

# New Feature Vectors Based Speaker Recognition Using Random Forest

Dr A.Nagesh

Professor, Department of IT, MGIT

**Abstract:** The speaker recognition is a process of identifying a speaker from a set of speakers using speech signal. The lot of research is going on last four decades using different kind of feature extraction techniques for speaker recognition. The traditional feature vectors are MFCC and LPC. In this paper, a new form of feature vectors is proposed for the of speaker recognition using Random Forest machine learning algorithm. The new form of feature vectors is extracted from MFCC features based on the principle frequency of occurrence of phone frequency is different between languages. Using this new form of feature vectors is input to Random forest algorithms which identify the recognized speaker. The new feature vectors are robust and noise free, which are important to improve the accuracy of the speaker recognition.

Keywords: MCC, Speaker Recognition, Feature Vectors.

## 1.Introduction

The speaker Recognition is are used in wide range of applications such as communications, security and smart technologies. It is commonly used to recognize person in cell phone, banking applications, and computer systems. It is used frequently in smart assistants and home appliances to recognize the person and provide user specific services. It also provide as audio evidence in forensic investigation to support law enforcement agencies. Further, speaker recognition is used in access control system, online examination system for verification purpose, surveillance systems, and human computer interaction.

The existing all the speaker recognition systems based on Mel Frequency Cepstral Coefficients (MFCC) or Lenier Predictive Coeff Picients (LPC) feature vectors. The feature vectors contain important characteristics of speech signal. The MFCC features represent the short term power spectrum of speech signal which represent the characteristics of human auditory system. This MFCC feature effectively capture spectral properties of speech signal and convert into compact numerical vector representation which is given as input to Random Forest algorithm for speaker recognition task. The LPC feature models characteristics of human vocal tract effectively. In the proposed new feature vectors based Speaker Recognition System Using Random Forest, it used new form of feature vectors. This new form of feature vectors, first from speech signal MFCC feature vectors are extracted and this feature vectors contain specker specific information. Then using these MFCC feature vector, new form of feature vectors derived using Gaussian probability distribution function.

The conventional speaker recognition systems based on Gaussian Mixture Model(GMM), Hidden Markov Model(HMM), and Support Vector Machine (SVM) approaches. In GMM

model, it models probability distribution of each speaker specific feature vectors using combination of Gaussian components. HMM model the temporal variations and sequential of patterns each speaker which represent dynamic patterns in speech signal. The SVM is supervised machine learning classifier using speaker feature vectors its construct an optimal boundary line between each speaker. Using this optimal boundary line easily speaker recognition task is performed. The conventional approaches widely used for speaker recognition applications due to its reliability to model feature vectors effectively and effective recognition performance.

In the recent times so many supervised machine learning algorithm are employed for Speaker recognition task. From the literature study it is found that much exploration is not done using Random Forest supervised machine learning algorithms. Due to this, in this paper a Speaker Recognition System using Random Forest is proposed. The Random Forest algorithm construct multiple decision trees to improve speaker recognition accuracy and reduce overfitting during training. The new form of feature vectors given as input to random forest algorithm to create speaker model.

## **2.Literature Survey**

The speaker recognition is widely explored in the fields of speech recognition, Pattern recognition, and machine learning. The researchers are proposed different approaches based on features extraction methods and classification algorithms for speaker recognition.

The starting speaker recognition systems mainly depend feature extraction methods such as MFCC and LPC. The MFCC features extract vocal tract and spectral information which in extracted from speech signal is used to differentiate between the speakers. The MFCC feature widely used in speaker recognition due to it closely models human auditory perception of sound.

The lot of speaker recognition systems are proposed using GMM, HMM and SVM over last tree decades. First Douglas A.Reynolds proposed speaker recognition using MFCC feature vectors technique [1]. This is first feature extraction method. The Juhani Saastamoinen, Evgeny Karpov, Ville Hautamaki, and Pasi Franti explored accuracy of speaker recognition using MFCC feature extraction for mobile phones on series 60 devices[2]. Asutosh Das, Manas Ranjan Jena, and Kalyan Kumar Barik shown effectiveness of MFCC feature extraction for improved accuracy and recognition performance [3]. P. Kumar and S. L. Lahudkar developed using MFCC and LPCC feature extraction techniques[4]. In this paper it is observed that performance of MFCC based speaker recognition system is better than LPCC based speaker recognition system. Further it also concluded that the fusion MFCC and LPCC feature extraction techniques is better performance than conventional MFCC and LPS based speaker recognition systems.

K. A. Lee, B. Ma, and H. Li developed speaker verification system based on MFCC feature vector using which captures spectral characteristics effectively from speech signal using GMM

modelling technique which capture distribution of specific features[5]. This proved MFCC feature vectors along GMM modelling technique provides reliable speaker verification task.

The SVM based SID system, which discriminative classifiers capable of maximizing the classification between speaker specific classes in high dimensional space [6]. Unlike GMM, SVM directly optimize decision boundary between speaker specific classes instead of modelling the distribution of acoustic feature in feature space [7]. The discriminative SVM based SID system achieved significant accuracy in classification performance using short duration test utterance compared to conventional generative GMM based SID systems accuracy performance[8][9][10].

The HMM based SID systems extended conventional acoustic feature modelling SID systems by incorporating temporal dynamics into SID systems. The HMM captures temporal dynamic of speech by modelling sequential patterns into features [11].

CNN based SID systems is the modern deep learning approaches for SID task [12]. These systems directly learn speaker specific discriminative feature from the speech signal in the form of spectrograms and filter bank energies [13] [14]. CNN automatically learns hierarchical acoustic patterns from the data. The performance of CNN based SID systems much superior when compared to GMM and HMM based SID systems.

In this paper, Random forest based SID system is proposed and this is machine learning supervised learning algorithm widely used for language and speaker identifications.

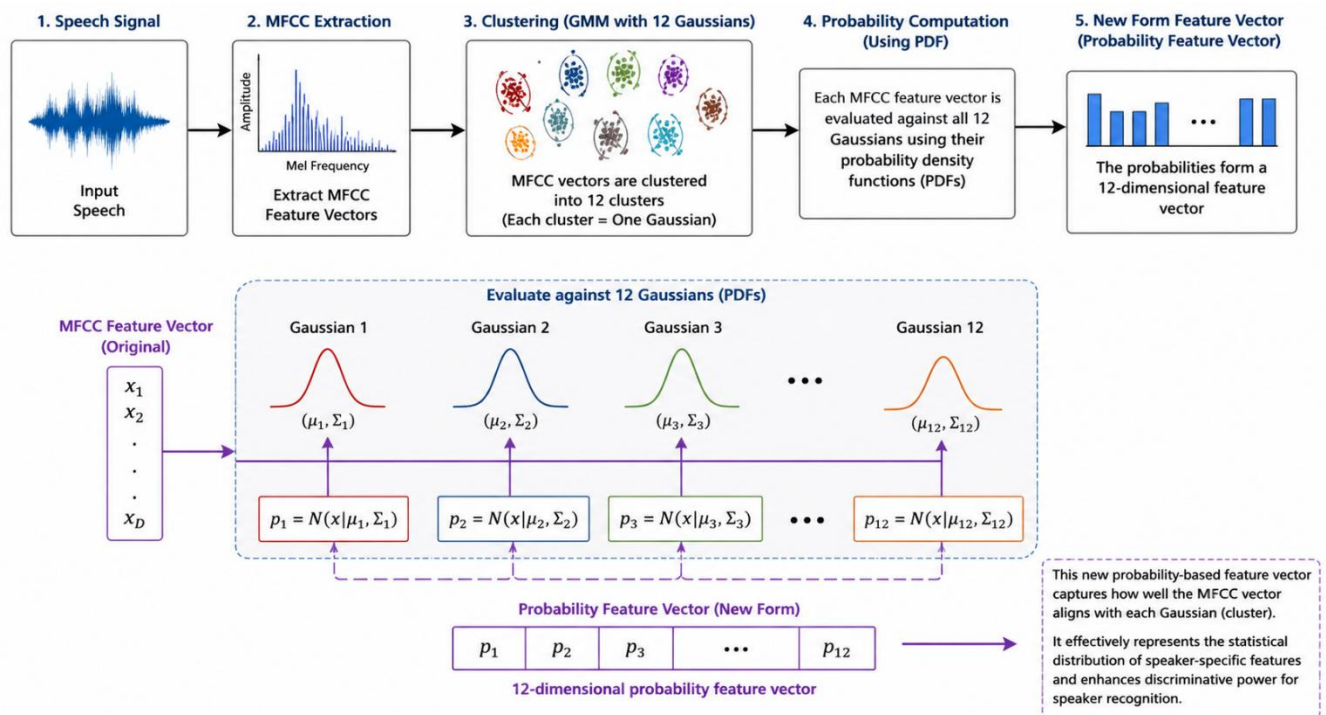
### **3.Methodology**

Here for speaker recognition task, random forest supervised explored due to its robustness and effectiveness. The random forest is ensemble learning algorithm which is has multiple decision trees and each tree is constructed using randomly selected subset of features from training data. The ensemble learning approach improves generalization by reducing overfitting.

The new feature vector based speaker recognition system using random forest consists of three stages, namely first new feature vectors extraction from speech signal, training speaker recognition model using new feature vectors and finally stage, recognizing speaker from test utterance.

#### **3.1 New feature vector extraction**

In first stage, new form of feature vectors extracted from speech signal. The MFCC feature vectors are extracted from the speech signal. Using these MFCC clustered into 12 clusters, which represents 12 Gaussians. Next these feature vectors pass through one Gaussian using probability density function (PDF), representing one coefficient in new form feature vector. This way feature vectors pass through 12 Gaussians creating 12 coefficients in the new form feature vector. This new form feature vector in probability feature vector form. These new feature vectors captures speaker specific features distribution effectively [15] [16].



**Fig1: New feature extraction vectors using MFCC feature vectors**

### 3.2 Random Forest Based Speaker Model

In the second stage using new form feature vectors are trained with random forest classifier supervised learning algorithm composed of an ensemble of decision trees. Here each decision tree is trained using different subset of training data new feature vectors with principles of bootstrap aggression(aggression). Each decision tree learns a different representation of undelaying data distribution and M set of decision trees forms a one speaker random forest model. This way training procedure is repeated all speaker under consideration and each speaker separate random forest model (set of decision trees) is created using its corresponding feature vectors. For N speaker N random forest models are created and each model captures statistical characteristics which are specific to each speaker.

### 3.3 Random Forest Based Classifier

During third stage(testing), from test speaker speech first 12 dimensional new feature vectors are extracted as explained above in training stage. The testing new feature vectors pass through every decision tree which is speaker specific to random forest model. Each decision tree produces one classification decision tree and the final decision model is obtained using majority voting among M decision trees[17] [18]. This procedure is repeated N speakers models which produces N decision trees. Here each speaker model outputs classification decision tree speaker model along with voting score. Which label of classification decision tree speaker model yields maximum score is classified as belongs to it [19] [20].

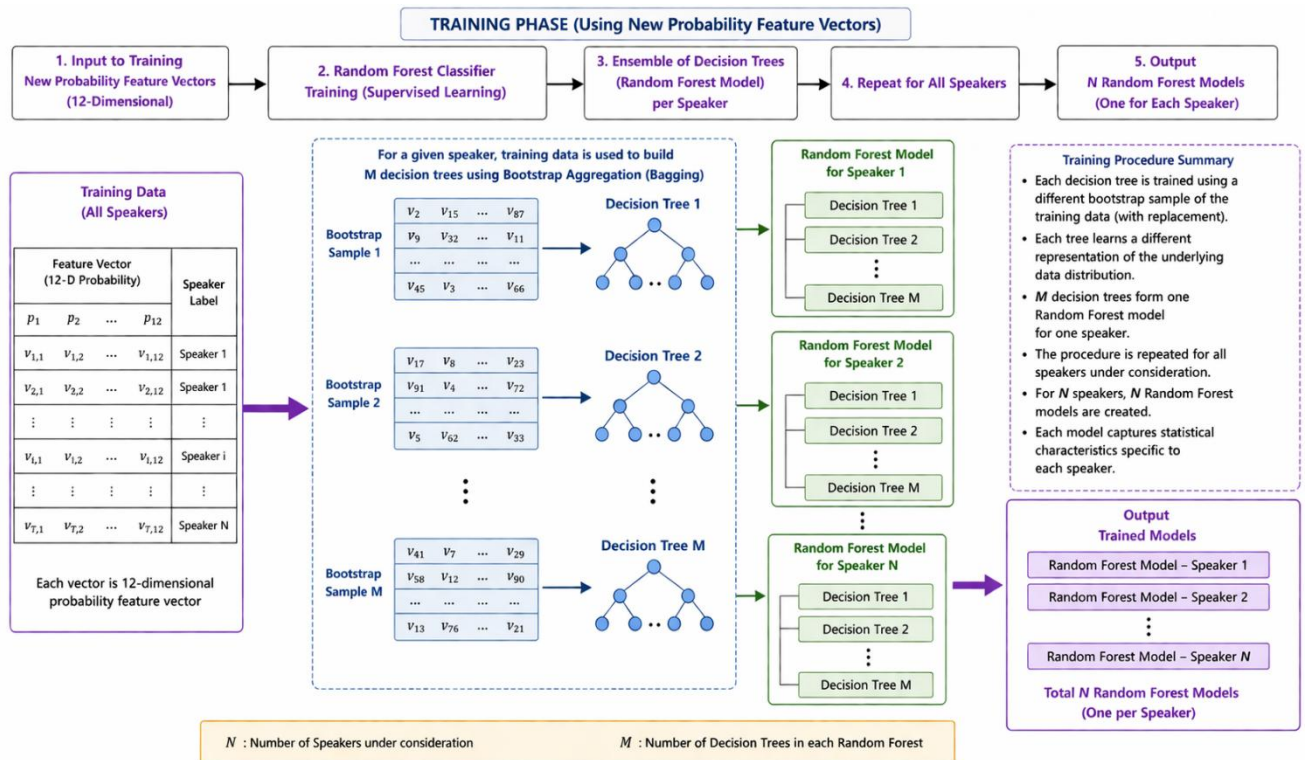


Fig2: Training new feature vector using Random Forest algorithm

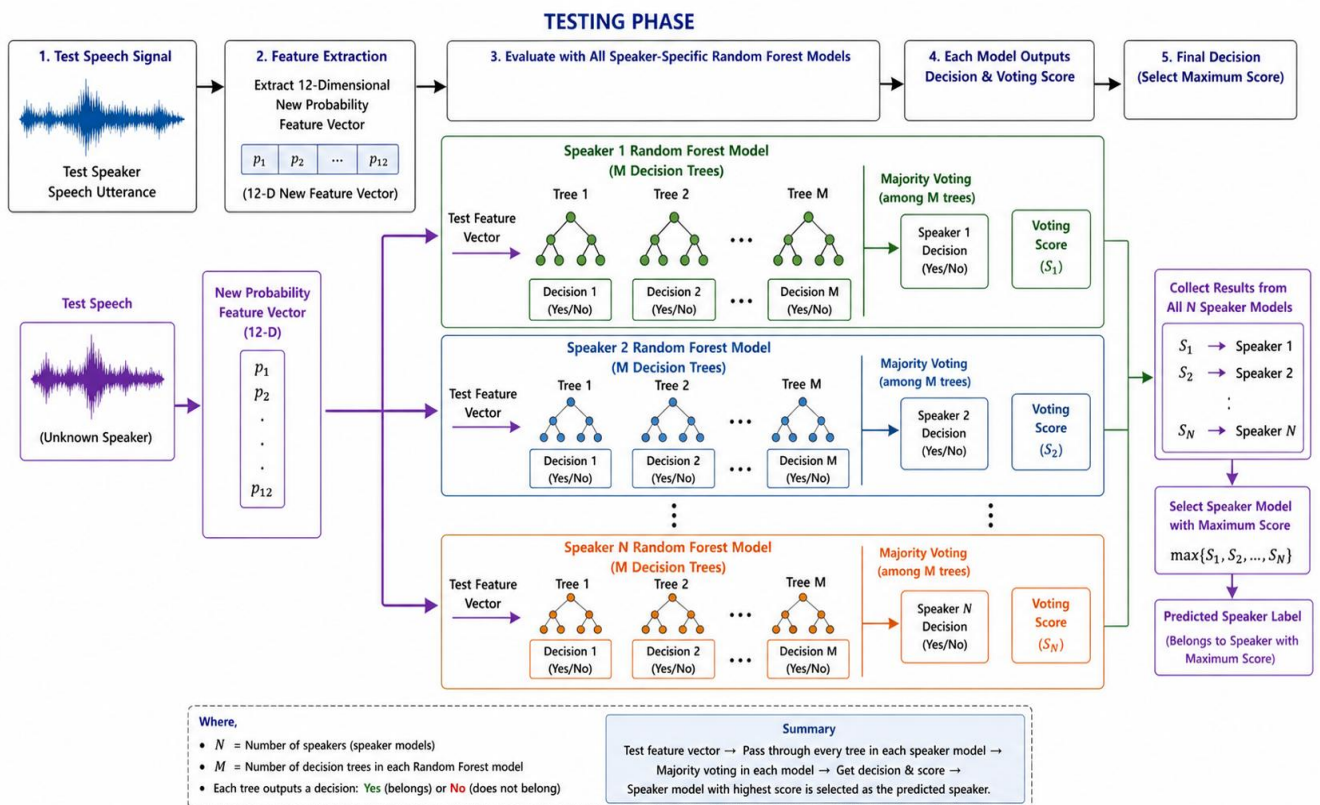


Fig3: Testing phase of speaker using New feature vectors

## 4. EXPERIMENTAL EVALUATION

### 4.1 Database used for the Study

TIMIT acoustic-phonetic corpus is used for training and testing. The data set composed of 630 speakers data. Here 200 speakers data used for the new feature vectors based speaker recognition system. For training 80% data and testing 20% data is used.

### 4.2 Result Analysis

The table1 shows comparison of two types speaker recognition systems performance, first one namely is speaker recognition system based on MFCC feature vectors and second one is proposed speaker recognition based on new feature vectors. The new feature vector-based speaker recognition system outperformed the traditional MFCC vectors-based speaker recognition system. This is because new form of feature vectors effectively captured speaker specific discrimination information effectively [21].

**Table 1. Comparing the two different types of feature vector-based speaker recognition system accuracy using Random Forest**

Speaker Recognition	Training accuracy	Testing accuracy
Random Forest using MFCC features	91%	87%
Random Forest using new probability features	100%	97%

### Conclusion

In this paper, new feature vectors based speaker recognition system using Random Forest is proposed. From speaker speech signal, MFCC feature vectors are extracted. From these MFCC feature vectors, new form feature vectors are extracted. These new feature vectors effectively capture speaker specific discriminative information.

By using these new feature vectors with random forest speaker recognition system achieved 97% accuracy. The performance of new feature vector based speaker recognition system is outperformed compare to conventional speaker recognition systems based on MFCC feature vectors. The newly proposed speaker recognition system is computationally efficient and reduce over fitting.

### Bibliography:

- [1] Douglas A. Reynolds, "Speaker recognition using Mel-frequency cepstral coefficients," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1995.
- [2] J. Saastamoinen, E. Karpov, V. Hautamäki, and P. Fränti, "Accuracy of MFCC-Based Speaker Recognition in Series 60 Device," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 17, pp. 2816–2827, 2005.

- [3] Asutosh Das, Manas Ranjan Jena, and Kalyan Kumar Barik, "Mel-frequency cepstral coefficient (MFCC) – A novel method for speaker recognition," *Digital Technologies*, 2015.
- [4] P. Kumar and S. L. Lahudkar, "Automatic speaker recognition using LPCC and MFCC," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, no. 4, pp. 2106–2109, 2015.
- [5] K. A. Lee, B. Ma, and H. Li, "Speaker verification using GMM and MFCC features," *IEEE Signal Processing Letters*, vol. 15, pp. 523–526, 2008.
- [6] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 19–41, 2000.
- [7] M. Rahul, S. K. Jha, A. Prakash, S. Verma, and V. Yadav, "Multilingual Speaker Recognition Using Mel-Frequency Cepstral Coefficients and Gaussian Mixture Model," 2024.
- [8] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308–311, 2006.
- [9] William M. Campbell, Douglas E. Sturim, Douglas A. Reynolds, and Alex Solomonoff, "SVM Based Speaker Verification Using a GMM Supervector Kernel and NAP Variability Compensation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006, pp. 97–100.
- [10] Alex Solomonoff, William M. Campbell, and Ian Boardman, "Advances in Channel Compensation for SVM Speaker Recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2005, pp. I-629–I-631.
- [11] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [12] Z. Bai and X.-L. Zhang, "Speaker recognition based on deep learning: An overview," *Neural Networks*, vol. 140, pp. 65–99, 2021.
- [13] Deep Neural Networks for Small Footprint Text-Dependent Speaker Verification, E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. Gonzalez-Dominguez, "Deep Neural Networks for Small Footprint Text-Dependent Speaker Verification," in *Proceedings of IEEE ICASSP*, 2014, pp. 4052–4056.
- [14] Z. Ji, G. Cheng, T. Lu, and Z. Shao, "Speaker Recognition System Based on MFCC Feature Extraction and CNN Architecture," *Academic Journal of Computing & Information Science*,
- [15] T. Kinnunen and H. Li, "An Overview of Text-Independent Speaker Recognition: From Features to Supervectors," *Speech Communication*, vol. 52, no. 1, pp. 12–40, 2010.

- [16] Karen Livescu, Karel Veselý, and Lukáš Burget, "Feature Learning for Speaker Recognition Using Deep Neural Networks," in *Proceedings of Interspeech*, 2011.
- [17] M. K. Nandwana, Jean-Luc Gauvain, and Laurent Lamel, "Investigating Random Forest for Speaker Verification in Diverse Conditions," in *Proceedings of Interspeech*, 2013.
- [18] S. R. M. Prasanna and P. V. S. Rao, "Speaker Identification Using Random Forest Classifier," *International Journal of Speech Technology*, vol. 15, no. 4, pp. 449–457, 2012.
- [19] R. Chakroborty and A. Roy, "Text-Independent Speaker Identification Using Random Forests," in *Proceedings of the National Conference on Communications*, 2011.
- [20] A. K. Paul, P. C. Shill, and M. A. Hossain, "Speaker Identification Using Mel Frequency Cepstral Coefficients and Random Forest Classifier," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 9, pp. 90–95, 2015.
- [21] Priyanka Bansal and Syed Akhtar Imam, "Performance of speaker recognition system using shifted MFCC and fuzzy techniques," *International Journal of Engineering and Technology*, 2018.