MFCC AND LPCC ANALYSIS OF SPEECH SIGNAL FOR DIFFERENT LANGUAGES

[1]Pooja Yadav, [2]Vinay Kumar Jain
[1]ME Scholar, [2]Associate Professor
[1]Communication, [2]Electronics and Telecommunication
[1]Faculty of Engineering and Technology, SSTC-SSGI, Bhilai, Chhattisgarh, India
[2]Faculty of Engineering and Technology, SSTC-SSGI, Bhilai, Chhattisgarh, India

Abstract—The listeners outperform Automatic speech recognition structures in each and every speech reputation task. Modern excessive-tech automated speech recognition systems carry out very well in environments, wherein the speech indicators are reasonably easy. In maximum of the instances popularity with the aid of machines degrades dramatically with mild adjustment in speech signals or talking environment, for this reason complicated algorithms are used to symbolize this unpredictability. So, the speech can be easily identified. Speech generation gives many possibilities for private identity this is herbal and non-intrusive. Besides that, speech era offers the capability to verify the identity of a person remotely over long distance by using an ordinary phone. In this paper, we proposed a technique to apprehend any words or speech thru the spectrogram analysis. This techniques is used to look at the ideas of speaker reputation in multiple languages and apprehend its uses in identification and verification systems and to assess the recognition capability of various voice functions and parameters to find out the technique this is appropriate for Automatic Speaker Recognition systems in phrases of reliability and computational efficiency.

Keywords: Speech Recognition, computational efficiency, speaker recognition MFCC, LPCC

I. INTRODUCTION

Speech is the number one mode of communication. It is a manner of sharing statistics, mind and feelings and also a way of shifting human intelligence from one person to each other.Listeners outperform Automatic speech reputation (ASR) systems in every and each speech reputation assignment. Modern excessive-tech automatic speech reputation systems carry out thoroughly in environments, in which the speech signals are fairly smooth. Currently there has been a developing body of studies in extending numerous speech popularity obligations. A complicated dating is discovered among physical speech sign and the corresponding phrases and can be very hard to understand [1]. The Very recognized programs of the stated systems consist of bodily get right of entry to access and wherein a long way off identification verification is vital. However, the emergence of elegant technology in special areas of ASR structures makes the relaxed operation of those structures sure. However, some areas of ASR [14] systems oppose the same reputation in phrases of possessing talented strategies or diffused techniques for solving many problems in the area. In maximum of the instances recognition with the resource of machines degrades dramatically with mild adjustment in speech indicators or speaking surroundings, consequently complex algorithms are used to symbolize this unpredictability [2]. The complex speech processing challenge has been divided into three alternatively less difficult classes.

(a) Speech recognition: that lets in the machines to recognize the phrases, sentences, terms spoken via the use of excellent audio system.
(b) Natural language processing: this shall us the system to apprehend the dreams of various speakers.
(c) Speech synthesis: proper right here the machines reply to the wishes of customers.

Speech era gives many opportunities for non-public identification this is herbal and non-intrusive. Besides that, speech era gives the functionality to verify the identification of a person remotely over lengthy distance by using the use of the use of a ordinary cellular phone. A communiqué among people carries a selection of information except actually the conversation of thoughts. Speech also conveys statistics which consist of gender, emotion, mindset, fitness scenario and identity of a speaker. The topic of this thesis deals with speaker reputation that refers to the project [13] of spotting human beings with the resource in their voices. Secure identification device requires someone to apply a cardkey (something that the character has) or to enter a pin (some factor that the patron is aware of) that allows you to gain get right of entry to the gadget. However, the 2 strategies mentioned above have some shortcomings because get access to control used can be stolen, misplaced, misused or forgotten.

II. OBJECTIVE

The main aim of this paper is to design and algorithm by which we can recognize the various language of people using Spectrum Analysis and comparison. In this paper we can be compare many voice signal in different languages with different user.We will be use Matlab platform for this system.

III. MEL FREQUENCY CEPSTRAL COEFFICIENTS FEATURE EXTRACTION

The first stage of the speech recognition system is to compress a speech signal into the streams of acoustic feature vectors, referred to as a speech feature vectors. The extracted vectors are be assumed to have sufficient information and to be compact enough for the efficient recognition [5]. The concept of the feature extraction is an actually divided into two parts: first is the transforming the speech signal into feature vectors; secondly is to choose the useful features, which are the insensitive to changes of environmental conditions and speech variation [6]. However, changes of environmental conditions and speech variations are crucial in speech recognition systems where accuracy has degraded massively in the case of their existence. As examples of changes of environmental condition: changes in the transmission channel, changes in properties of the microphone, cocktail effects, and the background noise, etc. Some examples of speech variations include accent differences, and male-female vocal tract difference. For developing robust speech recognition, speech features are required to be insensitive to those changes and variations. The most commonly used speech feature is definitely the Mel Frequency Cepstral Coefficients (MFCC) features, which is the most popular, and robust due to its accurate estimate of the speech parameters and efficient computational model of speech [7].

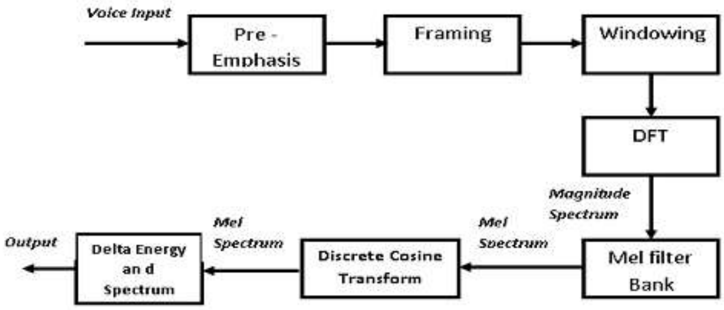

Figure 1: MFCC Block Diagram

As shown in Figure 3, MFCC consists of seven computational steps. Each step has its function and mathematical approaches as discussed briefly in the following:

Pre-emphasis

In this step, the signal spectrums are pre-emphasized, and the DC offset is removed, alow order digital system(generally a first order FIR filter) is applied to the digitized speech signal x(n) to spectrally flatten the signal in order to make It less usceptible tofind precision effects later in the signal processing

$$H(z) = 1 - az^{-1} \qquad 0.9 < a < 1$$

The most typical value of a is about 0.95 [7]. However, the signal spectrum is boosted approximately 20 dB/decade by pre-emphasis filter.

Framing

The speech signal is normally divided into small duration blocks, called frames, and the spectral analysis is carried out on these frames. This is due to the fact that the human speech signal is slowly time varying and can be treated as a quasi stationary process. The very popular frame length and frame shift for the speech recognition task are 20-30 ms and 10 ms respectively [8].

Windowing

After framing, each frame is multiplied by a window function prior to reduce the effect of discontinuity introduced by the framing process by attenuating the values of the samples at the beginning and end of each frame. The Hamming window is commonly used, it decreases the frequency resolution of the spectral analysis while reducing the side lobe level of the window transfer function

$$y(n) = x(n)w(n)$$

Hamming window is used for speech recognition task as

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi}{N-1})$$

Spectral Estimation

Spectral estimation is computed for each frame by applying Discrete Fourier Transform (DFT) to produce spectral coefficients. These coefficients are complex numbers comprising the two magnitude and phase information. Phase information is usually removed and only the magnitude of the spectral coefficients are extracted. Additionally, it is common to utilize the power of the spectral coefficients [6, 8]. DFT can be defined as:

$$X(k) = \sum_{n=0}^{N-1} y(n)e^{-\frac{j2\pi kn}{N}} \qquad 0 \leq n, k \geq N-1$$

Where *X (K)* are the spectral coefficients, and *y(n)* the framed speech signal

Mel Filtering

A group of triangle band pass filters that simulate the characteristics of the human's ear are applied to the spectrum of the speech signal. This process is called Mel filtering [10]. The human ears analyze the sound spectrum in groups based on a number of overlapped critical bands. These bands are distributed in a manner that the frequency resolution is high in the low frequency region and low in the high frequency region as illustrated in Figure

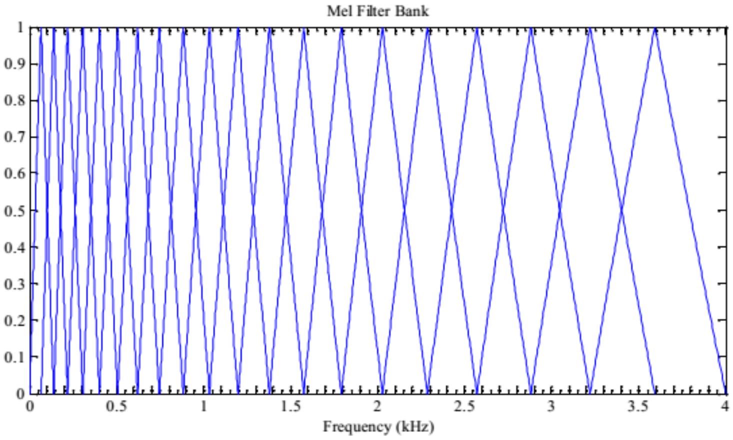

Figure 2: Mel Scale filter bank

This figure shows a set of triangular filters that are used to compute a weighted sum of filter spectral components so that the output of process approximates to a Mel scale. Each filter's magnitude frequency response is triangular in shape and equal to unity at the center frequency and decrease linearly to zero at center frequency of two adjacent filters [7, 8]. Then, each filter output is the sum of its filtered spectral components. After that the following equation is used to compute the Mel for given frequency f in HZ:

$$F(Mel) = [2595 * \log 10[1 + f]700]$$

Discrete Cosine Transform

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

IV. LINEAR PREDICTION CEPSTRAL COEFFICIENTS (LPCC)

LPCC represents the characteristics of certain speech channel, and the same person with dissimilar emotional speech will have dissimilar channel features, thereby extracting these feature coefficients to classify the emotions contained in speech. The computational process of LPCC is usually a repetition of computing the linear prediction coefficients (LPC)   LPC is one of the most powerful speech analysis techniques and is a useful method for encoding quality speech at a low bit rate. For estimating the basic parameters of a speech signal, LPCC has become one of the predominant techniques. . The basic theme behind this method is that one speech sample at the current time can be predicted as a linear combination of past speech samples,

LPCC is a technique that combines LP and cepstral analysis by taking the inverse Fourier transform of the log magnitude of the LPC spectrum for improved accuracy and robustness of the voice features extracted.

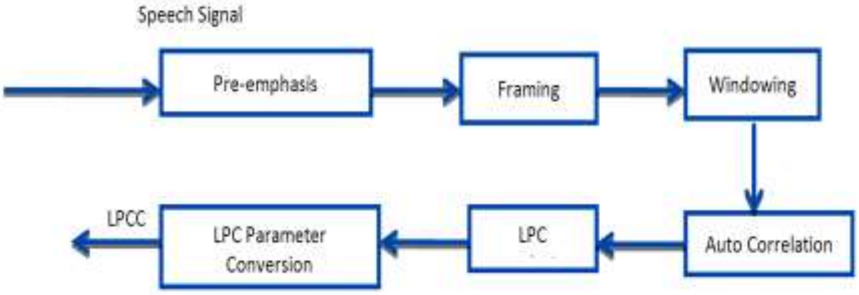

Figure 3: LPCC block diagram

LPC (Linear Predictive coefficient):

Linear prediction techniques are the maximum broadly used in, speech synthesis, speech coding, speech reputation, speaker recognition and verification and for huge speech garage. LPC strategies provide correct estimates of speech parameters, and do it extraordinarily successfully. The concept of Linear Prediction: present day speech pattern [16] can be intently approximated as a linear aggregate of the past samples. LPC is a method that offers a terrific estimate of the vocal tract spectral envelope and is critical in speech evaluation because of the accuracy and pace with which it can be derived. The characteristic vectors are calculated by way of LPC over each frame. The coefficients used to represent the frame typically tiers from 10 to twenty depending at [17] the speech sample, application and range of poles within the version. However, LPC also have dangers. Firstly, LPC approximates speech linearly in any respect frequencies that is inconsistent with the listening to notion of people. Secondly, LPC may be very susceptible to noise from the heritage which may additionally cause mistakes within the speaker modelling.

V. FORMANT FREQUENCY

A formant is a concentration of acoustic energy around a particular frequency in the speech wave. There are several formants, each at a different frequency, roughly one in each 1000Hz band. Or, to put it differently, formants occur at roughly 1000Hz intervals. Each formant corresponds to a resonance in the vocal tract.

Formants can be seen very clearly in a wideband spectrogram, where they are displayed as dark bands. The darker a formant is reproduced in the spectrogram, the stronger it is (the more energy there is there, or the more audible it is).

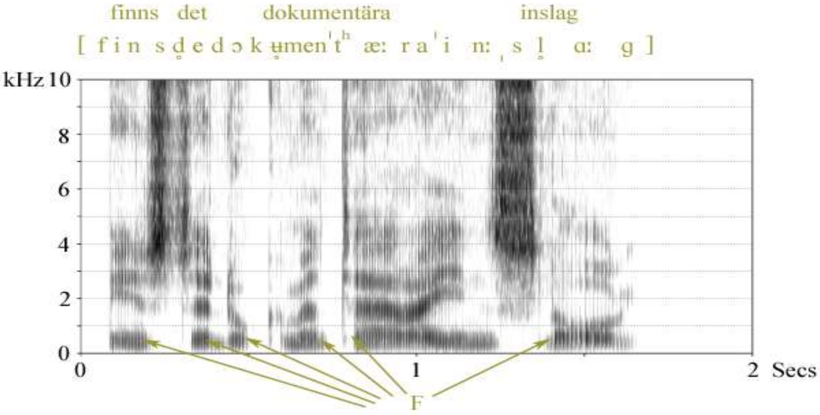

Figure 4: Formant frequency spectrum

The arrows at F on this spectrogram point out six instances of the lowest formant. The next formant occurs just above these, between 1 and 2 KHz. Then the next is just above that, between 2 and 3kHz.. And so on. When you look at a spectrogram, like this example, you will see formants everywhere, in both vowels and consonants. To understand why, you must recall the source-filter theory of speech production. The vocal tract filters a source sound (e.g. periodic voice vibrations or aperiodic hissing) and the result of the filtering is the sound you can hear and record outside the lips and show on a spectrogram.

VI. Testing and Implementation

In this system we design a schematic GUI using a Matlab platform. It's perform and handle easily by user. User can be add their voice using record and store a .wav file. In this system we have MFCC, LPCC and Formant Frequency analysis system. Which is used for recognition the user voice. We use three languages Hindi , Marathi and Rajasthani. We use speech processing toolbox to voice recognition. We try to find MFFC and LPCC of each voice.

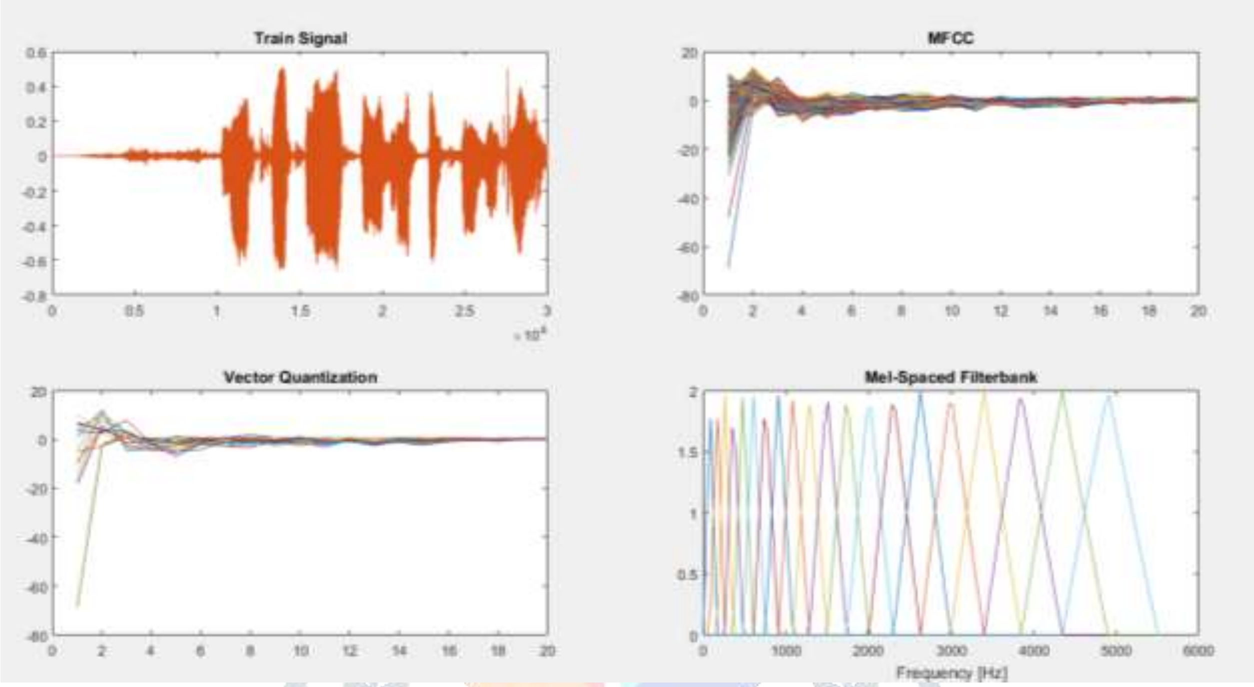

Figure 5: Voice component comparison using MFFC analysis

Using MFFC we can be find the vector quantization level and mel-spaced filter coefficient this method can fast response as compare to LPCC.

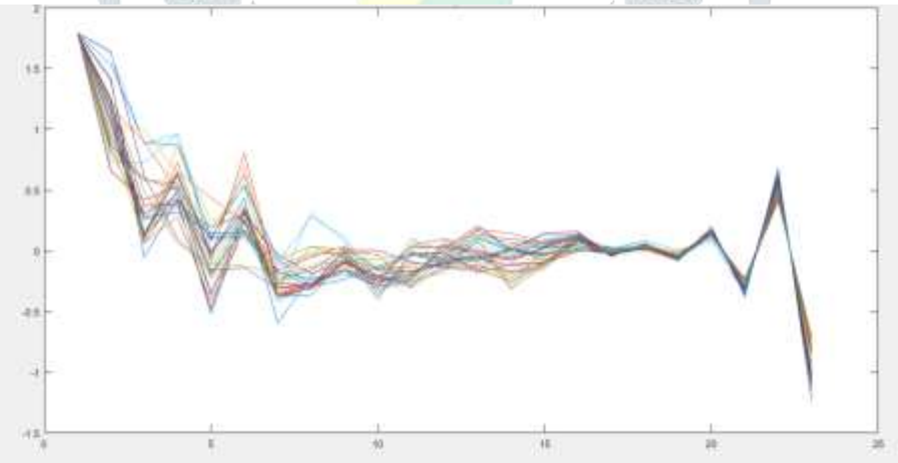

Figure 6: LPPC analysis of voice

VII. Results

Speech recognition System operate in two modes that is Enrollment mode and Recognition mode. The MFCC feature coefficient is used here for reasons stated earlier Euclidean distance is used to measure the distance between the feature vectors (Itakura-Saito).The major part of this work is the implementation and analysis of the Automatic Speech Recognition algorithms using MATLAB and observed their individual performance. The speech recognition system model has been developed for compare the two algorithms that is MFCC and LPCC. The performance has been evaluated by considering the sets of speech signal. It is analysis that MFCC used in Automatic speech Recognition system (ASP) provide 90 percentage accuracy where as LPCC used in Automatic Speech Recognition (ASP) given 83 percentages accuracy. Results and analysis show that MFCC algorithm gives better result than LPCC algorithm. From the simulation results we also add that MFCC algorithm, which require more calculation but perform better than LPCC in terms of efficiency and accuracy. We try to analysis many voice signal from different user, which shows in below table.

Table 1: Comparative analysis of MFCC and LPCC recognition rate

SENTENCE	Language	Recognition Rate	
		LPCC	MFCC
Speaker 1	HINDI	94.6%	98%
	MARATHI	97.7%	98.7%
	RAJASTHANI	96.7%	98.9%
Speaker 2	HINDI	95.5%	98.4%
	MARATHI	98.5%	98.6%
	RAJASTHANI	96.6%	98.8%
Speaker 3	HINDI	96.6%	98.4%
	MARATHI	94.7%	98.7%
	RAJASTHANI	96.8%	98.9%
Speaker 4	HINDI	94.2%	98%
	MARATHI	97.3%	98.9%
	RAJASTHANI	96.7%	97.9%
Speaker 5	HINDI	94.1%	98.8%
	MARATHI	97.9%	98.7%
	RAJASTHANI	96.4%	96.9%
Speaker 6	HINDI	97.3%	96%
	MARATHI	98.2%	99.7%
	RAJASTHANI	96.8%	98.9%
Speaker 7	HINDI	95.6%	96%
	MARATHI	97.8%	97.7%
	RAJASTHANI	96.9%	98.9%
Speaker 8	HINDI	97.6%	99%
	MARATHI	95.7%	98.7%
	RAJASTHANI	88%	98.7%
Speaker 9	HINDI	89.6%	99.6%
	MARATHI	88.9%	97.7%
	RAJASTHANI	87.5%	96.7%
Speaker 10	HINDI	85.7%	96.6%
	MARATHI	87.8%	97.7%
	RAJASTHANI	85.6%	96.7%

VIII. Discussions

We have use a decision based algorithm in our system which is presented in paper. In this algorithm we can be analysis and compare the speech signal. We can be design an automatic speech recognition system, where the user can be analysis their voice in different languages. We plot the graph of each speech signal. We can also compare the each signal.

IX. Conclusion

In this proposed methodology, a system is to be designed which can easily recognize any language and plots the respective spectrum as per the recognized language. The plotted curve will signify each word whatever is said by the speaker. Listeners outperform Automatic speech recognition systems in each and every speech recognition task. Modern high-tech automatic speech recognition systems perform very well in environments, where the speech signals are reasonably clean. In most of the cases recognition by machines degrades dramatically with slight adjustment in speech signals or speaking environment, thus this complex algorithms are used to represent this unpredictability. So, the speech can be easily recognized through the spectrogram.

X. Future Enhancement

We are considering an indoor environment(less noise). We want to see that the classification of sounds into global categories can be performed with very low calculation effort. For gender recognition algorithms that required the low cost and frequency domain features achieve results. We are seeing for the different categories (Global, Gender and in door sound classification),that use flow-cost algorithms can be equally effective for deployment indoors as the use of high-cost algorithms.

XI. Acknowledgement

Expression of giving thanks are just a part of those feeling which are too large for words, but shall remain as memories of wonderful people with whom I have got the pleasure of working during the completion of this work. I am grateful to “Shri Shankaracharya Technical Campus” which helped me to complete my work by giving encouraging environment. I would like to express my deep and sincere gratitude to my supervisor, “Associate Professor” “Vinay Kumar Jain”. His wide knowledge and his logical way of thinking have been of great value for me. His understanding, encouraging and personal guidance have provided a good basis for the present work.

References

[1] Taabish, G., Anand, S., Rajouriya, D.K. and Najma, F. 2014, A Systematic Analysis of Automatic Speech Recognition: An Overview, International Journal of Current Engineering and Technology, Vol.4, No.3

[2] Yuan, M. [2004], Speech Recognition on DSP: Algorithm Optimization and Performance Analysis.

[3] Saon, G. and Padmanabham, M. [2001],Data-driven approach to designing compound words for continuous speech recognition, IEEE Transactions on Speech and Audio Processing,Vol. 9,No.4, pp.327-332.

[4] InmaMohino-Herranz, Roberto Gil-Pita, Sagrario Alonso-Diaz and Manuel Rosa-Zurera [2014], “MFCC Based Enlargement Of The Training Set For Emotion Recognition In Speech”, Signal & Image Processing: An International Journal (SIPIJ) Vol.5, No.1. February.

[5] TSaon, G. and Padmanabham, M. [2001],Data-driven approach to designing compound words for continuous speech recognition, IEEE Transactions on Speech and Audio Processing,Vol. 9,No.4, pp.327-332.

[6] Han, Y., Wang, G.Y. and Yang, Y. [2008], Speech emotion recognition based on mfcc. Journal of Chongqing University of Posts and Telecommunications.

[7] Antoniol, G., Rollo, V. F., &Venturi, G. [2005]. Linear predictive coding and cepstrum coefficients for mining time variant information from software repositories. In Proceedings of the 2005 international workshop on mining software repositories.

[8] Hasnain, S.K., Maqsood, M., Shazad, M.A. and Bashir, S. [2008], development of speech recognition systems, TECHNOLOGY FORCES (Technol, forces) journal of engineering and science, Vol.2, No.1.

[9] Biing,H.J. and Sadaoki,F. [2000], Automatic recognition and understanding of Spoken launguage- A first step towards natural human-machine communication, Proceedings of the IEEE Vol.88.

[10] Taabish, G., Anand, S. And Vijay, S. [2014], An Improved Endpoint Detection Algorithm using Bit Wise Approach for Isolated, Spoken Paired and Hindi Hybrid Paired Words. International journal of computer applications, 0975-8887, Volume 92 – No.15.

[11] Hisashi, W. [1977], Normalization of Vowels by Vocal Tract Length and Its Applications to Vowel Identification, IEEE Transactions onAcoustics, Speech and Signal Processing, Vol. 25.

[12] Shasidhar G. Koolagudi, Reddy, R. ,Yadav, J. and Rao, K.S., [2011], IITKGP-SEHSC: Hindi speech corpous for emotion analysis, IEEE International Conference on Devices and Communications.

[13] Cowie, R. and Cornelius, R.R. [2003], Describing the emotional states that are expressed in speech, Speech Communication, Elsevier, Vol. 40.

[14] Cowie, R., [2000], Emotional states expressed in speech,” in Proc. of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research, pp. 224- 231.

[15] Picone, J. W. [2002]. Signal modeling techniques in speech recognition. Processings of IEEE, 81(9), 1215–1247.

[16] P. Kumar, M. Chandra [2011], “Hybrid of Wavelet and MFCC Features for Speaker (WICT), Verification”, IEEE World Congress on Information and Communication Technologies Mumbai, pp. 1150-1154, 11-14 December.

[17] S. Tripathi, S. Bhatnagar [2012], “Speaker Recognition”, IEEE Third International Conference on Computer and Communication Technology (ICCCT), Allahabad, pp. 283-287, 23-25 November.

[18] C. R. Jankowski Jr., H. H. Vo, and R. P. Lippman [1995], A comparison of signal processing front ends for automatic word recognition," IEEE Trans. Speech Audio Processing, vol. 3. pp. 286-293, Jul.

[19] T. R. JayanthiKumari, H.S. Jayana [2014], "Comparison of LPCC and MFCC features and GMM and GMM-UBM modeling for limited data speaker verification",IEEE International Conference on Computational Intelligence and Computing Research (ICCIC),ISBN: 978-1-4799-3975-6.

[20] SonghitaMisra, TusharKanti Das, ParthaSaha [2015],"Comparison of MFCC and LPCC for a fixed phrase speaker verification system, time complexity and failure analysis", IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), ISBN: 978-1-4799-7075-9.