International Journal for Modern Trends in Science and Technology
Volume: 03, Special Issue No: 01, February 2017
ISSN: 2455-3778
http://www.ijmtst.com

Some Studies on of Raaga Emotions of Singers Using Gaussian Mixture Model

B.Tarakeswara Rao[1] | P.Lakshmikanth[2] | E.Ramesh[3]

[1]Professor, Department of CSE, KHIT, Guntur, Andhra Pradesh, India.
[2]Assistant Professor, Department of CSE, KHIT, Guntur, Andhra Pradesh, India.
[3] Assistant Professor, Department of CSE, RVR&JC College of Engineering, Guntur, Andhra Pradesh, India.

To Cite this Article
B.Tarakeswara Rao, P.Lakshmikanth and E.Ramesh, "Some Studies on of Raaga Emotions of Singers Using Gaussian Mixture Model", International Journal for Modern Trends in Science and Technology, Vol. 03, Special Issue 01, 2017, pp. 149-153.

ABSTRACT

Raagas are the heart of Indian classical music. Raaga is one of the fundamental musical concepts in Indian music. Raaga plays an important role in Indian classical music. Identification of raaga emotions from speech has gained immense attention recently. With the increasing demand for human computer interaction, it is necessary to understand the raaga emotional state of the singer. In this paper an attempt has been made to recognize and classify the raaga emotions from singers Database. Here the classification is mainly based on extracting several key features like Mel Frequency Cepstral Coefficients (MFCCs) from the speech signals of those persons by using the process of feature extraction. For training and testing of the method, data collection is carried out from singers. It consists of acted speeches of one short raaga, emotionally biased sentence repeated 5 times with different states by 52 melakartha raagas for training and another 10 melakartha raagas for testing. The experiments were performed pertaining to singer raagas. Using a statistical model like Gaussian Mixture Model classifier (GMM) and features extracted from these speech signals. We build a unique identity for each raaga emotion that enrolled for raaga emotion recognition. Expectation and Maximization (EM) algorithm, an elegant and powerful method is used with latent variables for finding the maximum likelihood solution, to test the other raaga emotions against the database of all melakartha raagas in the database.

KEYWORDS: *Raaga emotion Recognition, Gaussian Mixture Model (GMM) classifier, Sequential Forward Selection, EM algorithm, Mel Frequency Cepstral Coefficients(MFCCs).*

I. Introduction

Identification of raaga emotion with a machine has been a part of active research in recent times. Effective raaga emotion identification system will help to make the interaction between human and computer more natural. It has its applications in many areas such as education, movies and cultural events. Every singer has his own style based on elements like articulation rate, pitch, energy, amplitude, diction etc. The temporal and spectral features of the individuals based on amplitude, pitch, formants, long term spectral features and short term spectral features are given as inputs to the classification algorithm [1,12]. The raaga emotion expressed by an individual depends on culture, language, social status, area, education and gender. A lot of research is projected towards emotion recognition based on Support Vector Machines (SVM) and Hidden Markov Models (HMM). However, there is no literature available to detect the raaga emotion from singers using Gaussian Mixture Models (GMM). Hence in this paper an attempt has been made to identify the raaga emotions of singers using Gaussian Mixture Model and Expectation Maximization algorithm to recognize the raaga emotional states viz., Anger, Surprise, Happiness, Sadness and Neutral. The singer dependent and independent experiments were conducted on raaga emotional speech database of singers.

II. RAAGA EMOTION IDENTIFICATION SYSTEM

It contains four modules: Raaga emotional speech database, feature Extraction, GMM model with EM algorithm and identified Raaga emotion output.

The Raaga emotional speech database used in this paper is from singers of Andhra Pradesh, India. The speech is expressed in five raaga emotional states: Anger, Surprise, Happiness, Sadness and Neutral. In this database, the wave file can hold compressed audio, the most common '.wav' format contains uncompressed audio in the pulse code modulation (PCM) **[2]** format.

2.1 Feature extraction

To identify a raaga emotion of a singer, features such as MFCCs **[3]**, rate of speech, pitch or some of the essential features, out of which, in our paper we have used MFCCs. In the process of my Research, to analyze raaga emotional speech, that indicates the fundamental frequency, energy and formant frequencies with amplitude are potentially effective parameters to distinguish certain raaga emotional states. In this study, five groups of short-term features that were extracted relate to fundamental frequency (F0), energy, the first four formant frequencies (F1 to F4), two Mel Frequency Cepstrum Coefficients (MFCC1, MFCC2). MELCEPST is used to calculate the Mel Cepstrum of a signal C=(S, FS, W, NC, P, N, INC, FL, FH).

200 features, including the five groups of features and their first and second derivatives, were extracted as the candidate input and these values were considered as the initial values which are given to the EM algorithm. The final estimates are obtained by using the EM algorithm which consists of two steps: 1) E-step and 2) M-step. The final estimates were obtained from the EM algorithm and given as inputs to the GMM model. Each of the extracted features was linearly scaled to the range of [0, 5000] to avoid having values too large or too small.

Spectral energy dynamics provides another possible indicator of the singer raaga emotional states. A novel parameter vector called the Mel Energy spectrum Dynamic Coefficients (MEDC) is proposed to distinguish the five raaga emotional states. It was extracted as follows: the magnitude spectrum of each speech utterance was estimated using FFT, then input to a bank of N filters equally spaced on the Mel frequency scale. The logarithm mean energies of the N filter outputs were calculated $(En(i), i=1, \ldots\ldots, N)$. Then, the first and second differences of $En(i), i=1, \ldots\ldots, N$ were computed.

$$\Delta En(i) = En(i+1) - En(i),\quad i = 1,\ldots,N-1 \qquad (1)$$

$$\Delta^2 En(j) = \Delta En(j+1) - \Delta En(j),\quad j = 1,\ldots,N-2 \qquad (2)$$

The final Mel Energy spectrum Dynamic Coefficients were then obtained by combining the first and second differences:

$$MEDC = [\Delta En(1) \ldots \Delta En(N-1) \quad \Delta^2 En(1) \ldots \ldots \Delta^2 En(N-2)] \qquad (3)$$

The value of N was set to 12 in this study, and the coefficients were linearly scaled to the range of [0, 1] before being input to the classifier.

Raaga emotional features like Anger, Surprise, Happiness, Sadness and Neutral states are extracted and the best feature among this is selected. Mat-lab7 [4,14] was used to train this feature. The trained data was given as input to GMM classifier. Finally to get the accuracy of singer raaga emotions were determined.

III. MEL-FREQUENCY CEPSTRAL COEFFICIENTS

MFCCs are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip ("spectrum-of-a-spectrum"). The difference between the cepstrum and the Mel-Frequency Cepstrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

MFCCs are to be derived by following steps, as below **[5]:**

1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the Mel scale, using triangular overlapping.
3. Take the logs of the powers at each of the Mel frequencies.
4. Take the discrete cosine transform of the list of Mel log powers, as if it were a signal.
5. The MFCCs are derived as the amplitudes of the resulting spectrum **[6].**

MFCC values are not very robust in the presence of additive noise, and so some researchers propose modifications to the basic MFCC algorithm to account for this example by raising the log-Mel-amplitudes to a suitable power (around 2 or 3) before taking the DCT (Direct Cosine Transform), which reduces the influence of low-energy components **[7].**

IV. Gaussian Mixture Model

Gaussian mixture model **[8]** is a type of density model which comprises a number of component Gaussian functions. These component functions are combined with different weights to result in a multi-modal density. Gaussian mixture models are a semi-parametric alternative to non-parametric histograms (which can also be used to approximate densities) and it has greater flexibility and precision in modeling the underlying distribution of sub-band coefficients.

Gaussian Mixture density is weighted sum of M component densities and can be expressed:

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \qquad (4)$$

Where $\vec{x}$ is D dimensional vector, p_i is the component weight, $bi(\vec{x})$-component densities, that can be written:

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} \left|\sum_i\right|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)' \sum_i^{-1} (\vec{x}-\vec{\mu}_i)} \qquad (5)$$

where μ_i - mean vector, $\sum_i$ - covariance matrix.

Mixture weights must satisfy constraint:

$$\sum_{i=1}^{M} p_i = 1 \qquad (6)$$

Gaussian mixture density is parameterized by the mean vectors, covariance matrices and mixture weights. All these parameters are represented by notation:

$$\lambda = \{p_i, \mu_i, \sum_i\} \quad i = 1,2,\ldots,M. \qquad (7)$$

Hence, each singer is represented by his/her GMM and is referred by his/her model λ.

The other task is to estimate the parameters of GMM λ, which best matches the distribution of the training feature vectors, given by speech of the singer. There are several available techniques for GMM parameters estimation **[9]**. The most popular method is maximum likelihood (ML) estimation **[10]**. The basic idea of this method is to find model parameters which maximize the likelihood of GMM. For a given set of T training vectors $X=\{\vec{x}_1,\ldots,\vec{x}_T\}$ GMM likelihood can be written:

$$p(x \mid y) = \prod_{t=1}^{T} p(\vec{x}_t \mid \lambda) \qquad (8)$$

V. EM Algorithm

An Expectation-Maximization [11,13] (EM) algorithm is used in statistics for finding maximum likelihood estimates of initial parameters (in probabilistic models, where the model depends on unobserved latent variables. EM alternates between performing an Expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and Maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E step. The parameters found on the M step are then used to begin another E step, and the process is repeated.

ML parameter estimates can be obtained iteratively using special case of Expectation-Maximization (EM) algorithm. There the basic idea is, beginning with initial model λ, to estimate a new model $\bar{\lambda}$, that $p(X \mid \bar{\lambda}) \geq p(X \mid \lambda)$. The new model then becomes the initial model for the next iteration. This process is repeated until some convergence threshold is reached.

On each iteration, following re-estimation formulas are used: mixture weights are recalculated

$$\bar{p}_i = \frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda) \tag{9}$$

Means are recalculated

$$\vec{\bar{\mu}} = \frac{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)\,\vec{x}_t}{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)} \tag{10}$$

Variances are recalculated

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)\,(x_t - \mu_i)^2}{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)} \tag{11}$$

VI. Results and Discussion

There are four main modules in this paper. They are extraction of raaga emotional features, and training the features using GMM classifier, training GMM through multiple raaga emotions data, testing the entire raaga emotions using both neutral and raaga emotional features. The results of classification obtained through both the features are combined to produce more accurate results. The regions commonly identified in both the classification results are now highlighted. In this paper, we discus the GMM classifier, a novel parameter vector called the Mel Frequency Cepstral Coefficients (MFCCs) that was implemented to distinguish the five raaga emotional states. After the values were obtained raaga emotion features like Anger, Surprise, Happiness, Sadness and Neutral dissimilarity were obtained. The various raaga emotional features that were extracted are shown below.

Classified values obtained from tested raagas :

Name of the Raaga	Recognized Raaga Emotions				
	Anger	Surprise	Happiness	Sadness	Neutral
Paripalayamam	Y	-	-	-	-
Pavana Thanaya	-	Y	-	-	-
Jhashakethana	-	-	Y	-	-
Palayamam	-	-	-	Y	
Vadamela	-	-	-	-	Y

Table1. Confusion matrix indicating different raaga emotions of singer.

Classified values obtained from tested raagas :

Stimulation	Recognized Raaga Emotions (%)				
	Anger	Surprise	Happiness	Sadness	Neutral
Anger	60	0	10	0	0
Surprise	0	56	20	0	0
Happiness	10	0	58	0	10
Sadness	10	10	0	60	10
Neutral	10	0	10	0	55

Table2. Confusion matrix indicating recognized raaga emotions of singer.

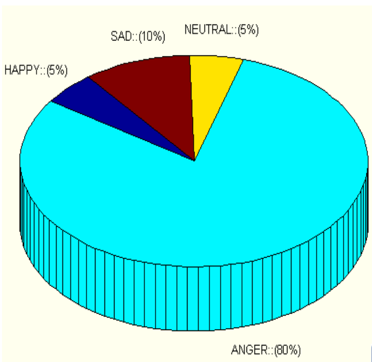

Figure1. Pie-Chart indicating different raaga emotional recognitions

VII. CONCLUSION

In this paper, we discuss the method based on Gaussian Mixture Model classifier and MFCCs as features for raaga emotion identification system that was developed. For this, we have considered the raaga emotions from singers, to detect 5 basic raaga emotions. The identified raaga emotions are presented in a confusion matrix table based on samples collected from the singers. As indicated in our experimental results, we achieved a high degree of accuracy of nearly 80% in identifying Anger raaga emotion and more than 60% accuracy in other raaga emotions. The GMM classifier achieved better performance in identifying Anger raaga emotion and differentiating the other raaga emotions perfectly.

REFERENCES

[1] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G., "Emotion recognition in human-computer interaction", IEEE Signal Processing magazine, Vol. 18, No. 1, pp. 32-80, Jan. 2001.

[2] Inger. S. Engberg, Anya. V. Hansen, "Documentation of the Danish Emotional Speech database DES", Aalborg, Sept 1996.

[3] Rabiner. L, Juang B. H, *Fundamentals of speech recognition*, Chap. 2,pp. 11-65, Pearson Education, First Indian Reprint, 2003.

[4] L. Arslan, Speech toolbox in MAT LAB, Bogazici University, http://www.busim.ee.boun.edu.tr/~arslan/

[5] Min Xu et al. (2004). "HMM-based audio keyword generation". in Kiyoharu Aizawa, Yuichi Nakamura, Shin'ichi Satoh. Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer.

[6] Fang Zheng, Guoliang Zhang and Zhanjiang Song (2001), "Comparison of Different Implementations of MFCC," J. Computer Science & Technology, 16(6): 582–589.

[7] V. Tyagi and C. Wellekens (2005), On desensitizing the Mel-Cepstrum to spurious spectral components for Robust Speech Recognition , in Acoustics, Speech, and Signal Processing, 2005. Proceedings (ICASSP '05). IEEE International Conference on, vol. 1, pp. 529–532.

[8] D. A. Reynolds and R. *C.* Rose,"Robust Text-Independent Speaker Using Gaussian Mixture Speaker Models", *IEEE Trans* Speech Audio Process, Vol. 3, No. 1. Jan. 1995, pp. 72 – 83.

[9] **McLachlan G.** Mixture Models. – New York: Marcel Dekker, 1988.

[10] **Dempster A., Laird N., and Rubin D.** Maximum likelihood from incomplete data via the EM algorithm // J. Royal Srar. Soc. – 1977. – Vol. 39. – P. 1–38.

[11] J.Kamarauskas, "Speaker Recognition using Gaussian Mixture Models", ISSN 1392 –1215, 2008. No. 5(85).

[12] B. Tarakeswara Rao et all, "Automatic Melakarta Raaga Identification Syste: Carnatic Music " IJARAI, Vol. 1, No. 4, 2012

[13] Prasad Reddy P.V.G.D, B. Tarakeswara Rao et all, "Automatic Raaga Identification System for Carnatic Musi Using Hidden Markov Model", Online ISSN: 0975-4172 & Print ISSN: 0975-4350 Volume 11 Issue 22 Version 1.0 December 2011.

[14] Prasad Reddy P.V.G.D, B. Tarakeswara Rao et all," K-Nearest Neighbour and Earth Mover Distance for Raaga Recognition", International Journal of Computer Applications (0975 – 8887) Volume 33– No.5, November 2011.