International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 10
ISSN: 2454-4248
78 – 85

Capstrum Coefficient Features Analysis for Multilingual Speaker Identification System

Vinay Kumar Jain
Research Scholar,
SSTC(SSGI), Bhilai
vinayrich_17@yahoo.co.in

Dr.(Mrs.) Neeta Tripathi
Principal,
SSTC (SSEC), Bhilai
neeta31dec@rediffmail.com

Abstract: The Capstrum coefficient features analysis plays a crucial role in the overall performance of the multilingual speaker identification system. The objective of the research work to investigates the results that can be obtained when you combine Mel-Frequency Cepstral Coefficients (MFCC) and Gammatone Frequency Cepstral Coefficients (GFCC) as feature components for the front-end processing of a multilingual speaker identification system. The MFCC and GFCC feature components combined are suggested to improve the reliability of a multilingual speaker identification system. The GFCC features in recent studies have shown very good robustness against noise and acoustic change. The main idea is to integrate MFCC & GFCC features to improve the overall multilingual speaker identification system performance. The experiment carried out on recently collected multilingual speaker speech database to analysis of GFCC and MFCC. The speech database consists of speech data recorded from 100 speakers including male and female. The speech samples are collected in three different languages Hindi, Marathi and Rajasthani. The extracted features of the speech signals of multiple languages are observed. The results provide an empirical comparison of the MFCC-GFCC combined features and the individual counterparts. The average language-independent multilingual speaker identification rate 84.66% (using MFCC), 93.22% (using GFCC)and 94.77% (using combined features)has been achieved.

Keyword*: Capstrum, Coefficients, GFCC, MFCC, Multilingual, Speaker.*

*****

I. INTRODUCTION:

Multilingual Speaker identification system refers to identifying persons from their speech of different languages. In India there are many peoples who are able to speak more than one language and hence the effect of multiple languages on a speaker identification system needs to be investigated When the Multilingual speaker identification system is being transferred to real applications, the need for greater adaptation in identification is required[13]. The performance of the monolingual speaker identification systems tends to decreases when speaker is speaking in another language. Therefore there is a need to make such systems which can work for multiple languages.

Languages are usually influenced by other languages that are present in the environment and by the speaker's mother tongue [2]. Multilingual speech processing (MLSP) is a distinct field of research in speech and language technology that combines many of the techniques developed for monolingual systems with new approaches that address specific challenges of the multilingual domain [8].

In order to find some statistically relevant information from speech signal, it is important to have mechanisms for reducing the information of each segment in the audio signal into a relatively small number of parameters, or features. Feature extraction is the first step for the multilingual speaker identification system. Many algorithms were developed by the researchers for feature extraction.

The MFCC are typically standard feature vector for any speaker identification systems because of their high accuracy and low complexity; however they are not very robust at the presence of additive noise. The GFCC features in recent studies have shown very good robustness against noise and acoustic change[17].

II. Literature Review:

W. Burgos [20] conducted the experiments on the Texas Instruments and Massachusetts Institute of Technology (TIMIT) and the English Language Speech Database for Speaker Recognition (ELSDR) databases, were the test utterances are mixed with noises at various SNR levels to simulate the channel change. The results provide an empirical comparison of the MFCC-GFCC combined features and the individual counterparts.

S.Sarkar et al.[1] reported the performance of multilingual speaker recognition systems on the IITKGP-MLILSC speech corpus. The standard GMM-based speaker recognition framework was used. The average language-independent speaker identification rate was 95.21% and an average equal error rate of 11.71%.

Nagaraja B.G.and H.S. Jayanna,[2] presented a paper in year 2013 on speaker identification in the context of mono,

 ISSN: 2454-4248
78 – 85

cross and multilingual using the two different feature extraction techniques, i.e., Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC) with the constraint of limited data. The languages considered for the study were English (international language), Hindi (national language) and Kannada (regional language). They reported that the standard multilingual database was not available, experiments were carried out on their own created database of 30 speakers in the college laboratory environment who speak the three different languages. As a result the combination of features gives nearly 30% higher performance compared to the individual features.

U Bhattacharjee and K.Sarmah[5] report the experiment carried out on a recently collected multilingual and multichannel speaker recognition database to study the impact of language variability on speaker verification system. The speech samples were collected in three different languages English, Hindi and a local language of Arunachal Pradesh. The collected database was evaluated with Gaussian Mixture Model based speaker verification system using universal background model (UBM) for alternative speaker representation and Mel- Frequency Cepstral Coefficients (MFCC) as a front end feature vectors. The impact of the mismatch in training and testing languages have been evaluated.

P. Kumar and S. L. Lahudkar[8] introduced a new method which combined LPCC and MFCC(LPCC+MFCC) using fusion output was proposed and evaluated together with the different voice feature extraction methods. The speaker model for all the methods was computed using Vector Quantization- Linde, Buzo and Gray (VQ-LBG) method. Individual modelling and comparison for LPCC and MFCC is used for the LPCC+MFCC method.

S.Sharma and P Singh [17] presented two working engines using both alluded alternatives by use of continuous BPNN and GFCC method. Their system has been fully implemented and tested for audio wave files. Result analysis was done using neural and GFCC tool. The emotional speech input to the system is the collection of speech data. After collection of database which is considered as the training samples, necessary features were extracted from the speech signal to train the system using GFCC and BPNN algorithm.

Zhao X. and Wang D.[18] tells about the various used techniques like GFCC i.e. Gamma tone Frequency Cepstral Coefficients as its speech detection algorithm and Gaussian Mixture Model (GMM) to estimate the Gaussian model parameters. Basically focuses on improvement of speech identification in noisy environment using Wavelet filter which are added to de-noise the speech signals. Experiment shows better results for stored database oriented system and gives 85% of the correct recognition rate i.e. CORR and 73% results are given when wavelet filter are not used.

III. Methodology:

Database Generation:

For multilingual speaker identification system, the database of different speakers has been recorded in three Indian languages i.e. Hindi, Marthi and Rajasthani. The sampling rate of recorded sentences is 16KHz.The sentences consist consonants i.e "cha", "sha" and "jha" for the recording. Total number of speakers involved are 100 including males and females. The recorded sentences are:

मुझे चाय पीना पसंद है।	चाय में शक्कर कम है।	तिरंगा हमारा झंडा है।
मला चाय पसंद आहे	चाय मध्ये शक्कर कमी आहे	तिरंगा अमच्चये झंडा आहे
मन्ने चाय पीनी पसंद है।	चाय मों शक्कर कम है।	तिरंगा मारा झंडा है।

Feature Extraction:

The Speaker identification mainly involves two modules namely feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the speaker`s speech signal that can later be used to represent the speaker.

The original speech signal of a speaker contains redundant information. For speaker identification eliminating such redundancies helps in reducing the computational overhead and also improve system accuracy. Therefore all most speech application involves the transformation of signal to set of compact speech parameter.

Speech feature extraction is the signal processing frontend which has purpose to converts the speech waveform into some useful parametric representation[15]. These parameters are then used for further analysis in multilingual speaker identification system. Here MFCC and GFCC features has been extracted from the speech signal of different languages of a speaker. MFCCs are one of the most popular feature extraction techniques used in speaker identification based on frequency domain using the Mel scale [8].Where as GFCC, based on equivalent rectangular bandwidth (ERB) scale, has finer resolution at low frequencies.

 ISSN: 2454-4248
78 – 85

MFCC features Extraction:

The complete process for obtaining MFC coefficients is shown in figure-1:

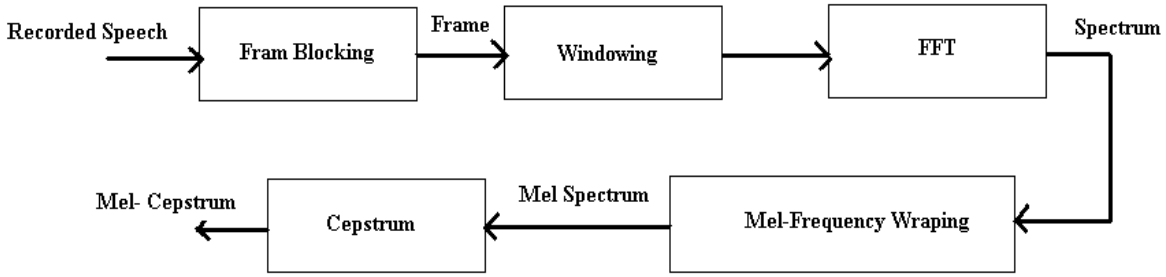

Figure-1: MFCCs feature extraction process.

In **Frame Blocking** the input speech signal is segmented into frames of 15~20 ms with overlap of 50% of the frame size. Usually the frame size (in terms of sample points) is equal to power of two in order to facilitate the use of FFT. Overlapping is used to produce continuity within frames. Through **windowing** technique each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame. Spectral analysis shows that different timbres in speech signals corresponds to different energy distribution over frequencies. Therefore **FFT** is performed to obtain the magnitude frequency response of each frame. When FFT is performed on a frame, it is assumed that the signal within a frame is periodic, and continuous when **wrapping** around. Since we have performed FFT, DCT transforms the frequency domain into a time-like domain called quefrency domain. The obtained features are similar to **cepstrum,** thus it is referred to as the **mel-scale cepstral** coefficients, or MFCC.

GFCC features Extraction:

In GFCC,Gammatone filter bank is applied to the raw speech signal to generate the respective cochleogram, which represents transformed raw speech signal in the time and frequency domain. The advantage of using a cochleogram over spectrogram is that the features of a cochleogram is based on ERB scale with finner resolution at low frequency than the Mel-scale used in spectrogram. Besides it allows more number of coeficients in comparision to MFCC. The Mel filter-bank for a power specrum is with 257 coefficients6 while the GFCC is with 512 coefficients.

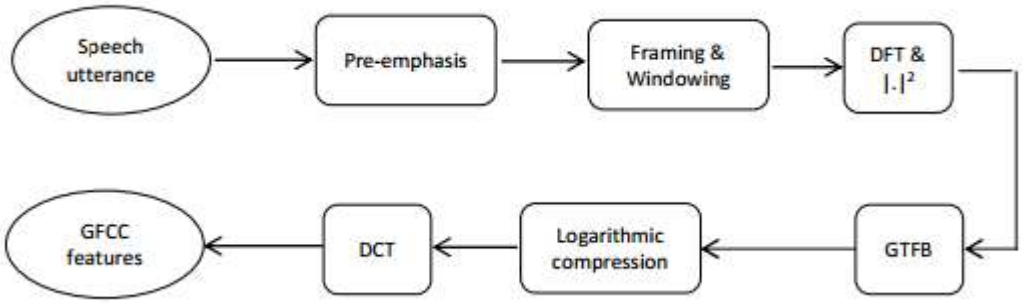

Figure-2: GFCCs feature extraction process.

The first step of the algorithm is **pre-emphasis**. The idea of pre-emphasis is to spectrally flatten the speech signal and equalize the inherent spectral tilt in speech [1,2]. **Pre-emphasis** is implemented by a first order FIR digital filter. The **windowing** function used is the Hamming window , which aims to reduce the spectral distortion introduced by windowing. After windowing, **Fast Fourier Transform (FFT)** is applied to the windowed speech frame. The 512point FFT spectrum of the speech frame is obtained as a result. The **Gammatonefilterbank** consists of a series of bandpass filters, which models the frequency selectivity property of the human cochlea. The next step of the algorithm is to apply **logarithm** to each filter output[19]. The aim of this procedure is to simulate the human perceived loudness given certain signal intensity. The last stage of the algorithm consists of correlating the filter outputs. For this the **Discrete Cosine Transform (DCT)** is applied to the filter outputs.

This system take the multilanguage speech samples as input, computes its GFCC, delta and double delta coefficients as a feature vector has been recorded. And accordingly all three matrices of 13 vectors are combined and are used as speech identification.

 ISSN: 2454-4248
78 – 85

Identification Process:

The identification process is broadly classified in following two phases: Training phase and Testing phase. During the **Training phase** a comprehensive database of speech feature vectors has been prepared. During the **Testing phase** speech is recorded and features are extracted. Compare features with the features of trained data using neural network. Once the database of speech feature is created then the next phase is to design appropriate neural network which can be trained by theses extracted feature[13]. In this algorithm, Feed forward neural network has been chosen for classification purpose. A feed forward neural network having 3-layers of neurons has been designed. The Modular programming has been developed in MATLab for feature extraction, training phase and testing phase.

The complete process of multilingual speaker identification system is illustrated in figure-3.

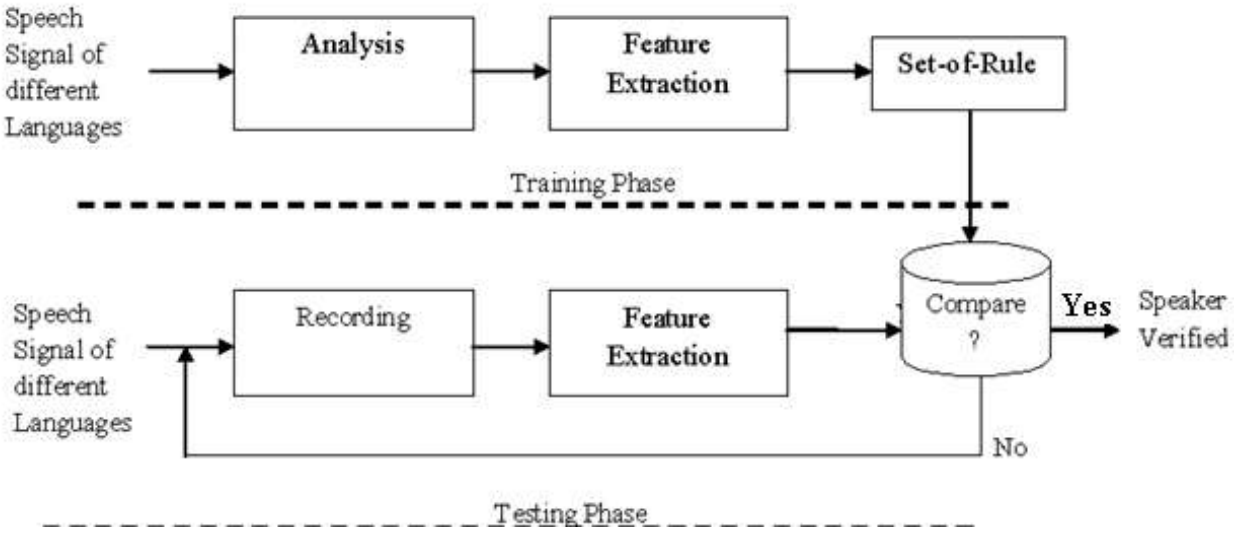

Figure-3: Two distinct phase of multilingual speaker identification system process.

IV. Results:

In this paper, investigation has been made for GFCC and MFCC features for male and female speakers in three languages Hindi and Marathi and Rajasthani. The analysis is done of three utterance 'cha' , ''sha' and 'jha' in three languages of 100 speakers.

On the basis of analysis there are two major differences has been observed in MFCC and GFCC features extraction techniques. One is the frequency scale. GFCC, based on equivalent rectangular bandwidth (ERB) scale, has finer resolution at low frequencies than MFCC (mel scale). The other one is the nonlinear rectification step prior to the DCT. MFCC uses a log while GFCC uses a cubic root.

MFCC Analysis:

The MFCCs feature extraction method has been implemented for multilingual speaker identification system. This system take the multilanguage speech samples as input, computes its MFCC, delta and double delta coefficients as a feature vector has been recorded.The analysis is done for the three utterance "cha", "sha" and "jha" in three Indian languages. Here report the experiment carried out on a recently collected multilingual speaker identification database to study the impact of language variability on speaker identification system. The effect of language on the features vector of a speaker has been observed.

Base of the analysis is observed the variation in mel frequency cepstral coefficients when speaker change the spoken language. The observations are, the minimum values and the maximum values of MFCCs for three languages are different. It is observed that the Rajasthani language has the larger values as compared to Hindi language and Marathi Language in minimum values of the feature vectors, where as Marathi Language has the larger values as compared to Hindi language and Rajasthani language in maximum values of feature vectors as shown in figure-4:

 ISSN: 2454-4248
78 – 85

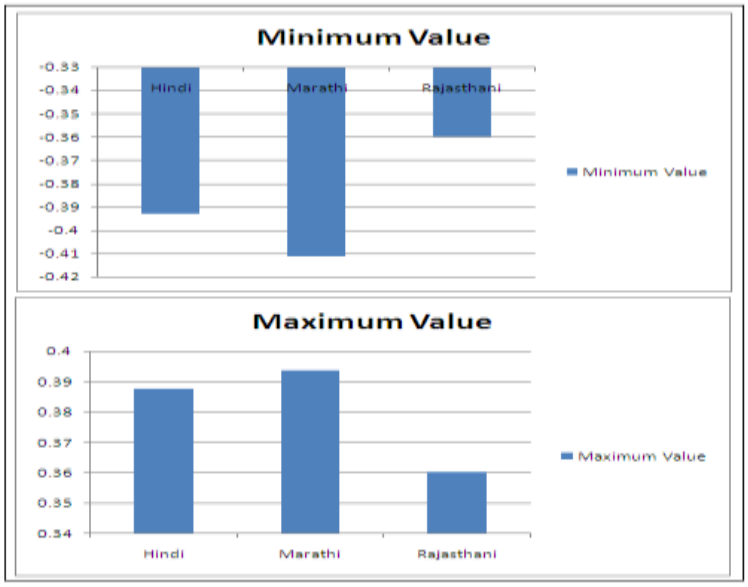

Figure-4: Variation in MFCC feature vectors of a speaker in multilingual environment.

GFCC Analysis:

The GFCCs feature extraction method has been implemented for multilingual speaker identification system. The analysis is done for the three utterance "cha", "sha" and "jha" in three Indian languages. The experiment carried out on a recently collected multilingual speaker identification database to study the impact of language variability on speaker identification system. This system take the multilanguage speech samples as input, computes its GFCC, delta and double delta coefficients as a feature vector has been recorded. And accordingly all three matrices of 13 vectors are combined and are used as speech identification

The effect of language on the features vector of a speaker has been observed. Base of the analysis is observed the variation in Gammatone Frequency Cepstral Coefficients when speaker change the spoken language. The observations are, the minimum values and the maximum values of GFCCs for three languages are different. It is observed that the Marathi language has the larger values as compared to Hindi language and Rajasthani Language in minimum values of the feature vectors, where as Rajasthani Language has the larger values as compared to Hindi language and Marathi language in maximum values of feature vectors as shown in figure-5.

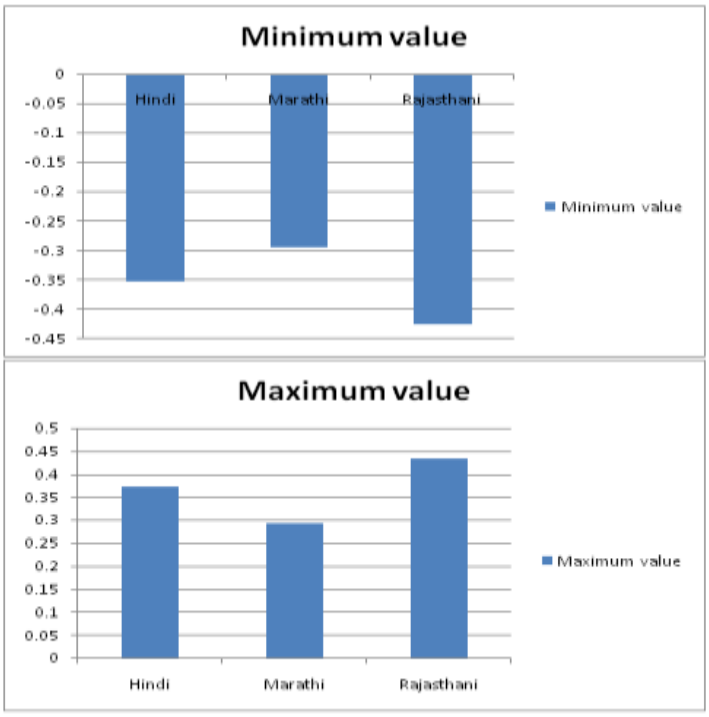

Figure-5: Variation in GFCC feature vectors of a speaker in multilingual environment

 ISSN: 2454-4248
78 – 85

Multilingual Speaker Identification Analysis:

Proposed multilingual speaker identification system is tested for three different languages i.e. Hindi, Marathi and Rajasthani. 300 samples of each language have been collected and 78 features are extracted out from each language and prepared a feature database. A feed forward neural network is designed with 78 input neurons, 10 hidden neurons and 3 output neuron (one for each language). A network is trained by providing the feature vector from the database to input layer and also the target vector. This experiment is performed in the matched and mismatched conditions for Hindi language, Marathi Language and Rajasthani language when training and testing with different databases. The Percentage rate of Multilingual Speaker Identification Systems are shown below in table-1.

Table -1: Percentage Accuracy of Multilingual Speaker Identification System.

No. of speech samples	No. of speech samples matched			% Identification Rate		
	Using MFCC	Using GFCC	Using (MFCC+GFCC)	Using MFCC	Using GFCC	Using (MFCC+GFCC)
Hindi Language(300)	266	285	289	88.66%	95%	96.33%
Marathi Language(300)	253	279	284	84.33%	93%	94.66%
Rajasthani Language(300)	243	275	280	81%	91.66%	93.33%
Average % Identification Rate				**84.66%**	**93.22%**	**94.77%**

This experiment is performed in the matched and mismatched conditions for Hindi language, Marathi Language and Rajasthani language when training and testing with different databases. From table-1 it is clear that if the speaker spoke the hindi language has the greater accuracy as compared to Marathi language and Rajasthani language in all three format.It is also clearly shown that GFCC give the better performance as compared to MFCC in multilingual speaker identification system. It is also observed that using combine features(MFCC + GFCC),the averages % identification rates slightly increases. This is shown in figure-6 and figure-7

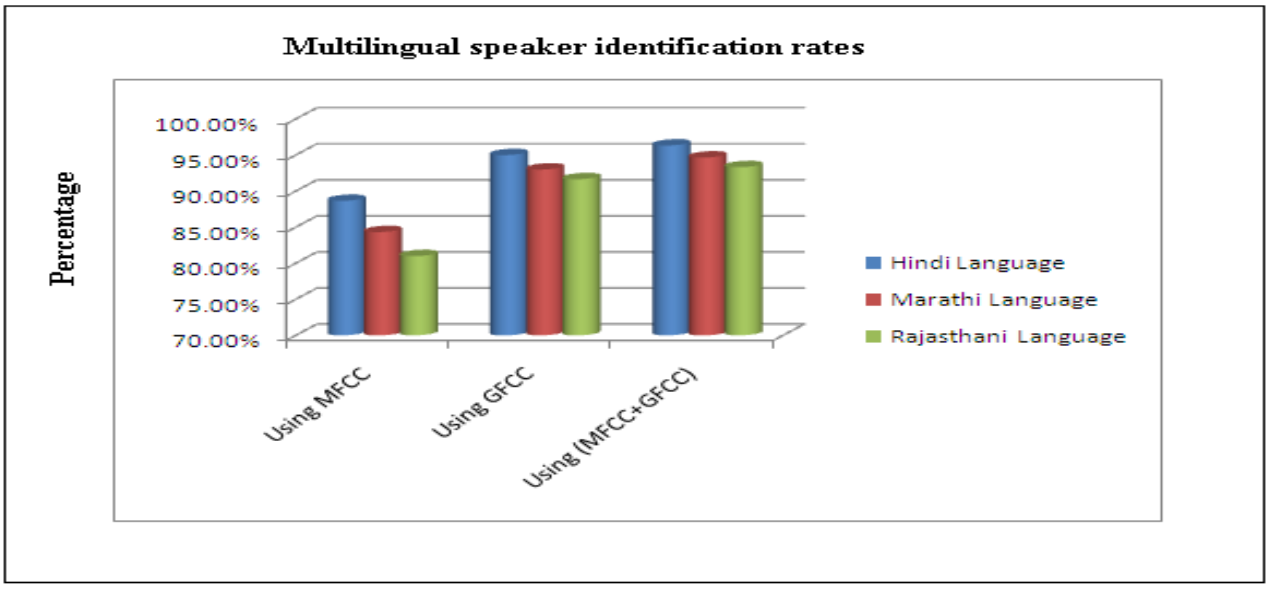

Figure-6 : Multilingual speaker identification rates.

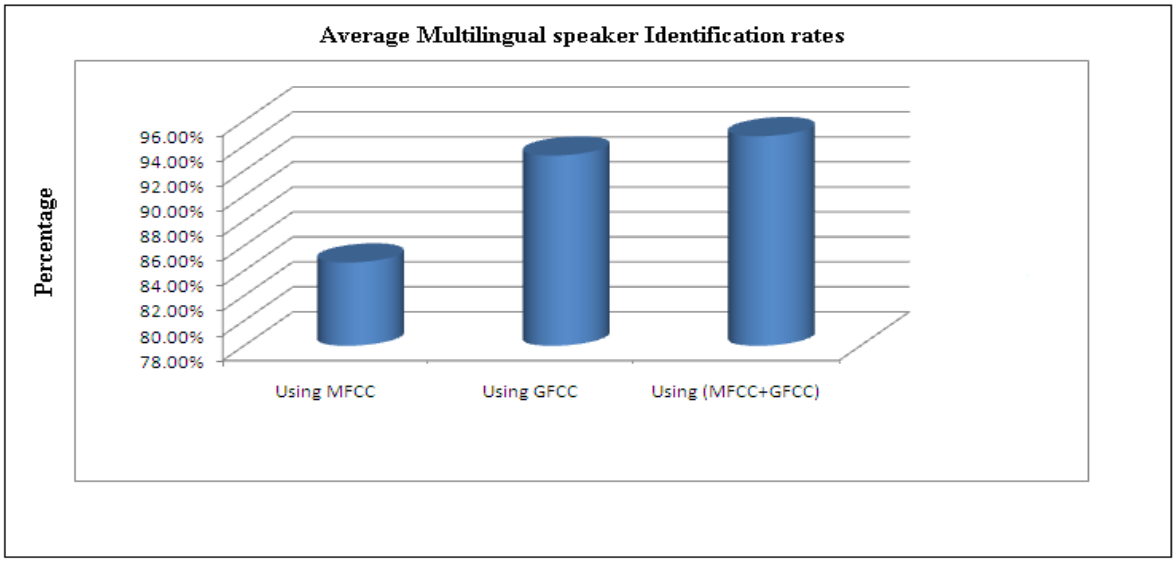

Figure-7 : Average Multilingual speaker Identification rates.

V. Conclusion:

For multilingual speaker identification system, MFCC and GFCC features has been successfully extracted. On the basis of experiments following conclusion comes out that the value MFCCs and GFCCs has changed when speaker change the spoken language. The identification rate of hindi lingual speakers are high as compared to other languages in all three cases. The results provide an empirical comparison of the MFCC-GFCC combined features and the individual counterparts. The MFCC and GFCC feature components combined are suggested to improve the reliability of multilingual speaker identification system.

References:

[1] S.Sarkar and et al. "Multilingual speaker recognition on Indian languages". 2013 Annual IEEE India Conference (INDICON),Mumbai,2013:1-5.

[2] Nagaraja B.G. and H.S. Jayanna, "Combination of Features for Multilingual Speaker Identification with the Constraint of Limited Data". International Journal of Computer Applications (0975 - 8887)Volume 70 - No. 6, May 2013:1-6.

[3] S. Agrawal, and et al. "Prosodic feature based text dependent speaker recognition using Machine learning althorithms". International Journal of Engineering Science and Technology.2(10), 2010:5150-5157.

[4] W.Bharti. and et al. "Marathi Isolated Word Recognition System using MFCC and DTW Features".ACEEE Int. J. on Information Technology, Vol. 01, No. 01, Mar 2011:21-24

[5] U.Bhattacharjee. and K. Sarmah, "A multilingual speech database for speaker recognition", IEEE International Conference on Signal Processing, Computing and Control (ISPCC), Waknaghat Solan,2012:1-5.

[6] U.Bhattacharjee. and K. Sarmah. "Development of a Speech Corpus for Speaker Verification Research in Multilingual Environment". International Journal of Soft Computing and Engineering (IJSCE).2(6):2012,443-446.

[7] U.Bhattacharjee. and K. Sarmah. "GMM-UBM Based Speaker Verification in Multilingual Environments". IJCSI International Journal of Computer Science Issues.9(6), 2012:373-380.

[8] P. Kumar and S. L. Lahudkar, "Automatic Speaker Recognition using LPCC and MFCC", International Journal on Recent and Innovation Trends in Computing and Communication Volume: 3 Issue: 4, April 2015,ISSN: 2321-8169: 2106 – 2109.

[9] R Ranjan, and et al,. "Text-Dependent Multilingual Speaker Identification for Indian Languages Using Artificial Neural Network".3rd International Conference on Emerging Trends in Engineering and Technology.Goa, India,2010:632-635.

[10] G.Kaur and H.Kaur, "Multi Lingual Speaker Identification on Foreign Languages Using Artificial Neural Network with Clustering". International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 5, May 2013 ISSN: 2277 128X:14-20.

[11] M.Ferras, and et al,. "Comparison of Speaker Adaptation Methods as Feature Extraction for SVM-Based Speaker Recognition". IEEE Transaction on Audio, Speech, and Language Processing.18(6), 2010:366-1378.

[12] H. A.Patil and et al "Design of Cross-lingual and Multilingual Corpora for Speaker Recognition Research and Evaluation in Indian Languages". International Symposium on Chinese Spoken Languages Processing(ISCSLP 2006),Kent Ridge, Singapore.

[13] V.K. Jain and N. Tripathi, "Multilingual Speaker Identification using analysis of Pitch and Formant frequencies". Published in IJRITCC Journal, Volume 4, Issue 2, February 2016. ISSN: 2321-8169:296-298.

[14] K. K. Sahu and V. Jain."A Novel Language Identification System For Identifying Hindi, Chhattisgarhi and English Spoken Language", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,Vol.-3 Issue -12, December-2014:728-731.

[15] N.Tripathi and et al. "Correlation Between Eyebrows Movement and Speech Acoustic Parameters. PCEA-IFToMM International Conference on *Recent Trends in Automation and its Adaptation to Industries,* to be organized at Nagpur on July 11-14, 2006.

[16] N. Tripathi, and et al. "A Close Correlation between Eyebrows Movement and Acoustic Parameter" *Journal of Acoustics Society of India (ISSN No.0973-3302),*. Vol. 35, No 4, Oct. 2008: 158-162.

[17] Shaveta Sharma , Parminder Singh," Speech Emotion Recognition using GFCC and BPNN", International Journal of Engineering Trends and Technology (IJETT) – Volume 18 Number 7 – Dec 2014, ISSN: 2231-5381.

[18] Xiaojia Zhao and DeLiang Wang,"ANALYZING NOISE ROBUSTNESS OF MFCC AND GFCC FEATURES IN SPEAKER IDENTIFICATION" ICASSP 2013(IEEE),978-1-4799-0356-6/13

[19] Sahil Arora and Nirvair Neeru," Speech Identification using GFCC, Additive White Gaussian Noise (AWGN) and Wavelet Filter", International Journal of Computer Applications (0975 – 8887) Volume 146 – No.9, July 2016.

[20] Wilson Burgos, "GAMMATONE AND MFCC FEATURES IN SPEAKER RECOGNITION" M.Tech Thesis,Melbourne, Florida November 2014.