ISSN No. 0976-5697

Volume 1, No. 3, Sept-Oct 2010

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Recognition of Melakartha Raagas with the Help of Gaussian Mixture Model

Tarakeswara Rao B*
Assistant Professor, School of Computing
Vignan University, Guntur, India
Tarak7199@gmail.com

Dr. Prasad Reddy P.V.G.D
Professor
Department of Computer Science and Engineering
Andhra University, Visakhapatnam, India
prasadreddy.vizag@gmail.com

Prasad A
Associate Professor, School of Computing
Vignan University, Guntur, India
prasadjkc@yahoo.co.in

Abstract: Recognizing Melakartha raagas from speech has gained immense attention recently. With the increasing demand for human computer interaction, it is necessary to understand the state of the singer. In this paper an attempt is made to recognize and classify the raagas from the singers database where 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 is collected from the existing database with due verification relating to melakartha raagas. The 72 melakartha raagas for training, of them, a few raagas were specifically selected and tested. Then it is found that all the tested raagas are well recognized. In another case the 52 melakartha raagas for training and another 20 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 that enrolled for raaga 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 raagas against the database of all singers who enrolled in the database.

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

I. INTRODUCTION

The Raaga system is a method of organizing tunes based on certain natural principles. Indian classical music is defined by two basic elements. It must follow a classical mode and a specific rhythm. Each Raaga has a musical form which is otherwise known as image. The feature of Raaga usually contains the “aarohanam”, “avarohanam” phrases and general usage notes. It is intended more for the performer than for the listener. In this connection one must remember that Note Transcription is the first step in Raaga identification.

The other significant factor of Raaga identification of Carnatic music, for music information retrieval. It can be thought of as a part of multimedia information retrieval. In this paper, we discuss music processing which could be used as the basis of music information retrieval and its characteristic features.

Today, the computational efficiency of computers permits the research community to deal with different aspects. Hence the musicologist has the semi-automated search of specific sound patterns with large number of stored sound files. These musical patterns have been shaped and categorized through practice and experience in several musical traditions. This study proposes a scheme for the recognition of such pre-defined musical patterns in a monophonic environment in the context of South Indian classical music.

Indian classical music is defined by two basic elements. They are primarily the Raaga which is classical mode and the Taala which is called rhythm. In otherwords, Raaga is a characteristic arrangement or progression of Notes.

A raaga is characterized by several attributes like its Vaadi – Samvaadi, Aarohana – avarohana and Pakad, besides the sequence of notes. The recognition scheme that we propose consists of three categories’ viz: the tempo-tracking stage, fundamental frequency algorithm. The output from the second category is given to a pre-trained Gaussian Mixture Models.

Recognizing raaga with a machine has been a part of active research in recent times. Effective raaga recognition 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 other cultural events. Every singer has his own style based on elements like articulation rate, pitch, energy, amplitude, speech rate, degree 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]**. The raaga expressed by an individual depends on his note, degree, pitch, and diction. A lot of research is projected towards raaga recognition based on Support Vector Machines (SVM) and Hidden Markov Models (HMM).

However, there is no literature available to detect the raagas from singers using Gaussian Mixture Models (GMM). Hence, in this paper, an attempt has been made to identify the raagas of singers using Gaussian Mixture Model and Expectation Maximization algorithm to recognize the process of raaga.

© 2010, IJARCS All Rights Reserved

II. RAAGA RECOGNITION SYSTEM

This system has four modules: Raaga database, feature Extraction, GMM model with EM algorithm and recognized Raaga output. The Raaga database used in this paper is from singers who rendered classical music . The raaga is expressed in various types of melakartha raagas. 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.

A. Feature extraction

To recognise a raaga 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. Research to analyze melakartha raagas that indicates the fundamental frequency, energy and formant frequencies with amplitude are potentially effective parameters to distinguish various types of raagas. 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 singers 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 raaga features . A novel parameter vector called the Mel Energy spectrum Dynamic Coefficients (MEDC) is proposed to distinguish between various types of raagas. It was extracted as follows: the magnitude spectrum of each raaga 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.

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

1

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

2

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

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

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 features like Note, Swara, Pitch, degree, Tala, and Diction are extracted and recognize the differences among raagas. Mat-lab7 **[4]** was used to train this feature. The trained data was given as input to GMM classifier. Finally, to get the accuracy of singer raagas were determined.

III. MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCCs)

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:

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

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:

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

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 raaga 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={$\vec{\boldsymbol{x}}_1,\ldots,\vec{\boldsymbol{x}}_T$ } GMM likelihood can be written:

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

An Expectation-Maximization **[11]** (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 the following re-estimation formulas are used:

Mixture weights are recalculated

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

Means are recalculated

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

Variances are recalculated

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

V. RESULTS AND DISCUSSION

There are four main modules in this paper. They are extracted indicators of raaga features, and training the features using GMM classifier, training GMM through multiple raaga data, testing the selected raagas by using the features of raaga. 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 discussed the GMM classifier, a novel parameter vector called the Mel Frequency Cepstral Coefficients (MFCCs) that was implemented to distinguish the various types of raagas.

We have trained 72 raagas. We have also tested the said 72 raagas all of them are recognized. Then, we have tested the raagas which are not in the train. After testing the entire process we got four different types of results. The same may be indicated under four categories.

Category I: A few raagas are recognized and it is upto 90%. However, the remaining 10% is difference between those recognized raagas in note, swara, degree, voice etc.

Category II: The raagas under second category, they are recognized only up to 80% and the remaining 40% are found to be at structural difference.

Category III: In respect of Category III, a few raagas are in the range of 60% to 79% and they also recognized. However, the left over raagas are still found to be at difference.

Category IV: The remaining are accounted for raagas of neutral category.

All the above said four category raagas were also recognized based on their note, degree, swara, pitch, diction etc.

The above mentioned results are given here under to elucidate its veracity.

Classified values obtained from Singer Raagas:

Table I. Confusion matrix indicating we have tested the raagas which are in the train.

Name of the Raaga	Recognized Raagas (%)				
	Rathangi	Nataka Priya	Keeravani	Pavani	Lathangi
Rathangi	**100**	84	60	65	70
Nataka Priya	60	**100**	70	75	80
Keeravani	68	72	**100**	85	75
Pavani	55	70	80	**100**	90
Lathangi	80	90	85	92	**100**

Table2. Confusion matrix indicating we have tested the raagas, which are not in the train.

List of Trained Raagas	Tested Raagas, Outside the Trained Set (%)				
	Kanakangi	Varunapriya	Sarasangi	Sucharitha	Rasikapriya
Rathangi	**90**	60	70	65	70
Nataka Priya	90	**70**	75	80	65
Keeravani	75	90	**80**	65	70
Pavani	80	85	80	**60**	64
Lathangi	70	73	90	75	**58**

VI. CONCLUSION

In this paper, we discuss the method based on Gaussian Mixture Model classifier and Mel Frequency Cepstral Coefficients as features for raaga recognition system that was developed. For this we have considered the raagas from singers that are under four categories. The recognized raagas are presented in a confusion matrix table based on samples collected from the singers. As detailed in our experimental results, we achieved a high degree of accuracy of nearly 95% considered to be as maximum recognized raagas. We find more than 70% accuracy in other raagas. The GMM classifier, thus, achieved a better performance in recognizing maximum raagas and differentiating the other raagas perfectly**.**

VII. 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] 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 Identification 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).

© 2010, IJARCS All Rights Reserved