Volume 2, Issue 4, April 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN

Kshamamayee Dash[*] , Debananda Padhi[2] , Bhoomika Panda[3], Prof. Sanghamitra Mohanty[4]
P.G. dept. Of CS&A,Utkal University , Vani Vihar, Bhubaneswar, Odisha,India
kshamamayee.dash@gmail.com*

***Abstract*— Speech processing is emerged as one of the important application area of digital signal processing. Various fields for research in speech processing are speech recognition, speaker recognition, speech synthesis, speech coding etc. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. Feature extraction is the first step for speaker recognition. Many algorithms are suggested/developed by the researchers for feature extraction. In this work, the Mel Frequency Cepstrum Coefficient (MFCC) feature has been used for designing a text dependent speaker identification system. BPNN is used for identification of speaker after training the feature set from MFCC. Some modifications to the existing technique of MFCC for feature extraction are also suggested to improve the speaker recognition efficiency. Information from speech recognition can be used in various ways in state-of-the-art speaker recognition systems. This includes the obvious use of recognized words to enable the use of text-dependent speaker modeling techniques when the words spoken are not given. Furthermore, it has been shown that the choice of words and phones itself can be a useful indicator of speaker identity. Also, recognizer output enables higher-level features, in particular those related to prosodic properties of speech.**

***Keywords*— Speaker identification, BPNN, MFCC, speech processing, feature extraction, speech signal**

I. Introduction

Human beings are able to recognize someone just by hearing him or her talk. So, few seconds of speech are sufficient to identify a familiar voice. Speaker recognition, which can be classified into identification and verification, is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services. We are basically dealing with the speaker identification task. Speaker Identification uses the acoustic features of speech in biometric point of view and that have been found to differ between individuals. Human speech contain several discriminative features. To identify the authenticate speaker by considering the features is the application of our task. In this paper we want to do the same Feature extraction using MFCC and our main aim to do the speaker identification task using Back propagated Artificial Neural Network.

The human speech contains numerous discriminative features that can be used to identify speakers. Speech contains significant energy from zero frequency up to around 5 kHz. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. The property of speech signal changes markedly as a function of time. To study the spectral properties of speech signal the concept of time varying Fourier representation is used. However, the temporal properties of speech signal such, as energy, zero crossing, correlation etc are assumed constant over a short period. That is its characteristics are short-time stationary .Therefore, using hamming window, Speech signal is divided into a number of blocks of short duration so that normal Fourier transform can be used. In this work, the Mel frequency Cepstrum Coefficient (MFCC) feature has been used for designing a text dependent speaker identification system. The extracted speech features (MFCC's) of a speaker are quantized to a number of centroids using vector quantization algorithm. These centroids constitute the codebook of that speaker. MFCC's are calculated in training phase and again in testing phase. Speakers uttered same words once in a training session and once in a testing session later. The Euclidean distance between the MFCC's of each speaker in training phase to the centroids of individual speaker in testing phase is measured and the speaker is identified according to the minimum Euclidean distance. The code is developed in the MATLAB environment and performs the identification satisfactorily.

In this paper , Requirement of SI & the Proposed system is given in section II, in section III the technique of SI, & in section IV & V details about MFCC and comparison of diff. Implementation of filter in MFCC is given. In section VI we discussed about BPNN and result calculation etc. Conclusion and future works are given in section VII.

II. Requirement of SI and Proposed SI System

A. *For Authentication*

Speaker identification for authentication allows the users to identify themselves using nothing but their voices. This can be much more convenient than traditional means of uthentication which require to carry remember a PIN. There are a few distinct concepts of using the human voice for authentication, i.e. there are different kinds of speaker identification systems for authentication purposes:

***Single pass phrase system** is* single pass phrase system lets the user chose a phrase that is uttered in enrolment as well as for authentication. Therefore, text dependent speaker recognition techniques can be used, which has the advantage that a good identification accuracy can be achieved with very little speech data in training as well as test. For a replay-attack of a system of this kind, an intruder needs a recording of the correct pass phrase uttered by the corresponding trained user of the system. If the user keeps his pass phrase secret, such a recording is difficult to obtain for the intruder unless he has the possibility to "steal" the user's voice during the training or authentication process.

Text prompt system: A text prompt system requires the user to utter a specific text which is generated individually for each authentication. As an example, a series of digits from "zero" to "nine" may be used. But also the generation of arbitrary phrases which are to be spoken by the person to be authenticated is conceivable. Depending on the kind of prompt, the speaker recognition technique may be text dependent as well as text independent.

Speaker verification integrated within a dialog system: If biometric authentication is desired in combination with a dialog system that performs automatic speech recognition, a third kind of speaker authentication system may be the most useful. In this case, the utterances of the user which are made in order to provide some kind of information to the system can be used for the authentication purpose as well.

B. Speaker Identification for Surveillance

Security agencies have several means of collecting information. One of these is electronic eavesdropping of telephone and radio conversations. As it results in high quantities of data, filter mechanisms must be applied in order to find the relevant information. One of these filters may be the recognition of target speakers that are of interest for the service.

C. Forensic Speaker Identification

Proving the identity of a recorded voice can help to convict a criminal or discharge an innocent in court. Although this task is probably not performed by a completely automatic speaker recognition system, signal processing techniques can be of use in this field nevertheless.

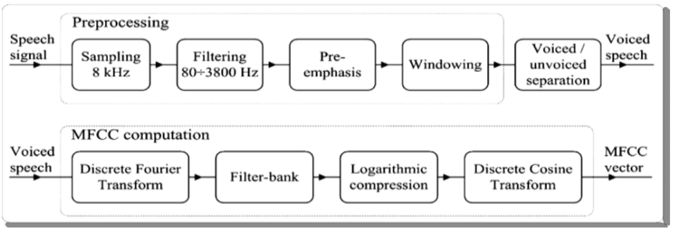

Figure 1: Diagram of the speech pre-processing and MFCC computation

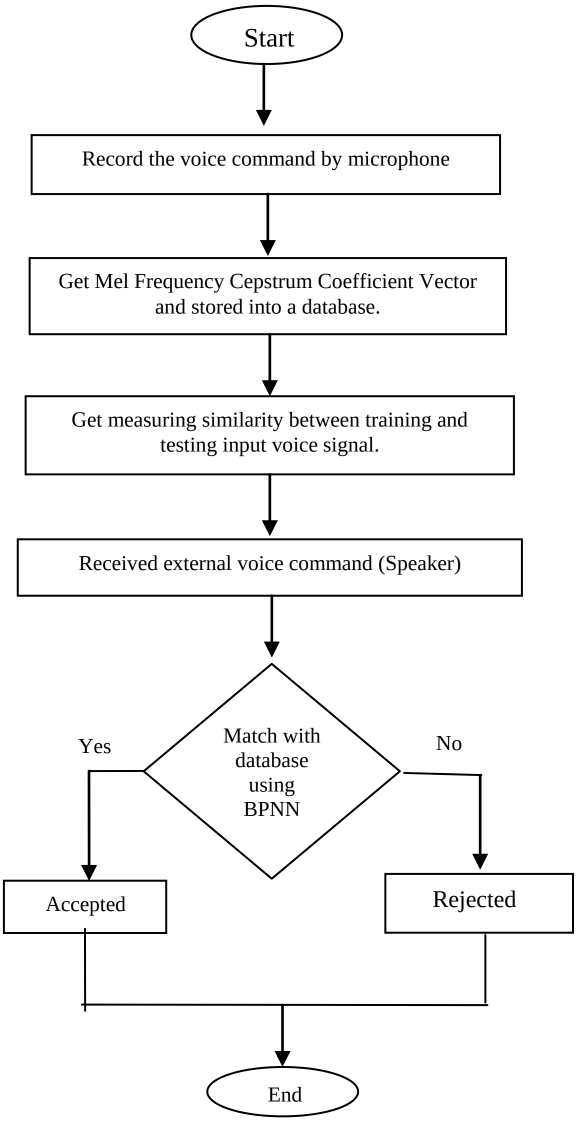

Figure 2: Block diagram of the proposed BPNN based SI System.

III. peaker Identification techniques

Speaker recognition concentrates on the identification task. The aim in speaker identification (SI) is to recognize the unknown speaker from a set of known speakers (closed-set SI).

A speaker recognition system is composed of the following modules:

A speaker recognition system is composed of the following modules:

1. Front-end processing - the “signal processing” part, which converts the sampled speech signal into set of feature vectors, which characterize the properties of speech that can separate different speakers. Front-end processing is performed both in training and testing phases.
2. Speaker modelling - this part performs a reduction of feature data by modelling the distributions of the feature vectors.

3. Speaker database - the speaker models are stored here.
4. Decision logic - makes the final decision about the identity of the speaker by comparing unknown feature vectors to all models in the database and selecting the best matching model.

The general methodology of audio classification involves extracting discriminatory features from the audio data and feeding them to a pattern classifier. Different approaches and various kinds of audio features were proposed with varying success rates. The features can be extracted either directly from the time domain signal or from a transformation domain depending upon the choice of the signal analysis approach. Some of the audio features that have been successfully used for audio classification include Mel-frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Local discriminant bases (LDB). Few techniques generate a pattern from the features and use it for classification by the degree of correlation. Few other techniques use the numerical values of the features coupled to statistical classification method. In our experiment we are taken only MFCC for feature extraction.

IV. Mel Frequency Cepstral Coefficient

MFCC is based on the human peripheral auditory system. The human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency t measured in Hz, a subjective pitch is measured on a scale called the 'Mel Scale' .The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.

As shown in the block diagram of figure 3, A compact representation would be provided by a set of mel-frequency cepstrum coefficients (MFCC)[4], which are the results of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale. The MFCCs are proved more efficient[4]. The calculation of the MFCC includes the following steps.

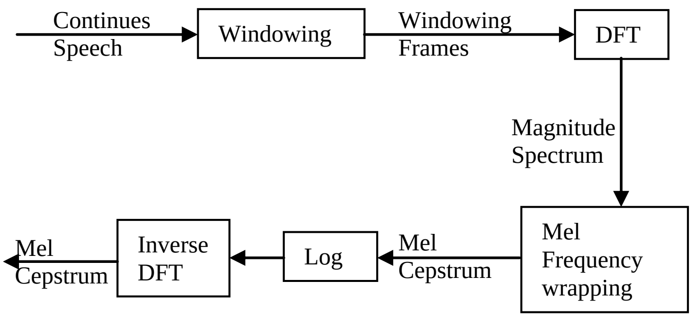

Figure 3 :Complete pipeline for MFCC.

A. *Calculation of MFCC*

According to psychophysical studies, human perception of the frequency content of sounds follows a subjectively defined nonlinear scale called the Mel scale. This is defined as,

$$f_{mel} = 25951 log_{10} \left(1 + \frac{f}{700}\right) \qquad \text{--(1)}$$

where f_{mel} is the subjective pitch in Mels corresponding to f, the actual frequency in Hz.

Let , $\{y(n)\}_{n=1}^{Ns}$ represent a frame of speech that is pre-emphasized and Hamming-windowed, where Ns is the total number of samples (here it is 160) present in a frame. First, $y(n)$ is converted to the frequency domain by an Ms-point *Discrete Fourier Transform* (DFT) which leads to the energy spectrum ,

$$|Y(k)|^2 = \left|\sum_{n=1}^{Ns} y(n).e^{\left(\frac{-j2\Pi nk}{Ms}\right)}\right|^2 \qquad \text{---(2)}$$

where, $1 <= k <= Ms$. This is followed by the construction of a filter bank with Q unity height triangular filters, uniformly spaced in the Mel scale (Eq. (1)). The filter response wi(k) of the ith filter in the bank is defined as,

$$\psi_i(k) = \begin{cases} 0 & \text{for } k < k_{b_{i-1}} \\ \frac{k - k_{b_{i-1}}}{k_{b_i} - k_{b_{i-1}}} & \text{for } k_{b_{i-1}} \leqslant k \leqslant k_{b_i} \\ \frac{k_{b_{i+1}} - k}{k_{b_{i+1}} - k_{b_i}} & \text{for } k_{b_i} \leqslant k \leqslant k_{b_{i+1}} \\ 0 & \text{for } k > k_{b_{i+1}}, \end{cases} \qquad \text{---(3)}$$

where $1 <= i <= Q$, Q is the number of filters in the bank, are the boundary points of the $\{k_{b_i}\}_{i=0}^{Q+1}$ filters and k $\{k_{b_i}\}_{i=0}^{Q+1}$ denotes the coefficient index in the Ms-point energy spectrum. The filter bank boundary points, are equally spaced in the Mel scale which is satisfied by the definition,

$$k_{b_i} = \left(\frac{M_s}{F_s}\right) \cdot f_{\text{mel}}^{-1}\left[f_{\text{mel}}(f_{\text{low}}) + \frac{i\{f_{\text{mel}}(f_{\text{high}}) - f_{\text{mel}}(f_{\text{low}})\}}{Q+1}\right] \qquad \text{--(4)}$$

where the function $f_{mel}(.)$ is defined in Eq. (1), Ms is the number of points in the energy spectrum (Eq. (2)), Fs is the sampling frequency, f_{low} and f_{high} are the low and high frequency boundaries of the filter bank and f^{-1}_{mel} is the inverse of the transformation in Eq. (1) defined as,

$$f = f_{\text{mel}}^{-1}(f_{\text{mel}}) = 700 \cdot \left[10^{\frac{f_{\text{mel}}}{2595}} - 1\right] \qquad \text{---(5)}$$

The sampling frequency Fs and the frequencies f_{low} and f_{high} are in Hz while f_{mel} is in Mels. For both the databases considered in this work, Fs is 8 kHz. Ms was taken as 256, $f_{low} = Fs/Ms$= 31:25 Hz while $f_{high} = Fs/2$= 4 kHz. Next, this filter bank is imposed on the spectrum calculated in eq. (2). The outputs $\{e(i)\}_{i=1}^{Q}$ of the Mel-scaled band-pass filters can be calculated by a weighted summation between respective filter response $\psi_i(k)$ and the energy spectrum $|Y(k)|^2$ as,

$$e(i) = \sum_{k=1}^{\frac{Ms}{2}} |Y(k)|^2 . \Psi_i(k) \qquad \text{---(6)}$$

Finally, Discrete Cosine Transform (DCT) is taken on the log filter bank $\{\log[e(i)]\}_{i=1}^{Q}$ energies and the final MFCC cefficients C_m can be written as,

$$C_m = \sqrt{\frac{2}{Q}} \sum_{l=0}^{(Q-1)} \log[e(l+1)] \cdot \cos\left[m \cdot \left(\frac{2l-1}{2}\right) \cdot \frac{\pi}{Q}\right] \qquad \text{---(7)}$$

where, $0<=m<=R-1$, and R is the desired number of cepstral features. Typically, $Q = 20$ and $10–30$ cepstral coefficients are taken for speech processing applications.

Here we took $Q = 20$, $R = 20$ and used the last 19 coefficients to model the individual speakers. Note that the first coefficient C_0 is discarded because it contains only a d.c. term that signifies spectral energy, not useful in speaker identification problem.

*B. **Steps for MFCC***

From the above assumptions the steps used in calculating the MFCC as shown in figure 3 as:

Step 1 : Frame Blocking

The process of segmenting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M (M<N). Typical values used are M = 100 and N= 256.

Step 2: Windowing

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines.

Step 3: Fast Fourier Transform (FFT)

To convert each frame of N samples from time domain into frequency domain. The Fourier Transform is to convert the convolution of the glottal pulse U[n] and the vocal tract impulse response H[n] in the time domain. This statement supports the equation below[4]:

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

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

Step 4: Mel-frequency Wrapping

The frequencies range in FFT spectrum is very wide and voice signal does not follow the linear scale. The bank of filters according to Mel scale as shown in figure 4 is then performed. This figure shows a set of triangular filters that are used to compute a weighted sum of filter spectral components so that the output of process approximates to a Mel scale. Each filter's magnitude frequency response is triangular in shape and equal to unity at the centre frequency and decrease linearly to zero at centre frequency of two adjacent filters [7, 8]. Then, each filter output is the sum of its filtered spectral components. After that equation 1 is used to compute the Mel for given frequency f in HZ.

Step 5: Cepstrum (Discrete Cosine Transform (DCT))

This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT)[18]. The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector.

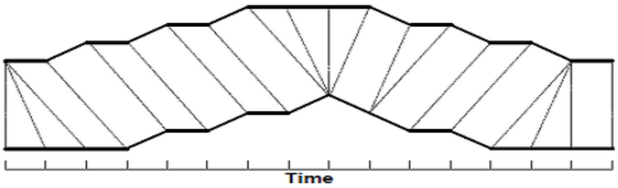

Figure 4. A Warping between two time series

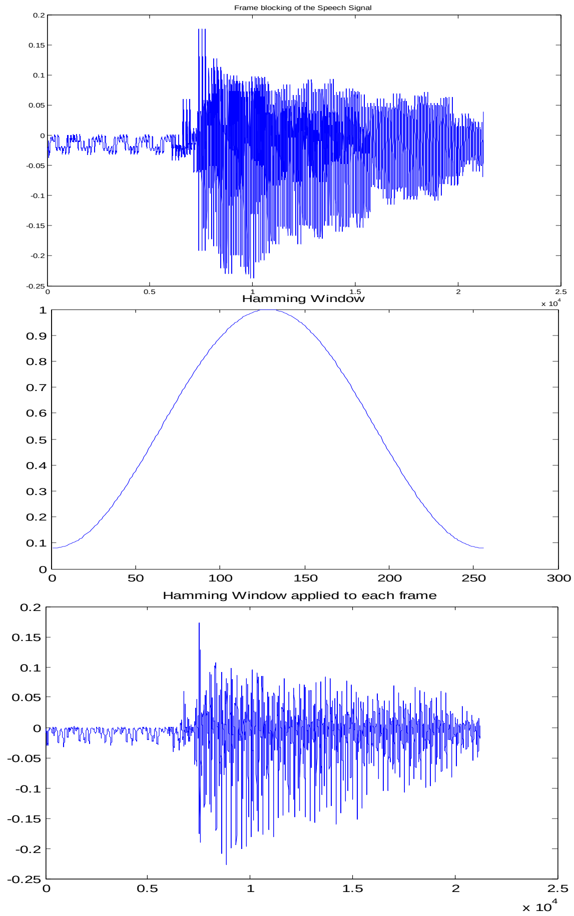

Figure 5 : Speaker1, top to bottom (a) Frame blocking of the speech signal (b) Hamming window (c) Hamming window applied to each frame

Table I
Database for SI

Items	**Descriptions**
Speech Utterances	Test data, Train data
Average words	10
Transducer	Noise cancelled microphone
Speakers	Male and Female , Age: 22-43
Speech Style	Reading text fragments
Environment	University Laboratory
Sampling Rate	16000HZ, 16bits, 1-channel

Table II
Feature sets for NN training from MFCC calculation of 5 speakers

Speakers / Values	Sp1	Sp2	Sp3	Sp4	Sp5
1	1.954	1.667	1.774	1.441	2.533
2	7.165	5.422	7.302	4.779	5.942
3	4.092	3.890	4.356	3.831	4.593
4	3.733	3.326	3.358	3.317	3.622
5	3.897	2.825	3.043	2.946	2.737
6	2.755	2.516	2.491	2.387	2.570
7	2.376	2.577	2.266	2.312	2.213
8	2.644	2.218	2.360	2.159	2.495
9	2.771	2.301	2.272	2.516	2.228
10	2.595	1.705	2.276	2.050	2.077
11	2.462	2.122	2.289	2.485	2.224
12	2.350	1.949	2.420	2.131	1.937
13	2.206	1.884	2.039	1.957	1.960

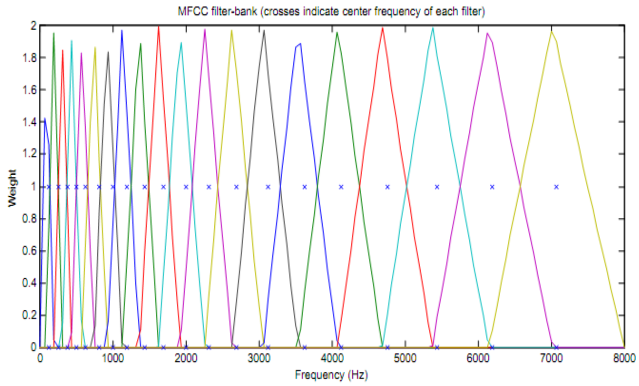

Figure 6: MFCC comparision between 5 speakers

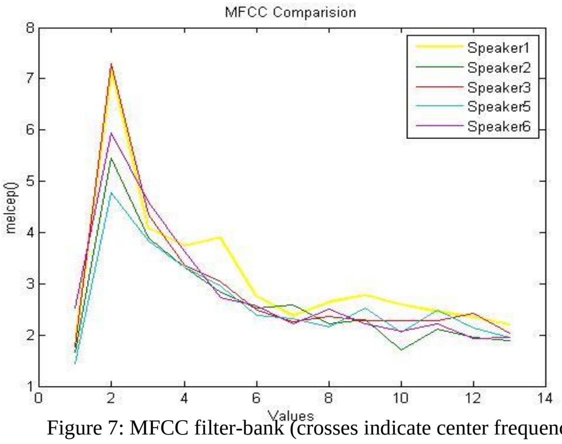

Figure 7: MFCC filter-bank (crosses indicate center frequency of each filter)

V. COMPARISION OF DIFFERENT IMPLEMENTATIONS OF MFCC

The performance of the Mel-Frequency Cepstrum Coefficients (MFCC) may be affected by (1) the number of filters,(2) type of window. In this paper, several comparison experiments are done to find a best implementation.

A. Effect of number of filters

Results of the speaker recognition performance by varying the number of filters of MFCC to 12, 22, 32, and 42 are given[7]. The recognizer reaches the maximal performance at the filter number K = 32. Too few or two many filters do not result in better accuracy. Hereafter, if not specifically stated, the number of filters is chosen to be K = 32.

Table III
MFCC with 12 Filters

Speaker	*No of attempts*	*False acceptance*	*False Rejection*
Sp1	4	0	0
Sp2	4	0	1
Sp3	4	0	2
Sp4	4	0	0
Sp5	4	0	2
Total	**20**	**0**	**5**

Threshold value of distance = 130, efficiency = 75%

Table IV
MFCC with 22 Filters

Speaker	*No of attempts*	*False acceptance*	*False Rejection*
Sp1	4	0	0
Sp2	4	0	2
Sp3	4	0	2
Sp4	4	0	0
Sp5	4	0	3
Total	**20**	**0**	**7**

Threshold value of distance = 150, Efficiency = 65%

Table V
MFCC with 32 Filters

Speaker	*No of attempts*	*False acceptance*	*False Rejection*
Sp1	4	0	0
Sp2	4	0	2
Sp3	4	0	1
Sp4	4	0	0
Sp5	4	0	2
Total	**20**	**0**	**3**

Threshold value of distance = 150, Efficiency = 85%

Table VI
MFCC with 42 Filters

Speaker	*No of attempts*	*False acceptance*	*False Rejection*
Sp1	4	0	0
Sp2	4	0	2
Sp3	4	0	1
Sp4	4	0	1
Sp5	4	0	0
Total	**20**	**0**	**4**

Threshold value of distance = 150, Efficiency = 80%

B. Effect of variation in type of window using 32 filters

Considering 32 filters as a standard number of filters we have changed the window type[7]. In this experiment we have used only the Hamming Window. Results show that efficiency is 75% while using the hamming window.

Table VI
Hamming Window

Speaker	*No of attempts*	*False acceptance*	*False Rejection*
Sp1	4	0	0
Sp2	4	0	2
Sp3	4	0	0
Sp4	4	0	3
Sp5	4	0	0
Total	**20**	**0**	**5**

Threshold value of distance = 150, Efficiency = 75%

VI. Back Propagated Artificial Neural Network

Artificial Neural Network (ANN) is an information processing paradigm that is inspired from biological nervous systems, such as the brain process information. It is composed of a large number of highly interconnected processing elements (neurons) working in union to solve specific problem. Basically, neural networks are built from simple units, sometimes called neurons or cells. Artificial Neural Network is information processing devices with the capability of performing computations similar to human brain or biological neural network. There are mainly three different types of layers presented in most ANNs. The first layer is called the input layer. Its main task is to receive input from the outside world.

Parameters used for the network.

Functions	***Description***
Network type	Feed-forward backpropagation
No. of layers	Four layers: input, two hidden and output
No. of neurons in layers	35 input, 20 hidden and 5 output
weight function	DOTPROD
training function	Levenberg–Marquardt backpropagation
Activation functions	Log-sigmoid
Performance function (mse)	10^5
No. of epochs	1003, 2009, 3006, 4000, 5013

Table VII
Error values of NN training at different Epochs

Epoch / Error	1003	2009	3006	4000	5013
1	- 1.9540	1.9540	1.9540	1.9540	1.9540
2	1.8784	1.9861	1.9694	1.9188	1.8957
3	-0.0756	0.0321	0.0154	-0.0352	-0.0583
4	1.6670	1.6670	1.6670	1.6670	1.6670
5	1.8784	1.9861	1.9694	1.9188	1.8957
6	0.2114	0.3191	0.3024	0.2518	0.2287
7	1.7740	1.7740	1.7740	1.7740	1.7740
8	1.8784	1.9861	1.9694	1.9188	1.8957
9	0.1044	0.2121	0.1954	0.1448	0.1217
10	2.2340	2.2340	2.2340	2.2340	2.2340
11	1.8784	1.9861	1.9693	1.9188	1.8957
12	-0.3556	-0.2479	-0.2647	-0.3152	-0.3383

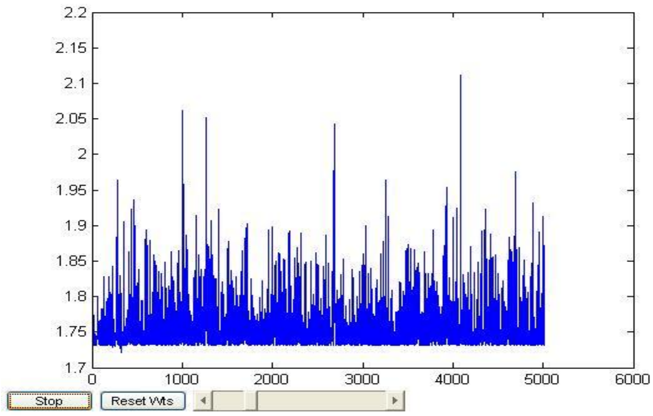

Figure 8: Neural network training for error calculation plot stoped at 5013 epoch

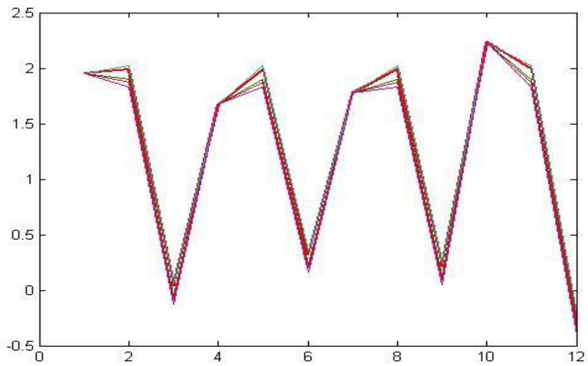

Figure 9: Error comparison for different epoch values

VII. Conclusion and Future Works

In this paper we generally discussed about MFCC, which is well known techniques used in speaker recognition to describe the signal characteristics, relative to the speaker discriminative vocal tract properties. The goal of this project was to create a speaker recognition system, and apply it to a speech of an unknown speaker. By investigating the extracted features of the unknown speech and then compare them to the stored extracted features for each different speaker in order to identify the unknown speaker. In our results we find that at 32 number of filters the efficiency is 85%. The error matrix we found from NN training and testing (table VII) at different epoch values shows that our algorithm is correct. The error range is set within -1 to 2.5, beyond this value is rejected . as we are taken only five text dependent speech, we are trying to collect 100 such speech and calculate the MFCC feature for NN training which will give more accurate figure of identification. In future we will try to improve this system to be a text independent speaker identification system as we have taken in this system as text dependent value that is “welcome to the world of speaker identification” spoken by five different female speakers.

ACKNOWLEDGMENT

To complete this work, we have got valuable suggestions and guidance from different experts of this field. The Book *Digital Processing* of *SpeechSignals.* by L.R. Rabiner and R.W. Schafer is very helpful for beginners' to start research in this field. I also thankful to all the researcher and authors of the manuscript from where I got valuable information for completing the work.

REFERENCES

[1] K.K. Paliwal and B.S. Atal, 'Frequency related representation of speech,' in Proc. EUROSPEECH, p.p.65-68 Sep. (2003).

[2] T. Fukuda, M. Takigawa and T. Nitta, "Peripheral features for HMMbased speech recognition," in Proc. ICASSP, 1: 129-132(2001).

[3] M. Pandit and J. Kittler, "Feature selection for a dtw-based speaker verification system, in Proceedings of IEEE Int.Conf. Acoust. Speech and Signal Processing, 2: 769-772 (1998).

[4] K.R. Aida–Zade, C. Ardil and S.S. Rustamov," Investigation of Combined use of MFCC and LPC Features in Speech Recognition Systems",World Science and engineering and Technology 19 2006

[5] S. Furui, "An overview of speaker recognition technology, in Automatic Speech and Speaker Recognition (C.H. Lee,F.K. Soong, and K.K. Paliwal,eds), ch.2 pp.31-56Boston :Kluwer Academic, (1996).

[6] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals. Englewood Cliffs, NJ: Prentice-Hall, (1978).

[7] Vibha Tiwari, "MFCC and its applications in speaker recognition", International Journal on Emerging Technologies 1(1): 19-22(2010), ISSN : 0975-8364

[8] Automatic speaker recognition by S.Khan, Mohd Rafibul lslam, M. Faizul, D. Doll. 3rd international conference on electrical and computer engineering (ICECE), 28-30th Dec. (2004), Dhaka, Bangladesh.

[9] Speaker recognition using MFCC by S. Khan, Mohd Rafibul lslam, M. Faizul, D. Doll, presented in IJCSES (International Journal of Computer Science and Engineering System) 2(1): 2008.

[10] Mohd Rasheedur Hassan, Mustafa Zamil, at al.. "Speaker identification using MFCC coefficients ",3rd international conference on electrical and computer engineering (ICECE), (2004).

[11] Goutam Saha and Malyaban Das,On Use of Singular Value Ratio Spectrum as Feature Extraction Tool in Speaker Recognition Application, CIT-2003, pp. 345-350, Bhubaneswar, Orissa, India, (2003).

[12] Premakanthan and W.B. Mikhael, Speaker verification/ recognition and the importance of selective feature extraction: Review, Proceedings of the 44th IEEE 2001, Midwest Symposium, 1: 14-17(2001).

[13] Molau, S, Pitz, M, Schluter, R, and Ney, H., Computing Mel-frequency coefficients on Power Spectrum, Proceedings of IEEE ICASSP-2001, 1: 73-76(2001).

[14] C.D. Bei and R.M. Gray. An improvement of the minimum distortion encoding algorithm for vector quantization. IEEE Transactions on Communications, October (1998).

[15] Lawrence Rabiner and Biing-Hwang Juang, Fundamental of Speech Recognition", Prentice-Hall, Englewood Cliffs, N.J., (1993).

[16] Todor D. Ganchev at al. "Generalized locally recurrent probabilistic neural networks with application to text-independent speaker verification", Neurocomputing 70 (2007) 1424–1438, www.elsevier.com/locate/neucom.

[17] Vibha Tiwari," MFCC and its applications in speaker recognition", International Journal on Emerging Technologies 1(1): 19-22(2010), ISSN : 0975-8364

[18] Md. Rashidul Hasan and Mustafa Jamil at.al ," speaker identification using mel frequency cepstral coefficients", 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh.

[19]. Khaled Daqrouq," Wavelet entropy and neural network for text-independent speaker identification", Engineering Applications of Artificial Intelligence 24 (2011) 796–802.

[20] Tomi Kinnunen and Ilja Sidoroff at. Al." Comparison of clustering methods: A case study of text-independent speaker modeling " , Pattern Recognition Letters 32 (2011) 1604–1617.

[21] Hesham Tolba, "A Comparative Study to Evaluate a Text-Independent Speaker Identification Engine for Arabic Speakers Using a CHMM-Based Approach", 978-1-4244-4530-1/09/$25.00 ©2009 IEEE.

[22] Melih hal and Erhan Biitiin at al "Comparison Of Linear Predictive Analysis Methods For Ann-Based Speaker Identification", 0-7803-5512-1/99/$10.00 Q 2000 IEEE.

[23] Mihailo S. Zilovic, Ravi P. Ramachandran, "Speaker Identification Based on the Use of Robust Cepstral Features Obtained from Pole-Zero Transfer Functions ",Ieee Transactions On Speech And Audio Processing, Vol. 6, No. 3, May 1998.

