# Rhythm and Timbre Analysis for Carnatic Music Processing

**Rushiraj Heshi, S.M. Suma, Shashidhar G. Koolagudi, Smriti Bhandari and K.S. Rao**

**Abstract** In this work, an effort has been made to analyze rhythm and timbre related features to identify *raga* and *tala* from a piece of Carnatic music. *Raga* and *Tala* classification is performed using both rhythm and timbre features. Rhythm patterns and rhythm histogram are used as rhythm features. Zero crossing rate (ZCR), centroid, spectral roll-off, flux, entropy are used as timbre features. Music clips contain both instrumental and vocals. To find similarity between the feature vectors T-Test is used as a similarity measure. Further, classification is done using Gaussian Mixture Models (GMM). The results shows that the rhythm patterns are able to distinguish different *ragas* and *talas* with an average accuracy of 89.98 and 86.67 % respectively.

**Keywords** Carnatic music · *Raga* · *Tala* · Gaussian Mixture Models · Rhythm · Timbre · T-test

R. Heshi (✉) · S. Bhandari
Walchand College of Engineering, Sangli, India
e-mail: rushiraj.heshi@walchandsangli.ac.in

S. Bhandari
e-mail: smriti.bhandari@yahoo.com

S.M. Suma · S.G. Koolagudi
National Institute of Technology, Karnataka Surathkal, India
e-mail: suma.cs13f02@nitk.edu.in

S.G. Koolagudi
e-mail: koolagudi@nitk.edu.in

K.S. Rao
Indian Institute of Technology, Kharagpur, India
e-mail: ksrao@sit.iitkgp.ernet.in



A. Nagar et al. (eds.), *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, Smart Innovation, Systems and Technologies 43, DOI 10.1007/978-81-322-2538-6_62

## 1 Introduction

Since large collection of multimedia data is available in digital form, we need to identify ways to make those collections accessible to users. Efficient indexing criteria makes retrieving/accessing an easy task. The techniques for indexing/accessing music are categorized into 3 types. The first two kinds are meta-data and text based access, use meta data and user provided tags for retrieval. However, these tags need not be correct always. Where as in content based Music Indexing and Retrieval (MIR), music signal is analyzed for deciding the genre, hence it is more effective than the previous two techniques. The fundamental concepts of Carnatic music are *raga* (melodic scales) and *tala* (rhythmic cycles). A *raga* in Carnatic music prescribes a set of rules for building a melody very similar to the western concept of mode [1]. *Raga* tells about the set of notes used and the way in which these notes are rendered. Technically a note is a fundamental frequency component of a music signal defined using starting and ending time [2]. A *tala* refers to a fixed time cycle or a metre, set for a particular composition, which is built from groupings of beats [1]. *Tala* has cycles of a defined number of beats and rarely changes within a song. Since *raga* and *tala* are the fundamental concepts, extracting these information from music signal helps in building efficient MIR systems.

In this work, an attempt has been made to analyze different features like rhythm and timbre for classification of *raga* and *tala*. Rhythm is the pattern of regular or irregular pulses caused in music by the occurrence of strong and weak melodic and harmonic beats [3]. Timbre describes those characteristics of sound, which allow the ear to distinguish sounds that have the same pitch and loudness which is related to the melody of the music [3].

Rest of the paper is organized as follows. A brief review of the past works and the issues are discussed in Sect. 2. The features extracted and classifier used are explained in Sect. 3. Section 4 explains the experiments conducted and the analysis of results. Section 5 concludes the work with some future research directions.

## 2 Related Work

In this section, different feature extraction approaches towards audio retrieval and classification have been discussed. Many of the works have used pitch derivatives as features for *raga* identification since pitch feature is related to melody of the music. In [4], Hidden Markov Model (HMM) is used for the identification of Hindustani ragas. The proposed method uses note sequence as a feature. Many micro-tonal variations present in the ICM make note transcription a challenging task even for a monophonic piece of music. Two heuristics namely Hillpeak heuristics and Note duration one try to overcome these variations. The limitation of this work is limited data set as it contains only two *ragas* and considerably lower accuracy of note transcription. Similar work has been carried out by Arindam et al. [5] using manual note transcription. The HMM evaluated for this sequence claimed to achieve 100 % accuracy, if the given note sequence is correct. However, it is difficult to achieve high accuracy in ICM transcription because of the micro-tonal variations and improvisations. P. Kirthika et al. introduced an audio mining technique based on *raga* and emphasized importance of *raga* in audio classification [6]. Individual notes are used as features. Pitch and timbre indices are considered for classification. Koduri et al. presented *raga* recognition techniques based on pitch extraction methods and KNN is used for classification [7]. Property of the tonic note with pitch of highest mean and least variation shown in pitch histogram, is used for identification of tonic pitch value [8]. Using Semi-Continuous Gaussian Mixture Model (SC-GMM), tonic frequency and *raga* of the musical piece are identified. Only 5 *Sampurna ragas* are used for validating this system. In [9], Rhythm patterns and Rhythm histogram are used as a feature for identifying and tagging songs. GMM is used for classification. In [10], timbre features such as spectral centroid, spectral roll-off, spectral flux, low energy features, MFCC and rhythmic features are used for classification of *raga* and *tala*.

From the literature, it is evident that many of the works have used features such as as pitch and its derivatives and set of note information for identifying *raga* and *tala*. In this work features other than pitch and note information are analyzed for classification.

## 3 Methodology

Figure 1 shown below represents the activities done while implementing the idea. Rhythm and timbre related features are extracted from each frame of the music clip. GMMs are trained using these features to model the training music clips on the basis if their rhythm and timbre. Further trained models are used to classify unknown test clips.

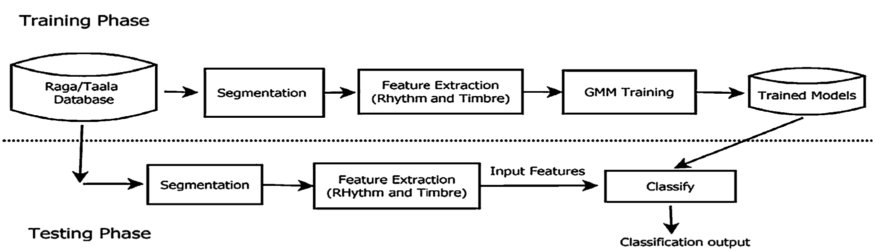


**Fig. 1** Schematic diagram to classify *raga* and *tala*

### 3.1 Features Extraction

**Rhythm Features**: Rhythm patterns and Rhythm histograms are extracted from the first 60 frames of the given music piece [3]. It results in a 24 × 60 matrix where 24 represents the critical bands of the Bark scale and 60 represents the number of frames (shown in Fig. 2a). The x-axis represents the rhythm frequency up to 10 Hz and the y-axis represents bark scale of 24 critical bands. From the rhythm patterns, rhythm histogram is obtained by adding the values in each frequency bin in the rhythm pattern. This results in a 60 dimensional vector representing the "rhythmic energy" of the corresponding modulation frequencies. In Fig. 2b x-axis represents the rhythm frequency up to 10 Hz and the y-axis represents the magnitude of respective frequency.

**Timbre Features**: Timbre related features such as ZCR, centroid, roll-off, flux, entropy are extracted from the signal. ZCR is the number of times signal crosses x-axis. Centroid determines frequency area around which most of the signal energy concentrates. Centroid is calculated using Eq. 1.

$$C_t = (M_t[n] * n)/M_t[n] \tag{1}$$

where $M_t[n]$ is the magnitude of the Fourier transform of frame $t$ and frequency bin $n$. Roll-off is used for finding out frequency such that certain fraction of total energy is contained below that particular frequency. The spectral roll-off is defined as the frequency $R_t$ below which 85 % of the magnitude distribution is concentrated and is calculated using Eq. 2.

$$M_t[n] = 0.85 * M_t[n] \tag{2}$$

The spectral flux is a measure of the amount of local spectral change. Flux is the distance between spectrum of two successive frames. It is calculated using Eq. 3.

$$F_t = (N_t[n] - N_t - 1[n]) \tag{3}$$

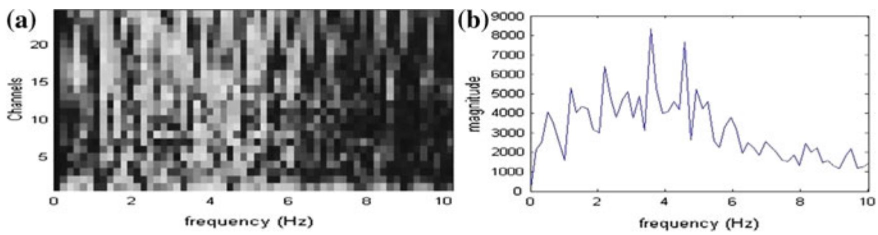


**Fig. 2** Rhythm features extracted from the music signal. **a** Rhythm patterns. **b** Rhythm histograms

where $N_t[n]$ and $N_t - 1[n]$ are the normalized magnitude of the Fourier transform of the current frame $t$, and the previous frame $t - 1$, respectively. Entropy is used to calculate randomness of the signal.

$$H(X) = -p(x_i) \log bp(xi) \tag{4}$$

where $p(x_i)$ is the probability mass function of outcome $x_i$. These features extracted from the signal are initially checked for similarity for each *raga* and *tala* class. Further are used for classification task.

### 3.2 Classifier

**T-Test:** Before developing a classifier model, the T-test is performed to determine whether the means of two groups are statistically different from each other. The result of test is 0 or 1. The output 0 implies T-Test is passed and there is no significant difference between two feature vectors. If output is 1 then T-test does not pass and there is significant difference between two feature vectors.

**GMM:** GMM is a mixture of Gaussian Distributions. Probability density function for mixture of Gausses is a linear combination of individual PDFs. A GMM is constructed for each class (*raga*/*tala*). Expectation Maximization algorithm is used for training GMM. In testing phase, the highest probability value (greater than 0.5) is used to decide the output class.

## 4 Experimentation and Results

### 4.1 Database

Two different audio datasets are collected for 10 *raga* and 10 *tala* considered for the study are given in Table 1. The music clips include both monophonic and polyphonic music and are rendered by different male and female singers. The dataset consists of 400 clips (20 clips in each type of *raga* or *tala*).

### 4.2 Performance Evaluation

6 sets of experiments are performed to evaluate the proposed method. Initial four experiments are conducted using T-test to validate rhythm and timbre on *raga* and *tala* datasets. Each music clip is compared with all the other music clips and the

**Table 1** Database: list of *raga*s and *talas* used

| Raga id | Raga name | Number of clips | Tala id | Tala name | Number of clips |
|---|---|---|---|---|---|
| 1 | Abhogi | 20 | 1 | Adi | 20 |
| 2 | Hamsadwani | 20 | 2 | Adi (2 kalai) | 20 |
| 3 | Hari Khambhoji | 20 | 3 | Desh Adi | 20 |
| 4 | Hindolam | 20 | 4 | Khanda Chapu | 20 |
| 5 | Kalyani | 20 | 5 | Khanda Ekam | 20 |
| 6 | Malyamarutha | 20 | 6 | Misra Chapu | 20 |
| 7 | Mayamalavagowla | 20 | 7 | Misra Jhampa | 20 |
| 8 | Mohanam | 20 | 8 | Rupakam | 20 |
| 9 | Nattai | 20 | 9 | Thisra Triputa | 20 |
| 10 | Shankarabharanam | 20 | 10 | Triputa | 20 |

similarity value (0 or 1) is recorded. The percentage of music clips that matches with the same class of clips is calculated. The values in the Table 2 show that rhythm features have better similarity than timbre features. Hence, rhythm features are considered for classification of *raga* and *tala*. Rhythm features of 60 dimensions from 14 music clips are used for GMM training and 6 music clips are used for testing from each *raga* and *tala*. The results in Table 3 show that the rhythm features are useful in classification and hence may be used as secondary features along with pitch related features for *raga* and *tala* identification.

**Table 2** Results of similarity test for rhythm and timbre features

| | | *Raga/tala* id | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Similarity for ragas (%) | Rhythm features | 100 | 66.6 | 66.6 | 87.5 | 88.8 | 66.6 | 66.6 | 50 | 50 | 66.6 |
| | Timbre features | 50 | 55.5 | 33.3 | 37.5 | 33.3 | 55.5 | 33.3 | 37.5 | 25 | 33 |
| Similarity for talas (%) | Rhythm features | 66.6 | 100 | 100 | 88.8 | 100 | 66.6 | 100 | 66.6 | 100 | 66.6 |
| | Timbre features | 33.3 | 50 | 33.3 | 33.3 | 100 | 33.3 | 50 | 55.5 | 100 | 33.3 |

**Table 3** Accuracy of classification of *raga* and *tala* using rhythm features

| | | *Raga/tala* id | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Accuracy of classification (%) | Raga | 100 | 66.6 | 100 | 66.6 | 66.6 | 100 | 100 | 100 | 100 | 100 |
| | Tala | 66.6 | 100 | 100 | 66.6 | 100 | 66.6 | 100 | 66.6 | 100 | 100 |

## 5 Summary and Conclusion

In this work, analysis of rhythm and timbre features for classification of *raga* and *tala* of Carnatic music has been done. From the experiments, it is found the rhythm features are able to distinguish *raga* and *tala* better than timbre features. The average accuracy of 89.98 and 86.67 % is achieved for classification of *raga* and *tala* respectively using GMM classifier. Even though the results obtained are promising, it cannot be generalized since it is validated using a small data set. As a future work, combination of rhythm and pitch related features shall be used for *raga* and *tala* classification. MIR systems for music recommendation shall be developed using these features.

## References

1. Agarwal, P., Karnick, H., Raj, B.: A comparative study of indian and western music forms. In: International Society for Music Information Retrieval (2013)
2. Klapuri, A., Davy, M.: Signal Processing Methods for Music Transcription. Springer, New York Inc., Secaucus (2006)
3. Orio, N., Piva, R.: Combining timbric and rhythmic features for semantic music tagging. In: International Society for Music Information Retrieval (2013)
4. Pandey, G., Mishra, C., Ipe, P.: Tansen: a system for automatic raga identification pp. 1350–1363 (2003)
5. Bhattacharjee, A., Srinivasan, N.: Hindustani raga representation and identification: a transition probability based approach. IJMBC **2**(1–2), 66–91 (2011)
6. P.Kirthika, Chattamvelli, R.: A review of raga based music classification and music information retrieval (mir). IEEE conference on signal processing (2012)
7. Koduri, G., Gulati, S., Rao, P.: A survey of raaga recognition techniques and improvements to the state-of-the-art. SMC (2011)
8. Ranjani, H.G., Arthi, S., Sreenivas, T.V.: Carnatic music analysis: shadja, swara identification and raga verification in alapana using stochastic models. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 29–32 (2011)
9. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. IEEE Transactions on Speech Audio process. **10**(5) (2002)
10. Christopher, R., Kumar, P., Chandy, D.: Audio retrieval using timbral feature. In: IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (2013)