Feature Extraction Techniques in Speech Processing: A Survey

Rekha Hibare
Department of Electronics &Telecommunication,
BMIT, Solapur (M.S.), India

Anup Vibhute
Department of Electronics & Telecommunication,
BMIT, Solapur (M.S.), India

ABSTRACT
Speech processing includes the various techniques such as speech coding, speech synthesis, speech recognition and speaker recognition. In the area of digital signal processing, speech processing has versatile applications so it is still an intensive field of research. Speech processing mostly performs two fundamental operations such as Feature Extraction and Classification. The main criterion for the good speech processing system is the selection of feature extraction technique which plays an important role in the system accuracy. This paper intends to focus on the survey of various feature extraction techniques in speech processing such as Fast Fourier Transforms, Linear Predictive Coding, Mel Frequency Cepstral Coefficients, Discrete Wavelet Transforms, Wavelet Packet Transforms, Hybrid Algorithm DWPD and their applications in speech processing.

Keywords
Feature Extraction; Fast Fourier Transform; Mel Frequency Cepstral Coefficients; Linear Predictive Coding; Discrete Wavelet Transforms; Wavelet Packet Transform; Hybrid Algorithm DWPD.

1. INTRODUCTION
The predominant mode of human communication for every day interaction is speech and it will also be the preferred mode for human-machine interaction [1].As speech signals are non-stationary in nature, speech recognition is a complex task due to the differences in gender, emotional state, accent, pronunciation, articulation, nasality, pitch, volume, and speed variability in people speak. Presence of background noise and other types of disturbances also makes a speech processing system complex and difficult. The performance of a speech processing system is usually measured in terms of recognition accuracy. Speech processing is useful for various applications such as like mobile applications, weather forecasting, agriculture, healthcare, automatic translation, robotics, video games, transcription, audio and video database search, household applications and language learning applications etc. [2]. Feature extraction and Classification are the two main stages of speech processing and among these stages; feature extraction is a key, because better feature is good for improving recognition rate. This survey focuses on details of various feature extraction techniques and its use by various researchers for speech processing.

2. FEATURE EXTRACTION TECHNIQUES

2.1 Fast Fourier Transform (FFT)
The Fourier Transform has countless applications ranging from telecommunications to crystallography, from speech recognition to astronomy, from Radar to mobile phones, from meteorology to archaeology. So it is one of the most useful mathematical techniques ever invented. For transforming the discrete-time signal from time domain into its frequency domain the FFT is nothing but the DFT but the difference is that the FFT is faster and more efficient on computation. So it is convenient to investigate FFT by firstly considering the N-point DFT equation which is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn} \quad (1)$$

Where $x(n)$ is the input and $W_N^{kn} = e^{jw_k n}$ is the phase factor. Also n, k are integers from 0 to N-1. Firstly separate $x(n)$ into two parts: $x(\text{odd})=x(2m+1)$ and $x(\text{even})=x(2m)$, where $m=0, 1,2,\ldots,N/2-1$. Then the N-point DFT equation also becomes two parts for each N/2 points which is given by

$$X(k) = \sum_{m=0}^{(N/2)-1} x(2m) W_N^{2mk} + W_N^k \sum_{m=0}^{(N/2)-1} x(2m+1) W_N^{2mk} \quad (2)$$

Where $m= 0, 1, 2... (N/2)-1$

As

$$(W_N^{kn})^2 = -W_{N/2}^{kn}$$

The N-point DFT equation finally given as

$$X(k) = \sum_{m=0}^{(N/2)-1} x_1(m) W_{N/2}^{mk} + W_N^{mk} \sum_{m=0}^{(N/2)-1} x_2(m) W_{N/2}^{mk} \quad (3)$$

$$= X_1(k) + W_N^k X_2(k) \text{ , } k=0, 1\ldots N/2.$$

$$X(k+\frac{N}{2}) = X_1(k) - W_N^k X_2(k) \text{ , } k=0, 1\ldots N/2.$$

So here the N-point DFT is separated into two N/2-point DFT.

For original N-point DFT Eq. (1), it has N^2 complex multiplications and N/2-point DFT Eq. (3) has $(N^2/2) + (N/2)$ multiplications. This is the process for reducing the calculations from N points to N/2 points. This signal for N point DFT is continuously separated until the final signal sequence is reduced to the one point sequence. So the total number of complex multiplications will be approximately reduced to $(N/2) \log_2 (N)$ [3].

Speaker-independent recognition system can be implemented using FFT. In [4], the experiments for speaker-independent recognition of 10 English vowels on isolated words which compared the use of an ear model were performed. Here FFT was used as a feature extractor and this FFT was done using a Mel scale and the same number of filters as for the ear model and Neural Networks were used as a classifier. The results obtained with recognition rate of 87% with the FFT preprocessing.

For noise-robust automatic speech recognition, an FFT-based companding algorithm for preprocessing speech was

described in [5]. To improve the performance of automatic speech recognition system in noisy environment, the algorithm which provided tone-to tone suppression and masking in the auditory system was used and results showed that this FFT-based companding was computationally efficient and it was also suited to digital implementations due to its use of the FFT.

A speaker-dependent, isolated-word speech recognition system was presented in [6]. This system was based on the use of the fast Fourier transform as a feature extractor to extract the features from the speech input and the classifier used for pattern matching was Dynamic Time Warping (DTW). The experimental results showed that the system has been successfully implemented and when tested using small dictionary, this system was also provided good results.

Recognition of musical instruments was presented in [7]. Here Feature Extraction was done using Fast Fourier Transform and K-NN classifier with cosine distance was used for pattern matching. Here FFT was applied instead of discrete Fourier transform because of its shorter time of calculations and concluded that the best results were obtained and for each sound of musical instrument efficiency of 100% was obtained for sound recognition.

To recognize the sound of a referee whistle in a noisy environment, an electronic system was described in [8]. Here Feature Extraction was done using Fast Fourier Transform (FFT) to obtain the signal spectrum in the frequency domain and concluded that FFT was an efficient algorithm to process DFTs and also due to lower calculation demanding of FFT, it was a better solution to be implemented on a microcontroller.

As the speed of speech recognition was enhanced without influencing the recognition rate by using integer FFT, so in [9], integer FFT was used to replace the floating FFT and also Artificial Neural Network (ANN) was used as a classifier for pattern matching. The experimental results revealed that from the FPGA platform, the speech recognition rate of the proposed hardware implementation methods was better than that in existing literatures. Also because of the FPGA chip, the speech recognition systems become applicable on the voice activated systems in toys, games, smart phones, office devices, vehicular communications, etc.

An automatic identification of bird calls without manual intervention which has been a meaningful research on the taxonomy and monitoring of bird migrations in ornithology was developed in [10]. Here for identification of bird calls, a new technique which computes the ensemble average on the FFT spectrum was proposed and also classification was done using Dynamic Time Warping (DTW) and Gaussian Mixture Modeling and this gave better results in identification of bird calls.

For physically challenged people who can use their voice only to register and attend the examination, one way to conduct online examination was offered in [11]. Here an efficient spectrum analysis to continuous-time signal based on DFT, the Fast Fourier Transform (FFT) method was adopted for authentication of voice and concluded that The Fast Fourier Transform (FFT) was a much faster mathematical algorithm which eliminates redundant calculations in the Fourier Transform and was therefore much speedier.

For Devanagari script and numerals, the frequency analysis of speech signals was presented in [12].This analysis was done by using FFT and the experimental results showed that for the spoken Devanagari Script and numerals, the Fourier descriptor feature was independent. By combining the Fourier transform and correlation technique commands which are used in MATLAB, a recognition system with high accuracy can be obtained.

2.2 Mel Frequency Cepstral Coefficient (MFCC)

The unique features of human voice can be extracted by using Mel Frequency Cepstral Coefficient (MFCC) and this MFCC also represents the short term power spectrum of human voice. To calculate the coefficients which represent the frequency Cepstral, MFCC is used and these coefficients are based on the linear cosine transform of the log power spectrum on the nonlinear Mel scale of frequency. The frequency bands are equally spaced in Mel scale and because of this it approximates the human voice more accurate. To convert the normal frequency *f* to the Mel scale *m*, Eq. (4) is used and it is given by

$$m = 2595\log_{10}(1+\frac{f}{100}) \quad (4)$$

By defining the pitch of 1000 Mel to a 1000 Hz tones, 40 db above the listener's threshold, Mel scale and normal frequency scale are referenced. Mel frequency is equally spaced on the Mel scale and this frequency is applied to linear space filters below 1000 Hz to linearize the Mel scale values and logarithmically spaced filter above 1000 Hz to find the log power of Mel scaled signal. Mel frequency wrapping is useful for the better representation of voice. By dividing the voice signal into frames, voice features are represented in MFCC and after that windowing process is carried out and finally Fourier transform of this windowing signal is carried out. To catch the changes between the different frames, Delta Cepstrum is used. It is also advantageous to have the time derivatives of (energy+MFCC) as new features, which shows the velocity and acceleration of (energy+MFCC). The equations to compute these features are given by

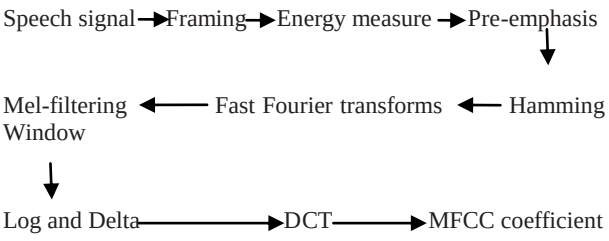
$$\Delta C_m(t) = [\sum \pi = -M^M C_m(t+\pi)\pi]/[\sum \pi = -M^M \pi^2] \quad (5)$$

In next step, DCT on the log energy *Ek* is carried out which is obtained from the triangular band pass filters to have L Mel scale cepstral coefficients. The formula for DCT is given by

$$C_m = \sum_k = 1^N \cos[m*(k-0.5)*\pi/N]*Ek$$

, *m*=1, 2... L (6)

Where *N* is the number of triangular band pass filters, L is the number of mel-scale cepstral coefficients. The value of M is usually set to 2.Since the process is of performance of FFT, DCT transforms the frequency domain into a time-like domain called quefrency domain and this obtained features are similar to cepstrum. Thus it is referred to as the Mel-scale cepstral coefficients or MFCC. Following figure shows the complete pipeline of Mel Frequency Cepstral Coefficients (MFCC) [13].



A technique for recognizing spoken letter in Bengali Language was presented in [13] and here feature was derived from spoken letter. Mel-frequency cepstral coefficient (MFCC) was used as a feature extractor to characterize a feature. To calculate the distance of an unknown letter with the stored ones, DWT was used and K-Nearest Neighbors algorithm was used as a classifier to improve accuracy in noisy environment. This system is useful for commands & control, data entry, mobile telephony and home automation task. Also a speaker independent Arabic speech recognition system was presented in [14]. Here this system was applied on the connected Arabic digits or numbers. The Mel Frequency Cepstral Coefficients (MFCC) technique was used for feature extraction and neural and support vector machines were used for classification. The performance of the system was found nearly 94%. This spoken digit recognition system is useful in various applications which are related with the use of numbers as input such as telephone dialing using speech, airline reservation, and automatic directory to retrieve or send information.

A voice recognition system for secure ATM was developed in [15] and this system used the technique of MFCC as a feature extractor to extract unique and reliable human voice feature pitch in the form of Mel frequency and also Hidden Markov Model (HMM) was used as a classifier for pattern matching. By using combination of MFCC and HMM, the system showed 86.67% results as correct acceptance and corrected rejections with the error rate of 13.33% and defined the MFCC as the unique and reliable feature extraction technique.

For human-machine interactions, automatic emotion recognition system in speech is a current research area and this system can be useful in a wide range of applications. To classify five emotional states such as anger, happiness, sadness, surprise and a neutral state, speech emotion recognition system was presented in [16]. Here a new method by using multi-class SVM which was a method based on SVM was used as a classifier and it was observed that recognition rate by using multi-class SVM was more than that by using SVM for Linear, Polynomial, RBF, and Sigmoid Kernel Function.

Marathi numerals recognition system was developed in [17]. Here Mel-Frequency Cepstral Coefficient (MFCC) was used for Feature Extraction and Distance Time Warping was used for feature matching which gave better results. This Numeral recognition system is useful in various applications related with reading postal zip code, passport number, employee code, postal mail sorting, job application form sorting, bank cheque processing automatic scoring of tests that contains multiple choice questions and video gaming etc.

For effective speech recognition system which were considered along with NRL speech evaluations in noise corpus entitled SPINE, in [18] various methodologies were discussed and also to improve speech recognition in noisy environments which was based on the SPINE corpus, various addressing trade-offs were focused. Here feature extraction was done by using MFCC which provided better results.

2.3 Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) is another method, which is used to obtain a frequency spectrum. This is the most successful method in widespread use today and it is a stronger method to analyze the coded voice files with better quality on low bit rate samples. There are various advantages for the use of LPC and they are: (a) LPC proves better approximation coefficient spectrum (b) LPC gives shorter and efficient calculation time for signal parameters and (c) LPC has been able to get important characteristics of the input signals [20].In LPC, the values of the signal can be expressed as a linear combination of the preceding values. That is, if s(i) is the amplitude at time i,

$$s(i) = a1*s(i-1) + a2*s(i-2) + ... + ap*s(i-p) \quad (7)$$

When the input data is filled in, this becomes a system of linear equations which can be solved to determine the values of a1 through ap. These values are useful to produce a signal which is free from noise and clearly identifies the formants. The typical values for p are 10-12 [19].

Gender recognition in two models, one for generating Formant values of the voice sample and the other for generating pitch value of the voice sample was presented in [20]. Here LPC technique was used for feature extraction which gave better results. For classification purpose the nearest neighbor method was used which calculates Euclidean distance from the Mean value of Males and Females. Gender recognition is useful in various applications which are related with biometric security, mobile and automated telephonic communication.

A research focused on analysis of matching process which gives a command for multipurpose machine such as a robot was presented in [21].Here Linear Predictive Coding (LPC) was used for feature extraction to analyze voice signals by giving characteristics into LPC coefficients and Hidden Markov Model (HMM) was used for classification. Such Biometric systems are commonly used for identification and verification of an individual to acquire the identity of the authorized individuals.

About implementation of speech recognition system on a mobile robot for controlling movement of the robot was described in [22]. Here LPC method used for extracting the features of voice signal and Artificial Neural Network is used for classification. By using these two techniques the highest recognition rate of 91.4% was obtained.

A Microprocessor Implementation that uses LPC as a feature extractor for isolated word recognizer which was useful in a module of dedicated hardware that used a microprocessor and programmable digital signal processing circuitry was developed in [23] and here pattern comparison was done by using dynamic time warping (DTW). Experimental results showed that the recognition algorithm and analysis parameters supported minicomputer simulations with greater processing speed, smaller size, and lower cost than array processor.

A low complexity but effective approach for speech/music discrimination, which exploits only one simple feature, called Warped LPC-based Spectral Centroid (WLPC-SC) was presented in [24] and a three-component Gaussian Mixture

Model (GMM) was used for classification. Experimental results revealed that the speech/music discriminator was robust and fast and it is also suitable for real-time multimedia applications.

An approach to recognize English words corresponding to digits Zero to Nine was proposed in [25]. A set of features consisting of a combination of Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), Zero Crossing Rate (ZCR), and Short Time Energy (STE) of the audio signal, were used to generate a feature vector, which was subsequently used for discrimination. Artificial neural networks (ANN) with feed-forward back-propagation architectures were used as a classifier. By the combination of

these features extraction techniques an accuracy of 85% was obtained which was more than that of by using a single feature extraction technique. The part of automatic speech recognition that recognizes phonemes using standard signal analysis methods such as DFT and LPC was implemented in [26]. A multilayer perceptron artificial neural network was used as a classifier and this gave better results.

Recognition of Vernacular Language Speech for Discrete Words using Linear Predictive Coding Technique for better interpretation of spoken words was presented in [27]. Here LPC coefficients as a feature extractor were used for compaction and learning the data for discrete spoken words and fuzzy neural networks were used as a classifier. The experimental results showed good precisions. The Bangla speech recognition system was presented in [28]. The feature extractor used a standard LPC cepstrum coder and pattern recognition was done by using artificial neural network (ANN) which provided better results.

Speech recognition in noisy car environment based on OSALPC representation which has shown to be attractive for speech recognition because of its simplicity and its high recognition performance than LPC was developed in [29] and concluded that cepstral representation based on linear prediction of One-Sided Autocorrelation sequence (OSALPC) provided excellent results in severe noisy car environment.

Performance Analysis of Optimization Tool for various values of the order of filters and various windows for Speech Recognition Using LPC was implemented in [30]. The speech coders taken for study were Linear Predictive Coder (LPC) and Cepstral coder. In this system they used the Levinson-Durbin and Time-Varying Lattice Filter blocks for low-bandwidth transmission of speech using linear predictive coding.

2.4 Wavelet Transform (WT)

The wavelet transform is very well suited for speech processing because of its similarity to how the human ear processes sound and it is a multi-resolutional and multi-scale analysis. A brief description of the three methods of wavelet transform used such as DWT, WPD and DWPD are given below.

2.4.1 Discrete Wavelet Transform (DWT)

Information about non-stationary signals like audio can be extracted by using DWT as it is a relatively recent and computationally efficient technique for feature extraction. of The wavelet transforms has a varying window size, being broad at low frequencies and narrow at high frequencies and thus leading to an optimal time–frequency resolution in all frequency ranges. To obtain a time-scale representation of the signals DWT and WPD use digital filtering. DWT is defined by following equation

$$W(j,k) = \sum_{j}\sum_{k} X(k)2^{-j/2}\Psi(2^{-j}n-k) \qquad (8)$$

Here Ψ (t) is the basic analyzing function which is called as a mother wavelet. In DWT, the original signal passes through two filters such as a low-pass filter and a high-pass filter and emerges as two signals. The output of a low pass filter is called as approximation coefficients and the output of high-pass filter is called as detail coefficients. In speech signals, the low frequency components characterize a signal more than its high frequency components and thus the low frequency components *h[n]* are of greater importance than that of high frequency signals *g[n]*. The successive high pass and low pass filtering of the signal is given by following equations

$$Y_{high}[k] = \sum_{n} x(n)g(2k-n) \qquad (9)$$

$$Y_{low}[k] = \sum_{n} x(n)h(2k-n) \qquad (10)$$

Where Y_{high} are the detail coefficients and Y_{low} are the approximation coefficients which are the outputs of the high pass and low pass filters obtained by sub sampling by factor 2. This filtering process is continued until the desired level is obtained according to Mallat algorithm [2]. The DWT decomposition tree is given in figure 1.

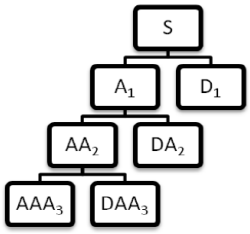


Fig 1: DWT Decomposition Tree

2.4.2 Wavelet Packet Decomposition (WPD)

WPD gives good time and frequency resolutions and thus it is useful in various fields of speech processing. WPD is nothing but a generalization of DWT and it is a more flexible and detailed method than DWT. In WPD, the signal is decomposed into low frequency components and high frequency components at each level like in DWT but there is a difference between DWT and WPD that the discrete wavelet transform is applied to the low pass result only whereas WPD applies the transform step to both the low pass and the high pass result. The decomposition tree for WPD is shown in figure 2.

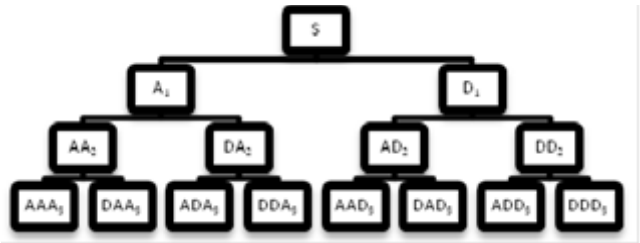


Fig 2: WPD Decomposition Tree

2.4.3 Hybrid Algorithm DWPD

A wavelet transform decomposes a signal into sub-bands. Here low frequency components contain the characteristics of a signal and high frequency components contain noise and disturbance in a signal. Removing of the high frequency contents retains the features of the signal and thus it causes reduction of the noise in the signal. But the disadvantage is that sometimes the high frequency components may contain useful features of the signal. It means that the main drawback related with DWT is that it cannot decompose the high frequency band into more partitions. Although WPD can achieve this decomposition, it is also applied to low frequency band signals and these low frequency band signals mainly includes the desired signals. So this causes unnecessary computational complexity which is one of the limitations. To overcome these limitations of DWT and WPD, there is a new algorithm for speech enhancement which combines the

features of both DWT and WPD. The outline of this DWPD algorithm is given below.

a) In DWPD, the speech signal is split into two bands namely a low frequency band signal and a high frequency band signal.

b) DWT is applied on the low frequency components whereas WPD are applied on the high frequency components.

c) The features obtained from both decompositions are combined together and this form the feature vector set [2].

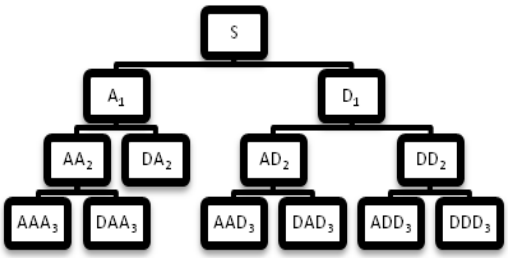


Fig 3: DWPD Decomposition Tree

The main advantage of this DWPD algorithm is that it can not only decompose high frequency band into more partitions but also save complexities in computation [2].

A phoneme recognition system based on Discrete Wavelet Transforms (DWT) as a feature extractor and Support Vector Machine (SVM) as a classifier for multi-speaker continuous speech environments was developed in [31]. Phonemes were divided into frames, and the DWTs were used to obtain fixed dimensional feature vectors and provided better accuracy.

An Automatic Emotion Recognition (AER) from speech was developed in [32]. DWT was used for the feature extraction and Artificial Neural Networks were used as classifier. Overall recognition accuracies of 72.05 %, 66.05%, and 71.25% could be obtained for male, female and combined male and female databases respectively.

Three novel noise robustness techniques for speech recognition based on DWT, which were wavelet filter cepstral coefficients (WFCCs), sub-band power normalization (SBPN), and low pass filtering plus zero interpolation (LFZI) were proposed in [33]. The proposed WFCC was found to provide a more robust c0 (the zeroth cepstral coefficient) for speech recognition.

A novel wavelet packet filter bank approach was presented in [34] to identify non-uniformly distributed dynamic characteristics of the speaker. Here Dynamic Time Warping (DTW) was used as a classifier. Evaluation results revealed that this method outperformed by incorporating the speaker-specific dynamic characteristics and also phase information of the speech signal.

An improved feature extraction method that was called Wavelet Cepstral Coefficients (WCC) was proposed in [35]. Comparisons with the traditional Mel-Frequency Cepstral Coefficients (MFCC) were done for further analyze the effectiveness of the WCCs and found that the WCCs showed some comparable results when compared to the MFCCs considering the WCCs small vector dimension when compared to the MFCCs.

An automatic speech recognition system for Polish continuous speech was demonstrated in [36]. Here to obtain a discrete wavelet power spectrum six levels dyadic decomposition procedure with discrete Meyer wavelet decomposition filters were applied to speech signal which provided better results.

A new hierarchical structure for Speech Recognition by units smaller than words was proposed in [37] and developed a recognition logic based on the production characteristics of phonemes in Brazilian Portuguese using Wavelet Packet Decomposition (WPD). Here Support Vector Machine (SVM) was used as a classifier. This combination of WPD and SVM provided good mean recognition rate of 98.16% for vowel recognition and 98.41% for consonant recognition and 96.82% for final total word recognition.

The idea using a hybrid approach of wavelet transforms tap 9/7 and MFCC was investigated in [38].This hybrid approach was used for feature extraction to recognize speech signals. Here neural networks were used for classification. The recognition rate of this hybrid approach gave better result than the recognizers based on MFCC or wavelet.

A new pre-processing stage based on wavelet denoising was presented in [39].This was proposed in the presence of additive white Gaussian noise to extract robust features. Here with the commonly used Mel frequency cepstral coefficients (MFCCs) with and without the preprocessing stage, performance of recognition was compared and it was found that using the proposed technique by 2 to 28% for signal to noise ratio in the range of 20 to 0 dB, the recognition accuracy of word was found to improve.

A Hybrid Wavelet-Fourier-HMM Speaker Recognition system was presented in [40]. This system was based on Wavelet-Fourier Transform (WFT) and more traditional HTK based system on Mel Frequency Cepstral Coefficients (MFCC) with voiced/unvoiced classes to improve recognition rate. The various methods of speaker recognition such as multivariate kernel density, Gaussian mixture models, artificial neural networks or support vector machine were used. Experimental results showed that DWFT branch alone gave recognition worse than 80% and also HTK-based system alone was not perfect having recognition rate of around 90%. But the hybrid system using both branches gave recognition rate of 100%.

A speech recognition system combining FFT and wavelet functions were developed in [41].Here neural networks are used for classification. When the speaker voice sample was distorted, either deliberately or by imperfections of the recording system, advantage of this system could be seen and the main disadvantage of this method was that when compared to some other procedures which were used for extraction of characteristics of a speaker's voice, it gave greater sensitivity to noise in the sample.

A Robust Speech Recognition Using Perceptual Wavelet-Packet Transform and Mel-frequency Product Spectrum Cepstral Coefficient was proposed in [42].Here Hidden Markov Model was used as a classifier. This proposed approach was compared with the MFCC-based conventional feature extraction method and showed that the proposed method improves recognition accuracy rate by 44.71 %, with an average value of 14.80 % computed on 7 SNR level for white Gaussian noise conditions.

A Robust Speech Recognition system using Gammatone Wavelet Cepstral Coefficients (GWCC) was developed in [43].Here the conventional Mel filter bank in MFCC was replaced with a Gammatone wavelet filter bank and results showed that when compared with MFFCs, the Gammatone based features yield a better recognition performance at low SNRs.

A speech recognition system for recognizing speaker-independent, isolated Words in Malayalam was developed in [44]. Feature extraction in the time- frequency domain was performed using Wavelet Packet Decomposition (WPD) and concluded that Wavelets were very much suitable for processing non stationary signals like speech because of its multi-resolution characteristics and efficient time frequency localizations. Here Artificial Neural Network (ANN) was used as a classifier. By using this hybrid architecture of WPD and ANN 87.5% recognition accuracy was obtained. Also a speech recognition system for recognizing isolated words in Malayalam was developed in [2] and used two wavelet based techniques namely Discrete Wavelet Transforms (DWT) and Wavelet Packet Decomposition (WPD) for extracting features from speech. Here Support Vector Machine (SVM) was used as a classifier. A recognition accuracy of 85.4 % was obtained using DWT and 83.2% for WPD. A new feature extraction method was proposed which uses the combined features of both DWT and WPD called Discrete Wavelet Packet Decomposition (DWPD). The feature vectors obtained from this hybrid method gave recognition accuracy of 87.8%. A speech recognition system using two different feature extraction techniques such as Linear predictive Coding (LPC) and Discrete Wavelet Transforms (DWT) was presented in [45] and a comparative study was carried out for recognizing speaker independent spoken isolated words and concluded that both the methods produce good recognition accuracy but discrete wavelet transforms were found to be more suitable for recognizing speech because of their multi-resolution characteristics and efficient time frequency localizations.

A modified voice identification system which was useful in business and consumer environment using over sampled Haar wavelets followed by proper orthogonal decomposition was developed in [46]. Results showed that when speech was collected through mouthpieces, this system gave consistently better performance, but when audio was collected through telephones, it gave comparatively poor performance. At the cost of higher computational time, the better performance was obtained.

Speaker Verification System using Discrete Wavelet Transform (DWT) and Logarithmic Power Spectrum Density (PSD) which were integrated for speaker accurate formants extraction was developed in [47] and the experimental results showed excellent performance with recognition rate of around 95%. This system can be useful in various applications such as in password, PINs identification, security system or mobile phones.

The performance analysis of voice activity algorithms (VAD) which were based on wavelet and AMR-WB (Adaptive Multi-Rate Wideband) speech codec was developed in [48] and HMM classifier was used for pattern matching. The experimental results showed that wavelet approaches provided good results in clean, noisy and reverberant environments with respect to speech clipping and also gave a much lower computational complexity. The performance of this system by using AMR VAD was improved upon by approaches which were based on wavelet.

The design of a hearing device based on fast wavelet transform and the difficulties in the use of wavelet transform for speech processing were proposed in [49] and showed that careful selection of wavelet coefficients was essential for the four major categories of-voiced speech, plosives, fricatives and silence to be identified and with the knowledge of these four categories showed that speech can be easily and effectively segmented.

A new method which was based on power fluctuations of the wavelet spectrum for a speech signal was presented in [50].Here discrete wavelet transform (DWT) was used as a feature extractor to analyze speech signals, the resulting power spectrum and its derivatives and concluded that DWT was efficient because some phonemes have power variations in the narrow band only and it was much easier to detect those analyzing DWT sub signals than the power of the whole signal.

A speech recognition system using Discrete Wavelets transformation using Daubechies wavelets was developed in [51]. This system was based on analysis of signals of regional dialects and gave similar pattern in the Regional Dialects of Demographic Region that provided a highly reliable way for recognizing speech. This system was also useful to find a new and effective technique which was helpful for analyzing signals of regional dialects and the current scenario of high performance computing.

Speech Recognition Using a Wavelet Transform to establish fuzzy interface system through Subtractive clustering was developed in [52]. This system used the combination of a feature extraction by wavelet transform, subtractive clustering and adaptive neuro-fuzzy inference system (ANFIS).The performance was evaluated by different samples of speech signals- isolated words- with added background noise and obtained a recognition rate of about 99%.

3. CONCLUSION

As feature extraction is a crucial phase in speech processing so in this paper some of the techniques for feature extraction were reviewed. A well-chosen feature set can result in quality recognition whereas a wrongly chosen feature set can result in poor recognition. In FFTs, MFCCs and LPCs the feature vector dimensions and computational complexity are higher to a great extent, also they have reduced accuracy and fixed window size because of which they are not suitable for non-stationary signals like speech. Using wavelet transforms the computational complexity and the feature vector size are successfully reduced and they have better accuracy, varying window size because of which they are suitable for non – stationary signals. Thus wavelet transform is an elegant tool for the analysis of non-stationary signals like speech. Hence different feature extraction techniques can be used for different kinds of applications. Though there are various common applications between these feature extraction techniques, among all these techniques wavelet transform gives better system accuracy.

Future works include experimenting with different numbers of coefficients, wavelet families and wavelet structure. In future experiments, the WCCs (Wavelet Cepstral Coefficients) should also be tested under different noisy conditions or environments to observe its robustness towards noisy speech or voice. Also there is a need to develop different new hybrid methods that will give better performance in robust speech processing area. To obtain better accuracy, in prosodic, text preprocessing and pronunciation fields there is still much work, research and improvements to be done.

4. REFERENCES

[1] Dr.Yousra F., Al-Irhaim Enaam Ghanem Saeed, "Arabic word recognition using wavelet neural network", *Scientific Conference in Information Technology*, November 2010.

[2] Sonia Sunny, David Peter S, K Poulose Jacob, "Design of a Novel Hybrid Algorithm for Improved Speech

Recognition with Support vector Machines Classifier", *International Journal of Emerging Technology and Advanced Engineering*, vol.3, pp.249-254, June 2013.

[3] Tingxiao Yang, "The Algorithms of Speech Recognition, Programming and Simulating in MATLAB", *University of Gavale*, pp.1-49, January 2012.

[4] Yoshua Bengio, Renato De Mori, Regis Cardin, "Speaker Independent Speech Recognition with Neural Networks and Speech Knowledge", *Department of Computer Science McGill University*, pp.218-225.

[5] Bhiksha Raj, Lorenzo Turicchia, Bent Schmidt-Nielsen, and Rahul Sarpeshkar, "An FFT-Based Companding Front End for Noise-Robust Automatic Speech Recognition", *EURASIP Journal on Audio, Speech, and Music Processing*, vol.2007,pp.1-13,2007.

[6] Greg Hopper, Reza Adhami, "An FFT-based speech recognition system", *Journal of Franklin Institute*, vol.329, no.3, pp.555-565, May 1992.

[7] Adam Glowacz, Witold Glowacz, Andrzej Glowacz, "Sound Recognition of Musical Instruments with Application of FFT and K-NN classifier with Cosine Distance "*,AGH university of Science and Technology, Work supported by European Regional Development Fund INSIGMA Project No.POIG.01.01.02-00-062/09*, 2010.

[8] Gil Lopes, Fernando Ribeiro, Paulo Carvalho, "Whistle Sound Recognition in Noisy Environment "*Universidade do Minho, Departamento de Electrónica Industrial, Guimarães, Portugal.*

[9] Shing-Tai Pan, Chih-Chin Lai and Bo-Yu Tsai, "The Implementation of Speech Recognition Systems on FPGA-Based Embedded Systems with SOC Architecture", *International Journal of Innovative Computing, Information and Control*,vol.7, no.11,pp.6161-6175, November 2011.

[10] Hemant Tyagi, Rajesh M. Hegde, Hema A. Murthy and Anil Prabhakar, "Automatic Identification of Bird calls using Spectral Ensemble Average Voice Prints", *Indian Institute of Technology Madras.*

[11] Dwijen Rudrapal, Smita Das, S. Debbarma, N. Kar, N. Debbarma, "Voice Recognition and Authentication as a Proficient Biometric Tool and its Application in Online Exam for P.H People", *International Journal of Computer Applications (0975 – 8887)*, vol.39,no.12, pp.7-12, February 2012.

[12] Umesh Kumar Gupta,Dr. R. K. Prasad, "Frequency Analisys of Speech Signals for Devanagari Script and Numerals Using FFT", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol.3.no. 5, pp. 471-477, May 2013.

[13] Asm Sayem, "Speech Analysis for Alphabets in Bangla Language:Automatic Speech Recognition", *International Journal of Engineering Research*, vol.3, no.2, pp.88-93, February 2014.

[14] Shady Y. EL-Mashed ,Mohammed I. Sharway , Hala H. Zayed, "Speaker Independent Arabic Speech Recognition using Support Vector Machin*", Department of Electrical Engineering, Shoubra Faculty of Engineering, Benha University, Cairo, Egypt.*

[15] Shumaila Iqbal, Tahira Mahboob and Malik Sikandar Hayat Khiyal, "Voice Recognition using HMM with MFCC for Secure ATM", *IJCSI International Journal of Computer Science Issues*, vol. 8, No 3, pp. 297-303, November 2011,

[16] Vaishali M. Chavan, V.V. Gohokar, "Speech Emotion Recognition by using SVM-Classifier", *International Journal of Engineering and Advanced Technology (IJEAT)*, vol.1, pp.11-15, June 2012.

[17] Siddheshwar S. Gangonda, Dr. Prachi Mukherji, "Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features", *International Journal of Engineering Research and Applications (IJERA)*, pp.218-222, March 2012.

[18] John H.L. Hansen, Ruhi Sarikaya, Umit Yapanel, Bryan Pellom, "Robust Speech Recognition in Noise: An Evaluation using the SPINE Corpus", *CSLR: Center for Spoken Language Research; Robust Speech Processing Laboratory*, 2001.

[19] David Wagner, "A Speech Recognition Project", http://www.cs.dartmouth.edu/~dwagn/aiproj/speech.html

[20] Kumar Rakesh, Subhangi Dutta and Kumara Shama, "Gender Recognition using Speech Processing Techniques in Lab View", *International Journal of Advances in Engineering & Technology*, vol.1, pp.51-63, May 2011,

[21] Wahyu Kusuma R., Prince Brave Guhyapati V., "Simulation Voice Recognition System for controlling Robotic Applications", *Journal of Theoretical and Applied Information Technology*,vol.39, no.2,pp. 188-196, May 2012.

[22] Thiang and Suryo Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot", *International Conference on Information and Electronics Engineering*, vol.6, pp.179-183, 2011.

[23] John G. Ackenhusen, L. R. Rabiner, "Microprocessor Implementation of an LPC-Based Isolated Word Recognizer", *Proc. IEEE*, pp.746-749, 1981.

[24] J.E. Munoz-Exposito, S. Garcia-Galan, N. Ruiz-Reyes, P. Vera-Candeas and F. Rivas-Pena, "Speech/Music Discrimination using a single Warped LPC-Based Feature", *Queen Mary University of Londan*, pp.614-617, 2005.

[25] Bishnu Prasad Das, Ranjan Parekh, "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers", *International Journal of Modern Engineering Research*, vol.2, pp.854-858, May-June 2012.

[26] Tobias Bengtsson, "Speech recognition using multilayer perceptron artificial neural network", *Department of Computer Science Lund University*.

[27] Omesh Wadhwani, Amit Kolhe, Sanjay Dekate, "Recognition of Vernacular Language Speech for Discrete Words using Linear Predictive Coding Technique", *International Journal of Soft Computing and Engineering*, vol. 1, pp.188-192, November 2011.

[28] Paul A.K., Das D., Kamal M.M., "Bangla Speech Recognition System Using LPC and ANN", *Proc. IEEE*, pp. 171-174, 2001.

[29] Javier Harnando, Climent Nadeu, "Speech Recognition in noisy car environment based on OSALPC representation and robust similarity measuring techniques", *Proc. IEEE*, 1994.

[30] Kadam V.K, Dr.R.C.Thool, "Performance Analysis of Optimization Tool for Speech Recognition Using LPC & DSK TMS3206711/13 Using Simulink & Matlab", *International Journal Of Computational Engineering Research*, vol.2, pp.1243-1248, September 2012.

[31] Gatt E., Grech I, Casha O., "Discrete wavelet transforms with multiclass SVM for phoneme recognition", *Proc. IEEE*, pp.1695-1700, 2013.

[32] Firoz Shah. A, Raji Sukumar. A and Babu Anto. P, "Discrete Wavelet Transforms and Artificial Neural Networks for Speech Emotion Recognition", *International Journal of Computer Theory and Engineering*, vol. 2, no. 3, pp.319-322, June 2010.

[33] Jeih-Weih Hung, Hao-Teng Fan , and Syu-Siang Wang, "Several New DWT-Based Methods for Noise-Robust Speech Recognition", *International Journal of Innovation, Management and Technology*, vol. 3, no. 5, pp.547-551, October 2012.

[34] Jagannath H Nirmal, Mukesh A Zaveri, Suprava Patnaik1 and Pramod H Kachare, "A novel voice conversion approach using admissible wavelet packet decomposition", *EURASIP Journal on Audio, Speech, and Music Processing*, 2013.

[35] T. B. Adam, M. S. Salam, T. S. Gunawan, "Wavelet Cesptral Coefficients for Isolated Speech Recognition", *International Islamic University Malaysia*, vol.11, no.5, pp.2731-2738, May 2013.

[36] Mariusz Ziołko, Jakub Gałka, Bartosz Ziołko, Tomasz Jadczyk, Dawid Skurzok, Jan Wicijowski, "Automatic Speech Recognition System Based on Wavelet Analysis", *IEEE Fourth International Conference on Semantic Computing*, pp.450-451, 2010.

[37] Adriano de Andrade Bresolin, Adriao Duarte Doria Neto e Pablo Javier Alsina, "A New Hierarchical Structure for Speech Recognition by units smaller than words, using Wavelet Packet and SVM", *UTFPR Brazil, UFRN Brazil*.

[38] Sozan Mahmood and Mihran Abdulrahim, "Hybrid Speech Recognition System based on Wavelet 9/7 and Mel-Frequency Cepstral Coefficient", *International Conference on Emerging Trends in Computer and Electronics Engineering*, pp.19-22, March 2012.

[39] S.Datta and Farooq O. "Wavelet-based denoising for robust feature extraction for speech recognition", *Proc. IEEE*, vol.39, pp.163-165, January 2003.

[40] Bartosz Ziolko, Wojciech Koz lowski, Mariusz Ziolko, Rafa Samborski, David Sierra, Jakub Ga lka, "Hybrid Wavelet-Fourier-HMM Speaker Recognition", *AGH University of Science and Technology Krakow, Poland*, July 2011.

[41] Sanja Grubesa, Tomislav Grubesa, Hrvoje Domitrovic, "Speaker Recognition Method combining FFT, Wavelet Functions and Neural Networks", *Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia.*

[42] Mohamed Cherif Amara Korba, Djemil Messadeg, Rafik Djemili, Hocine Bourouba, "Robust Speech Recognition Using Perceptual Wavelet Denoising and Mel-frequency Product Spectrum Cepstral Coefficient Features", *Informatica 32*, pp.283-288, 2008.

[43] Aniruddha Adiga, Mathew Magimai, Chandra Sekhar Seelamantula,"Gammatone Wavelet Cepstral Coefficients for Robust Speech Recognition".

[44] Sonia Sunny, David Peter S, K Poulose Jacob, "Development of a Speech Recognition System for Speaker Independent Isolated Malayalam Words", *International Journal of Computer Science & Engineering Technology*, vol. 3, no.4, pp.69-75, April 2012.

[45] Sonia Sunny, David Peter S, K Poulose Jacob, "Recognition of Speech Signals: An Experimental Comparison of Linear Predictive Coding and Discrete Wavelet Transforms", *International Journal of Engineering Science and Technology*, vol.4, no.4, pp.1594-1601, April 2012.

[46] Mohammed Anwer and Rezwan-Al-Islam Khan, "Voice identification Using a Composite Haar Wavelets and Proper Orthogonal Decomposition", *International Journal of Innovation and Applied Studies*, vol. 4, no. 2, pp.353-358, October 2013.

[47] Tariq Abu Hilal , Hasan Abu Hilal, RiyadQqQ El Shalabi and Khalid Daqrouq, "Speaker Verification System Using Discrete Wavelet Transform And Formants Extraction Based On The Correlation Coefficient", *International Multi Conference of Engineers and Computer Scientists*,vol.2, March 2011.

[48] Marco Jeub, Dorothea Kolossa, Ramon F. Astudillo, Reinhold Orglmeister, "Performance Analysis of Wavelet-based Voice Activity Detection", *NAG/DAGA-Rotterdam*, 2009.

[49] Beng T Tan, Robert lang, Hieko Schroder, Andrew Spray, Phillip Dermody, "Applying Wavelet Analysis to Speech Segmentation and Classification", *Department of Computer Science*.

[50] Bartosz Zioko, Suresh Manandhar, Richard C. Wilson and Mariusz Zioko, "Wavelet Method of Speech Segmentation", *University of York Heslington, YO10 5DD, York, UK.*

[51] Akhilesh Tiwari, Dr. A. S. Zadgaoankar, "Speech Signal Analysis through Wavelets and Finding Similar Patterns in Signals of Regional Dialects of Large Demographic Region", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol.3,pp.420-423, July 2013.

[52] Mohamed El-wakdy, Ehab El-sehely, Mostafa El-tokhy, Adel El-hennawy I.M., "Speech Recognition using a Wavelet Transform to establish Fuzzy Inference System through Subtractive Clustering and Neural Network (ANFIS)", *12th WSEAS International Conference on SYSTEMS, Heraklion, Greece*, pp.381-386, July 2008.