www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-6, Issue-1, February 2018
Research Paper

Int. J. Sc. Res. in Network Security and Communication

ISSN: 2321-3256

Speech Features Analysis and Biometric Person Identification in Multilingual Environment

V.K. Jain[1*], N. Tripathi[2]

[1*]Electronics &Telecommunication, SSTC (SSGI),CSVTU, Bhilai, India
[2]Electronics &Telecommunication, SSTC (SSEC),CSVTU, Bhilai, India

[*]*Corresponding Author: vinayrich_17@yahoo.co.in, Tel.: 09826195251*

Received: 28/Jan/2018, Revised: 06/Feb/2018, Accepted: 19/Feb/2018, Published:28/Feb/2018

Abstract**—**Biometric person identification systems are the important in the environments where security must be needed. In this respect, a Biometric person identification systems has been designed which identify the person by determining the authenticity by their voice in multilingual environment. The speech samples are recorded in three Indian languages Hindi, Marathi and Rajasthani for multilingual environment. Pitch, formant frequencies, MFCC and GFCC feature are extracted from the speech signals. For training and testing, neural network using radial basis functions are used. In this experiment accuracy of Biometric person identification 96.52% has been achieved.

Keywords**—***Biometric, Pitch, Formant, MFCC and GFCC*

I. INTRODUCTION

A biometric system is a identification system, which identify the person by determining the authenticity of a specific physiological or behavioral characteristics of a user. Biometrics system can be used as a form of identity access management and access control in many applications. Biometric characteristics can be divided into two main classes: Physiological and Behavioral. The human voice is a physiological characteristic because each person has different vocal tract, but biometric person identification is mainly based on the study of the way a person speaks, commonly classified as behavioral characteristic.

Biometric person identification system refers to identifying persons from their speech or voice of different languages. In India there are many peoples who are able to speak more than one language [1]. Therefore there is a need to identify the effect of languages on a biometric person identification system in multilingual environment. When the multilingual biometric person identification system is being transferred to real applications, the need for greater adaptation in identification is required [13]. The performance of the monolingual biometric person identification 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 generally affected by other languages that are present in the surrounding and by the speaker's mother tongue [2]. Multilingual speech processing for multilingual speaker identification is a field of research in speech signal processing and language identification which come together many techniques developed for biometric person identification in single language environment with new approaches that convert it to the multilingual environment [8].

To find some statistically relevant information from multilingual human speech signals, it is required to have system for reducing the information which is available in multilingual speech signals into a relatively small number of parameters, called Feature Vectors. Feature extraction is the first step for the Biometric person identification system. Many algorithms were developed by the many people for feature extraction.

II. LITERATURE REVIEW

W. Burgos [20] conducted the experiments on the TIMIT database and the English Language Speech Database, were the test utterances are mixed with noises at various SNR levels to simulate the channel change. The results of their experiment shows an empirical comparison of the MFCC-GFCC combined features with the individual features.

S.Sarkar et al.[1] reported the performance of multilingual speaker identification systems on the IITKGP-MLILSC speech database. They designed a GMM-based speaker recognition system. They reported that the average language-independent speaker identification rate were 95.21% and an average equal error rate of 11.71%.

B.G.Nagaraja and H.S. Jayanna,[2] presented a paper in year 2013 on speaker verification in the environment of mono, cross and multilingual using MFCC and LPCC with the constraint of limited database for languages: English, Hindi and Kannada. They reported that the combination of MFCC and LPCC features gives 30% higher performance compared to the individual MFCC and LPCC features.

U Bhattacharjee and K.Sarmah[5] were conduct the experiment on their collected Multilanguage database for study the effect of language discrepancy on multilingual speaker identification system. English, Hindi and a local language of Arunachal Pradesh were considered for their experiments. GMM and UBM model were used for the speaker identification system They used Mel- Frequency Cepstral Coefficients (MFCC) vectors. The effect of the language mismatch in training and testing has been reported.

P. Kumar and S. L. Lahudkar [8] introduced a new method i.e fusion output for combined LPCC and MFCC(LPCC+MFCC) features and assess together with the different speech feature extraction methods. Speaker verification model for all methods was evaluated by VQ-LBG method.

S.Sharma and P Singh [17] presented two working engines using both alluded alternatives by use of continuous BPNN and GFCC method. They designed a system which was used for training and testing of audio wave files. They used neural network and GFCC tool for their result. After collection of database of speech signals for the training, extracted necessary features and train the system using GFCC and BPNN algorithm.

III. METHODOLOGY

Database Generation:

For Biometric person identification system, there is a need of database. The database is generated by recording of the speech signal of different persons of three Indian languages i.e. Hindi, Marthi and Rajasthani. To record the voice samples we have used microphone. The recording and segmentation was done by Gold wave software. The sampling rate of recorded sentences was 16KHz.The sentences consist consonants i.e "cha", "sha" and "jha" for the recording. Total number of speakers involved were 100 including males and females.

Feature Extraction:

There are two modules in Biometric person identification system in multilingual environment i.e Feature Extraction and Feature Matching. A small amount of information from speech signal of a multilingual person has been extracted in Feature Extraction, where as in Feature Matching extracted features are compared to identify the person in multilingual environment.

The original speech signal of a person consist redundant information. For Biometric person identification system, eliminating such redundancies helps in reducing the computational overhead and also improve system accuracy. Therefore all most biometric system involves the transformation of signal to set of compact speech parameter that is called feature vectors.

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 Pitch, Formant Frequencies, MFCC and GFCC features has been extracted from the speech signal of different languages of a person.

Pitch and Formant Frequencies features Extraction:

The pitch is fundamental frequency F0 and it is determined by the vibratory frequency of the vocal folds. The standard range of pitch is from 75 Hz to 500 Hz for human voice. Formants are frequency peaks which have, in the spectrum, a high degree of energy [15].The features pitch and first three formant frequencies are extracted through the PRAAT software. After extracting the different features the database has been prepared.

MFCC features Extraction:

MFCCs are one of the most important feature extraction techniques used in *Biometric person* identification system based on frequency domain using the Mel scale [14]. The complete process for obtaining MFCC coefficients is shown in figure-1:

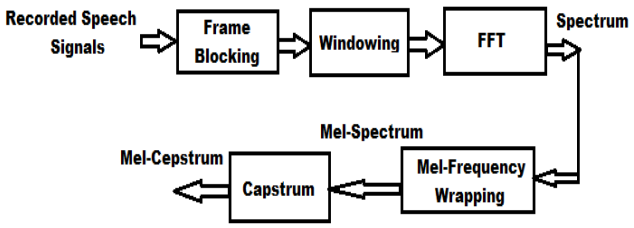

Figure-1: MFCCs feature extraction process.

In figure-1 recorded multilingual speech signals are segmented by Frame Blocking. The frame size of 15~20 ms with overlap of 50% is used. To keep the continuity of the first and the last points in the frame, each frame has to be multiplied with a hamming window using windowing technique. We get magnitude response and frequency response of each frame through Fast Fourier Transform (FFT). In Mel-Frequency Wrapping it is assumed that the signal within a frame is periodic, and continuous. Using Discrete Cosine Transform (DCT) we convert the log Mel spectrum into time domain is called Mel Frequency Cepstrum Coefficient(MFCCs). These set of coefficient is called acoustic vectors.

GFCC features Extraction:

In GFCC features extraction, for generate the cochleogram of the respective raw speech signal Gammatone filter bank is used, which represents transformed multilingual speech signal in the time domain and frequency domain. The cochleogram is based on ERB scale with finner resolution at low frequency Besides it allows more number of coefficients in comparison to MFCC. The Mel filter-bank for a power spectrum is with 257 coefficients while the GFCC is with 512 coefficients. The GFCC feature extraction process as shown in figure-2.

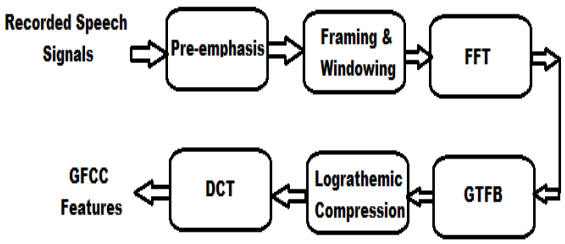

Figure-2: GFCCs feature extraction process.

In GFCC process the Pre-emphasis circuit is designed by a first order FIR digital filter for spectrally smooth the multilingual speech signals and make uniform the inherent spectral tilt in speech signals [1],[2]. Hamming window is used to reduce the spectral distortion. After that on windowed speech frame FFT is applied. As a result we obtained 512-point FFT spectrum of the speech frame. Gammatone filter bank(GTFB) which contains series of bandpass filters used to models the frequency selectivity property of the human cochlea. In order to simulate the human perceived loudness of given certain signal intensity logarithm is applied to each filter output [19]. The last stage of this process is consists of correlating the filter outputs using Discrete Cosine Transform (DCT).

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.

Classifier:

There are two major phases are involved to develop a multilingual speaker identification system i.e. 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.

Radial Basis Function network is one of the static feed forward neural network in which two layer, single hidden layer and one output layer. Gaussian or other kernel basis activation functions are used in hidden unit of Radial Basis Function network. Radial Basis Function neural network are much faster than the multi layer perceptron trained using back propagation algorithm. Radial Basis Function are less susceptible with non stationary input because speech is also non stationary type of signal, so Radial Basis Function is more appropriate classifier for Biometric person identification system.

IV. RESULTS

First of all the speakers voice (sentence) has been recorded in three Indian languages: Hindi, Marathi and Rajasthani, which are having ‘cha’ , ’’sha’ and ‘jha’ utterence in it are extracted from the sentences. Then pitch, formant frequencies, MFCC and GFCC features analysis has been done. Total 100 speakers including male and female are considered for the analysis. The result of features analysis and biometric identification system are discussed in following subsections:

Pitch and Formant frequencies analysis:

During the pitch and formant frequencies analysis, it has been observed that when speaker change the language, the value of pitch and formant frequencies has changed. The variation in the pitch value for utterance ‘cha’ in three languages as shown in figure-3.

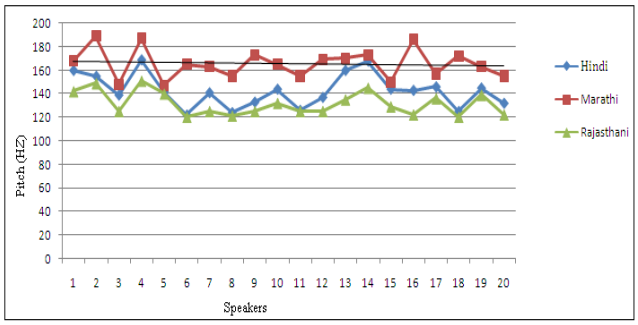

Figure-3: Variation of pitch for three languages of utterance ‘cha’.

The value of pitch and formant frequencies of Marathi language has more as compared to hindi and Rajasthani languages as shown in figure-3. The observation has been presented that the pitch of speaker are increases when speaker change the language from rajasthani to hindi to marathi. Similar analysis has been perform for utterance ‘sha’ and ‘’jha’ also. Same analysis has been performed on the first three formant frequencies F1,F2 and F3 and percentage deviation has been calculated. The percentage deviation shows that when the speaker change the language ,the percentage deviation in formant frequencies F1,F2 and F3 for Rajasthani and Marathi from hindi are positive and negative respectively for utterance ‘cha’.Similar analysis has been perform for ‘sha’ and ‘jha’ also.

MFCC Analysis

Here examine some detail analysis of Mel Frequency Cepstral Coefficients (MFCCs)-the domain features used for Biometric person identification system in multilingual environment. Their success has been due to their ability to represent the speech amplitude spectrum in a compact form. The analysis has been done for the three utterance "cha", "sha" and "jha" in three Indian languages. Base of the analysis is observed the variation in mel frequency cepstral coefficients when speaker change the spoken. 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.

GFCC Analysis

The GFCCs feature extraction method has been implemented for Biometric person identification system in multilingual environment. 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 a feature vectors.

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.

Biometric person identification system in multilingual environment:

Biometric person identification system in multilingual environment system has been designed. Features vectors database using MFCC and GFCC has been prepared The extracted features of the speech signals of multiple languages are observed. For training and testing RBF neural network are used. The experimental result are shown in table -1

Table -1: Percentage identification rates of Biometric person identification system in multilingual environment:

Language	% Identification Rate (Using)		
	MFCC	GFCC	MFCC + GFCC
Hindi	89.33	96	97.56
Marathi	86	95	96.25
Rajasthani	83.66	94	95.75
Average % identification rate	**86.33**	**95**	**96.52**

From table-1, it is clear that if the person spoke the hindi language has the greater identification rates 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 biometric person identification system in multilingual environment. It is also observed that using combine features(MFCC+GFCC), the averages % identification rates slightly increases as shown in figure-4.

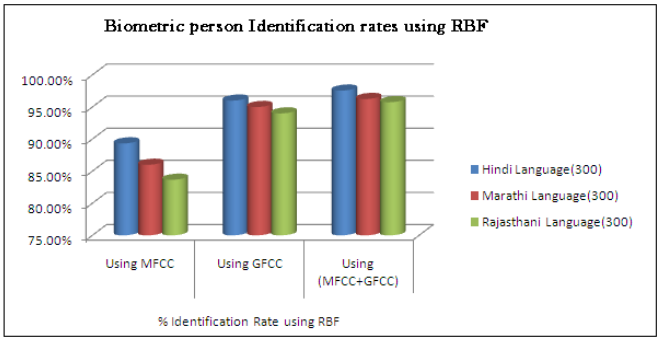

Figure-4: Biometric person identification using RBF

V. CONCLUSION

For Biometric person identification system in multilingual environment, Pitch, Formant Frequencies, MFCC and GFCC features has been successfully extracted. The value of features 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. A combined approach for feature extraction is suggested to improve the reliability of multilingual speaker identification system. The RBF neural network gives the better percentage identification rates than the other algorithms

REFERENCES

[1] S.Sarkar, K.S.Rao, D. Nandi and S.B.S.Kumar, "Multilingual speaker recognition on Indian languages",Annual IEEE India Conference(INDICON), Mumbai,2013:1-5.

[2] B.G Nagaraja. 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, A.K. Shruti and C. R. Krishna. "Prosodic feature based text dependent speaker recognition using Machine learning althorithms". International Journal of Engineering Science and Technology.2(10), 2010:5150-5157.

[4] B W. Gawali , S. Gaikwad , P. Yannawar and S. C.Mehrotra, "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, C. Leung, C. Barras, and J-L. Gauvain, "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, S. Ghosh, A. Si and T.K.Basu "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, A. S. Zadgaonkar and A. Shukla, "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, A. S. Zadgaonkar and A. Shukla, "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] S. Sharma and P. 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] X. Zhao and D. Wang, "Analyzing Noise Robustness Of MFCC And GFCC Features In Speaker Identification" ICASSP 2013(IEEE),978-1-4799-0356-6/13

[19] S. Arora and N. 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] W. Burgos, "Gammatone and MFCC features in speaker recognition" M.Tech Thesis,Melbourne, Florida November 2014.

ABOUT THE AUTHORS

Vinay Kumar Jain, pursuing Ph.D. degree from Chhattisgarh Swami Vivekanand Technical University, Bhilai. He has received M.E. degree in Electronics & Telecommunication with specialization in Communication with honors from Chhattisgarh Swami Vivekanand Technical University, Bhilai in 2008 and B.E degree in Electronics & Telecommunication from Pt. Ravi Shankar University Raipur in 2003. He is working as a associate professor in Shri Shankaracharya Technical Campus, Bhilai since 2003. His current research interests include speaker identification in multilingual environments, acoustic modeling, and machine learning for speech and audio applications.

Dr (Mrs) Neeta Tripathi, a young and dynamic Principal, is presently heading the Shri Shankaracharya Engineering College, Bhilai. She is first women and youngest Principal of Engineering Institution recognized by the C.S.V.T.U. Bhilai in state of Chhattisgarh. She has an excellent track of academic record. She graduated in Electronics and Telecommunication Engineering from Government Engineering College Jabalpur in 1986 and obtained M. E. degree in Electronics and Telecommunication with specialization in Communication Systems in 1992 from the same Institute. Pandit Ravishankar Shukla University Raipur has awarded her Ph. D. Degree in Electronics and Telecommunication in 2007 on "Study of Face and Speech Parameters and Identification of their relationship for Emotional status Recognition". She had started her career as Research and Development Engineer in NITEL in 1987 and worked there for Five years. In 1992, she had switched over from Industries to Teaching and serving this noble profession for more than twenty five years. She guided four Ph.D. candidates and five are in pipe line. She received two best paper awards. Her areas of research are *Emotion Recognition, Signals Processing, Mobile Computing, Communication Systems and Image Processing.*

