International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 02, Issue – 04, April – 2017, PP – 25-30

Singer Identification using MFCC and LPC and its comparison for ANN and Naïve Bayes Classifiers

Kumari Rambha Ranjan, Kartik Mahto, Dipti Kumari,S.S.Solanki

Dept. of Electronics and Communication
Birla Institute of Technology, Mesra
Ranchi, India

Abstract: Singer identification is one of the challenging tasks in Music information retrieval (MIR) category. Indian Bollywood has a rich culture of music and every movie on average consists of five songs sung by different singers. The revenue generated by music in India generates 4-5% of net revenue for a movie. This paper focuses on the singer identification using MFCC and LPC coefficients from Indian audio songs. The audio songs used are divided into segments each of 10 seconds and for each segments 13 Mel-frequency cepstral coefficients (MFCC) and 13 linear predictive coding (LPC) coefficients are computed. Classifier models are trained using Naive Bayes classifier and back propagation algorithm using neural network.

Keywords: LPC; MFCC; MIR; Neural Network

I. Introduction

An audience usually has a favorite singer and they might be interested in listening to only the songs sung by him. Hence, it would be desirable to create a system which will classify the given audio files as per singer's voice. The proposed paper works in this field of singer identification by extracting features from it to train a model. Such system can be used for automatic labeling of unknown CDs and records from a large database. It can also be used for copyright protections. The theory behind the paper is similar to that used in speaker recognition where a system is first trained with features of the entire speakers and then a test data is input to the model. Depending on different outcome parameters, results are generated. Both MFCC and LPC features are widely used in the field of speaker recognition and the same has been used to test its effectiveness in singer recognition [1,2,3,5]. For classification, the input data is tested with both ANN and Naïve Bayes Classifier and their results are compared.

II. Block Diagram

In this section, block diagram of the approach has been discussed.

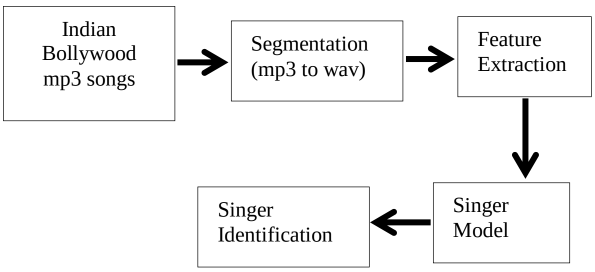

Fig.1. Block Diagram of System

Figure 1.shows the block diagram of the system. The first step was database preparation where 10 different songs of 10 different Bollywood singers were collected from various sources. Second step was to segment these audio files to small clips where the whole file was manually heard and corresponding portion of the song was saved which was sung by the relevant singers. 40 clips each of 10 seconds was created for each singers making a total of 400 samples for 10 singers. Out of the 400 samples, 300 samples were used to train the system and rest 100 samples were used for testing.

These clips were later used for feature extraction and the features were then used for modeling the system. Table I shows the list of singer used in the dataset.

III. Feature Extraction

A. MFCC

In sound processing, Cepstral feature is widely used. Among various Cepstral features, MFCC is the most effective one. MFCC is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. The Coefficients of MFC are collectively referred to as Mel-frequencyCepstral Coefficients (MFCCs). MFCCs are Cepstral coefficients used for representing audio in a way that mimics the physiological properties of the human auditory system.

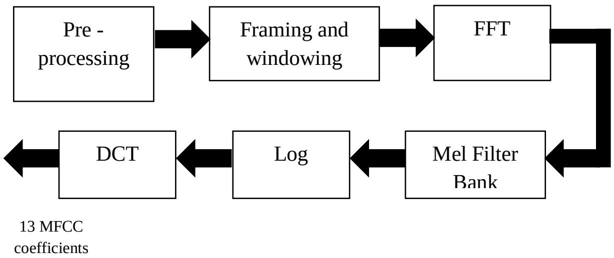

Fig.2. MFCC Feature Calculation

Block diagram of MFCC calculation is shown in figure 1.
Steps involved in its calculation are as follows:

- Preprocessing of the wav-file is the first step. It includes pre-emphasis, normalization and dc offset removal.
- The Fourier transform of a windowed and framed signal is calculated.
- Triangular overlapping windows are used to map the power of the spectrum obtained above onto the Mel scale.
- Logs of the powers at each of the Mel frequencies are found.
- Discrete cosine transform (DCT) of the list of Mel log powers is taken.
- The MFCCs are the amplitude of the resulting spectrum. First thirteen coefficients are saved.

Conversion from linear frequency scale to the Mel scale frequency M_f is achieved using the following equation

$$m_f = 2595\, log_{10}(1 + \frac{f}{700}) \qquad (1)$$

Where, f is frequency in hertz in linear scale.

B. LPC

LPC is a model for speech signal production based on the assumption that the speech signal is produced by a very specific model [5].All LPC variants are based on the same simple model of an excitation signal and a filter.LPC determines the coefficients of a forward linear predictor by minimizing the prediction error in the least squares sense. It has applications in filter design and speech coding.
A closer inspection of this system shows that speech can be modeled as a p^{th} order autoregressive process, where the present sample, x(k) depends on the linear combination of past p samples added with a stochastic or random component that represents noise. In other words, it is an all-pole FIR filter with Gaussian noise as input.

Where a_i are the linear prediction coefficients (LPCs) and u(k), the process noise, is a zero-mean Gaussian noise with variance σ^2_u .

Table I. List of Singers

Sl. No.	Singers
1	AshaBhosle
2	NehaKakkar
3	Manna Dey
4	ShreyaGoshal
5	Rafi
6	Kavita Krishnamurthy
7	Mukesh Kumar
8	Kishore Kumar
9	Sonu Nigam
10	Surraiyya

IV. Classfiers

A. Artificial Neural Network

Artificial Neural Network mimics the working of the neural network present inside our brains. There are three stages of ANN, input layer, hidden layer and output layer. Training data is fed to the input layer. During testing phase, weights of the neurons are adjusted to match the output. Back-propagation algorithm is used for weight adjustment where the error from output layer is slowly propagated to the input layer updating each weight values.

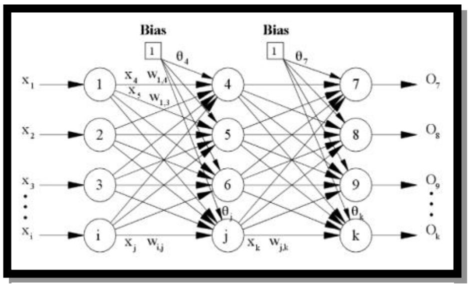

Fig. 3. Architecture of ANN

Simulation of ANN in MATLAB requires two input matrix: Feature Matrix (MFCC, LPC) and Target Matrix.

B. Naïve Bayes Classifier

Bayesian Classifier is based on probabilistic model where posterior probablity of an output is calculated given a set of observation when likelihood and prior probablity is known.

where $P(X_1,....,X_n/Y)$ is the likelihood probablity , $P(Y)$ is prior probablity and $P(Y/X_{1...},,X_n)$ is posterior probablity.

$$P(Y|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y)P(Y)}{P(X_1,\ldots,X_n)}$$

MATLAB Function “FitNaiveBayes” is used to create a model with all the prior probabilities and “Predict” function is used to test the model.

V. Simulation & Results

In this section, the simulation steps and results are discussed.
Following steps were involved in designing the ANN classifier on MATLAB.

- Create Input matrix (MFCC, LPC) and Target Matrix.
- Select number of hidden layer and type of ANN i.e. feed-forward or pattern network
- Select the percentage data for training and testing.
- Train the model and observe result for test data.

Following steps were involved in designing the Naïve Bayes classifier on MATLAB.

- Create Input matrix(MFCC,LPC) and Target Class.
- Use fitNaiveBayes command to train the model. The system will estimate the prior probabilities from the input matrix file and map it to particular target.
- Use “predict” function to input test data into the generated model and observe the classified output.

The system was trained and tested with different combinations of feature vectors and classifiers and different results were observed. The following combination wereused:

- LPC Feature + Naïve Bayes Classifier
- LPC Feature + ANN Classifier
- MFCC Feature + Naïve Bayes Classifier
- MFCC Feature + ANN Classifier

The efficiency of each of the combination is tabulated in Table IV and a comparative chart is shown in Figure 4.

Table IV : Efficiency for different Combinations of Feature and Classifier

Feature	Classifier	Efficiency
LPC Feature	Naïve Bayes	53%
LPC Feature	Neural Network	44%
MFCC Feature	Naïve Bayes	77%
MFCC Feature	Neural Network	71%

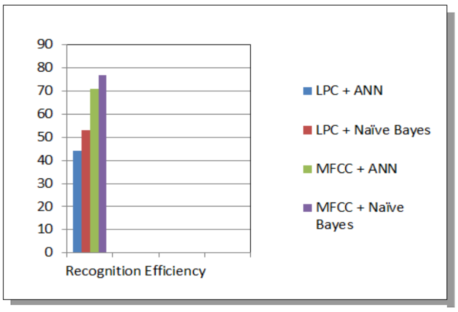

Fig. 4. Comparison Chart

Table I. LPC Feature + ANN Classifier(Hidden layer : 50)

	Asha	Kavita	KD	MD	Mukesh	NK	Rafi	SG	Sonu	Sur
Asha	5	1	2	0	1	0	0	0	0	1
Kavita	1	6	0	0	0	0	0	2	0	0
KD	2	0	3	1	0	0	3	1	0	0
MD	2	0	0	4	2	1	1	0	0	1
Mukesh	0	0	0	4	4	0	1	0	0	1
NK	0	0	0	0	0	5	0	3	2	0
Rafi	1	0	0	4	0	0	4	0	0	1
SG	0	1	0	0	2	0	1	5	1	0
Sonu	1	0	2	0	0	1	0	1	5	0
Sur	2	0	0	1	2	0	2	0	0	3

Table II. MFCC Feature + ANN Classifier(Hidden layer : 50)

	Asha	Kavita	KD	MD	Mukesh	NK	Rafi	SG	Sonu	Sur
Asha	**10**	0	0	0	0	0	0	0	0	0
Kavita	1	**7**	0	0	1	0	0	1	0	0
KD	0	0	**6**	0	0	0	2	0	2	0
MD	0	0	0	**7**	0	0	1	0	2	0
Mukesh	0	0	0	0	**9**	0	1	0	0	0
NK	2	0	0	0	0	**5**	0	2	1	0
Rafi	0	0	0	1	1	0	**6**	0	1	0
SG	1	2	0	0	0	0	0	**6**	1	0
Sonu	0	0	0	0	0	0	0	0	**10**	0
Sur	1	0	0	0	1	1	0	0	1	**6**

Table III. MFCC Feature + Naive Bayes Classifier

	Asha	Kavita	KD	MD	Mukesh	NK	Rafi	SG	Sonu	Sur
Asha	**6**	2	0	1	1	0	0	0	0	0
Kavita	0	**7**	0	0	0	0	0	3	0	0
KD	0	0	**8**	1	0	0	0	0	1	0
MD	0	0	0	**10**	0	0	0	0	0	0
\Mukesh	0	0	0	1	**8**	0	0	1	0	0
NK	0	0	1	0	0	**8**	0	0	1	0
Rafi	0	0	0	2	1	0	**7**	0	0	0
SG	0	1	0	0	0	1	0	**8**	0	0
Sonu	0	0	0	0	0	1	0	0	**9**	0
Sur	2	0	0	1	1	0	0	0	0	**6**

Table IV. LPC Feature + Naive Bayes Classifier

	Asha	Kavita	KD	MD	Mukesh	NK	Rafi	SG	Sonu	Sur
Asha	**2**	1	4	1	0	0	0	0	1	1
Kavita	1	**7**	0	0	0	1	0	1	1	0
KD	0	0	**4**	2	0	0	2	0	1	1
MD	0	0	0	**5**	5	0	0	0	0	0
Mukesh	0	0	0	1	**7**	0	0	0	0	2
NK	0	2	1	0	0	**5**	0	2	0	0
Rafi	0	0	1	3	3	0	**3**	0	0	0
SG	1	3	1	0	0	2	0	**2**	1	0
Sonu	0	0	2	0	0	1	0	0	**7**	0
Sur	0	0	0	0	0	0	1	0	0	**9**

Table I-IV shows the confusion matrix for singer identification for different combinations of features and classifiers. Each of the classifiers was trained with 30 samples of each singer making a total of 300 samples for training and was tested with 10 samples.

VI. Conclusion

The results obtained were observed and analyzed. MFCC features proved to provide a better result as compared to LPC for both the classifiers. Best results were observed for the combination of MFCC and Naïve Bayes Classifier with an identification percentage of 77%.

References

[1]. Salim, A., GowriSankar, "A novel method for Text-Independent speaker identification using MFCC and GMM," Proc. IEEE, Audio Language and Image Processing (ICALIP), Shanghai, Nov. 2010.

[2]. Ahmad Al Marashli, Dr. Oumayma Al Dakkak, "Automatic, Text-Independent, Speaker Identification and Verification System Using Mel Cepstrum and ANN," Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 3rd International Conference, 978-1-4244-1751-3, Apr. 2008.

[3]. "Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques," Proc. IEEE, 978-1-4577-1326-2, Cholula, Puebla, Feb. 2012.

[4]. O Lartiloot, P Toiviainen and T Eerola, "A Matlab Toolbox for Music Information Retrieval," Springer, pp.261-268, 2008.

[5]. B.S Atal. Speech analysis and synthesis by linear prediction of the speech wave. Journal of the Acoustical Society of America, 47(1):65, 1970.