Wavelet Based Noise Robust Features for Speaker Recognition

Vibha Tiwari *vibhatiwari19@gmail.com*
Asst. Professor/Electronics and Communication Department
Gyan Ganga Institute of Technology and management
Bhopal, India

Dr. Jyoti Singhai *j.singhai@gmail.com*
Associate Professor/ Electronics and Communication Department
Maulana Azad National Institute Of Technology
Bhopal, India

Abstract

Extraction and selection of the best parametric representation of acoustic signal is the most important task in designing any speaker recognition system. A wide range of possibilities exists for parametrically representing the speech signal such as Linear Prediction Coding (LPC) ,Mel frequency Cepstrum coefficients (MFCC) and others. MFCC are currently the most popular choice for any speaker recognition system, though one of the shortcomings of MFCC is that the signal is assumed to be stationary within the given time frame and is therefore unable to analyze the non-stationary signal. Therefore it is not suitable for noisy speech signals. To overcome this problem several researchers used different types of AM-FM modulation/demodulation techniques for extracting features from speech signal. In some approaches it is proposed to use the wavelet filterbanks for extracting the features. In this paper a technique for extracting the features by combining the above mentioned approaches is proposed. Features are extracted from the envelope of the signal and then passed through wavelet filterbank. It is found that the proposed method outperforms the existing feature extraction techniques.

Keywords: Speaker Recognition, Mel Frequency Cepstral Coefficients (MFCC), Amplitude Modulation (AM) Wavelet Filterbank.

1. INTRODUCTION

Speaker recognition is the process of automatically recognizing who is speaking on the basis of the information extracted from the speech signal. This technique is used to verify the identity of a person with the help of speaker's voice. There are many application areas that are using Speaker recognition for controlling access to services such as voice dialing, voice mail, banking by telephone, telephone shopping, database access service, security control for confidential information etc.

The first speaker recognition system was introduced by Pruzansky [1] in 1960. In his work he used filter banks and spectrogram correlators for speaker recognition. Li et. al. [2] further developed it by using linear discriminators. In some of the approaches the speaker specific features are extracted by using the statistical or predictive parameters. A wide range of possibilities exists for parametrically representing the speech signal such as instantaneous spectra covariance matrix [3], spectrum and fundamental frequency histograms [4], Linear Prediction coefficients (LPC) [5], Mel frequency Cepstrum Coefficients (MFCC) [6] and others. Among all the approaches mentioned above MFCC are the most popular choice for any speaker recognition system, though one of the shortcomings of MFCC is that the signal is assumed to be stationary within the given time frame and is therefore unable to analyze the non-stationary signal. There exist two different approaches to deal with this problem.
In first approach researchers used different types of AM-FM modulation/demodulation techniques for extracting features from speech signal. Qifeng Zhu and Abeer Alwan considerd AM

modulation for speech recognition [7] [8]. They used the Harmonic demodulation methods in computing MFCC. Different from previous studies Petros Maragos, and Alexandros Potamianos proposed AM-FM modulation model to represent the speech signal [9]. Using AM-FM model Dimitrios Dimitrriadis et.al. proposed robust AM-FM features for speech recognition[10],[11] . It has been found in the above proposed methods that the AM and FM modulation features characterize the very fine structure of speech and they improve noise robust speech recognition efficiency.

In second approach wavelet transform based methods are used for feature extraction. In particular, in feature extraction schemes designed for the purpose of speaker and speech recognition, wavelets have been used as an effective decorrelator instead of Discrete Cosine Transform in the feature extraction stage [12]. In [13] wavelet coefficients with high energy are taken as features but such methods suffer from shift variance. T. Kinnunen proposed use of wavelet subband energies instead of Mel filterbank subband energies [14]. Later, Sarikaya et.al. [15] ,[16] Farooq and dutta [17] proposed wavelet filterbanks that are close approximation of Mel filter bank.

But all these existing speaker recognition systems discussed above performs well only under clean speech conditions. Their recognition efficiency decreases in noisy and real time speech conditions. This noise is additive in nature and can be removed from the noise corrupted features. But this requires noise estimation which increases the processing and time complexity. This noise estimation can be avoided by extracting noise robust features. In this paper a noise robust speaker recognition system is proposed which extracts noise robust features by extracting the vocal tract transfer function (VTTF) and then passing this through wavelet filterbank. The extraction of VTTF avoids the spectral mismatch produced by noise at low signal to noise ratio areas are avoided and wavelet filterbank avoids the loss of energy in side lobes. Following will be the organization of the paper: Section 2, critically analyses popular feature extraction technique MFCC and discuss existing modifications of MFCC by using AM demodulation and wavelet filterbank. Section 3 discusses proposed algorithm. Section 4 gives the performance evaluation and comparison of proposed algorithm with existing algorithms. Section 5 concludes the algorithm.

2. FEATURE EXTRACTION TECHNIQUES FOR SPEAKER RECOGNITION

The purpose of feature extraction is to convert speech waveform, to a set of features for further processing. These features can be used to generate a pattern and then classification can be done by the degree of correlation. Few other techniques use the numerical values of the features coupled to statistical classification method. Different approaches for feature extraction that have been successfully used include, Linear discriminator bases [2], Linear predictive coding (LPC) [5] Mel-frequency cepstrum coefficients (MFCC) [6]. Among them MFCC is the most popular technique for feature extraction. In this section MFCC feature extraction technique is discussed with its limitations and then some existing modifications of MFCC are outlined.

2.1 MFCC

Mel-frequency cepstrum coefficients (MFCC) [6] are well known features used to describe speech signal. They are based on the known variation of human ear's critical bandwidths with frequency. In MFCC feature extraction filters are spaced linearly at low frequencies and logarithmically at high frequencies as the information carried by low frequency components are more important for human perception than high frequency components. Figure 1. shows the feature extraction technique using MFCC.

In MFCC feature extraction first step is windowing. The window used for MFCC feature extraction is the hamming window given by the equation (assuming a window that is L frames long):

$$hamming \qquad w[n] = 0.54 - 0.46\cos(2\pi n/L); \quad 0 \le n \le L-1$$

$$; \qquad \text{otherwise}$$

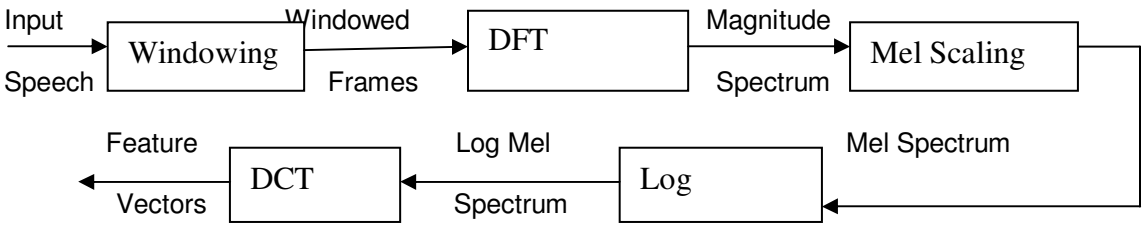


Figure1: Feature extraction using MFCC

The spectral information from the windowed signal is calculated by using DFT. The result after this step is often referred to as *spectrum* or *periodogram* and gives the information about the amount of energy at each frequency band. Since Human hearing is not equally sensitive at all frequency bands therefore the form of the model used in MFCCs is to warp the frequencies output by the DFT onto the Mel scale. The mapping between frequency in Hertz and the Mel scale is linear below 1000 Hz and the logarithmic above 1000 Hz. The Mel frequency *m* can be computed from the raw acoustic frequency as follows:

$$Mel(f) = 1127\ln(1+(f/700))$$

A bank of filters is created which collect energy from each frequency band, with 10 filters spaced linearly below 1000 Hz, and the remaining 10 filters are spread logarithmically above 1000 Hz [6]. This filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant Mel frequency interval. Then Log of the Mel spectrum is taken to incorporate the logarithmic human response to signal level. A set of Mel-frequency cepstrum coefficients is computed by taking discrete cosine transform of the logarithm of the short-term power spectrum expressed on a Mel-frequency scale.

MFCC is perhaps the best known and most popular feature extraction technique though it has somelimitations. In MFCC the signal is assumed to be stationary within a given time frame and may therefore lack the ability to analyze localized events accurately. Secondly a triangular filter bank is used whose frequency response is not smooth and therefore it may not be suitable for noisy speech data.

Some modifications are proposed by researchers to overcome the limitations of MFCC. In some approaches it is proposed to modify the signal under consideration before calculating MFCC [7] [8] [10] [18] [19]. In other approaches the MFCC is modified by using wavelet transform methods [13] [14] [15] [16] [17] [21]. These two approaches are discussed in section 2.2 and 2.3 respectively.

2.2 Feature Extraction Using Demodulated Speech Signal

The feature extraction using AM demodulation of speech signal was proposed by Zhu et. al.[8] in the year 2000. In [8] Zhu and Alwan performed the envelope detection of speech spectrum in frequency domain by performing the linear convolution between the speech spectrum and the characteristics of low pass filter. After extracting the envelope of the speech signal MFCC features are calculated from this envelope.

In [18] Fan-Gang Zeng, Kaibao Nie and others found that although AM from a limited number of spectral bands may be sufficient for speech recognition but FM significantly enhances speech recognition in noise, as well as speaker and tone recognition. Additional speech reception threshold measures revealed that FM is particularly critical for speech recognition with a competing voice and is independent of spectral resolution and similarity. These results suggest that AM and FM provide independent yet complementary contributions to support robust speech recognition under realistic listening situations.
In [19],Maragos, Kaiser and Quatieri proposed AM and FM models to represent time varying amplitude and frequency patterns in speech resonance and the total speech signal is superposition of such AM-FM signals. To demodulate a signal resonance, Energy separation

algorithm (ESA) is used by Maragos, Kaiser and Quatieri [10] using the nonlinear energy-tracking operator. The ESA estimates the instantaneous frequency and amplitude of the signals.

The energy separation methodology has led to several classes of algorithms for demodulating AM–FM signals. Dimitrios Dimitriadis et. al. proposed to extract speech features inspired by the AM-FM model [11]. The proposed features measure instantaneous amplitude and frequency model and these features when combined with the MFCC are robust to noise.

2.3 Feature Extraction Using Wavelet Transform Methods

Wavelet transform is a multi resolution transform that has a capability to process the non stationary signal as well. Recently this transform has been used to extract features for the purpose of speech and speaker recognition. Another advantage of using the wavelet transform is its compact support which avoids spilling of the energy of the side lobes. In particular, in feature extraction schemes designed for the purpose of speech and speaker recognition, wavelets have been used twofold. The first approach uses wavelet transform as an effective decorrelator instead of Discrete Cosine Transform in the feature extraction stage [21]. According to the second approach, wavelet transform is applied directly on the speech signal. In this case, either wavelet coefficients with high energy are taken as features [13], or wavelet subband energies are used instead of the Mel filter-bank subband energies. In particular, in the speech recognition area, the wavelet packet transform, employed for the computation of the spectrum, was first proposed in [14]. Later, wavelet packet bases were used by SBCC of Sarikaya & Hansen [15], WPP of Sarikaya & Hansen [16] , WPF of Farooq and dutta [17]. In these approaches a wavelet packet tree/filter that is close approximation of Mel frequency division is used i.e. low to mid frequencies more subbands are assigned so that their decomposition preserves a log like distribution of subband. Then the subband energy for each subband is calculated.

3. PROPOSED METHOD USING AM DEMODULATION AND WAVELET FILTERS

The feature extraction techniques discussed in section 2 performs well for the clean speech signals. In real time conditions, the speech data is noisy due to sensor, environment and channel conditions. This reduces the signal to noise ratio and therefore degrades the performance of speaker recognition system. Hence an efficient method is required which can consider non-stationarity of the speech signal and can enhance signal to noise ratio. In this paper a hybrid noise robust speaker recognition system is proposed which utilizes advantages of both AM demodulation techniques (discussed in section 2.1) as well as wavelet filterbank (discussed in section 2.2). It is proposed to extract noise robust features from the AM demodulated envelop of real time signal and avoids loss of energy in sidelobes by using wavelet filterbank to improve signal to noise ratio. The proposed training system architecture used for noise robust speaker recognition system is shown in figure 2. It consists of three stages Pre-Processing, Feature extraction and Speaker modeling. In the preprocessing stage pre-emphasis and framing operations are performed. Feature extraction stage is modified by taking the features from the envelop of the speech signal instead of complete signal thereby reducing the noise effects. This is performed by AM demodulation of the signal in the frequency domain. This AM demodulated signal is then passed through a wavelet filterbank that is a close approximation of Mel filter bank. Since wavelet transforms avoids spilling of energy of the side lobes and this helps in selecting a non overlapping window. Speech features are then obtained by taking the Log and discrete cosine transform of the signal at the output of the filterbank. The speech features thus obtained are then modeled by using vector quantization. Each stage is described in detail in following section:

3.1 Preprocessing

Pre-emphasis-Pre-emphasis is done to boost the amount of energy in the high frequencies. In spectrum for voiced signals there is more energy at the lower frequencies than the higher frequencies. Boosting the high frequency energy makes information from these higher formants more available to the acoustic model.

Framing- Spectral features are extracted from a small window of speech assuming that the signal is stationary within this region. The window used in feature extraction is the Hamming window, which shrinks the values of the signal toward zero at the window boundaries, avoiding discontinuities. The equation or the window is given as

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\Pi n/L) & ; \ 0 \le n \le L-1 \\ 0 & ; \ \text{otherwise} \end{cases}$$

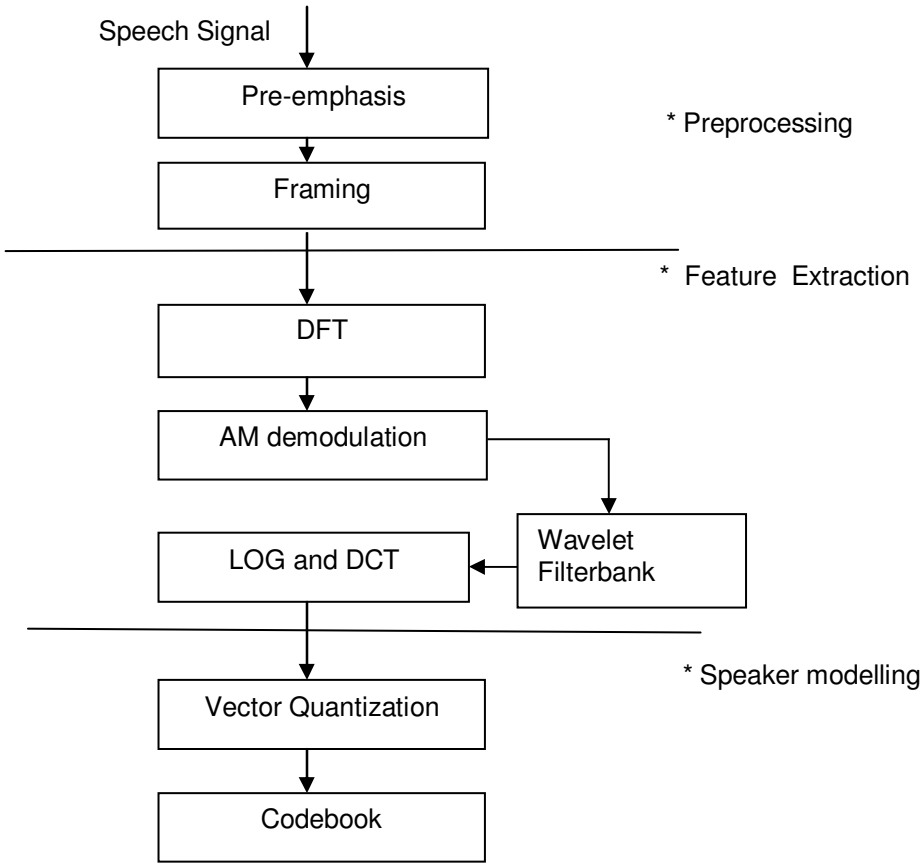


FIGURE 2: Training using proposed Algorithm.

3.2 Feature Extraction

In feature extraction stage it is proposed to extract the features from amplitude demodulated signal. Zhu et.al.[8] proposed that in frequency domain the speech signal is AM modulated signal with excitation spectrum as the carrier and Vocal tract transfer function (VTTF) as the modulating signal. They considered the speech waveform as the result of convolution between the excitation signal (which is either quasi –periodic, noise like, or a combination of the two) and the impulse response of the vocal tract transfer function (VTTF) as proposed by Fant [22]. Therefore in frequency domain the speech signal is obtained by multiplying the excitation spectrum (source spectrum) and the VTTF as shown in figure 3. If excitation spectrum is taken as the carrier and vocal tract transfer function as a modulating signal then by amplitude modulating the carrier with respect to the modulating signal in the frequency domain will give the results as shown in figure 3.

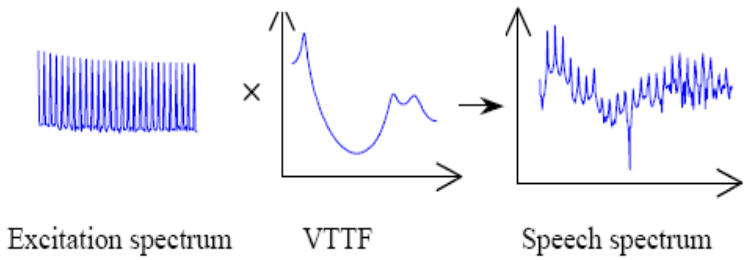


FIGURE 3: The linear source-filter model of speech production in the frequency domain. The x-axis is frequency [7].

Therefore in the proposed method, to AM demodulate the signal first the speech signal is transformed from time domain to frequency domain by using Discrete Fourier Transform (DFT). Then this signal is AM demodulated by using either Harmonic demodulation [8] or square law demodulator figure 4.

In Square law demodulator initially the signal is passed through the squarelaw device. Output of this is then passed through a low pass filter and then MFCC features are extracted. In this paper squarelaw demodulator is used for extracting the envelope.

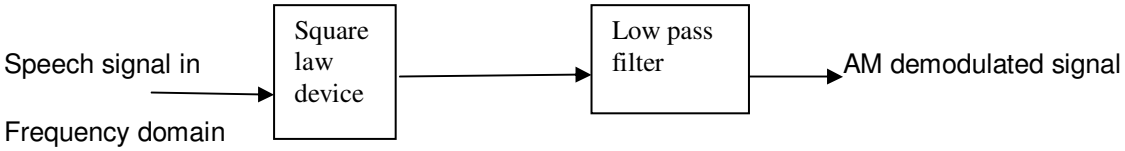
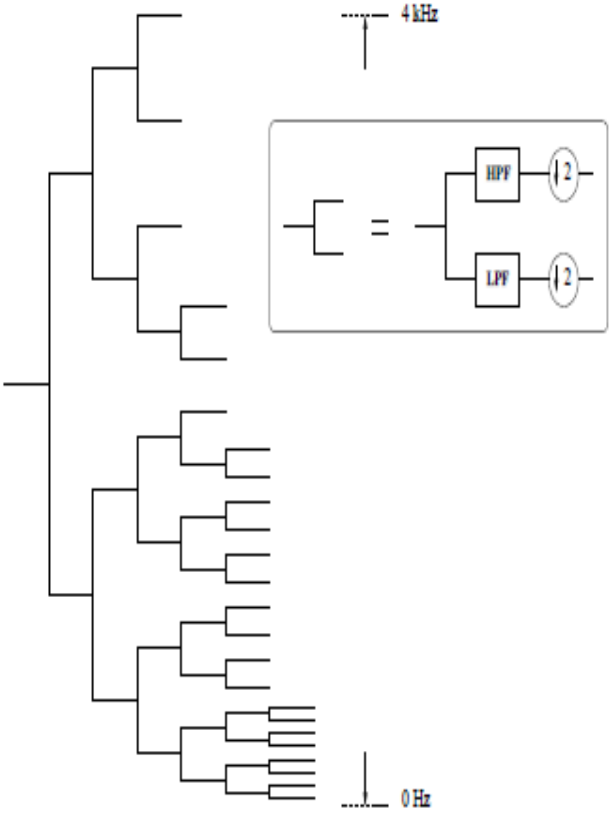


FIGURE 4: Block diagram of the square law demodulator

After extracting the envelop, instead of passing the signal through Mel filterbank, the speech signal is passed through the wavelet filterbank . For this filterbank suggested in [15] and given in figure 5 is used. Sarikaya & Hansen [15] performed a wavelet packet decomposition of the frequency range [0, 4] kHz such that the 24 frequency subbands obtained follow the Mel scale for the task of monophone recognition problem. The proposed analysis emphasizes low to mid frequencies assigning more subbands in these bands; overall, their decomposition preserves approximately a log-like distribution of the subbands across frequency. The wavelet packet decomposition is performed by using Daubechies'-32 wavelet filter. Finally, the log of each of the sub-band values is taken. In general the human response to signal level is logarithmic; humans are less sensitive to slight differences in amplitude at high amplitudes than at low amplitudes. In addition, using a log makes the feature estimates less sensitive to variations in input (for example power variations due to the speaker's mouth moving closer or further from the microphone). The last step in feature extraction is the computation of the Discrete cosine Transform. The outputs of the DCT stage are extracted features.

3.3 Speaker Modeling

The acoustic vectors extracted from input speech of each speaker provide a set of training vectors for that speaker. For each speaker there is a very large amount of data to store. The next step is therefore to condense this data. For clustering this data vector quantization is used. For that an algorithm, namely LBG algorithm [21], for clustering a set of *L* training vectors into a set of *M* codebook vectors is used. By using this LBG algorithm a set of vectors or codewords is created for each speaker. The codewords for a given speaker are then stored together in a codebook for that speaker. Each speaker's codebook is then stored together in a master codebook which is compared to the test samples during the testing phase.



Filter	SBCC central	MFCC
1	31	28
2	94	89
3	156	154
4	219	224
5	281	300
6	344	383
7	406	472
8	469	569
9	563	674
10	688	787
11	813	910
12	938	1043
13	1063	1187
14	1188	1343
15	1313	1512
16	1438	1694
17	1563	1892
18	1688	2106
19	1875	2338
20	2125	2589
21	2375	2860
22	2750	3154
23	3250	3472
24	3750	3817

FIGURE 5: 24-subband wavelet packet tree [15]

In the recognition phase, for making decision, features are extracted from an input utterance of an unknown voice using the proposed algorithm. The extracted features are then "vector-quantized". The *total VQ distortion* between each trained codebook and testing codebook is computed using Euclidean distance defined as

$$d(x, y_i) = \sqrt{\sum_{j=1}^{k} (x_j - y_{ij})^2}$$

where x_j is the jth component of the input vector, and y_{ij} is the jth is component of the codeword y_i. The speaker corresponding to the VQ codebook with smallest total distortion is identified as the speaker of the input utterance. The process is given in figure 6.

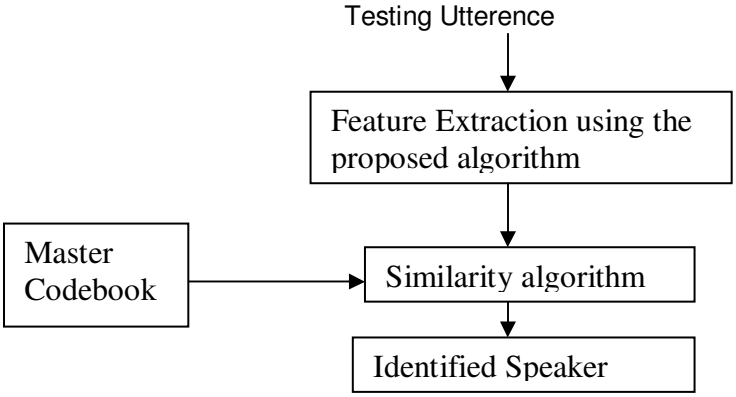


FIGURE 6: Testing using proposed Algorithm

4. RESULTS AND DISCUSSIONS

Performance of the proposed hybrid method is compared with feature extraction techniques using AM demodulation proposed by Zhu et.al. [8] [9] and named as HDMFCC, techniques using wavelet filterbank SBCC[15] , WPF [17] and with MFCC[6]. All these algorithms are implemented using MATLAB 7.4. These algorithms are tested for Hindi digits from 0 to 9 spoken by 80 male and 70 female speakers. The speech input is recorded at a sampling rate above 10000 Hz so that sampled signals covers all frequencies in human speech signal. Frame size of 30 millisecond is taken with 50% overlapping. DFT is performed by using 1024 point radix-2 FFT. A wavelet filterbank with 24 filters is used. The wavelet packet decomposition is performed by using Daubechies'-32 wavelet filter. A codebook of 64 codewords is used for performing vector quantization. The results obtained when tested for this real time data are given in Table 1 and figure 7.

Feature Extraction Technique Used	**% Recognition Efficiency**
MFCC [6]	63%
HDMFCC[7] [8]	65%
WPF [17]	68%
SBCC [15]	73%
Proposed Hybrid method	78%

TABLE 1: Recognition rates for MFCC, HDMFCC, SBCC, WPF, and AM-SBCC
Results: For Hindi digits from 0 to 9 spoken by 80 male and 70 female speakers

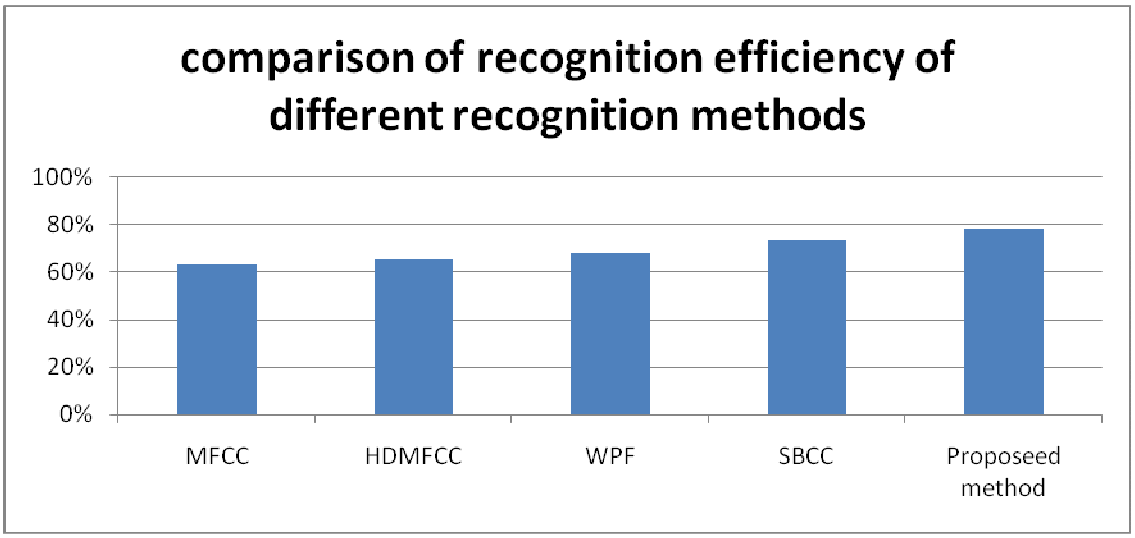


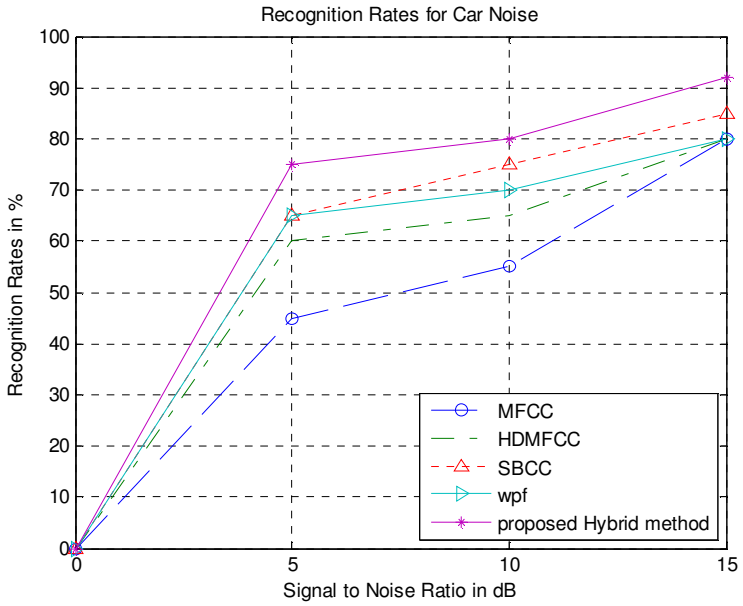
FIGURE 7: Comparison of recognition Efficiency of different recognition methods.

The results shows that the proposed hybrid method gives a recognition efficiency of 78% as compared to the 63% of MFCC for real time data. Compared with other methods also the recognition efficiency of the proposed method is higher. It gives improvement of 5 % as compared to SBCC[15] , improvement by 10 % as compared to WPF[17] and 13% when compared with HDMFCC[7] [8].

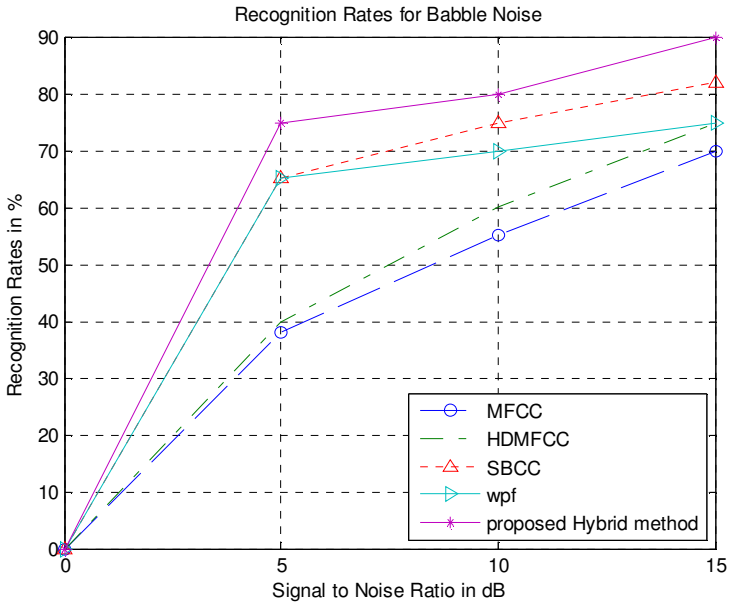
Experiments are also conducted by training the system with clean speech and testing their performance at different noise levels. Noisy speech corpus (NOIZEUS) is used [22]. For the experiments. This corpus is available to researchers freely and can be downloaded at http://www.utdallas.edu/~loizou/speech/noizeus/. In this speech corpora noise is artificially added to the speech signals. Noise signals were taken from the AURORA database [23] and included

the recordings from different places such as Babble (crowd of people), Car, Street, and Train. The noise signals were added to the speech signals at Signal to noise ratio of 0dB, 5dB, 10dB, and 15dB.

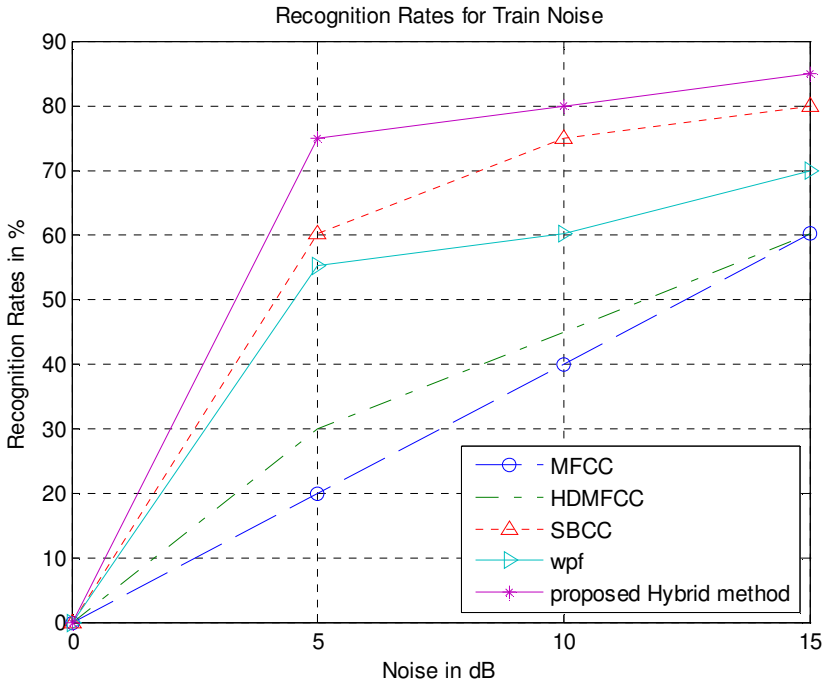
In the experiments the system under test is trained with the clean speech signal. During testing for recognition performance the same signal corrupted by different types of noise such as Car noise , Babble Noise , Street Noise and Train noise . The performance of the at signal to noise level of 0dB, 5dB, 10 dB, and 15 dB are plotted. The results obtained for the MFCC[6] HDMFCC[7] [8] , SBCC[15] , WPF [17] and proposed Hybrid method are given in figure 8. From the figure it is clear that recognition performance improves with the increase in signal to noise ratio for all the methods discussed. From the results it can be shown that proposed Hybrid method gives better recognition efficiency compared to other methods discussed at all signal to noise ratio .



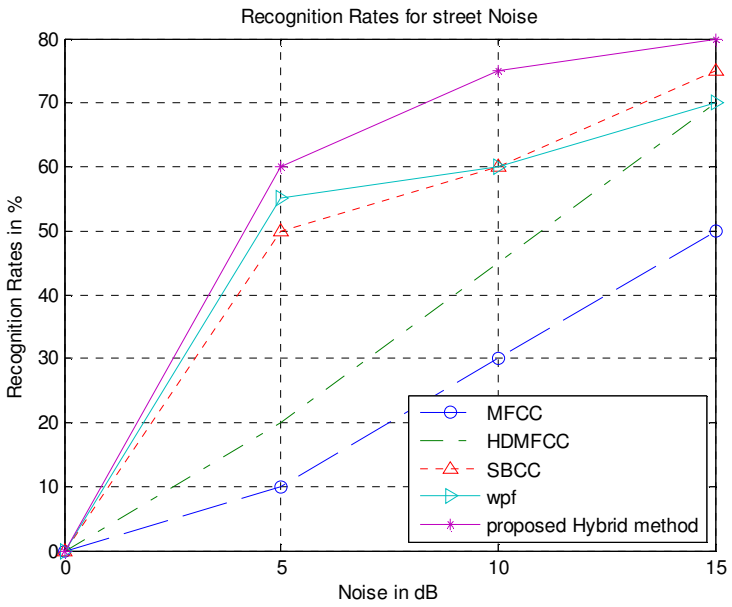
(a) Recognition Rates for Car noise



(b) Recognition Rates for Babble noise



(c) Recognition Rates for Train noise



(d) Recognition Rates for Street noise

FIGURE 8: Comparison of MFCC, HDMFCC, SBCC, WPF and, proposed method at different noise levels

5. CONCLUSION

The first drawback of MFCC is that the signal is assumed to be stationary within a given time frame and may therefore lack the ability to analyze localized events accurately. Hence it is proposed to extract the envelope of the signal. From results it is observed that the features extracted from envelope of the signal are more noise robust. The second limitation is, in MFCC triangular filterbank is used and frequency response of triangular filterbank is not smooth. Therefore it is proposed to use wavelet filterbank. Experimental results show that the proposed algorithm gives better recognition efficiency as compared to the other methods discussed. In future we will try to incorporate the Frequency modulation parameters for improving the speaker recognition performance

6. REFERENCES

[1] S. Pruzansky, "Pattern-matching procedure for automatic talker recognition", *J.A.S.A.*, 35, pp. 354-358, 1963.

[2] K. P. Li, et. al., "Experimental studies in speaker verification using a adaptive system, *J.A.S.A.*, 40, pp. 966-978, 1966.

[3] K. P. Li and G. W. Hughes, "Talker differences as they appear in correlation matrices of continuous speech spectra", *J.A.S.A.* , 55, pp. 833-837, 1974.

[4] B. Beek, et. al., "An assessment of the technology of automatic speech recognition for military applications", *IEEE Trans. Acoustics Speech and Signal Processing, ASSP*-25, pp. 310-322, 1977.

[5] M. R. Sambur, "Speaker recognition and verification using linear prediction analysis," .Ph. D. Dissert, M.I.T., 1972.

[6] P. Mermelstein and S. Davis, "Comparison of parametric representation for mono syllabic word recognition in continuously spoken sentences", *In IEEE Transactions on Acoustic Speech and Signal Processing*, Vol. 28, No. 4, pp. 357-366, 1980.

[7]. Q.Zhu and A. Alwan, "AM demodulation of speech spectra and its application to noise robust speech recognition" in proceedings *ICSLP*, 2000.

[8]. Q.Zhu and A. Alwan "Non linear feature extraction for robust speech recognition in stationary and non stationary noise" *Computer speech and Language* (17) ,pp. 381-402 , Elsevier Science Ltd. ,2003.

[9] A. Potamianos and P. Maragos "Speech analysis and synthesis using an AM-FM modulation model", *Speech Communication*, vol.28, (no.3), pp195-209, 1999.

[10] D. Dimitriadis, J.C. Segura, L. Garcia, A. Potamianos, P. Maragos and V. Pitsikalis "Advanced Front-end for Robust Speech Recognition in Extremely Adverse Environments", *Proc. of Intern. Conf. on Speech Communication and Technology - Interspeech 2007*, Antwerp, Belgium, Aug. 2007

[11] D. Dimitriadis, P. Maragos, and A.Potamianos "Robust AM-FM Features for Speech Recognition", *IEEE signal processing letters*, vol. 12, no. 9, pp. 621-624, Sep. 2005

[12] J.N. Gowdy, and Z. Tufekci, "Mel-Scaled Discrete Wavelet Coefficients for Speech Recognition," *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, pp 1351-1354, Jun 2000.

[13] Long and Dutta, "Wavelet based feature extraction for phoneme recognition", in the proceedings of *4th international conference of spoken language processing,USA* vol.1, 1996.

[14] T. Kinnunen, V. Hautamäki, P. Fränti "Fusion of spectral feature sets for accurate speaker identification", In Proc. 9th *Int. Conf. Speech and Computer* ,SPECOM ,2004

[15] Sarikaya et.al. "Wavelet packet transform features with application to speaker identification", in proceedings of the *IEEE Nordic signal processing symposium 1998*.

[16] R. Sarikaya and J. H. L. Hansen,"High resolution speech feature parameterization for monophone-based stressed speech recognition", *IEEE signal processing letters* vol7(7),pp. 182-185Jul 2000.

[17] O. Farooq and S. Datta "Mel Filter- Like Admissible Wavelet packet Structure for Speech Recognition" , *IEEE Signal Processing letters, Vol.8 , No. 7,pp 196-198 , Jul 2001*

[18] F.G. Zeng, K. Nie, G. S. Stickney, Y.Y.Kong, M. Vongphoe, A. Bhargave, C. Wei, and K. Cao "Speech recognition with amplitude and frequency modulations" *PNAS* vol. 102 , no. 7 ,pp 2293–2298, Feb., 2005

[19] P. Maragos, J. F. Kaiser and T. F. Quatieri , "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE transactions on signal processing*, vol. 41, no. 10, pp. 3024-3051 , Oct. 1993.

[20] F.Gunnar, " The acoustic theory of speech production" , S'Gravenhage , Mouton,1960.

[21] Y. Linde, A. Buzo, and R. M. Gray, ``An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications,* pp 84-95*,* Jan. 1980.

[22] Y. Hu, and P. Loizou, , "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, Elsevier, 49, pp 588-601, 2007.

[23] H. Hirsch, and D. Pearce , "The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions" , *ISCA ITRW ASR 2000*,
Paris, France, Sep 18-20, 2000.