Text Indpendent Speaker Recognition Using Wavelet Cepstral Coefficient and Butter Worth Filter

Sandeep Rathor[1], R. S. Jadon[2],
[1]Asst. Professor, Dept of CEA, GLA University, Mathura, Uttar Pradesh, India
[2]Professor, Dept of MCA, MITS, Gwalior, Madhya Pradesh, India.

***Abstract*-. In this paper an effective and vigorous method for text independent speaker identification is proposed to extract speech features. The objective of feature extraction is to extract features from speech and captures the unique characteristics of a individual speaker. The proposed method can be used in noisy environment with high degree of accuracy. The proposed method is based on the wavelet transform and the input speech signal is decomposed into various frequency channels. The purpose of Wavelet transform is to find the frequency spectrum while wavelet cepstral coefficient is used to capture the characteristic of the signal. It is more suitable than Fourier transform because it is restricted in both time and frequency whereas fourier transform is only restricted in frequency. The proposed method is capable to reduce the noise as well as also improves recognition effectively. Fuzzy rules are used for decision making. The proposed method is very useful in the field of forensic also. The performance of WCC is about 22% higher than mel- frequency cepstral coefficients.**

Keywords- Speaker Recognition, Wevelet transform, Cepstral Coefficient, Fuzzy Logic, Butter Worth Filter.

I. INTRODUCTION

Speaker recognition is a technique that is able to recognize the person; who is speaking based on individual information. Any speech has basic information about the words and the identity of the speaker like frequency, pitch, entropy etc and each speaker also has unique characteristics [1]. Speaker recognition systems are divided into two categories; text-dependent and text-independent. In text-dependent systems, a user has to speak some defined set of words, containing the same text as the training data while in text-independent system there is no limitation of text i.e. speaker can speak any word, therefore to recognize speaker in text independent mode is more challenging task. It can be used to verify a person and allow accessing various services like security controls, confidential data accessing through remote site, etc. Speech is the fundamental and most effective way of communication in real time systems. The research on speech has started in 18th century [2].

Automatic speaker recognition has a machine that is capable to recognize a person based on voice. Automatic speaker recognition includes two processes: speaker identification and speaker verification. The objective of speaker identification is to identify a person. In Speaker identification first speaker has to be enrolled in the system and then on the basis on feature extraction we determine which enrolled speaker has provided sound among a set of known speakers. Speaker identification is very useful in forensic and can also be used in applications that make our daily lives more convenient [3]. While in speaker verification the objective is to verify a person based on the test pattern. An important parts of a speaker recognition system is feature extraction because it converts the properties of speech signal that is used for pattern matching [1][4].

The speech can be represented in the simplest form with the help of spectrogram. The spectrogram is a grayscale image, whose pixel's intensities represent the energy content of the frequencies with respect to time [5]. Generally, speech recognition systems used Mel-Frequency Cepstral Coefficients (MFCCs) and fourier transform because it is useful for analysis of speech signals whose statistical properties are constant with respect to time or space However, wavelets represent non-stationary signals as sum of basic functions which are restricted in time. This can be derived from a single prototype function called the "mother wavelet". The basis functions or wavelets are formed by translating and dilation the mother wavelets therefore, in this paper we used Wevelet Cepstral Coefficient and result analysis shows that proposed approach is better than others exiting methods.

According to the literature, a text dependent speaker recognition system compares a sequence of features to a model of the user. So, for this purpose there are two methods that have been mainly used i.e. template based methods and statistical methods. Most popular template based model is Dynamic Time Warping (DTW) and Statistical based method Hidden Markov Models (HMMs) [6]. HMM is one of the most popular method because of flexibility; allow using speech units from sub-phoneme units to words. [7][8]. In this paper, a feature extraction algorithm is proposed. It is based on wavelet transform and combining both the wavelet transform and the wavelet cepstral coefficient at feature extraction phase. By the

wavelet decomposition multi-resolution features is extracted and then calculating the required coefficients.

A. Bhalla et. al., summarized the Growth of speaker recognition system during the last six decades along with proposed a Mel frequency based extraction model of speaker recognition [8]. it analyzed different parts of speech at same scale. A. Hussein explained the role of HMM the the field of speech recognition [9].

P. Meeelin et. al., proposed a method for speaker recognition. It consists of two sub process i.e. identification and verification of speaker on the basis of individual information. Speech waves are analyzed by using neural network. One important thing in this paper is the use of Genetic Algorithms. It is used to optimize the architecture of neural network [11].

K. Meena et. al., proposed a Gender classification in Speech Recognition Using Fuzzy Logic and Neural Network. In general gender is classified through a common feature i.e. pitch. The pitch value of female is higher than male. However in some specific cases pitch value of male may be higher than female. For these specific cases there are different features i.e. short time energy, zero crossing rate and energy entropy. Calculate these different values by using fuzzy logic and neural network and then finally, the result is obtained by calculating mean value. Using the threshold value this method identifies the gender of speaker [12]. It has high time complexity

G. Tsenov et. al., proposed speech recognition using neural networks. This paper presents investigation on speech recognition using different standard neural networks structures as a classifier. It also used the concept of Feed forward neural network with back propagation algorithm in this paper. The classification is done on th.e basis of spectrogram samples and cepstrum and Mel frequency is also calculated using the spectrograms [13].

II. PROPOSED SCHEME

In our paper, we used wavelet transform because it has the capability to construct a time-frequency representation of a signal that offers very good time and frequency localization. The continuous wavelet transform of a function x(t) is expressed by

$$X_w(a,b) = \frac{1}{|a|^{1/2}} \int x(t)\, \bar{\psi}\left(\frac{t-b}{a}\right) dt \qquad (1)$$

where $\bar{\psi}(t)$ is a continuous function in both the time domain and the frequency domain and a is a scale factor or dilation parameter and b is the time delay or translation variable. The scale factor a, governs its frequency content, the delay parameter b, gives the position of the wavelet $\bar{\psi}_{(a,b)}(t)$.

The DWT of a signal x can be calculated by passing it through a series of filters. First the samples are passed through low pass (butter worth filter) with impulse response g and then the signal is also decomposed simultaneously using a high-pass filter and band pass filter respectively, results are shown in figure 2 to 4 respectively. If we are going to apply two filters then it is important that the two filters must be related to each other.

In this paper, an effective method is proposed for text independent speaker recognition. It enhances the performance of text independent speaker identification with a capacity of handling in noisy environment. Feature extraction is the technique that extracts small information from the speech signal that is used to represent each individual speaker. The proposed scheme is based on the discrete wavelet transform. It is a type of signal processing tool. It is based on multi label resolution technique and can be used to obtain the frequency spectrum. It has more advantages over DFT because it can be analyze different part of speech at different scales and it is also restricted in both time and frequency while Fourier transform is only restricted in frequency. Here the input speech signal is decomposed into various speech features i.e. Energy Entropy (EE), Short Time Energy (STE), Zero Crossing Rate (ZCR), pitch, formants, spectrogram, sampling frequency etc. In the proposed scheme, Butterworth filter is used to remove background noise. The wavelet cepstral coefficient (WCC) is used to capture the characteristic of speech signal. The proposed method not only effectively reduces the effect of noise, but also improves the supremacy of recognition. Fuzzy rules are used for decision making on the basis of feature matching. It can distinguish between the unknown speaker and known speaker by comparing extracted features from their voice input.

Steps for Speaker Recognition

1. Apply Butter worth filter to remove background noise.
2. Wavelet transform is applied to decompose the speech signal into two different frequency channels
3. To capture the characteristics of the individual speakers, the Wavelet Ceptral Coefficient is applied.
4. Create a rule base database on the basis of characteristics as step 3.
5. Perform features matching by using step 4.
6. Speaker is recognized.
7. Stop

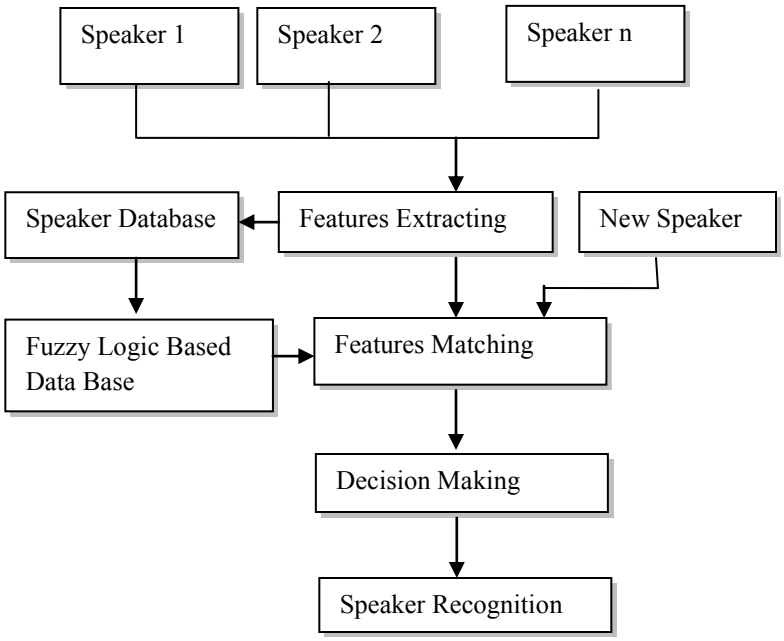


Figure. 1. A Process of Speaker Recognition.

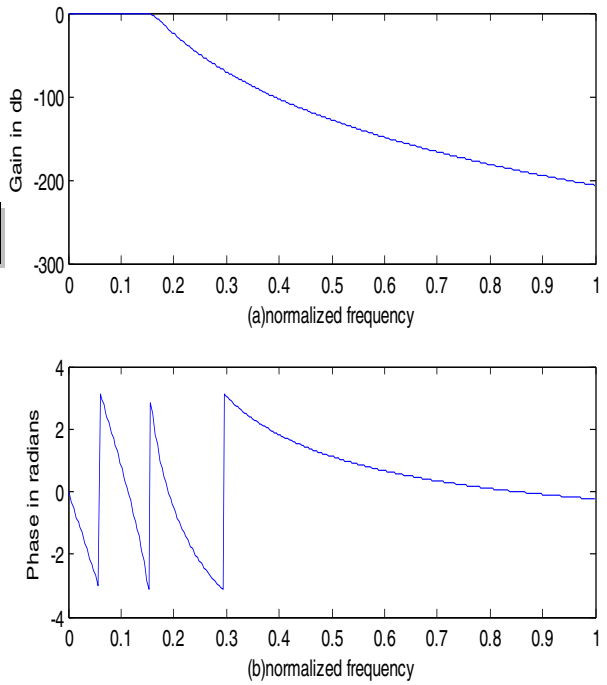


Figure 2. Amplitude Response, Phase Response, after butter worth Low pass filter

III.EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we used MatlabR2016a platform to execute the code and a database contains speech file of 27 different speakers. The speech of speaker is recorded through microphone. The butter worth filter is applied on receiving files to remove the noise from speech signals. In butter worth filter; a low pass filter is shown in figure 2, high pass filter is shown in figure 3 and figure 4 shows the band pass filter respectively. Figure 5 represents sinusoidal representation of speech signal, figure 6 shows the Frequency at which the attenuation begin and figure 7 shows represents a normalized signal. Features extraction is done by using wavelet cepstral coefficient after applying wavelet transformation. A fuzzy logic based database is used to verify the speaker.

Table 1 shows the recognition rate of speech in noisy environment along with the comparison with the existing methods. The result analysis shows that our proposed method has much accuracy or recognition rate with minimum cpu time.

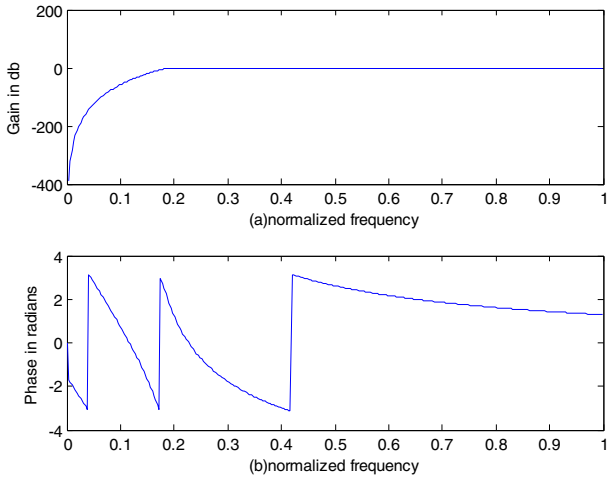


Figure 3. Amplitude Response, Phase Response, after butter worth High pass filter

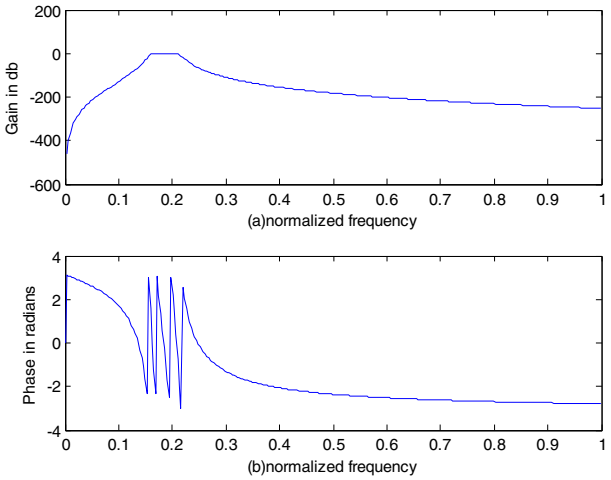


Figure 4. Amplitude Response, Phase Response, after butter worth band pass filter

e

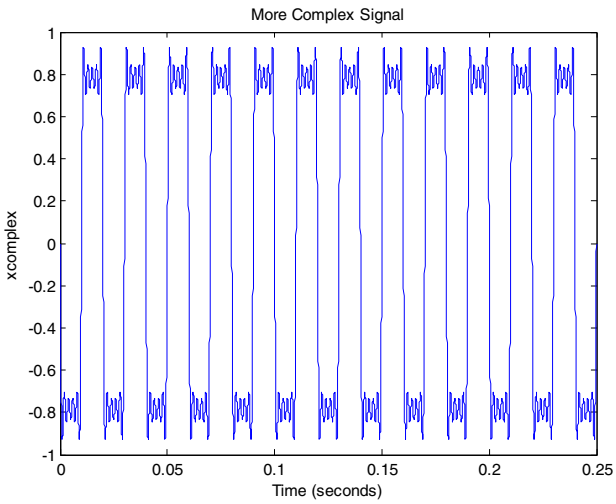


Figure 5. Ssinusoidal Representation of Speech Signal

After applying the filters, speech signals are refined. Multilevel wavelet transform reflect the meaningful data at the last lavel. The normalized speech signal is represented in figure 7. Figure 8 (a-e) show the wavelet coefficient at different bands i.e. first band , second band , third band , fourth band and fifth band respectively.

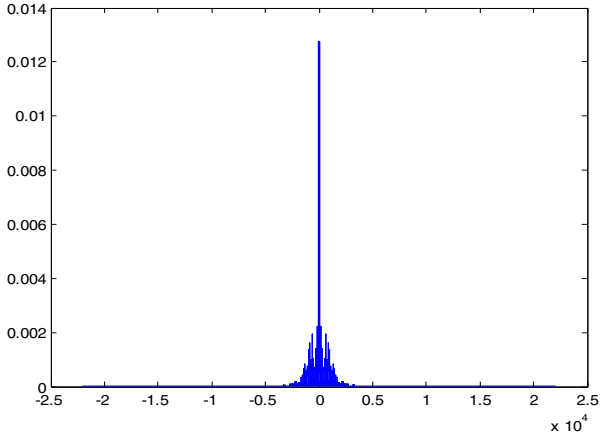


Figure 6.Frequency at Which the Attenuation Begin

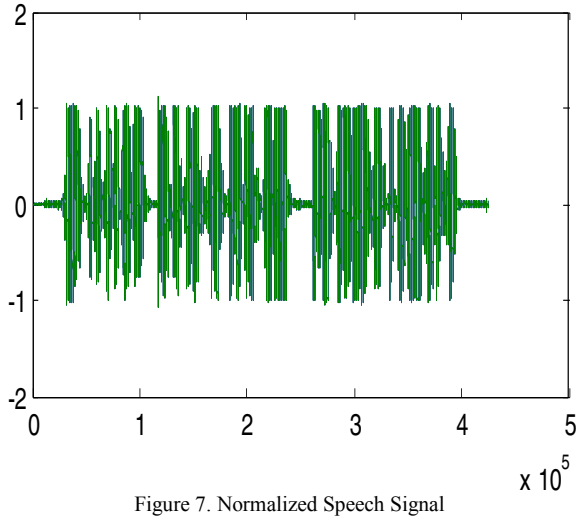
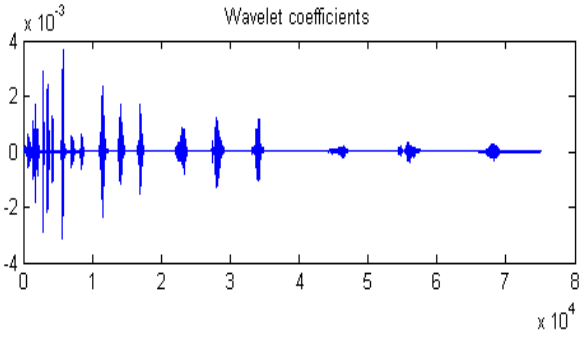


Figure 7. Normalized Speech Signal



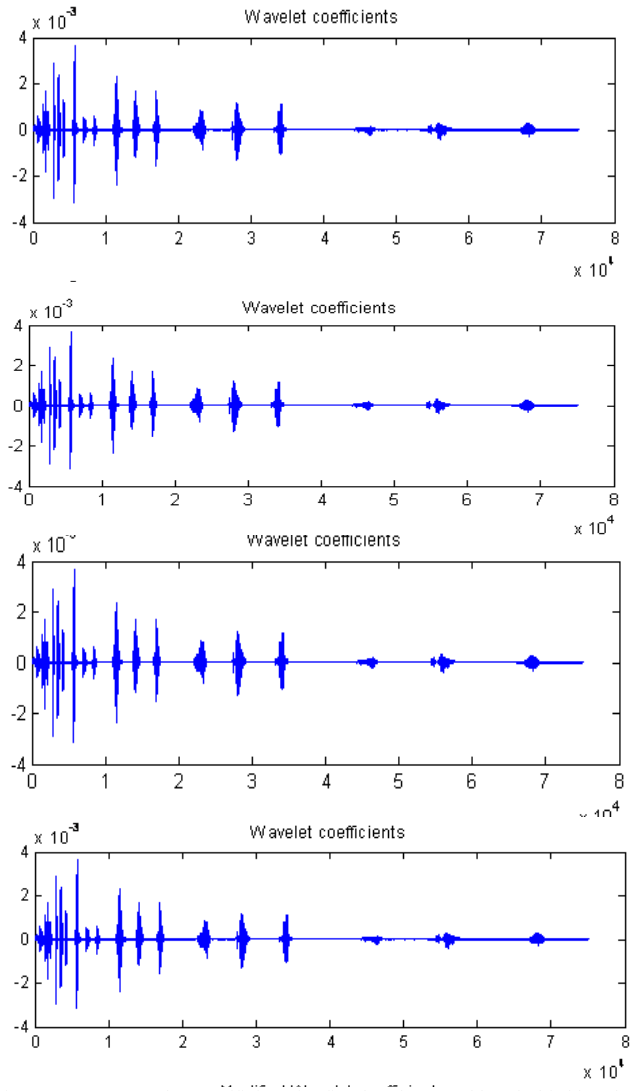


Figure 8(a-e). Wavelet Coefficient for first band, second band, third band, fourth band and fifth band of Speech Signal

Table 1. Recognition Result at Noisy Environment

S.No.	Method Used	Accuracy (%)
1	Mel Frequency Cepstral Coefficient &GMM	82.46
2	DFT based Probabilistic Neural network	84.3
3	Mel Frequency Cepstral Coefficient & DWT	86.8
4	Wavelet based Mel Frequency Cepstral Coefficient	92.9
5	**Proposed**	**98.5**

IV.CONCLUSION

The performance of proposed method is more accurate in compare to other methods proposed by the researcher. The proposed method can be apply to recognize speaker from noisy environment and can reflect the status of environment where speaker is speaking. The beauty of the proposed method is that it is text independent, means speaker is free to speak any word.

V. REFERENCES

[1] I. Mahmoud and S. Ali, "Wavelet-Based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models", Journal of Telecommunications, Vol. 1, Issue 2, March 2010.

[2] E. George. Dahl, Y. Dong, L. Deng and A. Acero, "Context- Dependent Pre-Trained Deep Neural Network for Large- Vocabulary Speech Recognition", IEEE transactions on Audio, Speech and Language Processing, Vol. 20, No. 1, Jan. 2012.

[3] J. Mikyong, K. Sungtak, K. Hoirin and Y. Ho-Sub, "Text-Independent Speaker Identification using Soft Channel Selection in Home Robot Environments", IEEE Transactions on Consumer Electronics ,Vol. 54, Issue: 1, February 2008.

[4] T. Scults and A. Waibel, "Language- Independent and Language-Adaptive Acoustic Modeling for Speech Recognition", Elsevier, Speech Communication 35 (2001) pp. 31-51.

[5] Y. Amit and A. Murua, "Speech Recognition Using Randomized Relational Decision Trees", IEEE transactions on speech and audio processing, Vol. 9, No. 4, may 2001.

[6] S. Bishnu and R. Rabiner,"A Pattern Recognition Approach to Voiced-Unvoiced Classificaton with Application to Speech Recognition", in proceedings of the IEEE International Conference on Volume 2, Issue 3, March 2012, Acoustic, Speech and Signal Processing (ICASSP'76), Pennsylvania, Vol. 24, No. 3, pp.201-212.

[7] F. Richardson and D. Reynolds,"Deep Neural Network Approaches to Speaker and Language Recognition," IEEE Signal Processing Letters, Vol. 22, No. 10, October 2015.

[8] A. Bhalla ,S. Khaparkar and M. Bhalla, "Performance Improvement of Speaker Recognition System, "International Journal of Advanced Research in Computer Science and Software Engineering", Volume 2, Issue 3, March 2012.

[9] A. Hussein, "Analysis of Voice Recognition Algorithms using MATLAB", International Journal of Engineering Research & Technology, ISSN: 2278-0181, Vol. 4 Issue 08, August-2015.

[10] A. Shukla, R. Tiwari and C. P. Rathore,"Intelligent Biometric System Using soft computing tools", Biomedical Engineering and Information Systems: Technologies, Tools and Applications, DOI: 10.4018/978-1-61692-004-3.ch014.

[11] P. Melin et. al., "Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithm", Engineering Letters, 13:2, EL_13_2_9 Advance online publication, Aug 2006.

[12] K. Meena, K. Subramaniam and M. Gomathy, "Gender Classification in Speech Recognition Using Fuzzy Logic and Neural Network", International Arab Journal of Information Technology, vol. 10, No. 5, September 2013.

[13] G. Tsenov and V. Mladenov, "Speech Recognition Using Neural Network", IEEE 10th symposium on Neural Network Applications in Electrical Engineering, NEUREL-2010, sep., 23-25, 2010.