International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

A ROBUST DEPENDENT AND INDEPENDENT SPEECH RECOGNITION BY USING ASR SYSTEM

N. Naresh Babu[1], N. Keerthana[2], G. Prachi[2], K. Sai Kirthana[2], G. Harshitha[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering
[1,2]Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad, Telangana, India

Abstract: It is proposed in this research to combine the discrete wavelet transform (DWT) with to feature extraction in order to enhance speech/speaker A Spectral Relative (ASR) detection. Different components of speech, such as speech recognition, speaker identification and speaker synthesis, all are connected to this. The goal of this research is to investigate the use of HMM with voice recognition. This method aims to improve performance besides introducing more features from either the speech signal including using parallel computations, which results in an increase between the recognition rate as well as computational speed both for the clean as well as noisy speech signals again for proposed method compared towards the HMM model.

Keywords: HMM Model, DWT, median filter RASTA-PLP.

1. INTRODUCTION

The most difficult part of building a voice recognition system was getting the recognition rate as accurate as possible while yet spending as little time as feasible on training and validation the system. We were able to solve these issues by selecting a decent feature extraction approach and implementing your neural network classifiers in parallel. Herein lay the nub of our project. Combining DWT with RASTA-PLP allows for feature extraction. Their multi-resolution, multi-scale analytical capabilities of both the DWT make it ideal for processing non-stationary data like voice. RASTA-key PLP's benefit is that they are resistant to noise. That method's goal is to improve the new method's performance with feature extraction, resulting in a greater detection accuracy than that achieved by combining DWT and MFCCs-based methods.

2. EXISTING SYSTEM

This same wavelet-based HMM model methodology is what we used to propose speech/speaker recognition throughout this present method. The voice signal is used to create the wavelet basis. Its primary purpose was to satisfy a time-and-bandwidth-consumption product. Even before FFT, it must have been employed to do the preprocessing. In just this case, some noise has been added.

Voice information may be utilised to identify individual speaker based on the anatomy of both the vocal tract, which really is unique to each individual. Speaker recognition was the procedure of identifying a person based on sound they make. Voice falls within the area called biometric identification because of the inherent distinctions inside the speaker's physical structure. There are several benefits to using one's voice as both a form of identification. Remote person authentication is indeed a big benefit. A speaker recognition system will have the same training and testing steps as every other pattern recognition system. In order to get a handle on how the system processes speech, training is necessary. Recognizing is done by testing. Figure 2.1 depicts the training period as just a block diagram. For develop the reference models, feature vectors reflecting the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

speaker's voice characteristics were retrieved using training utterances.
The degree to which the test utterance's related feature vectors match the others in the reference was determined utilising some matching approach during testing. The choice is made based on the level each match. This testing phase's block diagram can be seen here.

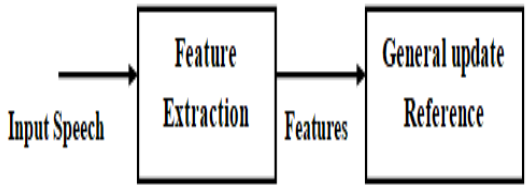

Fig.2.1: *The block diagram of training phase.*

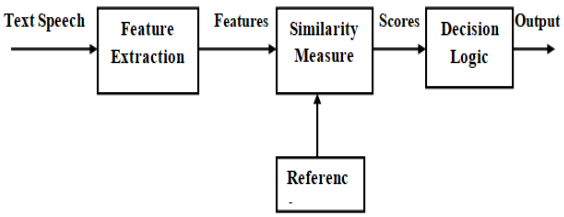

Fig.2.2:*The block diagram of the testing phase*

Every speech recognition system's extraction & selection of such an acoustic parametric description is important to its overall performance and accuracy. These MFCC (mel-frequency cepstrum coefficients) On either a mel-frequency scale, when are the cosine transforms of either the real logarithm of this with the short-term energy spectrum defined.

3. PROBLEM IDENTIFICATION

- Recognition process is less compare to mimicry voice
- Noise is very high

4. PROPOSED SYSTEM

Speech recognition was one of the ideas we floated. This RASTA-PLP (relative spectral perceptual linear prediction) approach is being used in the method.High pass filters throughout the filter bank output may be used by Rasta to eliminate the sluggish variations throughout the components of either the mg and mg issues caused by communication channels.

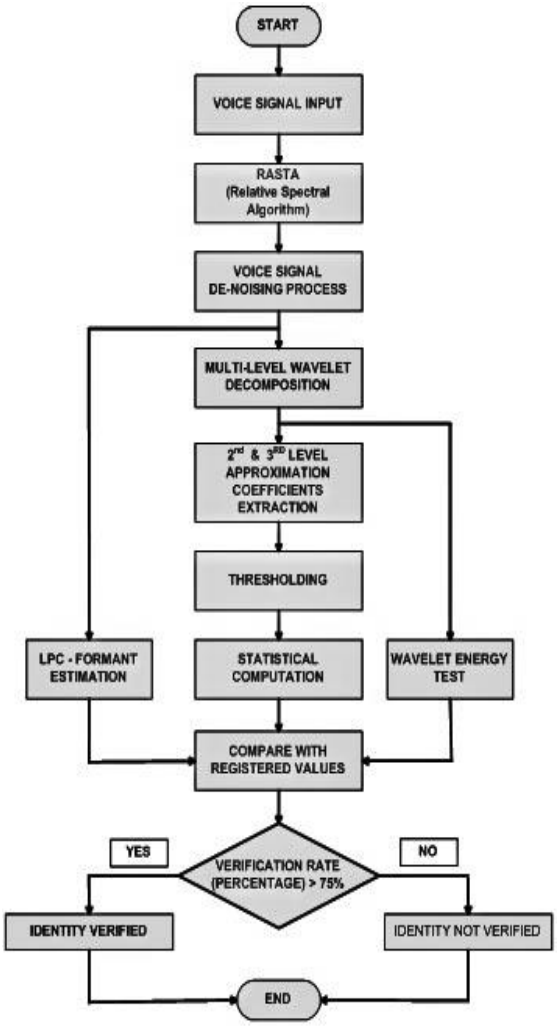

Fig4.1: *System flow chart*

4.1. Discrete Wavelet Transform

When it comes to the DWT, two separate scales and places are picked depending on their respective capabilities. Powers of 2 are used to rescale and dilate original mother wavelet, whereas integers are used to translate it. You might demonstrate f(t), which specifies the space

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

of square integrable functions, to be f(t) ∈ L2(R), by using the notation

$$f(t) = \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\, \Psi(2^{-j}t - k) + \sum_{k=-\infty}^{\infty} a(L,K)\emptyset(2^{-L}t - k) \ldots\ldots\ldots\ldots (1)$$

This function (t)φ is referred to as the scaling function, whereas the function φ(t) was referred to as one of the mother wavelet. Functions as a group

$$\{\sqrt{2^{-1}}\,\emptyset(2^{-j}t - l) | j \le L, j, k, L \in Z|\}$$

Using Z as an orthonormal basis of L2, we get L2 (R). For a scale L, the approximation coefficients were denoted by a(L,k), whereas the detail coefficients have been designated by a(J,k). The following is an expression for the approximations and detail coefficients:

$$d(L,K) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\, \emptyset(2^{-L}t - k) dt \ldots.(2)$$

$$d(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\, \Psi(2^{-j}t - k) dt \ldots(3)$$

These coefficients above may well be explained by the projection fl(t) on f(t) gives perfect calculation (in the sense of least error energy) of f(t).These coefficients a(L, k) might be used to generate this projection.

$$f_1(t) = \sum_{k=-\infty}^{\infty} a(L,K)\emptyset(2^{-L}t - k) \ldots\ldots (4)$$

Scale l lowers, and the approximations becomes finer, eventually reaching f(t) when l drops to zero. It is quite possible to explain the difference in approximation among scales of approximation, fl+1(t) and scales of approximation, fl(t), through using coefficients d(j,k).

$$f_{l+1}(t) - f_1(t) = \sum_{k=-\infty}^{\infty} d(l,k)\varphi(2^{-l}t - k) \ldots (5)$$

Using those relations, given a(L, k) & {d(j, k) | j ≤ L}, It is obvious that now the approximation could be constructed at any size. As a result, the wavelet transform produces a rough approximation of the original signal.fL(t) (given a(L, k)) and a number of layers of detail {fj+1(t)-fj(t)| j < L} (given by {d(j, k) | j ≤ L}). The closer the actual scale is to the next finer layer of information, the more accurate the final result becomes.

4.2. Vanishing Moments

This number of vanishing moments in either a wavelet demonstrates the smoothness and flatness of just a wavelet function like a frequency response, respectively (filters used to compute the DWT). P-vanishing-moment wavelets satisfy the following formula.

$$\int_{-\infty}^{\infty} t^m \varphi(t) dt = 0 \qquad for\ = 0, \ldots\ldots, p-1$$

Or equivalently,

$$\sum_k (-1)^k k^m c(k) = 0\ for\ m = 0, \ldots . p-1$$

The more vanishing moments that were in the representation given smooth signals, the faster the wavelet coefficients decayed. A wavelet with a high number of point sets effectively represents signals, which is beneficial for coding purposes. As the number of vanishing moments as well as the size of something like the wavelet filters rise, so does the computational difficulty of calculating the DWT coefficients. In the image below, we can see the low-frequency approximation

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

coefficients cA1 as well as the high-frequency detail coefficients cD1.

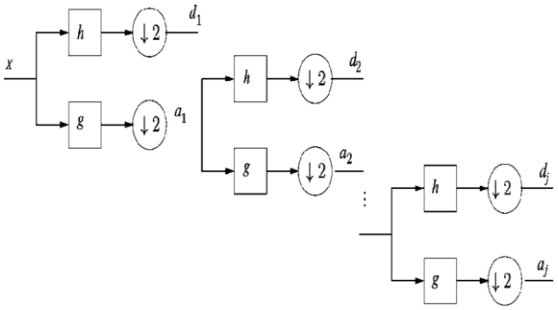

Fig.4.2: *Filtering operation of the DWT*

4.3. Implementation Using Filters

Fast Wavelet Transform method implementation is shown in the following formula:

$$\emptyset(t)=\sum_k c(k)\varphi(2t-k)\ldots\ldots\ldots(6)$$

$$\varphi(t)=\sum_k(-1)^k c(1-k)\emptyset(2t-k)\ldots.(7)$$

$$\sum_K C_K C_{K-2M}=2\delta_{0,m} \ldots\ldots\ldots\ldots(8)$$

The scaling function was defined by the twin-scale relation (also known as the dilation equation). Using the scaling function φ, the wavelet is expressed. In order therefore for wavelet to really be orthogonal to both the scaling function as well as its translating function, this third equation must be satisfied.

4.4. Multilevel Decomposition

In iterations, successive approximations may be deconstructed in turn to break down a single signal into several lower resolution components. This decomposition process called iterative. This same wavelet decomposition tree is indeed the name given to this structure.

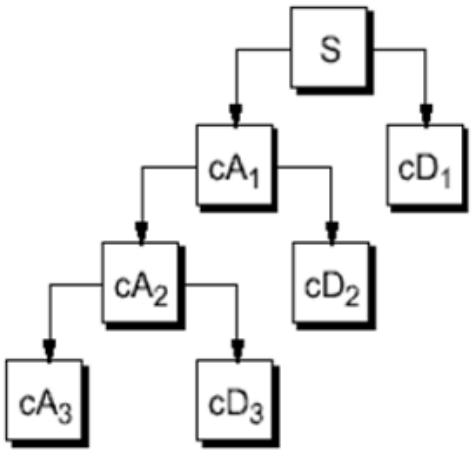

Fig4.3: *Decomposition of DWT co-efficient*

This is the wavelet decomposition of both the signals for level j: [CAj, CDj,..., CD1]. Information may be discovered by looking at such a signals wavelet decomposition tree.

4.5 Signal Reconstruction

This technique may be used to re-create or create a new signal from the old one (IDWT). Approximation coefficients cAj, CDj, are used to approximate CAj-1 before up sampling & filtering with both the reconstruction filters are used to rebuild CAj-1.

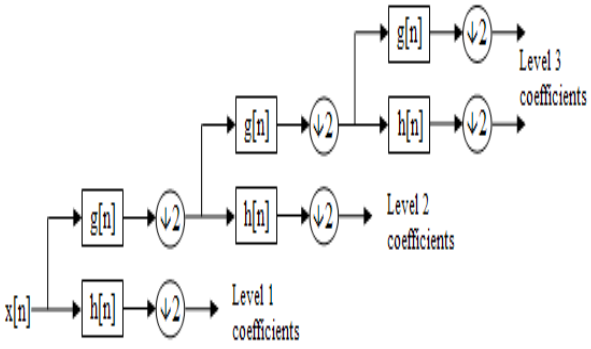

Fig4.4:*Filterlevels*

This aliasing caused inside this wavelet decomposition step is countered by the design of both the reconstruction filters. Low and high-pass decomposition filters

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

(Lo_ R and Hi_ R) are part of a quadrature mirror filters system (QMF).

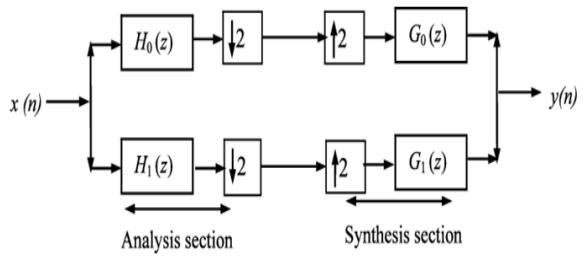

Fig4.5:*Quadrature mirror filters (QMF)*

4.6. Feature extraction: Wavelet Energy

The approximation as well as the detail each maintain a specific amount of energy whenever the signal was decomposed that use the wavelet decomposition technique. The wavelet accounting vector as well as the wavelet decomposition vector may be used to retrieve this energy. It is a ratio since it compares between original and decomposed signals, which yields the computed energy. This can only be discovered by a great deal of trial and error. Level 2 coefficients comprise the majority of the speech signal's associated data, therefore they are used to derive the wavelet decomposition coefficients.

5. RESULTS

This section gives the detailed analysis of simulation results. Further, the performance of proposed ASR system is compared with the state of art approaches using same dataset.

5.1 Dataset

The dataset contains the voice samples from five different persons. From each person 9 samples are collected with multiple phrases. Totally, 45 phrases are collected in entire dataset. Further, 80% of dataset is used for training, 20% of dataset is used for testing the ASR system.

5.2 Subjective Performance

Figure 5.1 shows the original audio with 2.5 seconds of time. Figure 5.2 shows the frequency of original audio, which is having the frequency of -3k to +3k. Figure 5.3 shows the spectrogram of MFCC. Figure 5.4 shows the spectral features of MFCC, which contain 8th order feature, cepstral feature, log-mfcc features. Figure 5.5 shows the recognized person details.

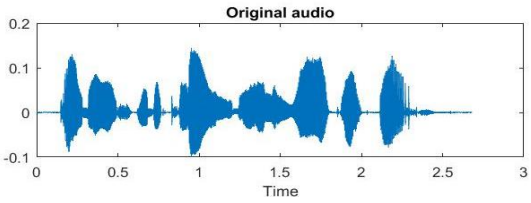

Figure 5.1. Original audio

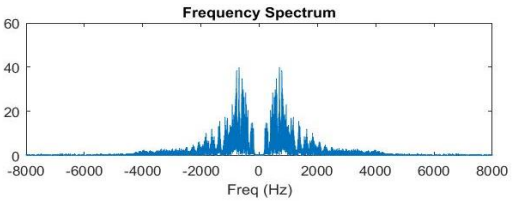

Figure 5.2. Frequency spectrum.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

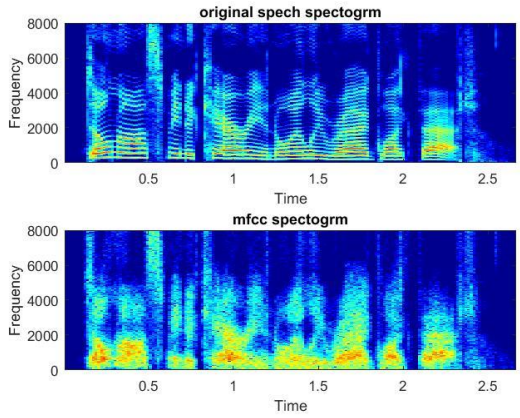

Figure 5.3. Spectrogram (a) original speech, (b) MFCC.

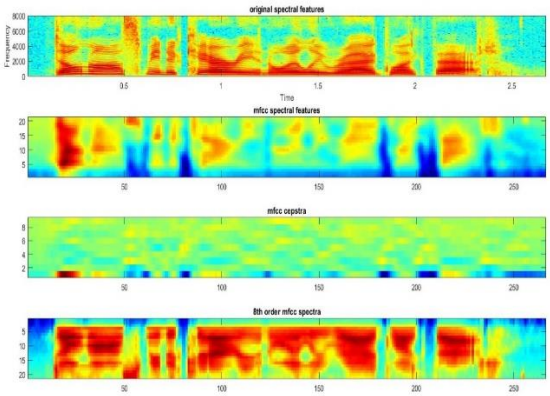

Figure 5.4. Spectral features of MFCC.

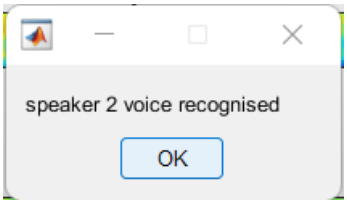

Figure 5.5. Recognized outcome.

5.3 Objective evaluation

Table 5.1. Performance comparison.

Metric	HDNN [1]	CNN [2]	DBN [7]	Proposed ASR
Accuracy (%)	92.550	94.750	94.200	99.98
Sensitivity (%)	90.307	91.267	95.715	99.973
Specificity (%)	94.770	94.104	95.669	99.83
F-measure (%)	93.466	92.825	95.257	99.67
Precision (%)	93.262	94.428	90.688	99.29
MCC (%)	92.947	95.241	90.583	99.45
Dice (%)	90.762	90.624	92.727	99.37
Jaccard (%)	93.565	94.364	91.614	99.46

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

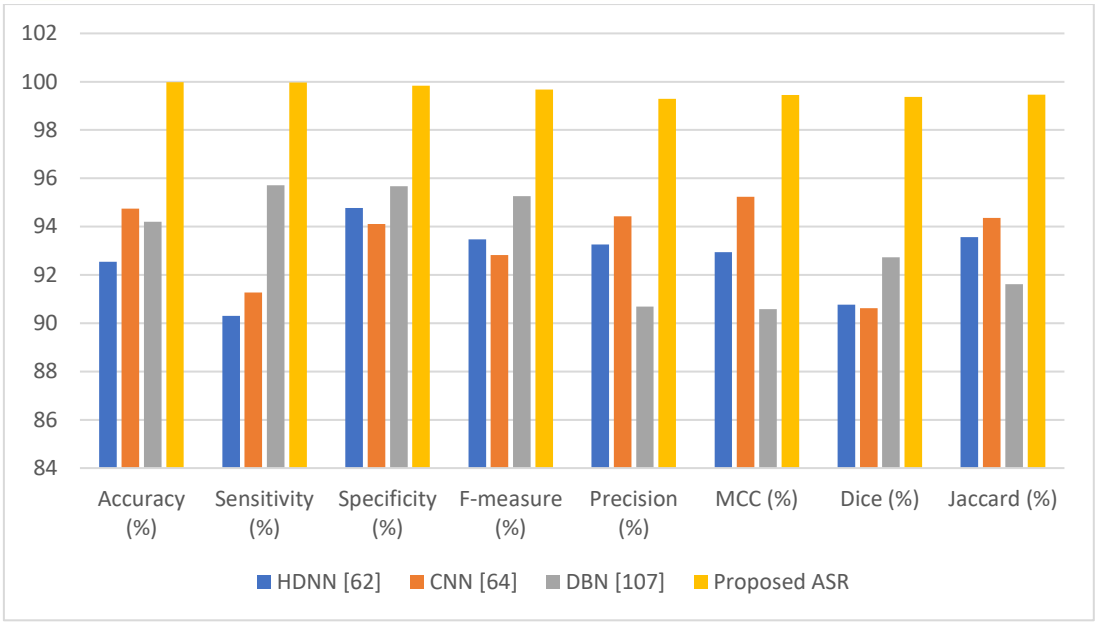

Figure 5.6. Performance comparison.

Table 5.1 shows the performance of proposed ASR system with conventional approaches such as HDNN [1], CNN [2], DBN [7]. The proposed DLCNN based ASR resulted in superior performance for all metrics. Figure 5.6 shows the graphical representation of Table 5.1. Figure 5.7 and Figure 5.8 shows the graphical representation of GUIS.

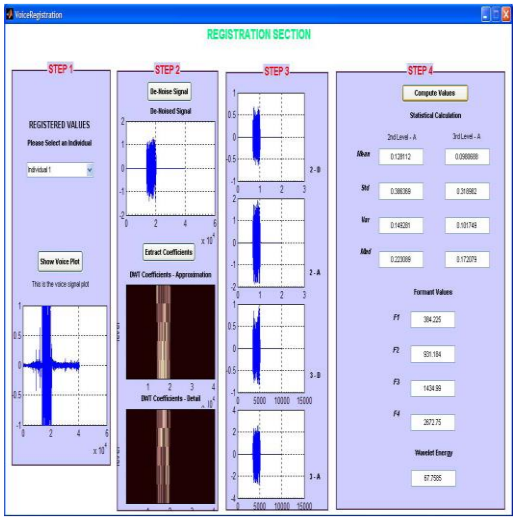

Fig5.7: Voice Registration GUI

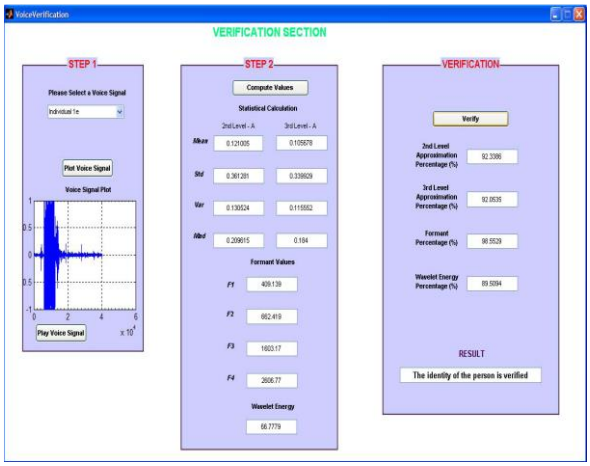

Fig 5.8: Voice Verification GUI

Applications

1)Education System

2) Industrial

3) ATM and Banking

4)Electronic (Laptop,Mobile)

6. CONCLUSION

Methods for pre-processing the voice signal vary. The most accurate feature for word identification was found using wavelet transformations, as shown in the paper.

Respectively clean and noisy voice signals have shown this to be true. Wavelets are employed to test for dependent and independent variables in noisy speech, as well as the RASTA-PLP approach produces superior findings. verification tests were conducted and a suggested algorithm was found to have an accuracy rate of roughly 90% Comparison with the HMM model was unmistakable.

FUTURE SCOPE

Our future work concentrate on increasing the accuracy of the model and also to complete the system as a whole for the railway system we took dataset for. Our model works only for trained users and the future work is to extent by using neural networks is provide better security

REFERENCES

[1]. SD.B.Paul,"Speech Recognition usingHidden Markov Model.International Magazine of Computers and Technology, No. 8, 2014 Vol. 13.

[2].M.A.Anusuya , S.K.Katti"Speech Recognition by Machine: A Review" International journal of computer science and Information Security 2009.

[3]. L.R.Rabiner and B.H.jaung ," Fundamentles of Speech Recognition Prentice-Hall, Englewood Cliff, New Jersy,1993.

[4]. Mahdi Shaneh, and Azizollah Taheri,"Voice Command Recognition System Based on MFCC and VQ Algorithms", World Academy of Science, Engineering and Technology 57 2009

[5] H. Combrinck and E. Botha, "On the mel-scaled cepstrum,"Video Communications, SPIE Electronic Imaging, UCRL Conf. San Jose, vol.200706, Jan 2004.

[6] Electronic Engineering, University of Pretoria.,Journal of Computer Science 3 (8): 608-616, 2007 ISSN 1549-3636.

[7] Ahmad A. M. Abushariah,Teddy S. Gunawan, Othman O. Khalifa"English Digits Speech Recognition System Based on Hidden Markov Models", International Islamic University Malaysia, International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[8] AnjaliBala,ABHIJEETKumar,Nidhika Birla,"Voice command recognition System Based on MFCC and DTW",International Journal of Engineering Science and Technology,Vol.2(12),2010

[9] Ibrahim Patel,Dr.Y.Srinivasa Rao, , "Speech recognition using Hidden Markov Model With MFCCSubband Technique." 2010 International Conference on Recent Trends in Information,Telecommunication and Computing.

[10] ShumailaIqbal,Tahira,Mehboob,Malik ,"Voice Recognition using HMM with MFCC for secure ATM",IJCS Vol.8,Issue 6 Nov 2011

[11] Vimala C, Dr.V.Radha, "A Review on Speech Recognition Challenges and Approaches", World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 1, 1-7, 2012

[12] Lawrence R. Rabiner, Fellow, IEEE 'A Tutorial On Hidden Markov Model And Selected Applications In Speech Recognition, Proceedings Of TheIEEE, Vol. 77, No. 2, February 1989.