SCALE INDEPENDENT RAGA IDENTIFICATION USING CHROMAGRAM PATTERNS AND SWARA BASED FEATURES

Pranay Dighe, Parul Agrawal, Harish Karnick

Department of Computer Science and Engineering
Indian Institute of Technology Kanpur, India
{pranayd,parulag,hk}@cse.iitk.ac.in

Siddartha Thota[1],Bhiksha Raj[2]

1. NIT Surathkal, India
2. Carnegie Mellon University, USA
sddhrthrt@gmail.com, bhiksha@cs.cmu.edu

ABSTRACT

In Indian classical music a raga describes the constituent structure of notes in a musical piece. In this work, we investigate the problem of scale independent automatic raga identification by achieving state-of-the-art results using GMM based Hidden Markov Models over a collection of features consisting of chromagram patterns, mel-cepstrum coefficients and timbre features. We also perform the above task using $1)$ discrete HMMs and $2)$ classification trees over *swara* based features created from chromagrams using the concept of *vadi* of a raga.On a dataset of 4 ragas- darbari, khamaj, malhar and sohini; we have achieved an average accuracy of $\sim 97\%$. This is a certain improvement over previous works because they use the knowledge of scale used in the raga performance. We believe that with a more careful selection of features and by fusing results from multiple classifiers we should be able to improve results further.

Index Terms— Raga, Chromagram, Hidden Markov Models, Gaussian Mixture Models, *Swara*

1. INTRODUCTION

There has been a lot of work on content analysis of western classical music[1] in terms of genre detection[2], music retrieval[3] and instrument/singer identification etc. to name some. Though Indian Classical is also considered to be a major form of music, the current literature related to it is very limited, in comparison to its western counterpart. Indian classical music is known for its technical soundness and its well defined structure. It is defined in terms of *ragas*, each of which in itself has layers of additional finer structure, yet these aspects have been under-utilised by the music analysis community. In this paper we address the analysis of Indian music; in specific, we attempt to develop techniques to identify ragas, incorporating information about its structure into our procedures. A key challenge to address is that unlike western music where notes have absolute pitch, notes in Indian music are defined only in terms of their relation to one another, and inference must be made from patterns of sounds, rather than their absolute frequency structure.

While an expert in Indian Classical music can identify a raga just by noticing the unique patterns of aspects such as *swaras, arohan, avarohan* and *pakad* in the performance (we explain these terms later in the paper), developing computational models for the same has been a challenging task for music researchers. The independence that Indian classical music provides to an artist to give his own personal flavour to a raga makes it difficult for a novice to identify two different performances of the same raga.

2. RELATED WORK

The importance of Raga identification in Indian music cannot be overstated – any analysis must begin with identifying the underlying raga. Consequently, several approaches have been proposed for automatic raga identification. One early work is by Sahasrabuddhe[4](1992) who modelled a raga as a finite state automaton based on the *swara* patterns followed in it.

Pandey et al.(2003) extended this idea of *swara* sequence in the "Tansen" raga recognition system [5] where they worked with Hidden Markov Models on *swara* sequences extracted using a heuristics driven note-segmentation technique. They employed a novel *pakad* matching algorithm that improved the HMM based results.

Chrodia and Rae [6](2007)derived Pitch-class Distributions(PCD)and Pitch-class Dyad Distributions(PCDD) from Harmonic Pitch Class Profiles (HPCP) and used these distributions for classification using SVMs. They achieved state-of-the-art results with accuracies of 78% for PCDs and 97.1% for PCDDs. Sridhar and Geetha [7](2009) used template matching for the task of raga identification.

Inspired by the use of HMMs over swara sequences and PCDs and PCDDs in [5] and [6], we propose a novel approach of raga identification using Chroma based features. Similar to HPCP features, we extract feature vectors from chromagram patterns(in detail later), but instead of learning probability distributions and using SVMs on their parameters for classification as done in [6], we look at the sequences of these feature vectors over time and employ a Gaussian Mixture Model based HMM test for identification of ragas.

Another novelty in our approach is performing Raga identification without the knowledge of the scale of the performance. For this, we employ *Swara* based features extracted from the chromagram using the concept of *vadi* (explained later) and perform discrete HMM based and Classification Tree based experiments on the data. We have achieved results comparable to the state-of-the-art using the above approaches.

In next section we discuss the raga theory. In section 4, we discuss extraction of chroma and swara features. Section 5 and 6 discuss the experiments performed and results achieved.

3. RAGA THEORY AND RELATED CONCEPTS

Indian classical music has a comprehensive theory, set of rules and ideas that govern it. *Ragas*, along with *taal* form the two major macro-level concepts which are based upon finer concepts of *swara,*

Table 1. A Sample Raga Dictionary (Underline-*Komal*, *-*Tivra*, ‘-Upper Octave)

Raga	Thaat	Time	Arohan	Avrohan	Vadi
Darbari	Asavari	Midnight	Sa Re $\underline{\text{Ga}}$ Ma Pa $\underline{\text{Dh}}$ $\underline{\text{Ni}}$ Sa‘	Re‘ $\underline{\text{Ni}}$ Sa‘ $\underline{\text{Dh}}$ $\underline{\text{Ni}}$ Pa, Ma Pa $\underline{\text{Ga}}$ Ma Re Sa	Re
Khamaj	Khamaj	Late Night	Sa Ga Ma Pa Dh Ni Sa‘	Sa‘ $\underline{\text{Ni}}$ Dh Pa Ma Ga Re Sa	Ga
Miya Ki Malhar	Kafi	Rains	Sa Ma Re Pa, Ma Pa $\underline{\text{Ni}}$ Dh Ni Sa‘	Sa‘ Dha $\underline{\text{Ni}}$ Pa Ma Pa Ga Ma Re Sa	Ma
Sohini	Marva	Bef. Sunrise	Sa Ma* Ga, Ma* Dh Ni Sa‘$\underline{\text{Re‘}}$ Sa‘	Sa‘ Ni Dh Ma* Ga $\underline{\text{Re}}$ Sa	Dh

shruti, alankar etc. While *taal* is a construction over time, *raga* is a construction over sounds. Thus, broadly a raga is a melodic piece of music which is characterised by the underlying units of sounds called *swaras*. The collection of *swaras* used in a raga, the patterns over these *swaras* and the presence or absence of certain *swaras* are some of the major discriminating features of a raga. Moreover, Indian classical music allows the artist to improvise over the definitions of a raga to create his own personal performance of the raga. Due to this, two performances of the same raga by different artists may sound strikingly different to the novice ears, though they still maintain the defining qualities of that raga. Here we discuss some raga related terminologies.

3.1. *Swara*

Swaras (or notes) are symbols used for sets of frequencies. They are used as finer units in a raga and thus they act as an alphabet. They are closely related to the solfege in western music. Basically, there are 7 *swaras* in Indian classical music: *Shadja(Sa), Rishab(Re), Gandhara(Ga), Madhyama(Ma), Panchama(Pa), Dhaivata(Dh), and Nishad(Ni)*. These *swaras* are related to each other by the fixed ratio of absolute frequency values they denote. Similar to notes in western music, we get the same *swara* one octave above or below the present *swara*. An artist always tunes his whole raga performance around a fixed tonic frequency which he chooses according to his own comfort. Note that this tonic frequency defines the scale in which the raga is being performed. Irrespective of the absolute value of this tonic frequency, it is termed as the *swara* Sa. The rest of the *swaras* are spread around *Sa* as per the ratios governing them. For example the *swara Pa* is always $\frac{3}{2}$ times *Sa* in terms of actual frequencies. *Sa* and *Pa* only have the *Shudhdha*(pure) form, while the rest have variants like *Tivra*(sharp) and *Komal*(soft) also. Often, a 12-note system is used with *Sa* as 1^{st} note and *Ni* as 12^{th}.

3.2. *Arohan* and *Avrohan* and *Pakad*

Though two different ragas might have the same constituent *swaras*, the basic differentiating characteristic among them is the unique ascending and descending sequence of *swaras* termed *arohan* and *avarohan* respectively. While ascending or descending in frequency, the artist skips certain *swaras* to form the *arohan* or *avarohan* of that particular raga.

Pakad is a small sequence of *swaras* in a raga that acts as a signature for the raga and an artist often visits and revisits the *pakad* over a performance. It is a major clue for human raga recognition. . For some ragas, the *pakad* might be simply the *avrohan* wrapped over *arohan* and for some others, it might be a totally different pattern of the constituent *swaras*.

3.3. *Vadi* and *Samvadi*

Vadi is the most prominent *swara* in a raga and is often described as the King of *swaras* for that raga. The *swara* second to *vadi* in importance is called the *samvadi*. An artists stays at the *vadi* and *samvadi* for significant durations and emphasises them in a performance. The *swaras* other than *vadi* and *samvadi* which constitute the raga are called *anuvadi swaras* whereas the *swaras* which are totally absent are called *vivadi*. We use the concept of *vadi* later for extraction of *swaras* from the dataset. Table 1 lists *pakad*, *arohan*, *avarohan*, *vadi* and *samvadi* of the ragas Darbari, Khamaj, Malhar and Sohini.

3.4. Different Ragas

The literal *Sanskrit* meaning of raga is color and it is said that each raga can color the listener's mood (or emotional state) to that of the raga. Ragas have also been classified according to the seasons or the time of the day that they are supposed to be performed, so as to induce the relevant mood or emotional state in the listener. We have different ragas for sunrise(*lalit, bilaval, asavari*), noon(*multani, sarang, patdip*), night(*kalyan, bihag, khamaj*) and rainy season(*malhar*).

Ragas are also characterized by the *thaats* from which they originate. There are 10 *thaats* which are actually 10 parent ragas. The number of *swaras* used in a raga also classifies ragas into three categories: audav(5 *swaras*) shadav(6 *swaras*) and sampoorna(7 *swaras*). All 10 *thaat* parent ragas are sampoorna.

4. EXTRACTION OF FEATURES

4.1. Chromagram and Chroma Features

A chromagram is a visual representation of energies in the 12 semitones(or chromas) of the musical octave namely C,C#,D,D#,E,F,F#,G,G#,A,A# and B. Since the semitones get repeated each octave above and below, we compute energy for each semitone(chroma) by adding it up over different octaves. This generates a 12-dimensional vector for each frame of music analysed, and can be visually represented as a 2D image using different colors for different energy values.

Figure 1 depicts a chromagram generated from an *arohan* of raga Bhairavi. The *Sa swara* of *Bhairavi* coincides with the semitone G and the rest of the *swaras* in the *arohan* get aligned with the semitone pattern in the chromagram. These observations and previous use of chromagrams in audio analysis[8] and chord recognition[9] motivates us to use the chromagram as a source of features.

We analyse each sample of a raga framewise using windows of length 0.2 seconds and a frame shift of 0.025 seconds. The frame size has been taken as large as 0.2 seconds because for shorter

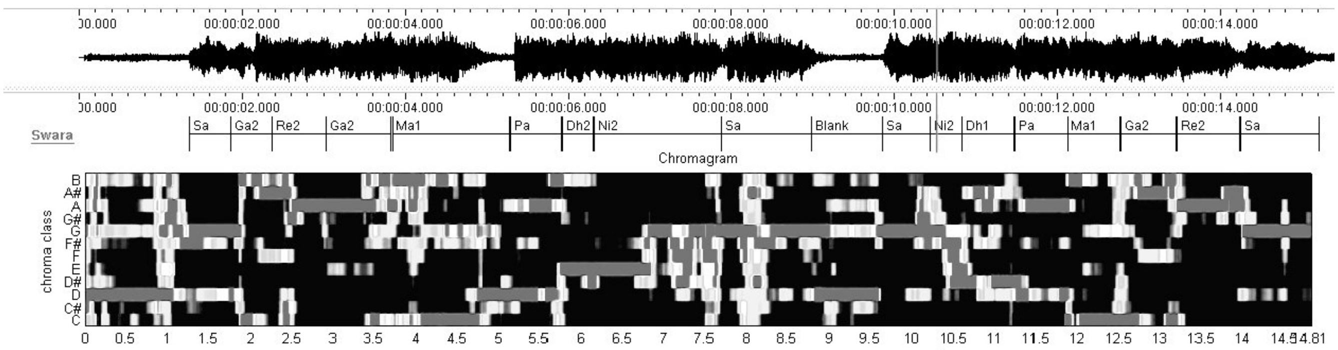

Fig. 1. Chromagram and Swara sequence for a *arohan* Piece of Raga Bhairavi

frames, we lose information about semitone energies in lower frequencies. We do a logarithm and discrete cosine transform (DCT) operations on the 12-dimensional chroma data. The DCT has an energy compaction property and also decorrelates the features. This gives a modified 12-dimensional vectors for each frame. In reported experiments, we further augment the chroma features with MFCC and Timbre features to improve results. For the chromagram experiments, the MIRToolbox[10], an open source toolbox for musical feature extraction was used to extract features.

4.2. Extracting *Swara* Features from Chromagram

From chromagram, we extract the semitone with maximum energy in each frame and get a sequence of semitones for the raga. Though these sequences of semitones might have some identifying information about the raga, using them for raga identification is not suggested because ragas are defined over patterns of *swaras* which in turn might get associated with different semitones depending on the tonic frequency selected by the artist for a particular performance.

In our approach, we assume that we do not have information about the tonic frequency of the raga performance. We must therefore find the mapping from the absolute frequency scale employed by the chromagram to the relative scale of the music, such that the *swara* sequence may be identified. To do so, we use the concept of *vadi* discussed earlier to convert the semitone sequence to a *swara* sequence. We compute the most frequently occurring semitone from the semitone sequence and associate it with the *vadi* of a raga which is known for each raga. For eg. *vadi* for raga Khamaj is *swara* “Ga”. In an audio of Khamaj, if the semitone C# is most prominent, we label *swara* “Ga” at semitone C# and then convert the rest of the semitone sequence into a *swara* sequence.

The above procedure is raga specific, *i.e.* the conversion from semitone sequence to *swara* sequence utilizes the identity of the raga-specific *vadi* note, Assume that we are building the system for n ragas and the raga for the test audio is not known, then for the same test audio, we must consequently compute separate *swara* transcriptions for each of the n ragas. These are combined into a single unified representation as follows:

- From a raga audio, we create 3 minutes long snippets
- Compute the snippet's *swara* transcription S_i for each raga R_i where $i = 1..n$
- From each transcription S_i, we create a 12-dimensional normalized frequency vector v_i that is basically a histogram of *swaras* over the 3 minute snippet
- Concatenate all such v_i's to get k-dimensional feature vector where $k = 12 \times n$. We call this feature the swara histogram and employ classification trees over it.
- We also use the swara sequences directly in a discrete output based HMM system for raga identification

5. DATASET AND EXPERIMENTS

Since a standard dataset for Raga identification is still not available, the data for this work was collected from online Youtube videos. We have performed the experiments on 4 ragas- Darbari, Khamaj, Malhar and Sohini taking 14 performances of each. The average duration for each performance is around 10 minutes.

We perform four experiments discussed below:

5.1. Expt 1: GMM based HMMs on Chroma Features

Hidden Markov Models are trained on chroma features (section 4.1) with 12 hidden states to correspond to 12 notes and the state emissions were 12 dimensional chroma data modelled by Gaussian mixtures with 4 gaussians. 80% of the snippets are randomly chosen for training and the rest are used for testing. The emissions of the 12 states were randomly initialized in both experiments and the states need not represent exactly the underlying *swaras*.

5.2. Expt 2: GMM based HMMs on Chroma + MFCC + Timbre Features

Here we combined the following features: 12-chroma features, 13 MFCC, 13 delta-MFCC, 13 delta-delta-MFCC, 11 Timbre features (zero-crossing rate, rolloff, brightness, centroid, spread, skewness, kurtosis, flatness, entropy, roughness and irregularity) to get a 62 dimensional vector and performed a similar HMM experiment with 8-Gaussian GMMs for state output densities.

5.3. Expt 3: Classification Tree over Swara Histograms

A categorical classification tree was employed on swara histograms and a 10-fold cross validation experiment was done.

Table 2. Confusion Matrices for (a) Chroma Data, (b) Chroma, MFCC and Timbre Features, (c) Swara Histograms and (d) Swara Sequences

(a) Expt. 1

	Darbari	Khamaj	Malhar	Sohini
Darbari	**76.9**	0	15.4	7.7
Khamaj	0	**100.0**	0	0
Malhar	0	21.4	**78.6**	0
Sohini	0	0	12.5	**87.5**

(b) Expt. 2

	Darbari	Khamaj	Malhar	Sohini
Darbari	**97.0**	0	3.0	0
Khamaj	0	**99.0**	0.3	0.7
Malhar	0	0	**96.1**	3.9
Sohini	1.4	0	0	**98.6**

(c) Expt. 3

	Darbari	Khamaj	Malhar	Sohini
Darbari	**84.6**	5.4	8.5	1.5
Khamaj	6.0	**85.0**	6.0	3.0
Malhar	8.9	7.2	**77.4**	6.4
Sohini	3.1	3.1	6.2	**87.3**

(d) Expt. 4

	Darbari	Khamaj	Malhar	Sohini
Darbari	**70.4**	3.7	22.2	3.7
Khamaj	9.5	**66.6**	4.8	19.0
Malhar	0	3.8	**92.4**	3.8
Sohini	0	0	0	**100.0**

5.4. Expt 4: Discrete output-HMMs over *swara* sequences

From training data, we compute HMM models for all the ragas. For a test audio, different *swara* transcriptions are first computed according to *vadis* for different ragas, and then, these transcriptions are tested in the corresponding HMM model. Raga associated with the HMM model giving maximum posterior probability is declared as the result.

6. CONCLUSIONS AND FUTURE WORK

We achieved average accuracies of 86.67%, 97.68%, 83.53% and 83.57% for experiments 1,2,3 and 4 respectively. These are the best current results for scale independent raga identification and compare closely with the results by Chordia[6].

The confusion between Darbari and Malhar in all experiments can be traced to the fact that they are nearly similar in the constituent *swaras* and *arohan-avrohan*. Raga Sohini seems to have performed better because it is only raga among the three which employs only 6 *swaras*) (a class of ragas called *shadav*) whereas the rest employ some variant of all 7 notes (a class of ragas called *Sampoorna*).

Overall, experiment 2 with a feature set consisting of chroma, MFCC and timbre data outperforms the other three experiments. In comparison, swara based features give decent but lower accuracies, because they work only on the highest energy component of the chromagram whereas the chroma feature takes into account the entire chromagram pattern. Still, the fact that we achieved good accuracies in experiments which exclude common audio features like MFCC, supports our initial hypothesis that the finer swara based substructure contains the defining information about ragas.

Competing with expert human efficiency is the final goal in raga identification. So the future work lies in fusing the above approaches to achieve more efficiency and expanding our dataset to many more ragas. We also wish to investigate structure discovery through minimum entropy or maximum-structure learning methods. Chroma and swara features can also be tested for classification of genres and *taals*. Some more creative extensions could be designing Indian Classical music tutoring systems and automatic cataloguing and indexing systems.

7. REFERENCES

[1] Martin, Keith D., Eric D. Scheirer, and Barry L. Vercoe. "Music content analysis through models of audition." Proc. ACM Multimedia Workshop on Content Processing of Music for Multimedia Applications. 1998.

[2] Tzanetakis, G.; Cook, P., "Musical genre classification of audio signals," Speech and Audio Processing, IEEE Transactions on , vol.10, no.5, pp.293,302, Jul 2002.

[3] G. Poliner, D. Ellis, A. Ehmann, E. Gmez, S. Streich, B. Ong, "Melody Transcription from Music Audio: Approaches and Evaluation," IEEE Transactions on Audio, Speech, Language Processing 2006.

[4] H. Sahasrabuddhe and R. Upadhye, "On the computational model of raag music of india." in Workshop on AI and Music: European Conference on AI, 1992.

[5] G. Pandey, C. Mishra, and P. Ipe, "Tansen: A system for automatic raga identification" in Proc. of Indian International Conference on Artificial Intelligence , 2003, pp. 13501363.

[6] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions," in Proc. of ISMIR, 2007, pp. 431436.

[7] R. Sridhar and T. Geetha, "Raga identification of carnatic music for music information retrieval," International Journal of Recent trends in Engineering, vol. 1, no. 1, 2009, pp. 571574.

[8] Xiaoqing Yu; Jing Zhang; Junwei Liu; Wanggen Wan; Wei Yang, "An audio retrieval method based on chromagram and distance metrics," International Conference on Audio Language and Image Processing (ICALIP), Nov. 2010.

[9] K.Lee, "Automatic Chord Recognition from Audio Using Enhanced Pitch Class profile," in Proc. of the International Computer Music Conference, 2006.

[10] Olivier Lartillot, Petri Toiviainen, "A Matlab Toolbox for Musical Feature Extraction From Audio", International Conference on Digital Audio Effects, Bordeaux, 2007.

[11] G. Koduri, S. Gulati and P. Rao,"A Survey Of Raaga Recognition Techniques And Improvements To The State-Of-The-Art," Sound and Music Computing, 2011.