JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 481 - 486
https://publishoa.com
ISSN: 1309-3452

New Feature Vector based on GFCC for Language Recognition

Dr. B. Chandrasekaram

Associate Professor
Department of Computer Science
National Sanskrit University
Tirupati

Abstract

Here a new form of feature vectorbased on Gammatone Frequency Cepstral Coefficients (GMFF) for language recognition is proposed. The major battle neck in degradation of language recognition(LR) performance is the presence of noise and mismatchedenvironment present in the speech signal.For any language recognition, the default feature vectors are MFCC, but the performance degradesin the presence of noise and mismatch conditions. From the literature, it is observed that GFCC has very good robustness against additive noise. In this work, a new feature vector using GFCC is introduced for language recognition tasks. The new feature vector based on GFCC for the GMM LR system task showed superior performance when it is compared tothe conventional MFCC feature vector-based GMM LR system.

Keywords: MFCC, GFCC, GMM, LR

1.Introduction

Themethod of identifying thelanguage from the short duration of the speech signal is termed as language recognition.The language recognition process comprises three stages, namely: first feature extractionstage, second stage training is performed,and in the final stage,testing is done. In the feature extraction phase, the language-related features are extracted, which are independent of the speaker, environment, and language. The identification performance of the LR was affected severely due to the presence of noise in the speech signal. The major factors that influence the accuracy of the language recognition system are noise and cross-channel utterances in an environment. Currently, researchers are exploring this area to overcome this problem.

The feature vectors used for language and speaker recognition are MFCC, GFCC, and perceptual linear prediction (PLP). These feature vectors represent the acoustics characteristics of the speech signal [1]. The MFCC features are commonly used for the speaker and language recognition system. In this method,the first windowed speech segment is extracted from the speech signal, and from this segment, the spectral information is extracted, which is given to the language recognition system for the recognition task. But these spectral feature vectors are efficient in a clean speech environment and are sensitive in a noisy environment.

The MFCC feature vectors use triangular filters for modeling the auditory critical bands. Due to the less performance of MFCC feature vectors in noisy or mismatch conditions, the Gammatone filters are used instead of triangular filters for modeling the auditory critical bands. The features extracted using Gammatone filters are called the Gammatone frequency cepstral coefficients (GFCCs). The main difference between MFCC and GFCC is scale,whereas MFCC is based on Mel scale, while GFCC is based on an equivalent rectangular bandwidth scale.The visual difference between the MFCC and GFCC feature vector is shown in figure.1From the figure.1, it is clearly evident that the energy components are very clear in Cochleagram(GFCC spectrogram) when compared to MFCC spectrogram[2].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 481 - 486
https://publishoa.com
ISSN: 1309-3452

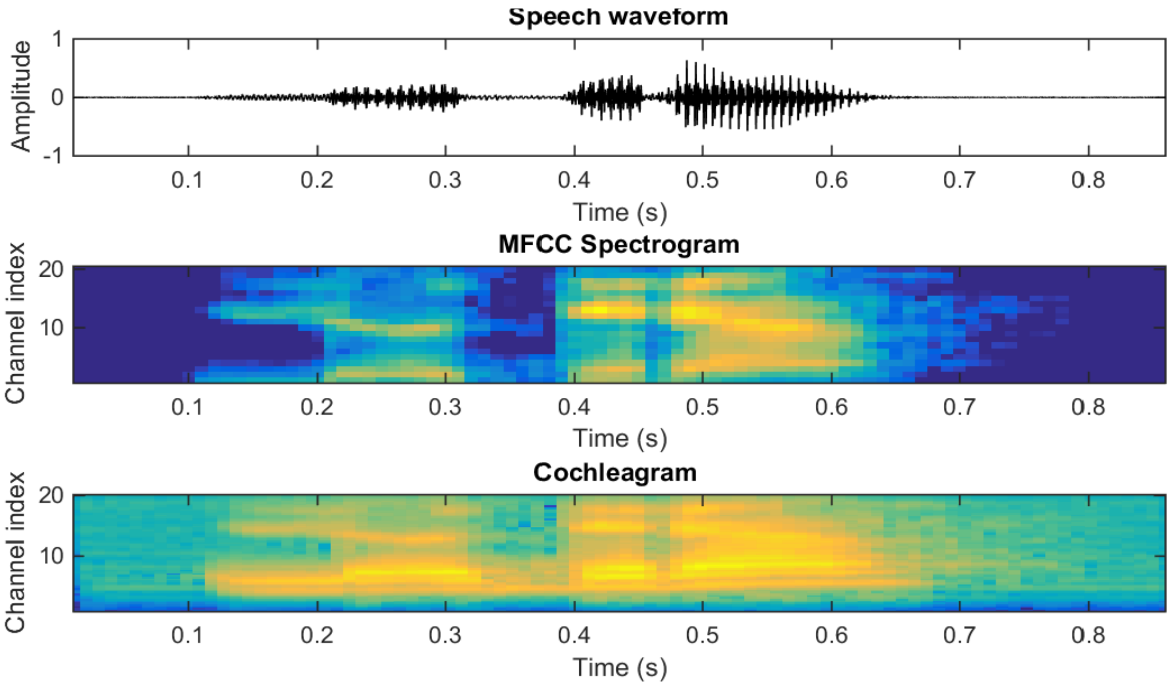

Fig.1: The visual difference between MFCC and GFCC feature

In the present work, a new form of GFCC feature vectors is obtained from GFCC features. This new form of GFCC features is used during the training and testing stages. Using these features, a new language recognition system is developed[3][4].

The work is described as follows. The first section relates tothe extraction of feature vectors from the speech signal. The second section explains the extraction of new feature vectors from GFCC. Section three deals with modeling these new forms of feature vectors using Gaussian mixture modeling. Finally, section 4describes the performance analysis of the new feature form of vectors LR task.

1.GFCC Feature Vector Extraction

The GFCCdepends on the Gammatone filter bank. The GFCC features are extracted as shown in figure.2. The feature extraction of GFCC is similar to the feature extraction of MFCC[5]. First, the speech signal is pre-emphasized to improve the signal. Next,the speech signal is divided into frames of 5ms to 10ms, and then to remove the spectral distortion, the signal is windowed. For this windowed signaloutput is the frame,then FFT is applied to obtain the FFT spectrum [6]. Then the Gammatone filter bank was used to model the frequency selectivity property of the human cochlea[7].

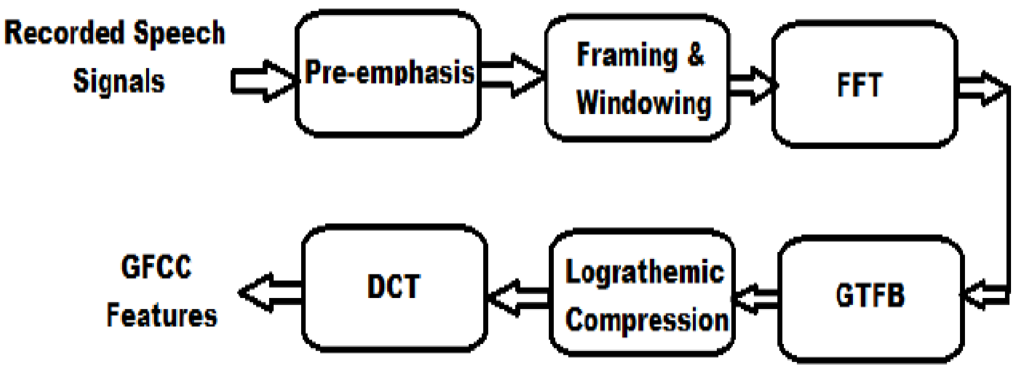

Figure-2: GFCCs feature extraction process.

2. New Feature Vector Extraction Using GFCC

In this section, the method of new feature vector extractionfrom GFCC is described. Here new form feature vectors are obtained from the GFCC feature vectors. The new feature vectors are expressed as a probability vector instead of a scalar value so that the performance of the language recognition system improves. The probability of the feature vector is calculated using the Gaussian probability density (pdf) function against each Gaussian, which is formed using GFCC training feature vectors using the GMM model.

First, from the speech signal, GFCC feature vectors of 12 dimensional are extracted. Then this new form of feature vectors isgrouped using a K-means clustering algorithm of 14, as depicted in fig.3.

Fig.3: R clusters for language.

Here each cluster represents one Gaussian. For each GFCC feature vector in training data, we calculate the probability using the probability density function against each Gaussian; this is represented as P_1. This way, we calculate 14 coefficients of the new feature vector, represented as ($P_{1,}$ $P_{2,.........,}P_{14}$). This way, the GFCC training data of size N is converted into a new feature vector of 14 dimensional of size N, as shown in fig.4. The new form of feature vectors is fed into the GMM language recognition system for identification purposes [8].

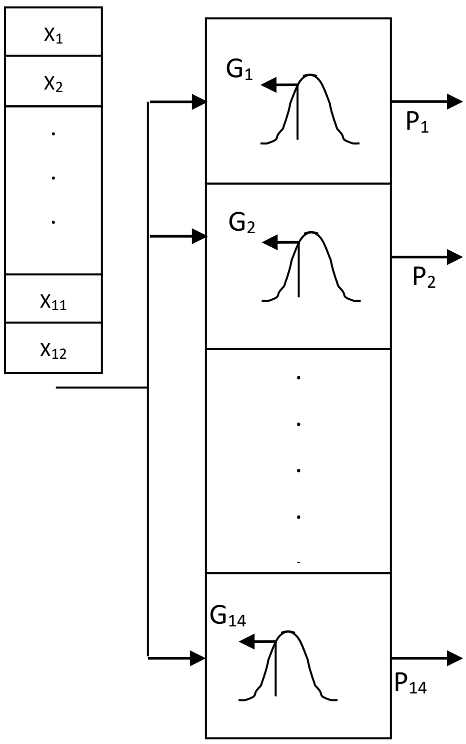

Fig.4: Converting GFCC feature vector into new feature vector P

3. Modelling of New Feature Vector-Based Language Recognition System

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 481 - 486
https://publishoa.com
ISSN: 1309-3452

One of the most important models for language recognition is the Gaussian mixture model (GMM). The GMM is a Gaussian distribution representing one-dimensional X. Here the X is the random variable, which is defined as a vector described using mean and variance. The Gaussian density(mixture) for the feature vector x, is calculated as

$$p(X\mid\lambda) = \sum_{i=1}^{M} w_i p_i(X)$$

The mixture density is expressed as a weighted linear combination of Gaussian density p(x)

$$p_i(X) = \frac{1}{(2\pi)^{D/2}\ |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}$$

The parameters of GMM is collectively represented as $\lambda = \{p_i, \bar{\mu}_i, \Sigma_i\}$. In the language recognition system, each language is represented by one individual GMM and represented by a language model λ_i.

3.1 New GFCC Feature Vector-Based Language Recognition GMM Based System

The training of the new feature vector-based language recognition system comprises of two stages, as shown in figure.5. They are training and testing. During the training, the new feature vectors are modeled using the GMM model. From training speech data, the 14-dimensional GFCC features vectors are extracted. These GFCC feature vectors are converted into new feature vectors, as explained in the previous section[9][10]. Now, these new form of feature vectors is modeled using the GMM model. For every individual language, one GMM model is created, and it is represented by L_i. This way, for each language under consideration, one GMM model is obtained[11][12].

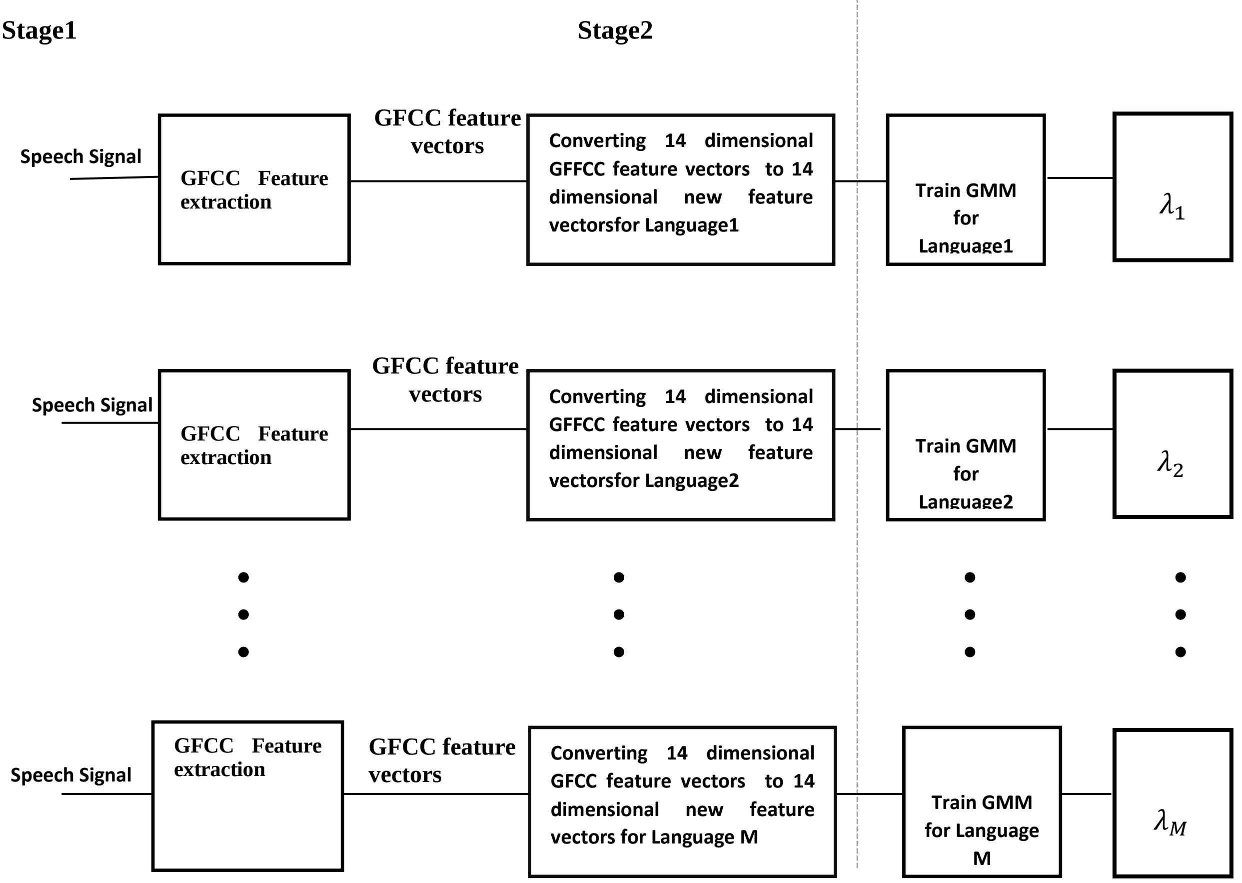

Fig.5: GMM training for new feature vector-based language recognition system.

During testing, from the test speech sample, GFCC features are extracted. Now, these GFCC feature vectors are converted into the new form of feature vectors. Using newly obtained GFCC new form of feature vectors, the likelihood values are calculated against each language GMM model. Which model yields the maximum likelihood value is identifiedby the type of that language.

3.2 Experimentation and Discussion

The new feature vector language recognition system based on GFCC was carried out using the OGMITS language corpus. The different speech data is used for both in training stage and testing stage. The LR system is modeled using GMM with the Gaussian mixtures such as 32, 64 and 128. The testing is carried out using speech utterance duration of 1 sec, 2 sec, and 3sec.

S.No	No. of Gaussian Components	Language Recognition based on new feature vectors using GFCC			Language Recognition basedonGFCC feature vectors		
		1 Sec	3Sec	5Sec	1Sec	3Sec	5Sec
1	4	77	80	84	62	66	70
2	8	79	81	86	63	68	72
3	16	80	83	88	65	69	73
4	32	82	84	91	66	71	76
5	64	83	86	92	69	74	77
6	128	84	89	94	70	76	80

Conclusion: The new set of feature vectors based on GFCC for language recognition tasks is introduced. The new set of featuresis robust against noise and environment mismatch conditions, so the performance of the new form of feature vectors language recognition system improved. To obtain new feature vectors from the speech signal, first, the GFCC feature vectors are extracted from speech, and these feature vectors are passed through Gaussians using the probability density function to get the new form of a set of feature vectors. The language recognition performance of new feature vectors is superior when compared with the conventional GFCC-based language recognition system.

References

1.E. B. Tazi, A. Benabbou, and M. Harti, "Efficient Text Independent Speaker Identification Based on GFCC and CMN Methods" ICMCS 2012, pp. 90-95.
2. 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.
3. Ambikairajah, E., 2007. Emerging features for speaker recognition. In: Proc. Sixth Internat. IEEE Conf. on Information, Communications & Signal Processing, Singapore, December 2007, pp. 1–7.
4.11. 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.
5. 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.
6. W. Burgos, ―Gammatone and MFCC features in speaker recognition‖ M.TechThesis,Melbourne, Florida November 2014.
7. 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.
8. M.S. Sinith, A.Salim, K. Gowri Shankar ,S. Narayanan, and V. Soman, "A novel method for Text-Independent speaker identification using MFCC and GMM".Audio Language and Image Processing (ICALIP), International Conference on. IEEE,Shanghai, 2010,Vol.5, pp.292-296.

9. U. Shrawankar, and V.M. Thakare, "Techniques for Feature Extraction In Speech Recognition System: A Comparative Study."International Journal Of Computer Applications In Engineering, Technology And Sciences, Vol. 2, No.5,2010, pp. 412-418.
10.Mannepalli, K., Sastry, P.N. and Suman, M.: MFCC-GMM based accent recognition system for Telugu speech signals. International Journal of Speech Technology.19(1), 87- 93 (2016).
11. Wilson Burgos, "Gammatone and MFCC features in speaker recogniton" M.TechThesis,Melbourne, Florida November 2014
12. V. Hubeika, L. Burget, P. Matejka, and P. Schwarz. Discriminative training and channel compensation for acoustic language recognition. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, pages 301–304, January 2008.
13. Deshpande, P. S., and J. S. Chitode. "Transformation coding for emotion speech translation: a review." International Journal of Electrical and Electronics Engineering Research (IJEEER) 6.1 (2016): 1-12.
14. Lanjewar, RAHUL B., and D. S. Chaudhari. "Comparative analysis of speech emotion recognition system using different classifiers on berlin emotional speech database." International Journal of Electrical and Electronics Engineering Research (IJEEER) 3.5 (2013): 145-56.
15. VYOMAKESISRI, TIPPABHOTLA. "CHALLENGES IN LEARNING ENGLISH AS SECONDARY LANGUAGE." International Journal of English And Literature (IJEL) 7 (2017): 21-24.
16. Parsania, Pankaj S., Nischal M. Chavda, and Krunal C. Kamani. "Information and Communication Technology & Its Impact in Improving the Teaching and Learning of English Language." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 5.3 (2015): 1-6.
17. Garg, Hitesh, R. K. Bansal, and Savina Bansal. "Improved Speech Compression using LPC and DWT Approach." International Journal of Electronics, Communication & Instrumentation Engineering Research and Development (IJECIERD) 4.2 (2014): 155-162.
18. VIJAYASRI, BV, and I. SANTIPRABHA. "IMPLEMENTATION OF A NOVEL TRANSFORMATION TECHNIQUE TO IMPROVE SPEECH COMPRESSION RATIO." Development (IJECIERD) 3.3 (2013): 99-106.