Journal of Computer Applications, Vol. 50 No.6, pp. 27-31, July 2012.