CSVTU International Journal of Biotechnology, Bioinformatics and Biomedical

Vol. 6(2), 33-39 2021 ISSN: 2455-5762

Investigation of Spoken-Language Detection in Multilingual Environment

Dr. Vinay Kumar Jain

Associate Professor, Shri Shankaracharya Technical Campus, Shri Shankaracharya Group of Institution, Junwani, Chhattisgarh 490020, Bhilai, India

vinayrich_17@yahoo.co.in

Received June 18, 2021; received in revised form August 15, 2021; accepted August 2021; Available online August 2021

Research Article

Abstract

Spoken language contains lot of information such as information about the content of a message and information about the speaker of that message. Content is composed of several levels of linguistic information like phonological information, morphological information, syntactic information, and the semantic information. For Present study, Multilingual Speech Processing database of different speakers has been recorded in three Indian languages, i.e., Hindi, Marathi, and Rajasthani. The sentences consist of consonants, i.e., "Cha", "Sha" and "Jha". Total numbers of speakers involved are 30 including males and females. The basic features of the speech signal: Pitch and first three Formant F1, F2 and F3 are calculated through PRAAT software whereas cepstral features Mel- Frequency Cepstral Coefficients (MFCC) and Gammatone Frequency Cepstral Coefficients (GFCC) has been extracted from MATLAB software. A model is proposed to identify the speaker by multi language speech signal of a speaker using MFCC, GFCC and combine features as acoustic features. For training and testing, is performed on using neural network function Resilient Back Propagation Algorithm and Radial Basis Functions and results are compared. In this experiment accuracy of spoken language identification is 94.77% using BPA and 96.52% using RBF neural network.

Keywords: Pitch, Formant, MFCC, GFCC, Multilingual.

1. Introduction

Multilingual speech processing for multilingual speaker identification is a field of research in speech signal processing and speaker identification which comes together many techniques developed for multilingual speaker identification in single language environment with new approaches that convert it to the multilingual environment [1]. In Multilingual Speech Processing, Multilingual Speaker Identifications are covered in two sections, first Processing and second one is Identification. In Processing, various features of the Multilingual speech signal is analyzed where as in Identification unknown language is identified through the speech signal from a speech database of multiple language. A text dependent language identification i.e., speaker will have to speak predefined sentences for identification of spoken language has been designed. The identity of a multilingual speaker can be determined from the information contained in the Multilanguage speech signal through speaker identification [2]. The Multilingual Speaker identification is concerned with identifying unknown speakers from a speech database of multilingual speaker models previously enrolled in the system [3]. The demand for multilingual speaker identification system increases for countries like India where many people are able to speak more than one language. The main objective of the multilingual speech processing is to observe the effect of the multiple languages spoken by a speaker on the spectral features of the speech signal. Another objective of this research work to observed the variations in the values of the spectral features when speaker change the spoken language. After multilingual speech processing and various observations of the features, the multilingual speaker identification system is designed to extract, analyze, characterize and recognize information about the speaker identity in multilingual environment.

2. Literature Review

Using average pitch and formant analysis of the speech signal, a closed-set text-independent Speaker Identification system was designed by Bashar and et al., in year 2014 [4]. The average pitch of speech signals was calculated from autocorrelation function and formant analysis from power density function and found that the designed speaker identification system based on pitch and formant method was superior to others [4].

Vimala and Radha [5] mainly targeted in research work to evaluate the performance of gammatone frequency capstral coefficient feature extraction technique with the others conventional feature extraction techniques. They designed and developed a speaker independent isolated speech identification system for Tamil language using various feature extraction and pattern matching techniques where the GFCC feature extraction techniques performed better than the conventional feature extraction techniques and achieved better results. In this work, highest word recognition accuracy had been achieved with GFCC features for both training and testing data [5].

Pahwa and Aggarwal [6] work involves the extraction of one of the most dominant and most researched up on speech feature, Mel coefficients and its first and second order derivatives. They extracted 13 values for each of these from a data-set 46 speech samples containing the Hindi vowels and trained them using a combined model of SVM and neural network classification to determine their gender using stacking. The results obtained showed the accuracy of 93.48% after taking into consideration the first Mel coefficient [6].

The speaker identification system using the GFCC features and GMMs has been developed and analyzed using TIMIT and NTIMIT databases. Moinuddin and Kanthi, [7] compared the performance of the system with the baseline system using the traditional MFCC features and concluded that the GFCC features has a good recognition performance not only in clean speech environment, but also in noisy environment.

The comparison of two feature extraction techniques of the speech signal i.e. MFCC and GFCC were presented in a paper in the year 2013. According their results the GFCC, exhibits superior noise robustness as compared to the MFCC. Analysis had been done to understand the noise robustness of GFCC as compared to MFCC. They designed speaker identification system with experimental set to systematically analyze their differences and similarities. This study reveals that the nonlinear rectification accounts for the noise robustness differences primarily [8].

Sarkar et al. 2013 [9] reported the performance of multilingual speaker identification systems on the IITKGP-MLILSC speech corpus for their designed GMM-based speaker recognition system where average language-independent speaker identification rate 95.21% and an average equal error rate of 11.71% was achieved.

3. Methodology Adopted

Methodology adapted for present work is divided into two sections as Multilingual Speech Signal Processing as shown in Fig.1 and Multilingual Speaker Identification as shown in Fig.2.

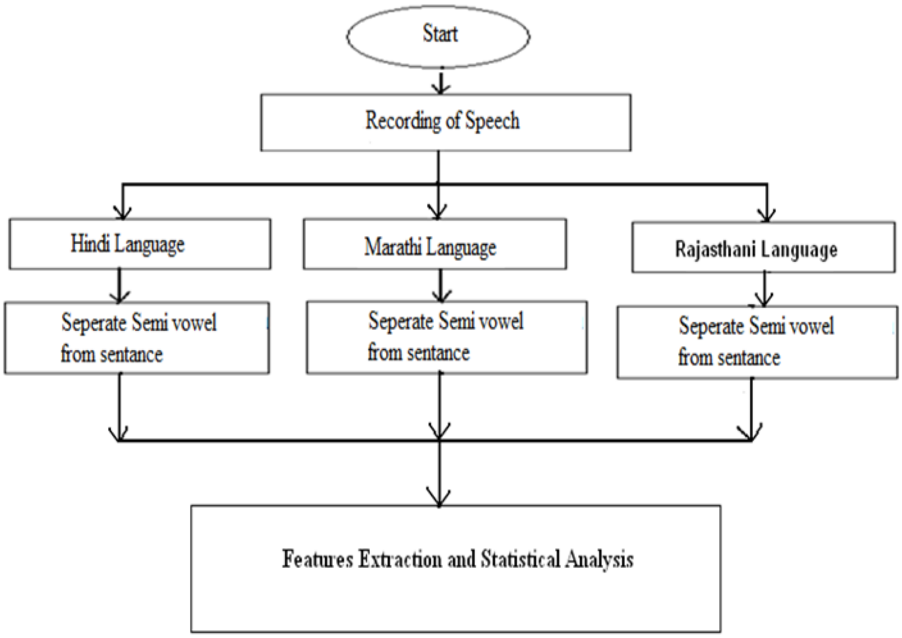

Fig.1 Multilingual Speech processing.

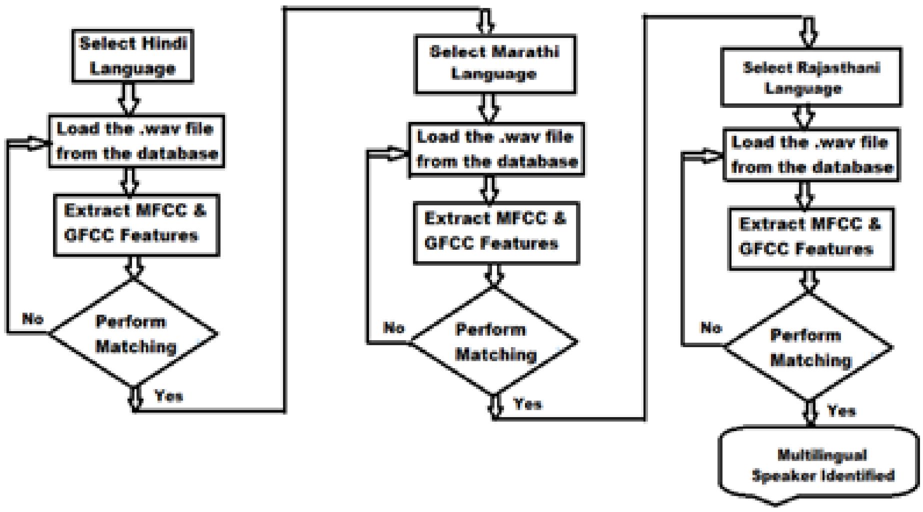

Fig.2 Spoken language Identification System.

For Multilingual speech processing, the database of different speakers in different languages has been recorded. Short clip of vowel and semi vowel are separated from recorded speech signals. Short clips of speech signals are used for further analysis. Vocal tract characteristic feature and other features of speech are extracted from short clip for statistical analysis. The basic features of the speech signal: Pitch and First three Formant Frequencies are calculated from PRAAT software whereas cepstral features: MFCC [10] and GFCC have been extracted from MATLAB software.

Process for multilingual speaker identification system adapted for present work is shown in Fig.2. MFCC and GFCC features are extracted from the multilingual speech signals of different speakers for identifying the variation of speech features in multilingual environments which are used to design multilingual speaker identification system [11,12]. For training and testing, neural network using resilient back propagation algorithm and radial basis functions are used.

4. Result and Discussion

The pitch values are obtained through PRAAT software of all 30 speakers in Hindi, Marathi and Rajasthani language for utterance "Cha" "Sha" and "Jha" and recorded. A statistical analysis is performed on recorded value and percentage deviation from Hindi language is calculated. A sample in pitch variation and percentage deviation for utterance of "Cha" are presented in Fig.3 and Fig.4 respectively.

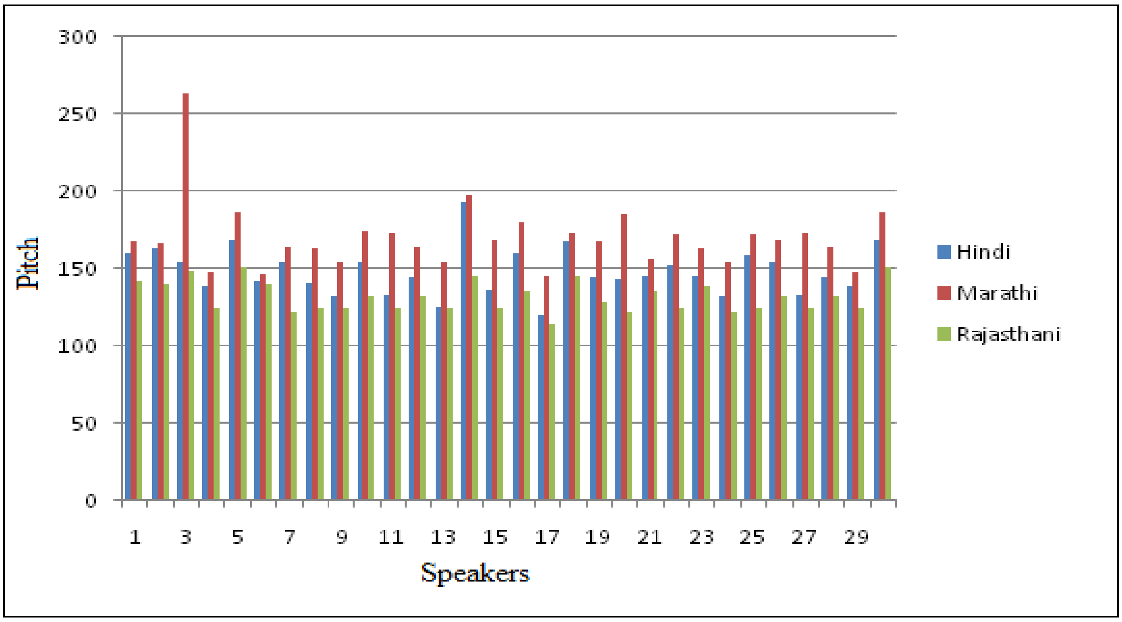

Fig.3 Pitch Analysis for utterance "Cha".

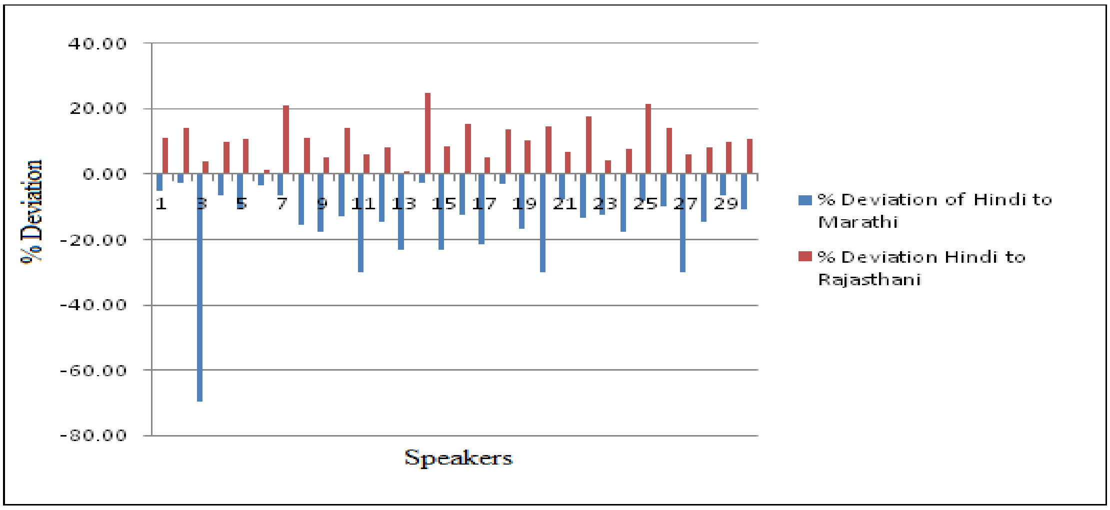

Fig.4 % Deviation of Pitch for utterance "Cha".

From Fig.3 and Fig.4, it is cleared that same speaker, same sentence but in three different languages has change the speech parameters. It has been observed that the percentage deviation in Pitch for Rajasthani and Marathi from Hindi are positive and negative respectively.

Similar analysis has been performed for utterance “Sha” and “Jha” also and same result has been obtained.

Presented observation, the Formant F1 of the speakers has changed when the speaker changes the spoken language. The Formant F1 analysis and percentage deviation of Marathi and Rajasthani language from Hindi language for utterance “Cha” are shown in Fig.5 and Fig.6 respectively.

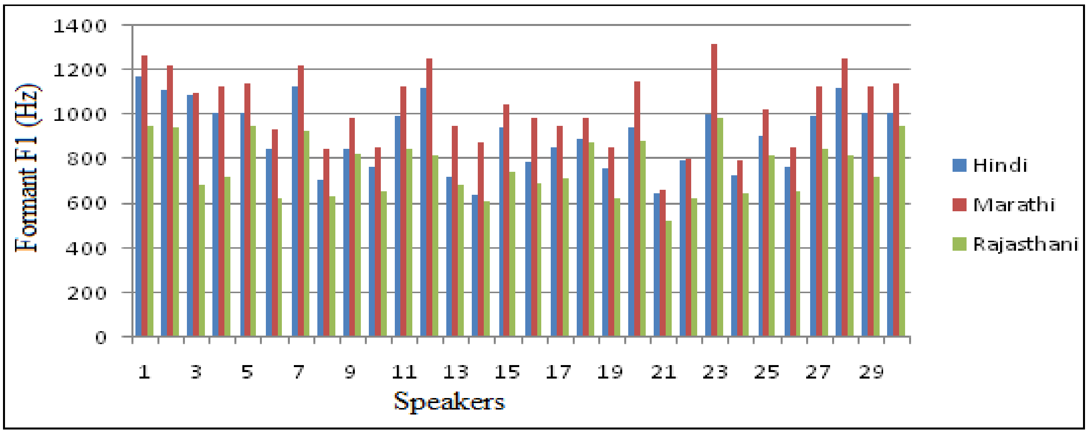

Fig.5 Formant F1 analysis for utterance ‘Cha”.

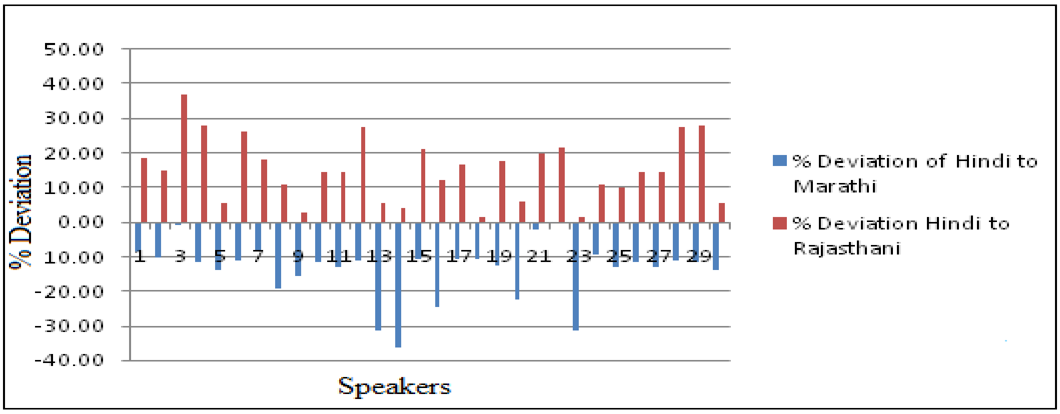

Fig. 6 % Deviation of Formant F1 for utterance “Cha”.

In Fig.5, the formant F1 of the speech signal has more when speaker speaks in Marathi language as compared to Hindi and Rajasthani language for utterance “Cha”. From Fig.6, it has been observed that the percentage deviation in Formant F1 for Rajasthani and Marathi from Hindi are positive and negative respectively for utterance “Cha”. Similar analysis has been performed for utterance “Sha” and “Jha” also. Formant F2 and F3 analysis has been performed in similar manner, and same result obtained.

Spoken Language identification system has been designed using MFCC, GFCC and combined features. These experiments are 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 and table-2, it is clear that if the speaker spoke the Hindi language has the greater identification rates as compared to Marathi language and Rajasthani language in all three formats.

Table-1: % Identification Rates using BPA.

Language	% Identification Rate (Using)		
	MFCC	**GFCC**	**Combine Features (MFCC+ GFCC)**
Hindi	88.66	95	96.33
Marathi	84.33	93	94.66
Rajasthani	81	91.66	93.33
Average % Identification Rate	**84.66**	**93.22**	**94.77**

Table-2: % Identification Rates using RBF

Language	% Identification Rate (Using)		
	MFCC	**GFCC**	**Combine Features (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 and table-2, the average identification rates of the spoken language for using MFCC, GFCC and combine (MFCC + GFCC) feature extraction techniques are 84.66%, 93.22% and 94.77% respectively with BPA neural network, whereas 86.33%, 95% and 96.52% respectively with RBF neural network. It is also observed that GFCC give the better performance as compared to MFCC in spoken language identification system.

Conclusion

For Multilingual speech processing, Pitch, Formant, MFCC and GFCC speech features has been successfully extracted from the multilingual speech signal. The values of the features have changed when speaker change the spoken language. The value pitch and first three Formant F1, F2 and F3 of the speech signal of a speaker are more when speaker, speaks in Marathi language as compared to Hindi and Rajasthani language. It has been observed that the percentage deviation in pitch and first three Formant F1, F2 and F3 for Rajasthani and Marathi from Hindi are positive and negative respectively. It is observed that GFCC give the better performance as compared to MFCC in multilingual speaker identification system. Using Combine features (MFCC+GFCC), the averages percentage identification rates slightly increase as compared their individual counterparts. It is observed that RBF neural network give the better percentage identification rates than the BPA neural network.

Conflicts of interest: Authors have no conflict of interest to declare.

REFERENCES

[1] Rathore P.S. and Tripathi N. 2014. *Multilingual Person Identification*, International Journal of Engineering Trends and Technology ,10(1):1-3.

[2] Bourlard H. And Dines J. 2011. *Current trends in multilingual speech processing*, Indian Academy of Sciences, Sadhan 36(5): 885–915.

[3] Tripathi N.,2006. *Study of face and speech parameters and identification of their relationship for emotional status recognition,* Ph.D thesis, NIT Raipur.

[4] Bashar M. A., Ahmed M. T, Syduzzaman M., Ray P. J. and Islam A. Z. M. T. 2014.*Text-Independent Speaker Identification System Using Average Pitch And Formant Analysis.* International Journal on Information Theory (IJIT), 3(3):23-30.

[5] Vimala.C and Radha.V, 2014. *Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words.* International Journal of Computer Science and Information Technologies, 5 (1): 378-383.

[6] Pahwa A. and Aggarwal G. 2016. *Speech Feature Extraction for Gender Recognition.* I.J. Image, Graphics and Signal Processing, 9, :17-25

[7] Moinuddin M. and Kanth A. N. 2014. *Speaker Identification based on GFCC using GMM,* International Journal of Innovative Research in Advanced Engineering (IJIRAE), 1(8):224-232.

[8] Zhao X. and Wang D.,2013. *Analyzing Noise Robustness of MFCC and GFCC Features In Speaker Identification,* IEEE-ICASSP: 7204-7208.

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

[10] Sharma S., Shukla A. and Mishra P., 2014. *Speech and Language Recognition using MFCC and DELTA-MFCC,* International Journal of Engineering Trends and Technology, 12(9):449-452.

[11] Burgos W. 2014. *Gammatone And Mfcc Features In Speaker Recognition,* Mtech thesis, Florida Institute of Technology Melbourne, Florida.

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