

A Survey of Raaga Recognition Techniques and Improvements to the State-of-the-Art

Gopala Krishna Koduri

IIT Hyderabad

koduri@research.iiit.ac.in

Sankalp Gulati

IIT Bombay

sankalpg@ee.iitb.ac.in

Preeti Rao

IIT Bombay

prao@ee.iitb.ac.in

ABSTRACT

Raaga is the spine of Indian classical music. It is the single most crucial element of the melodic framework on which the music of the subcontinent thrives. Naturally, automatic raaga recognition is an important step in computational musicology as far as Indian music is considered. It has several applications like indexing Indian music, automatic note transcription, comparing, classifying and recommending tunes, and teaching to mention a few. Simply put, it is the first logical step in the process of creating computational methods for Indian classical music. In this work, we investigate the properties of a raaga and the natural process by which people identify the raaga. We survey the past raaga recognition techniques correlating them with human techniques, in both north Indian (Hindustani) and south Indian (Carnatic) music systems. We identify the main drawbacks and propose minor, but multiple improvements to the state-of-the-art raaga recognition technique.

1. INTRODUCTION

Geekie [1] very briefly summarizes the importance of raaga recognition for Indian music and its applications in music information retrieval in general. Raaga recognition is primarily approached as determining the scale used in composing a tune. However the raaga contains more information which is lost if it is dealt with western methods such as this. This information plays a very central role in the perception of Indian classical music.

In this work, we shortly discuss various properties of a raaga and the way the trained musicians recognize it using cues from the properties of a raaga. Further, we present a brief survey of various methods used by researchers based on such well defined rules of a raaga. We identify shortcomings in those methods and then, we present our system addressing a few of them. We discuss and compare it with the previous systems. We hope this work would be of help to Indian and non-Indian readers in understanding various properties of the raaga for computational purposes or otherwise.

Our work primarily concerns with Carnatic music, but most of the discussion applies to Hindustani music as well,

unless mentioned otherwise.

2. PROPERTIES OF A RAAGA

Matanga, in his epic treatise Brihaddeshi, defines raaga as "that which colors the mind of good through a specific *swara*¹ and *varna* (literally color) or through a type of *dhvani* (sound)"[2]. A technically insightful definition is given by Chordia [3] and Krishnaswamy [4, 5]. It says, "Raaga is a collection of melodic atoms and a technique for developing them. These melodic atoms are sequences of notes that are inflected with various micro pitch alterations and articulated with expressive sense of timing. Longer musical phrases are built by knitting these melodic atoms together". The notion that raaga is not just a sequence of notes is important in understanding it, for developing a representation of raaga for computational purposes. For a westerner, the notion of raaga as put by Harold S Powers might be helpful in understanding what a raaga is. It says, "A raaga is not a tune, nor is it a 'modal' scale, but rather a continuum with scale and tune as its extremes"[6].

Not surprisingly, these definitions coincide in what they try to convey. Though a given raaga has characteristic melodic phrases, they are neither limited nor given. It can be understood from the fact that even after a given raaga is used to tune numerous compositions by various people, it is always possible that a new tune can be composed using that raaga. On the other hand, it is neither just a set of notes, because different raagas have same set of notes yet sound very different. This is due to various properties of raaga like the order of the notes (called *arohana/avarohana* which mean ascending/descending patterns respectively), the way they are intonated using various movements (called *gamakas*), their relative position, strength and duration of notes (i.e., various functions of swaras). We'll now see these various aspects of raaga in detail.

2.1 Arohana and Avarohana: The Ascending and Descending Progressions of a Raaga

Typically a raaga is represented using the ascending (*arohana*) and descending (*avarohana*) progressions of notes. There are certain observations (or rules) that are necessary while reciting a raaga with regards to the transitions between notes. The transitions generally occur with a note that is near to the current note in *arohana/avarohana*. There are several other heuristics characteristic of a raagas aro-

¹ Swara refers to one the seven notes in the octave.

hana and avarohana, which are not always strictly followed. We have heard multiple viewpoints about such heuristics.

2.2 Gamakas

There is a reason why Indian classical music does not have a strongly followed notation system like the western classical tradition. Consider a note, a fixed frequency value. The rapid oscillatory movement about the note is one of the several forms of movements, which are together called as gamakas. Another form of gamaka involves making a sliding movement from one note to another. Like this, there are number of ways to move around or move between the notes. There are various ways to group these movements. But the most accepted classification speaks of 15 types of gamakas [7, 8]. Apart from gamakas, there are *alankaras* (ornaments) which are patterns of note sequences which beautified and instilled some kind of feeling when listened to.

Owing to the gamakas tremendous influence on how a tune sounds, they are often considered the soul of Indian classical music. Though gamakas are used in both Carnatic and Hindustani [9], the pattern of usage is very distinct. We would like to highlight the point that gamakas are not just decorative items or embellishments, they are very essential constituents of a raaga.

2.3 Characteristic Phrases

Each raaga, just like it has a set of notes, also has few characteristic phrases. These phrases are said to be very crucial for conveying the *bhava* or the feeling of the raaga. Typically in a concert, the artist starts with singing these phrases. These are the main clues for the listeners to identify what raaga it is.

2.4 Various Roles Played by the Notes

In a given raaga, not all the swaras play the same role. As very well put by [10], just like various checkers in the game of chess, various notes in the raaga have different functions. Certain swaras are said to be important than the rest. These swaras bring out the mood of the raaga. These are called *Jeeva* swaras. The musical phrases are built around the Jeeva swaras. The note which occurs at the beginning of the melodic phrases is referred to as *Graha* swara. *Nyasa* swaras are those notes which appear at the end of such musical phrases. *Dirgha* swaras are notes that are prolonged. A swara that occurs relatively frequently is called *Amsa* swara, and that which is sparingly used is called *Alpa* swara. Though two given raagas have the same set of constituent notes, the functionality of the constituent swaras can be very different, leading to a different feeling altogether.

In addition to these above discussed properties, Hindustani classical music also emphasizes the time and season, a raaga should be used in. They seem less relevant in Carnatic music today.

That said, a raaga is an evolution phenomenon. It continually takes place over time; no existing raaga was perceived the way it is today. The properties which enhance

the characteristic nature of a raaga are retained and others are done away with. This process happens continually over decades and centuries. The raaga takes its shape and sets a unique mood depending on these properties.

Now, we'll discuss the way listener and a musician identify a raaga from a composition.

3. HOW DO PEOPLE IDENTIFY A RAAGA

Though there are no rules of thumb in identifying a raaga, usually there are two procedures by which people get to know the raaga from a composition. It normally depends on whether the person is a trained musician or a rasika, the non-trained but knowledgeable person. People who have not much knowledge of raagas cannot identify them unless they memorize the compositions and their raagas.

3.1 Non-trained Person or The *Rasika's* Way

In a nutshell, the procedure followed by a rasika typically involves correlating two tunes based on how similar they sound. Years of listening to tunes composed in various raagas gives a listener enough exposure. A new tune is juxtaposed with the known ones and is classified depending on how similar it sounds to a previous tune. This similarity can arise from a number of factors - the rules in transition between notes imposed by arohana and avarohana, characteristic phrases, usage-pattern of few notes and gamakas.

This method depends a lot on the cognitive abilities of a person. Without enough previous exposure, it is not feasible for a person to attempt identifying a raaga. There is a note worthy observation in this method. Though the people cannot express in a concrete manner what a raaga is, they are still able to identify it. This very fact hints at a possible classifier, that can be trained with enough data for each raaga.

3.2 The Trained Musician's Way

A musician tries to find few characteristic phrases of the raaga. These are called *pakads* in Hindustani music and *swara sancharas* in Carnatic music. If the musician finds these phrase(s) in the tune being played, the raaga is immediately identified. But at times these phrases might not be found or, are too vague. In this case, the musicians play the tune on an instrument (imaginary or otherwise) and identify the swaras being used. They observe the gamakas used on these swaras, locations of various notes within the music phrases and the transitions between swaras. They use these clues to arrive at a raaga.

This method seems to use almost all the characteristics a raaga has. It looks more programmatic in its structure and implementation. If the current music technology can afford to derive various low level features which can be used to identify such clues, the same procedure can be implemented computationally with almost perfect results!

These methods used by trained musicians and non-trained listeners are both important which are to be used for implementing a raaga recognition system. As we will see, the existing systems try to mimic them as much as possible.

4. AUTOMATIC RAAGA RECOGNITION

In this section, we present a survey of previous systems which dealt with raaga recognition. We discuss the different approaches, implementations and results. In the next section, we outline the shortcomings of these systems. Later we present our raaga recognition method which seeks to address some of these.

Past approaches to computer-based raaga recognition have based themselves on the properties of raga such as pitch class distributions or pitch sequence information as captured by note bi-grams or HMMs (Hidden Markov models) or swara intonations. The needed inputs are obtained by the pitch tracking of usually monophonic audio signals of an unaccompanied instrument or voice, optionally followed by a step of note segmentation and labeling.

4.1 Scale Matching

Sridhar and Geetha [11] have followed an approach where the scale used in the tune is estimated, and compared with the scales in the database. The raaga corresponding to that scale in database which matches which the estimated scale is output by the system. Their test data consisted of 30 tunes in 3 raagas sung by 4 artists. They use harmonic product spectrum algorithm [12] to extract the pitch. The tonic is manually fed. The other frequencies in the scale are marked down based on the respective ratio with the tonic. The set of notes which are used are matched against several sets of notes stored in the database for various raagas. Note that this is not the same as pitch-class profile. Here, the comparison is between the scale intervals, and not the pitch-class distribution. The results thus obtained are shown in Figure 1. A similar approach based on detecting the swaras used in arohana and avarohana to find the raaga is presented by Shetty and Achary [13].

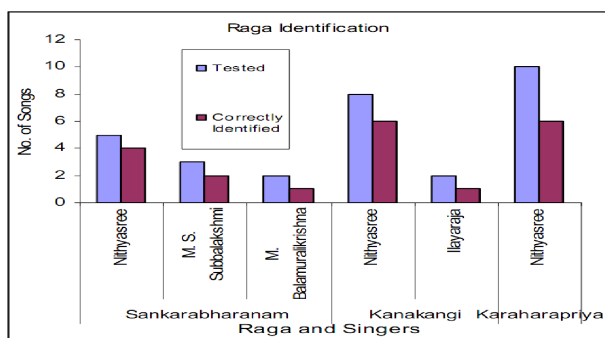


Figure 1. Results of Sridhar & Geeta’s raaga identification method [11]

4.2 Statistical Modeling and Pakad Matching

Sahasrabudde and Upadhye [14] modeled the raaga as a finite automaton based on the rules set by the properties of each raaga. This idea is used to generate a number of note sequences for a raga composition, which were technically correct and indistinguishable from human compositions. Inspired by this, Pandey et al. [15], used HMMs to

Raaga	Test Samples	Accurately Identified	Accuracy
Yaman Kalyan	15	12	80%
Bhupali	16	12	75%
Total	31	24	77%

Table 1. Results of Plain Raaga Identification in Tansen [15]

Raaga	Test Samples	Accurately Identified	Accuracy
Yaman Kalyan	15	12	80%
Bhupali	16	15	94%
Total	31	27	87%

Table 2. Results of Raaga Identification with Pakad Matching in Tansen [15]

capture the note transitions in their “Tansen” raaga recognition system. The rules to form a melodic sequence for a given raaga are well defined and the number of notes is finite. So, HMM model of a raaga proved to be good at capturing those rules in note transitions engraved by arohana and avarohana patterns of the respective raaga. They have complemented this system with scores obtained from two pakad matching modules. In one such module, pakad is identified with substring matching algorithm. In the other one, it is identified by counting the occurrences of n-grams of frequencies in the pakad.

The other important contributions of [15] include two heuristics to improve the transcription of Indian classical music - the hill peak heuristic and the note duration heuristic. Unlike western music, Indian music has a lot of micro tonal variations which makes even monophonic note transcription a challenging problem. The two heuristics try to get through these micro tonal fluctuations in attaining a better transcription. The hill peak heuristic says that a significant change in the slope or the sign reversal in slope is closely associated with the presence of a note. The note duration heuristic assumes that a note is played for at least a certain constant span of time.

Tansen is built to classify two raagas. The results are shown below. Table 1 shows the results obtained using HMM models. Table 2 shows the results obtained by complementing HMM models with pakad matching.

The central idea in this approach, which is to model a raaga as HMM, was also used in [16]. The same idea is used in an attempt to automatically generate Hindustani classical music [17], but with less success.

4.3 Pitch-class Profiles and Note Bi-grams

Chordia [3] has used the pitch class profiles and the bi-grams of pitches to classify raagas. The dataset used in his system consists of 72 minutes of monophonic instrumental (sarod) data in 17 raagas played by a single artist. The HPS algorithm is used to extract the pitch. Note onsets are detected by observing the sudden changes in phase

Classifier	Accuracy
Multi Variate Normal	94%
FFNN	75%
K-NN Classifier	67%
Tree-based Classifier	50%

Table 3. Raaga recognition accuracies with various classifiers in Chordia’s system [3]

and amplitude in the signal spectrum. Then, the pitch-class profiles and the bi-grams are calculated. It is shown that bi-grams are useful in discriminating the raagas with the same scale. He uses several classifiers combined with dimensionality reduction techniques. Using just the pitch class profiles, the system achieves an accuracy of 75%. Using only the bi-grams of pitches, the accuracy is 82%. Best accuracy of 94% is achieved using a multivariate Normal classifier, together with principal components analysis (PCA) to reduce the feature vector size from 144 (bi-grams) + 12 (pitch profile) to 50. Performance across classifiers is shown in Table 3.

We have run a preliminary experiment to see if pitch-class distribution features (PCDs) that have been reported to be showing high accuracy in identifying Hindustani raagas [3] work in the context of Carnatic raagas. Figure 2 shows distributions of PCDs of two tunes each for two raagas. It is quite evident from the figure that there is a strong intra-raaga consistency and inter-raaga variance.

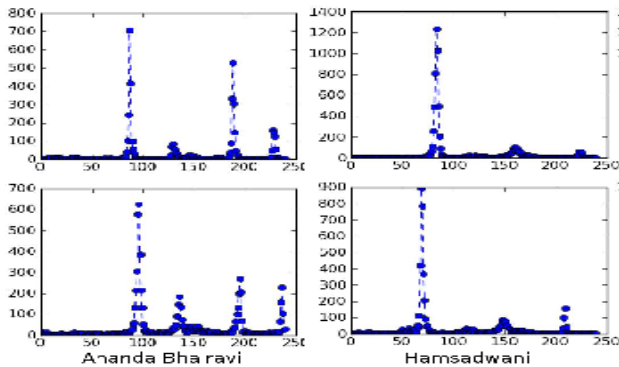


Figure 2. Pitch-class distribution for four tunes each in two raagas. Plots on the left side correspond to Ananda Bhairavi and those on the right side correspond to Hamsadwani. X-axis denotes the bin numbers. Y-axis denotes the count for the respective bin.

4.4 Swara Intonation

It is often said that, in Indian classical music, a swarasthana² does not correspond to a fixed frequency value (with its octave equivalents). It is a region [18]. So, although two raagas share the same scale, the precise intonation of specific notes can vary significantly. Belle *et al* [19] have used this

² the position held by a particular scale degree note with respect to the tonic

clue to differentiate raagas that share the same scale intervals. They evaluated the system on 10 tunes, with 4 raagas evenly distributed in 2 distinct scale groups. They showed that the use of swara intonation features improved upon the accuracies achieved with straightforward pitch class distributions.

In all the above attempts, we see that most of the approaches which we have mentioned in the beginning of the section, have been made use of. Ideally speaking those approaches should be capable of building a perfect raaga recognition system. In the the following section, we identify few problems that make this task difficult.

5. PROBLEMS THAT NEED TO BE ADDRESSED

5.1 Gamakas and Pitch Extraction for Carnatic Music

An appropriate pitch extraction module is that which can accurately represent the gamakas. It has not been a severe problem for the classification systems that were not depending on gamakas of a note for classification. If there is such a pitch extraction system in place, gamakas can be used as an additional feature to improve the accuracies of existing systems. Gamakas assume a major role when the number of raaga classes is high in the dataset.

5.2 Skipping tonic detection

The manually implemented tonic (the base frequency of the instrument/singer) identification stage needs to be eliminated if possible. Since the tonic identification itself involves some amount of error, this could adversely impact the performance of a raaga recognition system. Neither the Carnatic nor Hindustani systems adhere to any absolute tonic frequency, therefore it makes sense to build a system that can ignore the absolute location of the tonic.

5.3 Resolution of pitch-classes

Though 12 bins for pitch-class profiles look ideal to the Western eye, we hypothesize that a more continuous model can capture more relevant information related to Indian classical music. Dividing an octave into n bins where $n > 12$ can help us model the distribution with better resolution. Gamakas (the micro tonal variations) play a vital role in the perception of Indian music, and this has been confirmed by several accomplished artists. The transitions involved in a gamaka and the notes through which its trajectory passes are two factors that need to be captured. We hypothesize that this information can be obtained, at least partially, using a higher number of bins for the first-order pitch distribution.

5.4 A Comprehensive Dataset

The previous datasets which are used for testing have several problems. In Tansen, and the work by Sridhar and Geeta, the datasets had as few as 2 or 3 raagas. The dataset used by Chordia has all the data played on a single instrument by a single artist. The test datasets were constrained to some extent by the requirement of monophonic audio

(unaccompanied melodic instrument) for reliable pitch detection. In the present work, we investigate raaga recognition performances on a more comprehensive dataset with more raaga classes with significant number of tunes in each across different artists and different compositions. This should enable us to obtain better insight into the raaga identification problem.

With these issues about the raaga recognition in mind, we have implemented a system which addresses some of the challenges described. The following sections introduces our method, and presents a detailed analysis and discussion of the results.

6. OUR METHOD

As mentioned earlier, we propose to address some of the issues described in the previous section. We have taken a diverse set of tunes to include in the dataset. The use of amply available recorded music necessitates a pitch detection method that can robustly track the melody line in the presence of polyphony. The obtained sequence of pitch values converted to cents scale (100 cents = 1 semitone) constitutes the pitch contour. The pitch contour may be used as such to obtain a pitch-class distribution. On the other hand, given the heavy presence of ornamentation in Indian music, it may help to use identified stable note segments before computing the pitch-class distribution. We investigate both approaches. Finally, a similarity measure, that is insensitive to the location of the tonic note, is used to determine the best matched raaga to a given tune based on available labeled data. Each of the aforementioned steps is detailed next.

6.1 Pitch Extraction

Pitch detection is carried out at 10 ms intervals throughout the sampled audio file using a predominant pitch detection algorithm designed to be robust to pitched accompaniment [20]. The pitch detector tracks the predominant melodic voice in polyphonic audio accurately enough to preserve fast pitch modulations. This is achieved by the combination of harmonic pattern matching with dynamic programming based smoothing. Analysis parameter settings suitable to the pitch range and type of polyphony are available via a graphical user interface thus facilitating highly accurate pitch tracking with minimal manual intervention across a wide variety of audio material. Figure 3 shows the output pitch track superimposed on the signal spectrogram for a short segment of Carnatic vocal music where the instrumental accompaniment comprised violin and mridangam (percussion instrument with tonal characteristics). While the violin usually follows the melodic line, it plays held notes in this particular segment. Low amounts of reverberation were audible as well. We observe that the detected pitch track faithfully captures the vocal melody unperturbed by interference from the accompanying instruments.

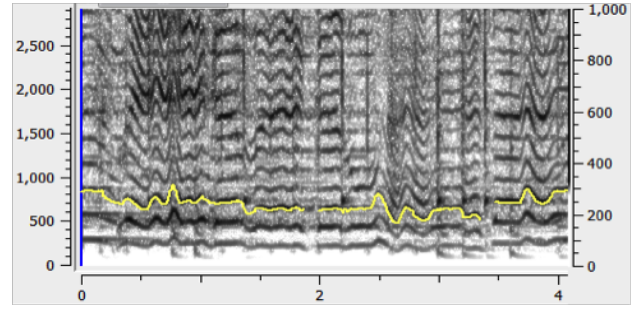


Figure 3. Screenshot from the melodic pitch extraction system of [20] showing the detected pitch superimposed on the signal spectrogram. The axis on the right indicates pitch value (Hz).

6.2 Finding the Tuning Offset

The pitch values obtained at 10 ms intervals are converted to the cents scale by assuming an equi-tempered tuning scale at 220 Hz. All the pitch values are folded into a single octave. The finely-binned histogram maximum of the deviation of the cents value from the notes of the equi-tempered 12-note grid provides us the underlying tuning offset of the audio with respect to 220 Hz. The tuning offset is applied to the pitch values to normalize the continuous pitch contour to standard 220 Hz tuning by a simple vertical shift but without any quantization to the note grid at this point.

6.3 Note Segmentation

As we observe in Figure 3, the pitch contour is continuous and marked by glides and oscillations connecting more stable pitch regions. The stable note regions too are marked by low pitch modulations. As described in Sec. 2, melodic ornamentation in Indian classical music is very diverse and elaborate. For our investigation of pitch class profiles confined to stable notes, we need to detect relatively stable note regions within the continuously varying pitch contour. The local slope of the pitch contour can be used to differentiate stable note regions from connecting glides and ornamentation.

At each time instant, the pitch value is compared with its two neighbors (i.e. 10 ms removed from it) to find the local slope in each direction. If either local slope lies below a threshold value of 15 semitones per second, the current instant is considered to belong to a stable note region. This condition is summarized by the Eq. 1.

$$(|(F(i-1) - F(i))| < \theta) \parallel (|(F(i+1) - F(i))| < \theta) \quad (1)$$

where $F(i)$ is the pitch value at the time index i and θ being the slope threshold. To put the selected threshold value in perspective, a large vibrato (spanning a 1 semitone pitch range) at 6 Hz pitch modulation frequency has a maximum slope of about 15 semitones per second. All instants where the slope does not meet this constraint are considered to belong to the ornamentation.

Finally, the pitch values in the segmented stable note regions are quantized to the nearest available note value in

the 220 Hz equi-tempered scale. This step smoothes out the minor fluctuations within intended steady notes. Figure 4 shows a continuous pitch contour with the corresponding segmented and labeled note sequence superimposed. We note several passing notes are detected which on closer examination are found to last for durations of 30 ms or more.

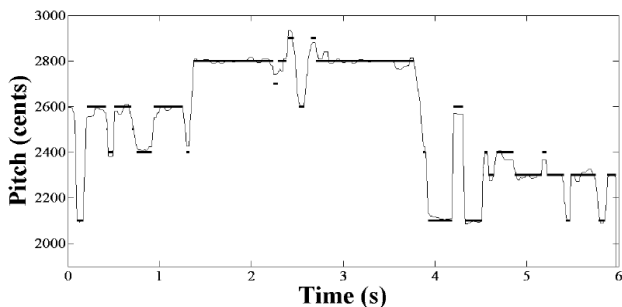


Figure 4. Note segmentation and labeling. Thin line: continuous pitch contour; Thick line: detected stable note regions.

6.4 Pitch-class Profiles

We investigate various approaches to deriving the pitch class profile. The first of two broad approaches corresponds to considering only the stable notes, segmented and labeled in the previous step. The pitch class profile is then a 12-bin histogram corresponding to the octave-folded note label values. There are two choices for weighting the note values for histogram computation. We call these P_1 and P_2 , where P_1 refers to weighting a note bin by the number of instances of the note, and P_2 refers to weighting by total duration over all instances of the note in the music piece.

A second broad approach is ignore the note segmentation step and to consider all pitches in the pitch contour irrespective of whether they correspond to stable notes or ornamentation regions. We call this P_3 . Further, the number of divisions of the octave is varied representing different levels of fineness in pitch resolution. The investigation of varying quantization intervals is motivated by the widely recognized microtonal character of Indian music.

6.5 Distance Measure

In order to compare pitch-class profiles computed from two different tunes, it is necessary that the distribution intervals are aligned in terms of the locations of corresponding scale degrees. This can be ensured by the cyclic rotation of one of the distributions to achieve alignment of its tonic note interval with that of the other distribution. Since information about the tonic note of each tune is not available a priori, we consider all possible alignments between two pitch class profiles and choose the one that matches best in terms of minimizing the distance measure. This is achieved by cyclic rotation of one of the distributions in 12 steps with computation of the distance measure at each step.

As for choosing the distance measure itself, we would like it to reflect the extent of similarity between two tunes

in terms of shared raaga characteristics. We choose the Kullback-Leibler (KL) divergence measure as a distance measure suitable for comparing distributions. Symmetry is incorporated into this measure by summing the two values as given below [19].

$$D_{KL}(P, Q) = d_{KL}(P|Q) + d_{KL}(Q|P) \quad (2)$$

$$d_{KL}(P|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (3)$$

where i refers to the bin index in the pitch class profile, and P and Q refer to pitch class distributions of two tunes.

7. EXPERIMENT AND RESULTS

We describe a raaga classification experiment and present results on the comparative performances of the various types of pitch-class profiles for different classifier settings. A suitable dataset is constructed from commercially available CD audio recordings. To make the best use of available data, we use leave-one-out cross validation with a k-NN (k Nearest Neighbors) classifier to evaluate the performance of our system. The details of the experiment are provided next.

7.1 Dataset

There are a few observations worth mentioning in connection with the design of a test dataset for our raaga recognition system. During preliminary trials of our system, we observed a performance bias in available datasets arising from the fact that several popular compositions in Carnatic music originate in the 17th, 18th and 19th centuries. Some of these compositions sung by several artists lead to the occurrence of several sets of near identical tunes in the dataset resulting in very similar pitch profiles for supposedly different pieces of music. This prompted us to exercising due care in selecting music pieces for our test dataset. We have been careful not to include different versions of the same composition in the dataset. For instance, a tune which renders the kruti 3 nanu brOvamani chepavE is not included if a tune based on that kruti already existed in the dataset. However, since alapanas are not pre-composed, and are purely based on the artists virtuosity, we have included them. To get a bigger dataset we considered the complete Raagam-Taanam-Pallavis of various artists besides shorter krutis. This expanded the list of options from which it is possible to extract a clip to be included in the dataset. The clips were extracted from the live performances and CD recordings of 31 artists, both vocal (male and female) and instrumental (veena, violin, mandolin and saxophone) music. The dataset consisted of 170 tunes from across 10 raagas with at least 10 tunes in each raga (except Ananda Bhairavi with 9 tunes) as summarized in Table 4. The duration of each tune averages 1 minute. The tunes are converted to mono-channel, 22.05 kHz sampling rate, 16 bit PCM. The dataset can be considered very representative of the Carnatic classical music, since it includes artists spanning several decades, male and female, and all the popular instruments.

Raaga	Total tunes	Avg. duration in seconds	Composition of Tunes
Abheri	11	61.3	6 vocal, 5 instrumental
Abhogi	10	62	5 vocal, 5 instrumental
Ananda Bhairavi	9	64.7	4 vocal, 5 instrumental
Arabhi	10	64.9	8 vocal, 2 instrumental
Atana	21	56.75	12 vocal, 9 instrumental
Begada	17	61.17	9 vocal, 8 instrumental
Behag	14	59.71	12 vocal, 2 instrumental
Bilahari	13	61.38	10 vocal, 3 instrumental
Hamsadwani	41	57.07	14 vocal, 27 instrumental
Hindolam	24	60	15 vocal, 9 instrumental

Table 4. Description of the dataset across 10 raagas.

Pith-class profile	k=1	k=3	k=5	k=7
P_1 (12 bins, weighted by number of instances)	55.9	56.5	57.1	59.4
P_2 (12 bins, weighted by duration)	71.2	73.5	76.5	76.5
P_3 (12 bins)	73.5	70	74.7	75.3
P_3 (24 bins)	72.4	72.9	75.3	74.1
P_3 (36 bins)	68.2	72.4	72.9	74.1
P_3 (72 bins)	67.7	68.2	69.4	68.2
P_3 (240 bins)	65.3	68.2	66.5	65.9

Table 5. Performance of weighted-k-NN classification with various pitch-class profiles

7.2 Classification Experiment

A k-NN classification framework is adopted where several values of k are tried. In a leave-one-out cross-validation experiment, each individual tune is considered a test tune in turn while all the remaining constitute the training data. The k nearest neighbors of the test tune in terms of the selected distance measure are considered to estimate the raaga label of the test tune. The distance measure used is the symmetric KL distance presented in the previous section. Since there are in all a minimum of 9 tunes per raaga, we consider values of k=1, 3, 5 and 7. Since the number of classes is high (10 raagas), it is more appropriate to consider a weighted-distance k-NN classification rather than simple voting to find the majority class. Weighted k-NN classification is described by the equations below. The chosen class is C^* ,

$$C^* = \arg \max_c \sum_i w_i \delta(c, f_i(x)) \quad (4)$$

where c is the class label (raaga identity in our case), $f_i(x)$ is the class label for the i^{th} neighbor of x and $\delta(c, f_i(x))$ is the identity function that is 1 if $f_i(x) = c$, or 0 otherwise. The weights are given by,

$$w_i = \frac{1}{d(x, y)} \quad (5)$$

where $d(x, y)$ is the symmetric KL distance between two pitch-class profiles x and y (e.g. its i^{th} neighbor).

The results in terms of percentage accuracy in raaga identification, obtained on the test dataset, appear in Table 5. Two important points emerge from the comparison of accuracies across the different types of pitch-class profiles.

For all values of k, except k=1, in the k-NN classification, we see that P_2 (the note segmented, duration weighted pitch-class profile) yields the highest accuracies. This implies that note durations play an important role in determining their relative prominence for a particular raaga realization. This is consistent with the fact that long sustained notes like dirgha swaras play a major role in characterizing a raaga than other functional notes which occur briefly in the beginning, the end or in the transitions. The benefit of note segmentation is seen in the slightly superior performance of P_2 over P_3 (12 bin). P_2 does not consider those instants that lie outside detected stable note regions. The second important point emerging from Table 5 is the decreasing classification accuracy with increasing bin resolution. Although the reverse might be expected in view of the widely held view that the specific intonation of notes within micro-intervals are a feature peculiar to a raaga, a more carefully designed, possibly unequal, division of the octave may be needed to observe this.

The overall best accuracy of 76.5%, which value is much higher than chance for the 10-way classification task, indicates the effectiveness of pitch-class profile as a feature vector for raaga identification. It is encouraging to find that a simple first order pitch distribution provides considerable information about the underlying raaga although the complete validation of this aspect can be achieved only by testing with a much larger number of raaga classes on larger dataset. Including the ornamentation regions in the pitch-class distribution did not help. As mentioned before, the gamakas play an important role in characterizing the raaga as evidenced by performance as well as listening practices followed. However, for gamakas to be effectively

exploited in automatic identification, it is necessary to represent their temporal characteristics such as the actual pitch variation with time. A first-order distribution which discards all time sequence information is quite inadequate for the task.

8. CONCLUSIONS

A brief but comprehensive introduction to the raaga and its properties is presented. Previous raaga recognition techniques are surveyed with a focus on their approach and contributions. Key aspects that need to be addressed are outlined and a method which deals with a few of them is discussed. Apart from these contributions of our work, we have also highlighted details such as the composition of the testing dataset, and provided insights into the post-processing steps involved with pitch extraction procedure for Carnatic music. This is the first work, to the best of our knowledge, that uses polyphonic audio recordings in the raaga recognition task.

The transitions in gamakas are discarded in the method explained, or are not fully utilized. A higher number of bins in the pitch distribution proved to be not necessarily useful. Future raaga recognition techniques can take into account the other properties of a raaga. Most important of these are the characteristic phrases and gamakas which suggest that temporal properties may be usefully exploited in future work.

9. REFERENCES

- [1] G. Geekie, "Carnatic ragas as music information retrieval entities," in *Proc. of ISMIR*, 2002, pp. 257–258.
- [2] P. Sharma and K. Vatsayan, *Brihaddeshi of Sri Matanga Muni*. South Asian Books, 1992.
- [3] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions," in *Proc. of ISMIR*, 2007, pp. 431–436.
- [4] A. Krishnaswamy, "Melodic atoms for transcribing carnatic music," in *Proc. of ISMIR*, 2004, pp. 345–348.
- [5] —, "Multi-Dimensional Musical Atoms in South Indian Classical Music," in *Proc. of the International Conference of Music Perception & Cognition*, 2004.
- [6] H. S. Powers, "The Background of the South Indian Raaga-System," Ph.D. dissertation, Princeton University, 1959.
- [7] S. R. Janakiraman, *Essentials of Musicology in South Indian Music*. The Indian Music Publishing House, 2008.
- [8] P. P. Narayanaswami and V. Jayaraman, 2004. [Online]. Available: <http://ibiblio.org/guruguha/ssp.htm>
- [9] Pratyush, "Analysis and Classification of Ornaments in North Indian (Hindustani) Classical Music," Master's thesis, University of Pompeu Fabra, 2010.
- [10] T. Viswanathan and M. H. Allen, *Music in South India*. Oxford University Press, 2004.
- [11] 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. 571–574.
- [12] K. Lee, "Automatic chord recognition from audio using enhanced pitch class profile," in *Proc. of the International Computer Music Conference*, 2006.
- [13] S. Shetty and K. Achary, "Raga Mining of Indian Music by Extracting Arohana-Avarohana Pattern," in *International Journal of Recent trends in Engineering*, vol. 1, no. 1. Acamey Publisher, 2009, pp. 362–366.
- [14] 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.
- [15] 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. 1350–1363.
- [16] M. Sinith and K. Rajeev, "Hidden Markov Model based Recognition of Musical Pattern in South Indian Classical Music," in *IEEE International Conference on Signal and Image Processing, Hubli, India*, 2006.
- [17] D. Das and M. Choudhury, "Finite State Models for Generation of Hindustani Classical Music," in *Proceedings of International Symposium on Frontiers of Research in Speech and Music*, 2005.
- [18] A. Datta, R. Sengupta, N. Dey, and D. Nag, *Experimental Analysis of Shrutis from Performances in Hindustani Music*. Scientific Research Department, ITC Sangeet Research Academy, 2006.
- [19] S. Belle, R. Joshi, and P. Rao, "Raga Identification by using Swara Intonation," *Journal of ITC Sangeet Research Academy*, 2009, vol. 23.
- [20] V. Rao and P. Rao, "Vocal melody extraction in the presence of pitched accompaniment in polyphonic music," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 8, 2010, pp. 2145–2154.