Journal of New Music Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/nnmr20

Rāga Recognition based on Pitch Distribution Methods

Gopala Krishna Koduri [a] , Sankalp Gulati [a] , Preeti Rao [b] & Xavier Serra [a]
[a] Universitat Pompeu Fabra, Spain
[b] Indian Institute of Technology Bombay, India

To cite this article: Gopala Krishna Koduri , Sankalp Gulati , Preeti Rao & Xavier Serra (2012) Rāga Recognition based on Pitch Distribution Methods, Journal of New Music Research, 41:4, 337-350, DOI: 10.1080/09298215.2012.735246

To link to this article: http://dx.doi.org/10.1080/09298215.2012.735246

Journal of New Music Research
2012, Vol. 41, No. 4, pp. 337–350

Rāga Recognition based on Pitch Distribution Methods

Gopala Krishna Koduri[1], Sankalp Gulati[1], Preeti Rao[2] and Xavier Serra[1]

[1]Universitat Pompeu Fabra, Spain; [2]Indian Institute of Technology Bombay, India

Abstract

Rāga forms the melodic framework for most of the music of the Indian subcontinent. Thus automatic rāga recognition is a fundamental step in the computational modelling of the Indian art-music traditions. In this work, we investigate the properties of rāga and the natural processes by which people identify it. We bring together and discuss the previous computational approaches to rāga recognition correlating them with human techniques, in both Karṇāṭaka (south Indian) and Hindustānī (north Indian) music traditions. The approaches which are based on first-order pitch distributions are further evaluated on a large comprehensive dataset to understand their merits and limitations. We outline the possible short and mid-term future directions in this line of work.

1. Introduction

There are two prominent art-music traditions in the Indian subcontinent: Karṇāṭaka in the Indian peninsular and Hindustānī in north India, Pakistan and Bangladesh. Both are orally transmitted from one generation to another, and are heterophonic in nature (Viswanathan & Allen, 2004). Rāga and tāla are their fundamental melodic and rhythmic frameworks respectively. However, there are ample differences between the two traditions in their conceptualization (Narmada (2001) does a comparative study of popular rāgas in both traditions).

There is an abundance of musicological resources that thoroughly discuss the melodic and rhythmic concepts (Bagchee, 1998; Sambamoorthy, 1998; Clayton, 2000; Viswanathan & Allen, 2004). However, only a few from a computational perspective are available (Levy, 1982; Krishnaswamy, 2004; Chordia & Rae, 2007; Subramanian, 2007). In this work, we discuss the past work on the rāga recognition task, and attempt to understand the impact of various sources of information (such as tonic, intonation analysis) and parameters involved in different methods based on first-order distributions, on the rāga recognition task.

In Section 2, we present the properties of rāga in detail, and in Section 3, we discuss the way trained and non-trained people recognize it. Section 4 presents a brief survey of previous computational rāga recognition approaches and a discussion of their contributions. In Section 5, we present the details of experiments which are formulated to better understand the merits and limitations of rāga recognition methods based on first-order pitch distributions. Finally in Section 6, we present the results of these experiments conducted on a comprehensive dataset consisting of both Karṇāṭaka and Hindustānī recordings, discussing them in detail. We hope this work will be of help to Indian and non-Indian readers in understanding rāga for computational purposes or otherwise. Music theory aspects referred to in this article primarily concern Karṇāṭaka music, but most of the content applies to Hindustānī music as well, unless mentioned otherwise.

2. Properties of rāga

Mātanga, in his epic treatise Bṛhaddēśī, defines rāga as 'that which colours the mind of good through a specific svara and varṇa (literally colour) or through a type of dhvani (sound)' (Sharma & Vatsayan, 1992). Each rāga

Correspondence: Gopala Krishna Koduri, Universitat Pompeu Fabra, Music Technology Group, 55.308, 138, Roc Boronat, Barcelona, 08018 Spain. E-mail: gopala.koduri@upf.edu

http://dx.doi.org/10.1080/09298215.2012.735246

therefore, can be thought of as a musical entity that leaves an impression on the minds of listeners which is shaped by the properties of constituent svaras. A technically insightful definition given by Chordia and Rae (2007), and Krishnaswamy (2004) says, 'Rāga is a collection of melodic atoms and a technique for developing them. These melodic atoms are sequences of svaras 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'. Therefore, the notion that rāga is more than a sequence of discrete svaras is important in understanding it, in order to develop a representation of rāga for computational purposes.

2.1 Svaras and their functions

The seven solfege symbols used are termed as svaras. Except for Sa and Pa svaras, the rest of them have two or three variants, which result in what are known as svarastānas, which literally mean the positions of svaras[1]. Table 1 gives the list of svarastānas with their Karṇāṭaka and Hindustānī names and the ratios they share with tonic (Shankar, 1983). Although there are 16 svarastānas in all, four of them share ratios with others (in Table 1, the svarastānas sharing the ratios are indicated with the same *Position* value). Tonic frequency is chosen according to the singer's comfort, and all the accompanying instruments are tuned accordingly. Note that the transposition of a set of svaras, i.e. shifting all of them linearly by a given interval, does not change the rāga. But making another svara Sa can result in a different rāga (see Section 7.2).

Table 1. The list of svarastānas used in Karṇāṭaka and Hindustānī music, along with the ratios shared with tonic. Note that the positions 3, 4, 10 and 11 are shared by two svarastānas each.

Symbol	Position	Ratio	Karṇāṭaka/Hindustānī name
Sa	1	1	Ṣaḍjama
R1	2	16/15	Śuddha/Kōmal Riṣabha
R2	3	9/8	Chatuśr̥ti/Tīvra Riṣabha
G1	3	9/8	Śuddha Gāṇdhāra
G2	4	6/5	Sādāraṇa/Kōmal Gāṇdhāra
R3	4	6/5	Ṣaṭśr̥ti Riṣabha
G3	5	5/4	Aṇtara/Tīvra Gāṇdhāra
M1	6	4/3	Śuddha/Kōmal Madhyama
M2	7	64/45	Prati/Tīvra Madhyama
Pa	8	3/2	Pañchama
D1	9	8/5	Śuddha/Kōmal Daivata
D2	10	5/3	Chatuśr̥ti/Tīvra Daivata
N1	10	5/3	Śuddha Niṣāda
N2	11	16/9	Kaisiki/Kōmal Niṣāda
D3	11	16/9	Ṣaṭśr̥ti Daivata
N3	12	15/8	Kākali/Tīvra Niṣāda

As aptly put by Viswanathan and Allen (2004), just like various checkers in the game of chess, svaras in rāga have different functions. Certain svaras are said to be more important than the rest. These svaras bring out the mood of the rāga. They are called the *jīva* svaras. The svara which occurs at the beginning of melodic phrases is referred to as *graha* svara. *Nyāsa* svaras are those svaras which appear at the end of melodic phrases. *Dīrgha* svaras are svaras that are prolonged. A svara that occurs relatively frequently is called *aṁsa* svara, and that which is sparingly used is called *alpa* svara, and so on. Therefore, even if two given rāgas have the same set of constituent svaras, their functions can be very different.

Rāgas can be loosely categorized into two kinds: those defined by their phraseology, and those which came into existence owing to theoretical organizations (Krishna & Ishwar, 2012). Mēḷakarta system in Karṇāṭaka and tāṭ system in Hindustānī are the most popular among them (Sambamoorthy, 1998).

In the Mēḷakarta system, there are 72 parent rāgas which are obtained through combinations of the 12 svarastānas with following conditions in place: only one variant of a svara is allowed in a combination, all the svaras are to be represented, two given combinations differ by at least one svarastāna, and the svaras should be straightly ordered with no *vakra*[2] pattern. The rāgas thus obtained are called janaka rāgas, literally the parent rāgas. The others, called janya/child rāgas are, in theory, derived from them. However, several janya rāgas predate the janaka rāgas by centuries indicating that the Mēḷakarta system serves mainly the academic and theoretical purpose of organizing rāgas. Such janya rāgas are phraseology based. The combinations of svarastānas which did not exist prior to the organization, evolved as new rāgas, which are primarily progression based (see Section 2.2). In Hindustānī, tāṭs provide a basic categorization system for the most prevalent rāgas and this doesn't include a comprehensive schema for their organization. They are considered as parent rāgas from which the other rāgas are derived. There are 10 tāṭs, each consisting of seven svaras, formed by the different combinations of the variants of the svaras. However there are many borderline rāgas that do not strictly come

[1]Svara and svarastāna are normally used interchangeably in this article, as elsewhere, except when the distinction is necessary.

[2]Vakra in Sanskrit literally means twisted. In this context, it means the order of the svaras is twisted/abnormal

under one tāṭ and can be classified into either of the two tāṭs.

2.2 Arōhaṇa and avarōhaṇa

A rāga is typically represented using the ascending (arōhaṇa) and descending (avarōhaṇa) progressions of the constituent svaras. The order of the svaras in both progressions determine the usage of the svaras in building melodic phrases. The svaras in ascending progression can only be used in melodic phrases which are ascending in nature, and vice versa. This seems to be especially important if the rāga has either differing sets of svaras in the progressions (e.g. Bhairavi rāga in Karṇāṭaka), or there is a vakra pattern in any of them (e.g. Saurāṣṭraṁ rāga in Karṇāṭaka). In the first case, it is imperative that the differing svaras are either used only during ascents or descents. In the latter case, the twisted svara positions allow few transitions which otherwise would not be possible. However, it has been noted that these progressions are not so relevant to the phraseology-based rāgas (Krishna & Ishwar, 2012).

2.3 Gamakas

Given a svara, rapid oscillatory movement about it is one of the several forms of movements, which are together known as gamakas. Another form of gamaka involves making a sliding movement from one svara to another. There are a number of such movements discussed in musicological texts (Narayanaswami & Jayaraman, 2012). Further, there are also various ways to classify these movements. The most accepted classification speaks of 15 types of gamakas (Janakiraman, 2008; Narayanaswami & Jayaraman, 2012). There are a few constraints for a svara to be sung with gamaka. For instance, if the given rāga has R1 and G1 svaras, R1 can not take a gamaka since G1 is very close and it is difficult to sing a gamaka on R1 without touching G2, which would result in violating rāga's properties[3]. Further, Sa and Pa do not take any gamakas, excepting glides (they are hence called achala/immovable svaras).

Gamakas bear a tremendous influence on how a tune is perceived. They are often considered the soul of these art-music traditions. Though gamakas are used in both Karṇāṭaka and Hindustānī, the pattern of usage is very distinct. Besides gamakas, there are alankāras (literally ornamentations) which are patterns of svara sequences which beautify and enhance the listening experience. On this note, we would like to emphasize that gamakas are not just decorative patterns or embellishments (whereas alankāras are), they are very essential to the definition of rāga. Krishna and Ishwar (2012) discuss various manifestations of the most important gamaka in Karṇāṭaka music, called Kaṁpita.

Unlike several other music traditions where music notation is a crucial source of information during learning as well as performing, notation is very sparingly used. Indeed, it is considered just a memory aid. One of the multitude of possible reasons can be the difficulty in notating gamakas, owing to the complexity of movements.

2.4 Characteristic phrases

If is often noted by musicians and musicologists that a rāga can only be learnt by getting familiar with several compositions in it. Any given rāga has a repertoire of characteristic phrases, each of which encapsulates its properties. Typically in a concert, the artist starts with singing these phrases. These are also the main clues for listeners to identify rāga. This pool of phrases for a rāga keeps evolving over time. Often the new phrases come from popular compositions in the rāga.

In addition to the properties of rāga which are discussed, Hindustānī music also emphasizes the time and season a rāga should be used in. They seem less relevant in Karṇāṭaka music today.

That said, a rāga is an evolutionary phenomenon. It continually takes place over time; no existing rāga was perceived the way it is today. For instance, Mēchakaḷyāṇi rāga was supposedly a less common rāga with less scope for improvisation in the past. However, today it is one of the most common rāgas of Karṇāṭaka music. The properties which enhance the characteristic nature of a rāga are retained and others are done away with. This process happens continually over decades and centuries. The rāga takes its shape and sets a unique mood depending on these properties.

3. Recognition of rāga by humans

Given these intricate properties, the task of identifying a rāga can seem overwhelming. But the seasoned rasikas[4] identify rāga within a few seconds of listening to a performance. Though there are no rules of thumb in identifying rāga, expert musicians believe that broadly there are two procedures by which people identify it from a composition. This normally depends on whether the person is a trained musician or a rasika. People who have not much knowledge of rāgas cannot identify them unless they memorize the compositions and their rāgas.

[3]A Karṇāṭaka musician and trainer explains this taking an example from Karṇāṭaka music in this podcast episode: http://raagarasika.podbean.com/2008/09/30/episode-15-featured-raaga-sivaranjani/

[4]A term often used for a seasoned Karṇāṭaka music listener, which literally means *the one who enjoys art*.

3.1 Intuitive approach based on listening (rasikas)

In a nutshell, the procedure followed by rasikas typically involves correlating two tunes based on how similar they sound. Years of listening to tunes composed in various rāgas gives a listener enough exposure. A new tune is compared 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: rules in transition between svaras imposed by ārōhaṇa and avarōhaṇa, characteristic phrases, usage-pattern of a few svaras and gamakas.

This process depends heavily on the cognitive abilities of a person. Without enough previous exposure, it is not feasible for a person to attempt identifying rāga. There is a noteworthy observation on this approach. Though many people cannot express in a concrete manner what the properties of rāga are, they are still able to identify it. This very fact hints at a possible supervised classifier, which can take advantage of the properties of rāga.

3.2 Analytical approach based on training (musicians)

A musician tries to identify the characteristic phrases of rāga. These are called svara sañchāras in Karṇāṭaka and pakaḍs in Hindustānī. If the musician finds these phrase(s) in the tune being played, rāga is immediately identified. In some cases, musicians play the tune on an instrument (imaginary or otherwise) and identify the svaras being used. They observe the gamakas used on these svaras, locations of various svaras within the melodic phrases and the transitions between svaras.

This latter process seems to use almost all the characteristics of rāga. It looks more systematic in its structure and implementation. The procedures used by trained musicians and non-trained listeners provide useful insights for implementing a rāga recognition system. As we will see, the existing approaches try to mimic them as much as possible. They can broadly be classified as example-based or knowledge-based or both. The example-based approaches correspond to the intuitive approach used by rasikas to identify a rāga, such as matching similar phrases. The knowledge-based approaches reflect the analytic approach which is employed by the trained musicians, such as identifying the svaras, their roles and gamakas.

4. Survey of rāga recognition approaches

Given several attributes of a rāga that serve to distinguish it from all the other rāgas, and the fact that rāgas are learnt by listening and imitation rather than by an analytical application of rules, there appear to be no clear-cut guidelines for the machine recognition of rāgas. The lay listener's intuitive approach suggests a loosely constrained machine learning strategy from a large rāga-labelled audio database of compositions.

Chakravorty, Mukherjee and Datta (1989) proposed the machine recognition of rāgas from notation. They use a kind of scale-matching i.e. set of permitted svaras (or rather, set of forbidden svaras) for a first-level identification. Next a knowledge-based approach is used via a lexicon of phrases of each rāga. The input notation is segmented into approximate ārōhaṇa-avarōhaṇa sections (for each candidate rāga considered). Then lexical matching is carried out, first with exact sequences, then with a coarse search allowing partial matching. The system is evaluated on 75 rāgas distributed over 45 scale classes[5] in all.

Inspired by Upadhye and Sahasrabuddhe (1992) on the use of a finite automaton to generate rāga svara sequences, Pandey, Mishra and Ipe (2003) used a generative statistical model in the form of a hidden Markov model (HMM) for each rāga. A sequence of svaras was automatically extracted from solo vocal recording by applying a heuristics driven note-segmentation technique. The individual svaras form the states. The HMM (actually, just MM since nothing is 'hidden') that best explained the observed svara sequence was the detected rāga. Thus sequential information is exploited subject to the limitations of a first-order Markov model. The same work also proposed phrase matching expressed as an approximate substring search for the pakaḍ (catch phrase) of the rāga. In another method, the rāga was identified by counting the occurrences of n-grams of svaras in the pakaḍ. The evaluation was restricted to discriminating two rāgas. The central idea in this work, which is to model a rāga as HMM, was also used by Sinith and Rajeev (2006). The same idea was used in an attempt to automatically generate Hindustānī music (Das & Choudhury, 2005), but with less success.

Chordia and Rae (2007) used pitch-class profiles to represent the distributions and hence the relative prominences of the different svaras. They also used svara bi-gram distributions to capture some sequential information. Using just the pitch-class profiles to classify 17 rāgas (142 audio segments of 60 s each), the system achieves an accuracy of 78%. Using only the bi-grams of pitches, the accuracy is 97.1%. Using both was reported to give an almost perfect result. Gedik and Bozkurt (2010) present a similar approach for makam classification in Turkish-makam music, by matching pitch histograms with higher bin resolution.

Sridhar and Geetha (2009) have followed an approach where the set of svaras used in an audio recording is estimated, and compared with the templates in the database. The rāga corresponding to the best matched template is selected as the class label. Their test data

[5]Rāgas in each scale class have identical svaras but different phrases.

consisted of 30 tunes in three rāgas sung by four artists, out of which 20 tunes are correctly labelled by the system. The tonic is manually input, and the other svaras are identified based on the respective ratio with the tonic. A similar approach based on detecting the svaras used in ārōhaṇa and avarōhaṇa to find the rāga is presented by Shetty and Achary (2009).

In Indian art music, a svarastāna does not correspond to a fixed pitch class. It is a region (Datta, Sengupta, Dey, & Nag, 2006; Krishna & Ishwar, 2012). Therefore, though two given rāgas share the same scale, the precise intonation of specific svaras can vary significantly. Belle, Joshi and Rao (2009) have used this information to differentiate rāgas that share the same scale intervals. They evaluated the system on 10 tunes, with four rāgas evenly distributed in two distinct scale groups, and showed that the use of svara intonation features improved the accuracies achieved with pitch-class profiles alone.

The approaches surveyed can be broadly categorized into one of the following: first-order distributions, higher-order distributions and phrase detection. Of these, the first-order distributions are the most explored. In Section 5, we discuss a few experiments which fall into this category, testing the impact of different sources of information and also the effect of changing different parameters specific to each approach. In Section 6, we report the results of these experiments conducted on a very comprehensive dataset compared to the ones on which they were tested in past literature. This will help us to better understand the merits and limitations of each approach. Later, we discuss the future direction to this work.

5. First-order pitch-distribution based approaches

The datasets used in the surveyed rāga recognition approaches are not representative enough for several reasons. Pandey et al. (2003) and Sridhar and Geetha (2009) used datasets which had as few as two or three rāgas. The datasets were also constrained to some extent by the requirement of monophonic audio for reliable pitch detection. The dataset used by Chordia and Rae (2007) is also quite limited. The data available on a popular commercial online music portal such as raaga.-com (> 500 performers, > 300 rāgas)[6], shows that there is scope to improve the quality and size of the data used for the task. Therefore the conclusions drawn from the existing experiments can not be claimed to be general.

In this section, we discuss different approaches to obtain rāga-specific measurements from a first-order distribution of the continuous pitch The results are reported and discussed in Section 6.

5.1 Template matching

In this approach, the set of svaras identified in a given recording are matched against the rāga templates and then the rāga corresponding to the template that scores the highest is output as the class label. We use two different methods to determine the svaras present in a given recording and to obtain the rāga templates.

In both the methods, from a given recording, pitch is extracted and a high-resolution octave-folded histogram (1200 bins) aligned to the tonic of the recording is obtained. In the first method (A_{th}), we consider the set of svaras of a rāga as defined in theory as the template for the respective rāga. From the histogram of a given recording, we obtain the values of bins at locations corresponding to the 12 just intonation intervals. The top seven are taken as the svaras used in the recording. In the second method (A_{de}), the rāga template is obtained by averaging tonic-aligned histograms corresponding to individual recordings in the rāga, and picking the most salient peaks, a maximum of seven. The svaras used in an individual recordings are also inferred in the same manner.

5.2 Distributions constrained to ‘steady regions’

The pitch contour obtained from the recording may be used as such to obtain a pitch-class distribution. On the other hand, given the heavy ornamentation in Indian art-music (see Figure 1), computing pitch-class distributions using only the stable pitch regions in the melody may be more beneficial.

In order to determine a stable region in pitch contour, the local slope of the pitch contour is used to differentiate stable svara regions from connecting glides and ornamentation (Pandey et al., 2003). At each time instant, the pitch value is compared with its two neighbours to find

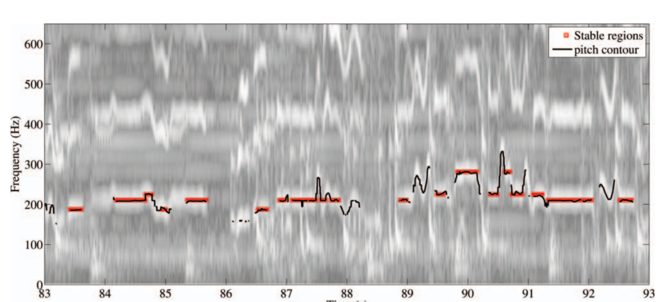


Fig. 1. The pitch contour is shown superimposed on the spectrogram of a short segment from a Karṇāṭaka vocal recording along with the identified stable pitch regions.

[6]Observations made on 25 May 2012.

the local slope in each direction. If the magnitude of either of the local slopes lies below a threshold value T_{slope}, the current instant is considered to belong to a stable svara region:

$$|F(i-1) - F(i)| < T_{\text{slope}} \quad \text{or} \quad |F(i+1) - F(i)| < T_{\text{slope}} \tag{1}$$

where F is the pitch contour converted to cent scale. All the instances where the slope is beyond T_{slope} are discarded as they don't belong to the stable regions. Finally, the pitch values in the segmented stable svara regions are quantized to the nearest available svara value in the just intonation scale using the tonic. This step helps to smooth out the minor fluctuations within intended steady svaras. Figure 1 shows a continuous pitch contour with the corresponding segmented and labelled svara sequence superimposed.

A pitch-class profile is computed using only the stable svaras thus obtained, and hence is a 12-bin histogram corresponding to the octave-folded values quantized to just intonation intervals. There are two choices of weighting for histogram computation. We call the pitch-class profiles corresponding to those two choices as $P_{\text{instances}}$ and P_{duration}, where former refers to weighting a svara bin by the number of instances of the svara, and the latter refers to weighting by total duration over all instances of the svara in the recording. In each case, results are reported for different values of T_{slope}. Further, we also experimented setting a minimum time threshold (T_{time}) to pick the stable regions.

5.3 Distributions obtained from full pitch contour

In this approach, we consider the whole pitch contour without discarding any pitch values. We call this $P_{\text{continuous}}$. In this case, we consider different bin resolutions for quantization in constructing the histogram to observe its impact. This step is motivated by the widely discussed microtonal character of Indian art music (Krishnaswamy, 2004).

For all the classification experiments of Sections 5.2 and 5.3, we need a distance measure and a classifier to perform rāga recognition. A good distance measure for comparing distributions should reflect the extent of similarity between their shape. Further, we would also like to observe the impact of adding tonic information. Therefore, we conduct experiments twice: with tonic and without it. To facilitate this, the distance measure should also facilitate comparing pitch-class profiles in the absence of tonic information. We choose the Kullback–Leibler (KL) divergence measure as a suitable measure for comparing distributions. Symmetry is incorporated into this measure by summing the two values as given below (Belle et al., 2009).

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

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

where i refers to the bin index in the pitch distribution, and P and Q refer to the pitch distributions of two tunes. In the cases where tonic information is not available, we consider all possible alignments between P and Q, and choose the one that scores best in terms of minimizing the distance measure.

The k-NN classifier is used in conjunction with the selected distance measure. Results are reported over several values of k. In a leave-one-out cross-validation experiment, each individual tune is considered a test tune while all the remaining constituted the training data. The class label of the test tune is estimated by a simple voting method to determine the most recurrent rāga in the k nearest neighbours. The selection of the class label C is summarized in the following equation:

$$C = \arg\max{}_c \sum_i \delta(c, f_i(x)) \tag{4}$$

where c is the class label (rāga identity in our case), $f_i(x)$ is the class label for the ith neighbour of x and $\delta(c, f_i(x))$ is the identity function that is 1 if $f_i(x) = c$, or 0 otherwise.

Intonation of a given svara varies from rāga to rāga based on several factors which emerge from the properties of the rāga. Therefore, it constitutes another important measure to distinguish rāgas. Gamakas play a vital role in the perception of Indian art music (Viswanathan & Allen, 2004; Janakiraman, 2008). As described in Section 2.3, relative positions of svaras play an important role in determining which svaras can take gamakas. Depending on the form of gamaka, the perceived intonation of a svara on which the gamaka is sung might change. We hypothesize that this information can be obtained, by parameterizing svaras with a continuous-time, continuous-value pitch distribution as the underlying representation. The roles of different svaras influence the distribution around each svara. For instance, the svara which is only touched upon in transitions but never elongated will have a distribution around it which is different as compared to the distribution around the svara which is often rested upon. We designed an exploratory rāga classification experiment in order to test this, in which three rāgas are distinguished using the features of just a single peak.

5.4 Common computational steps

In all our experiments, we use pitch contour, tonic information and histogram analysis. Here we briefly explain these computational steps.

5.4.1 Pitch extraction

To accurately mark the F0 of the stable pitch regions, the estimation errors that are generated by all F0 detection methods need to be minimized. In many portions of the vocal recording used, the accompanying violinist fills the short pauses of the vocalist, and also very closely mimics the vocalist with a small time lag. This is one of the main problems we encountered when using pitch tracking algorithms like YIN (Cheveigné & Kawahara, 2002): violin was also being tracked in a number of portions. As it is usually tuned an octave higher, this resulted in spurious pitch values. To overcome this, we use predominant melody extraction (Salamon & Gómez, 2012) based on multi-pitch analysis. But this has an inherent quantization step which does not allow high bin resolutions in histogram analysis. So we use a combination of both. In each frame, we transform the estimated pitch value from both methods into one octave and compare them. In those frames where they agree within a threshold, we retain the corresponding YIN pitch transforming it to the octave of the pitch value from multi-pitch analysis. We discard the data from frames where they disagree with each other. On average, data from 53% of the frames is retained. Though it is a computationally intensive step, this helps in obtaining clean pitch tracks, which have less F0 estimation errors. The frequencies are then converted to cents. We use tonic information as the base frequency when it is available, otherwise we use 220 Hz. The octave information is retained.

Figure 1 shows the output pitch track superimposed on the signal spectrogram for a short segment of Karṇāṭaka vocal music where the instrumental accompaniment comprised violin and mr̥daṅgaṁ (percussion instrument with tonal characteristics). We observe that the detected pitch track faithfully captures the vocal melody unperturbed by interference from the accompanying instruments.

5.4.2 Tonic identification

Tonic is the base pitch chosen by a performer that allows one to fully explore the vocal (or instrumental) pitch range in a given rāga exposition. This pitch serves as the foundation for melodic tonal relationships throughout the performance and corresponds to Sa svara of rāga. All the accompanying instruments are tuned in relation to tonic of the lead performer. The artist needs to hear tonic throughout the concert, which is provided by the constantly sounding drone that plays in the background and reinforces tonic. The drone sound is typically provided by Tāmpura (both acoustic and electric), sr̥tī box or by sympathetic strings of instrument such as Sitār or Vīṇa.

For computational analysis of Indian art music, tonic identification becomes a fundamental task, a first step towards many melodic/tonal analyses including rāga recognition, intonation analysis and motivic analysis. There has not been much research done in the past on tonic identification. However recent studies have reported encouraging results. Ranjani, Arthi and Sreenivas (2011) explores culture-specific melodic characteristics of Karṇāṭaka music that serve as cues for tonic (Sa svara). A more general approach applicable to both Karṇāṭaka and Hindustānī music is proposed by Salamon, Gulati and Serra (2012), which takes advantage of the presence of drone sound to identify tonic. However each of these approaches have there own limitations and requirements. We followed the approach proposed by Salamon et al. (2012), which is based on multi-pitch analysis of the audio data, and automatically learned set of rules (decision tree) to identify the tonic pitch. We evaluated our implementation on the same database that the authors had used and achieved nearly the same results (93% accuracy for 364 vocal excerpts of Hindustānī and Karṇāṭaka music).

5.4.3 Histogram computation

The pitch contour in cents is folded to one octave. Given the number of bins and the choice of weighting, the histogram is computed:

$$H_k = \sum_{n=1}^{N} m_k, \tag{5}$$

where H_k is the kth bin count, $m_k = 1$ if $c_k \leq F(n) \leq c_{k+1}$ and $m_k = 0$ otherwise, F is the array of pitch values and (c_k, c_{k+1}) are the bounds on the kth bin. If the histogram is weighted by duration, N is the number of pitch values. If it is weighted by the number of instances, the pitch contour is first segmented and N corresponds to the number of segments.

6. Data and experiments

A well annotated and comprehensive database is of fundamental value for this field of research. The Karṇāṭaka and Hindustānī datasets, taken from the growing collections of the CompMusic project (Serra, 2011), provide a convenient mechanism of organization and retrieval of audio and metadata (Serra, 2012). Statistics reported in Table 2 are indicative of their comprehensiveness. We use full-length recordings, which range in length from 2 to 50 min. The dataset encompasses all the possible improvisational and compositional forms of Karṇāṭaka music, and Dr̥pad and Khayāl[7] genres of Hindustānī music.

[7]The other genres in Hindustānī include Ghajal, Ṭhumrī, Kavvāli etc., which are classified as semi-classical in nature.

Table 2. Current size of Karṇāṭaka and Hindustānī collections in CompMusic database.

Collection	Recordings	Rāgas	Artists
Karṇāṭaka	982	200	60
Hindustānī	433	156	56

Table 3. Different datasets derived from CompMusic collections.

Task	Collection	Recordings/rāga	Rāgas
Template matching	Karṇāṭaka	7	14
	Hindustānī	4	17
Pitch-distribution based approaches	Karṇāṭaka	5	43
		5	12
		10	12
	Hindustānī	5	16
		5	8
		8	8

6.1 Template matching

The template matching approaches rely on just the svara positions. To evaluate such approaches, it is necessary to have a dataset which is representative of 72 Mēḷakarta rāgas for Karṇāṭaka music and 10 tāṭ families in Hindustānī music. In the Mēḷakarta scheme of rāgas, each one differs by a svara from its neighbours. Hence, having several pairs of rāgas in the dataset such that they are also neighbours in the Mēḷakarta scheme contributes to the dataset's completeness. However, it is difficult to find recordings for most of the Mēḷakarta rāgas. The Karṇāṭaka and Hindustānī datasets we chose for this purpose consist of 14 and 17 rāgas respectively (see Table 3). The rāgas are chosen such that they differ in constituent svaras.

Figure 2 shows the confusion matrices for both the methods (A_{th}, A_{de}, Section 5) over Karṇāṭaka and

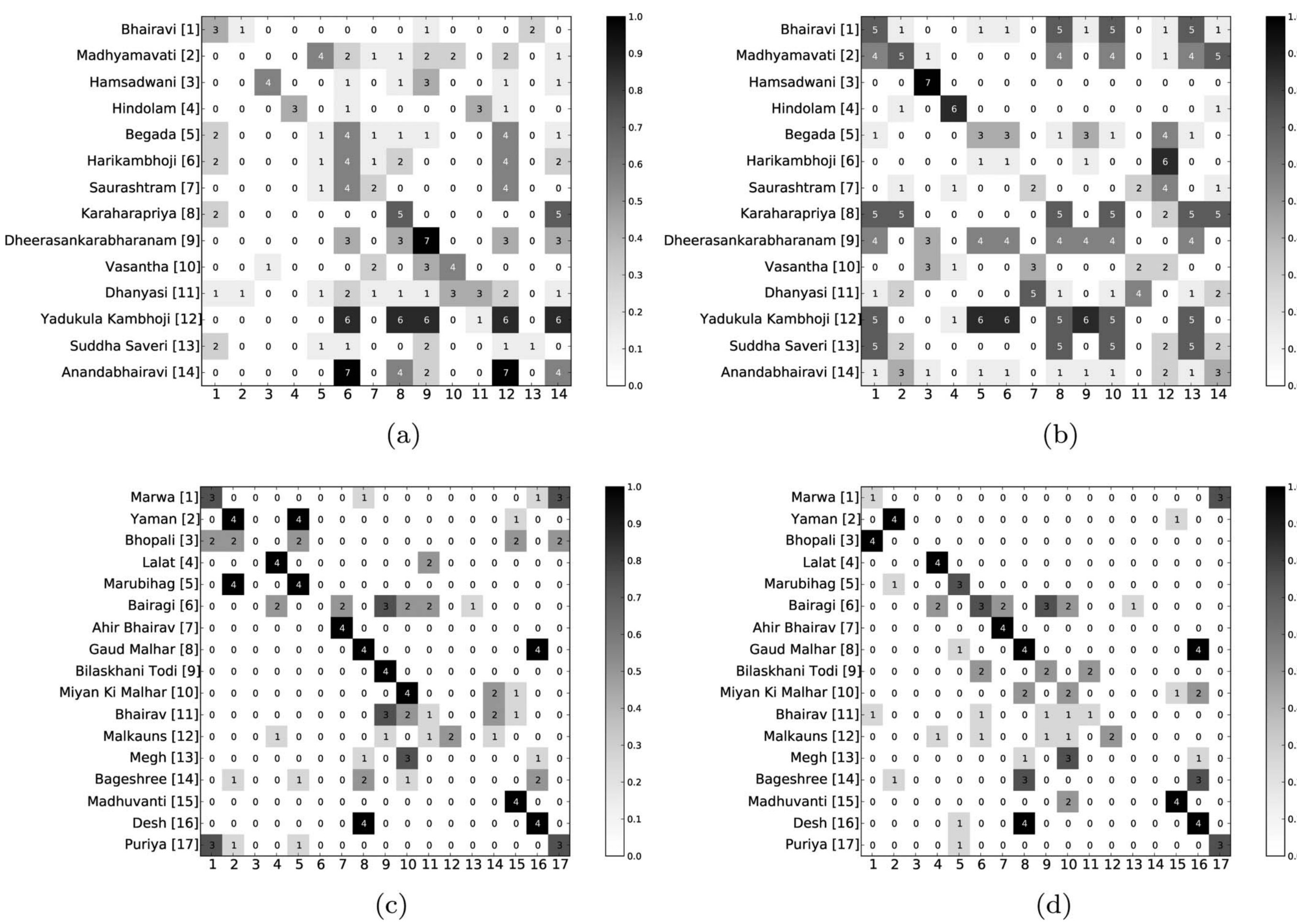


Fig. 2. Confusion matrices for the two template matching methods (A_{th} and A_{de}) on Karṇāṭaka and Hindustānī datasets. The greyness index of the (x, y) cell is proportional to the fraction of recordings in class y labelled as class x. (a) A_{th} on Karṇāṭaka dataset. (b) A_{de} on Karṇāṭaka dataset. (c) A_{th} on Hindustānī dataset. (d) A_{de} on Hindustānī dataset.

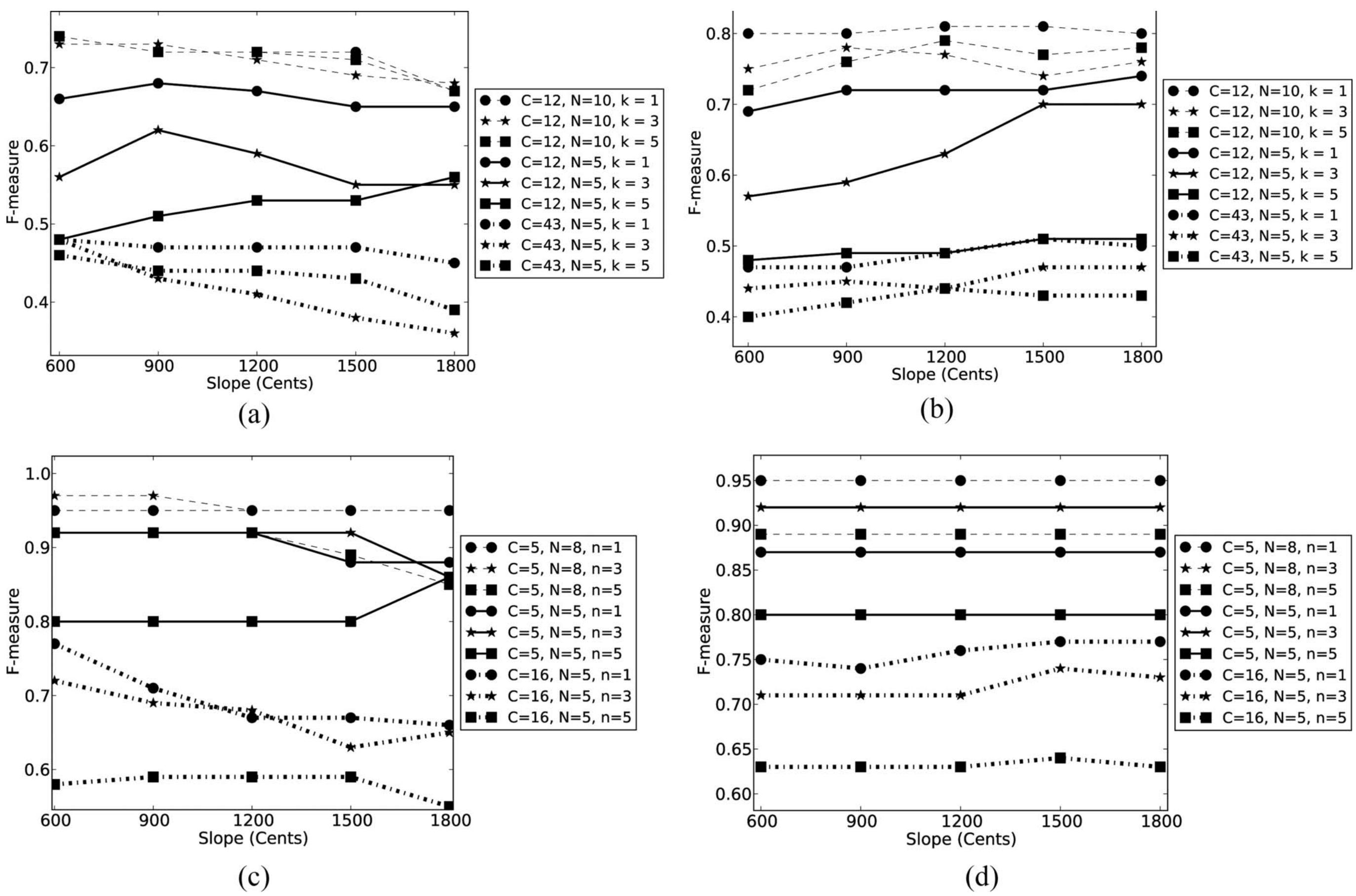


Fig. 3. F-measures for performances of $P_{\text{instances}}$ and P_{duration} on Karṇāṭaka and Hindustānī datasets, with T_{time} set to 0 and T_{slope} varied between 600 to 1800. C and N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbours in the k-NN classification. (a) Karṇāṭaka datasets with P_{instance}. (b) Karṇāṭaka datasets with P_{duration}. (c) Hindustānī datasets with P_{instance}. (d) Hindustānī datasets with P_{duration}.

Hindustānī datasets. F-measures for (A_{th}, A_{de}) over Karṇāṭaka and Hindustānī datasets are (0.26, 0.23) and (0.35, 0.39) respectively. In both the methods, there are cases where the actual rāga is correctly matched, and also the cases where there is another rāga which scores equal to the actual rāga.

It is interesting to note that A_{th} suits Karṇāṭaka and A_{de} suits Hindustānī in comparative terms though the difference in their performances is marginal. This can be a reflection of the fact that svaras in Karṇāṭaka music rarely occur without gamakas, which influence the peak characteristics of svaras in the histogram. As a result, the peaks corresponding to such svaras often appear as slides with a little bump, which can not be identified using a conventional algorithm to find local maxima and have a good probability to be accounted for only in A_{th}; whereas in Hindustānī music where the svaras are held relatively steady, A_{de} performs marginally better than A_{th}. On the other hand, it is to be noted that the methodologies which we have adopted to obtain the svaras used in a given recording are not the best and can be improved. A further step ahead would be to obtain histograms from stable pitch regions. In order to quickly test if this helps, from the stable pitch regions obtained, we picked the most recurrent intervals from each recording and matched those against templates obtained in A_{th} and A_{de}. The accuracies obtained are not very different from those reported in Figure 2. This observation reinforces our belief that rāga classification using a template matching approach alone cannot be scaled to classify, say, even just the Mēḷakarta rāgas.

6.2 Distributions constrained to 'steady regions'

When the number of rāga classes also include janya rāgas of a few of the Mēḷakarta rāgas in the dataset, we will require additional information beyond svara positions. Pitch-class profiles contain the information about relative usage of svaras besides their positions. Though several of such Mēḷakarta-janya rāga groups can possibly be distinguished using the template-matching approaches, for most cases we expect the additional information from pitch-class profiles to contribute for a better classification.

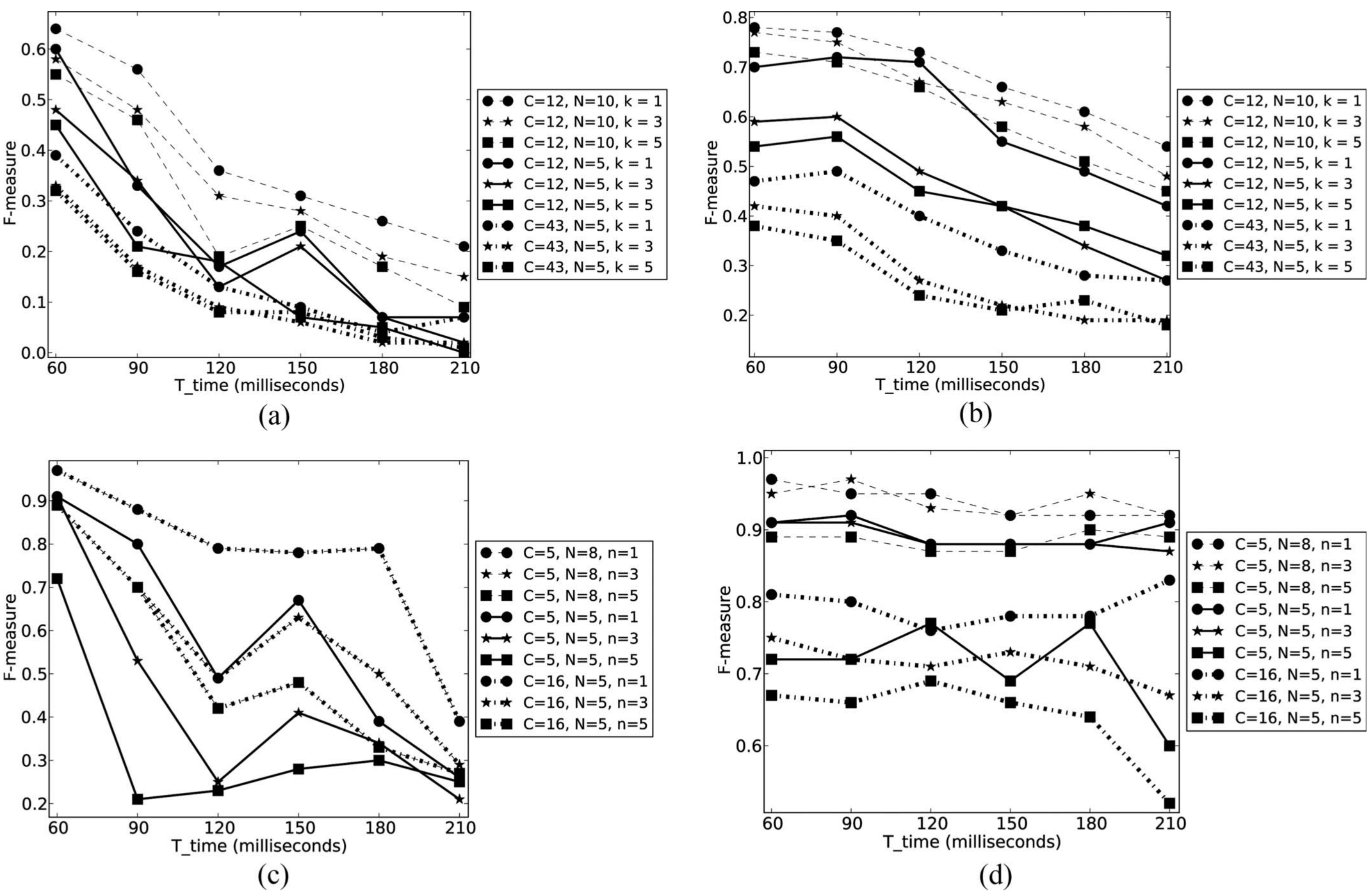


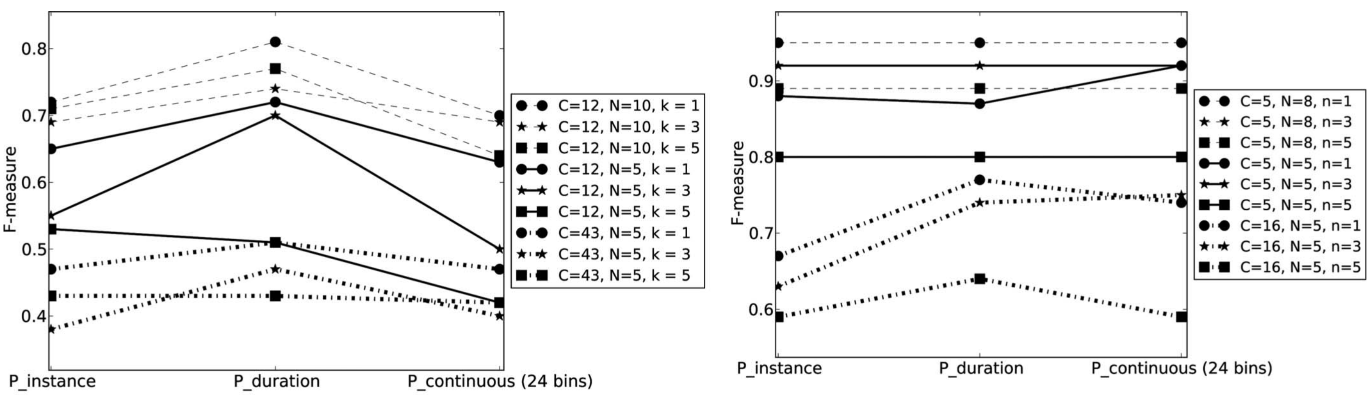
Fig. 4. F-measures for performances of $P_{\text{instances}}$ and P_{duration} on Karṇāṭaka and Hindustānī datasets, with T_{slope} set to 1500 and T_{time} varied between 60 to 210. C and N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbours in the k-NN classification. (a) Karṇāṭaka datasets with P_{instance}. (b) Karṇāṭaka datasets with P_{duration}. (c) Hindustānī datasets with P_{instance}. (d) Hindustānī datasets with P_{duration}.

There are two crucial factors in determining stable svara regions: slope threshold and time-duration threshold (T_{slope} and T_{time}, Section 5.2). We conducted three experiments to check their impact on the performances of P_{instance} and P_{duration}. The datasets chosen for this task are listed in Table 3. The datasets are chosen to facilitate observation of the impact of number of rāgas, and number of recordings per rāga in all the experiments.

In the first experiment, T_{time} is set to 0 and T_{slope} is varied from 600 to 1800 cents in steps of 300. Figure 3 shows the results. The performance of P_{duration} stays the same while that of P_{instance} slightly degrades with increasing T_{slope}. With lower T_{slope}, we observed that a svara sung with even a slight inflection is divided into multiple segments, which primarily effects P_{instance}. Better performance of P_{duration} over P_{instance} in general, also explains the slight increase in the performance of P_{instance} at lower T_{slope}.

In the second experiment, T_{slope} is set to 1500 and T_{time} is varied from 60 to 210 ms, in steps of 30. Figure 4 shows the results. With increasing T_{time}, the amount of pitch data shrinks drastically for Karṇāṭaka recordings. This taxes the classification performance heavily (see Figures 4(a) and (b)). Further, the effect is even more pronounced on the performance of P_{instance}. On the other hand, these observations are not as strong for the results over Hindustānī datasets (see Figures 4(c) and (d)). This can be explained by the presence of long steady svaras in Hindustānī music, which aligns with our observations in Section 5.1.

In the third experiment, we set T_{slope} to 1500 and T_{time} to 0, and classified Karṇāṭaka and Hindustānī datasets using $P_{\text{instances}}$ and P_{duration} and $P_{\text{continuous}}$ (24 bins) to compare their performances. Figure 5 shows the results. P_{duration} outperforms the other two, which is more evident in the classification of Karṇāṭaka rāgas. This implies that svara durations play an important role in determining their relative prominence for a particular rāga realization. This is consistent with the fact that long sustained svaras like dīrgha svaras play a major role in characterizing a rāga than other functional svaras which occur briefly in the beginning, the end or in the transitions. The benefit of identifying stable svara regions is seen in the superior performance of P_{duration} over $P_{\text{continuous}}$ (24 bins).



(a) (b)

Fig. 5. Comparison of the performances of different pitch class profiles ($P_{\text{instances}}$, P_{duration}, $P_{\text{continuous}}$ (24 bins)) on Karṇāṭaka and Hindustānī datasets. C and N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbours in the k-NN classification. (a) Karṇāṭaka. (b) Hindustānī.

From Figure 5(b), it can be seen that as the number of classes increase, the performance of P_{duration} is less affected compared to others. It is possible that a fall in number of samples per class causes it, but this argument can be ruled out as there is no noticeable difference between the results based on the three pitch distributions when we keep the number of classes at five for Hindustānī and 12 for Karṇāṭaka, and vary the number of samples per class (see Figure 5). Further, a considerable rise in f-measures with an increase in the number of samples per class, for all three distributions, indicate that there is a large diversity in the performances of a rāga. This also falls in line with the general notion that one has to become familiar with a number of compositions in a rāga in order to learn it.

6.3 Distributions obtained from full pitch contour

In this experiment, we would like to see the impact of bin resolution and tonic information on classification performance. We vary bins from 12 to 1200, and evaluate the performance of pitch distributions thus obtained, with and without tonic information. Figure 6 shows the results. For both Karṇāṭaka and Hindustānī datasets. when the tonic information is available, pitch distributions with a higher number of bins performed better. But it is also clear that beyond 24 bins, the accuracies for each combination of k and dataset, more or less remain saturated. In the case where tonic information is not given, however, there is a slight but comparatively steady increase in the accuracies with increasing number of bins.

However, the tonic identification method we used has an error rate of 7%. The results reported in Figures 6(a) and (c), therefore carry this error too. In order to realize the impact of tonic information on the performance of the system, we have analysed the cases where the systems with and without tonic information failed. In the cases where a correct class label is output, let T_s and N_s be the set of cases where the tonic information is used and not used respectively. Then, $|T_s - N_s|/N$, where N is total number of recordings in the dataset, is the proportion of cases where the availability of tonic information has helped in improving the accuracy. This comes out to be 5% for most configurations run with $P_{\text{continuous}}$ (12 bins). As there is a 7% inherent error in the tonic identification method, we expect this proportion to go up further.

A further additional source of information to distinguish rāgas is the intonation of svaras, which to a limited extent, can be inferred from continuous pitch distributions. In this case, the allied rāgas, which are the rāgas that share the same set of svaras, will make an ideal contribution to the dataset. Because, it is often the differences in intonation and gamakas of the svaras that they are distinguished with, our dataset consists of 42 recordings distributed in three allied rāgas from Karṇāṭaka music.

We parametrized the peak corresponding to each svara, and have shown that this information indeed is helpful in distinguishing a selected subset of rāgas (Koduri, Serrà, & Serra, 2012). The method consists of the following steps. The prominent vocal segments of each performance are extracted using a trained support vector machine (SVM) model. Then, the pitch corresponding to the voice is extracted using multi-pitch analysis. Using all the performances of each rāga, an average pitch histogram for every rāga is computed and its prominent peaks detected. In the following step, we compute the pitch histogram for each single performance, detecting the relevant peaks and valleys using information from the average histogram of the corresponding rāga. Each peak is characterized by using the valley points and an empirical threshold. Finally, each of the peak distributions are characterized by parameters such as peak position, height, mean, skewness and kurtosis.

In an exploratory rāga classification task in which the three allied rāgas are distinguished using the features of just a single peak, the results indicate that intonation

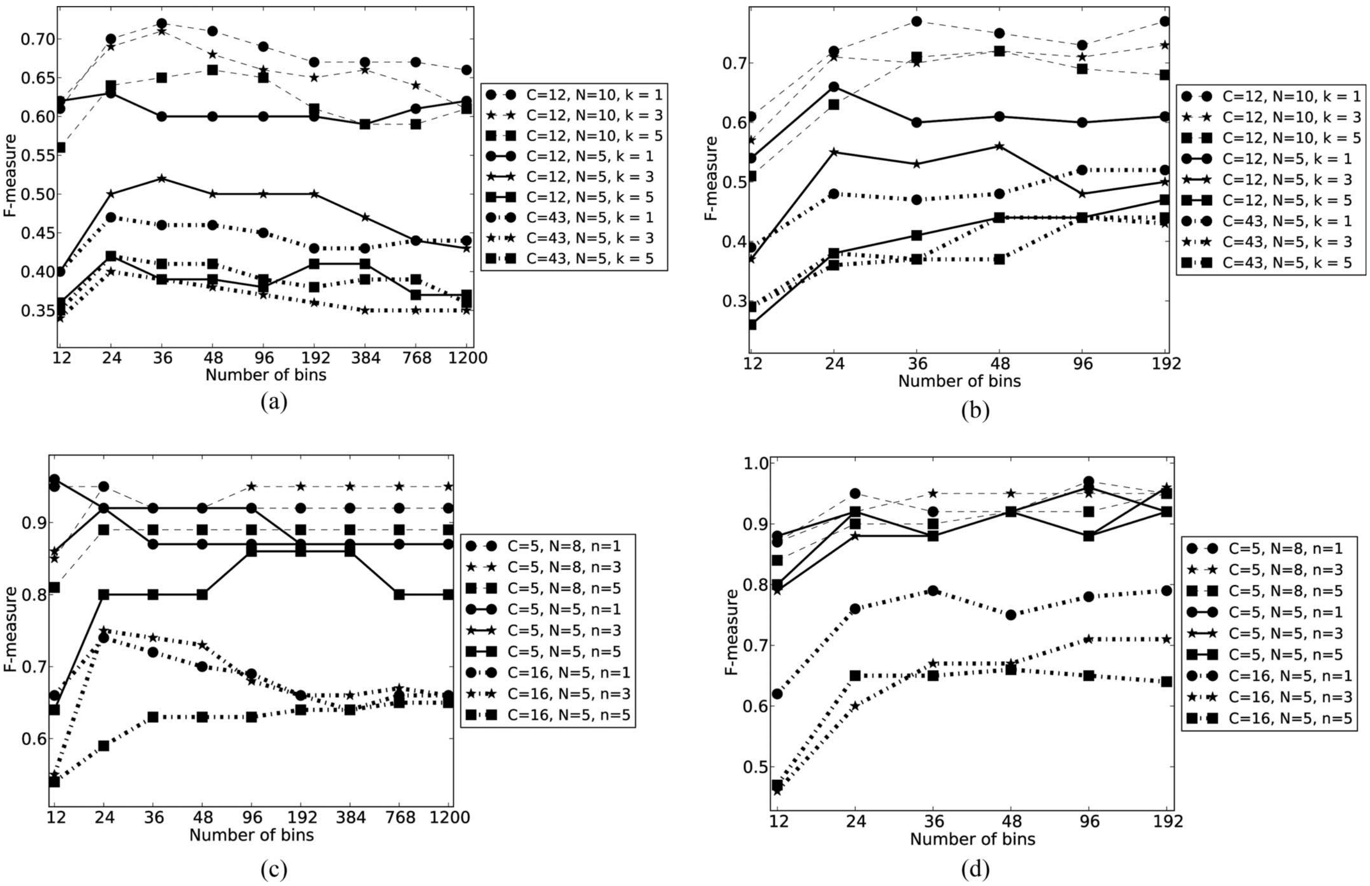


Fig. 6. Comparison of the performances of $P_{\text{continuous}}$ with different bin resolutions on Karṇāṭaka and Hindustānī datasets. C and N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbours in the k-NN classification. (a) Karṇāṭaka datasets with tonic information. (b) Karṇāṭaka datasets without tonic information. (c) Hindustānī datasets with tonic information. (d) Hindustānī datasets without tonic information.

information improves the accuracy considerably. Also the distribution of peak parameters obtained through this approach for allied rāgas are indicative of the usefulness of the parameters.

7. Future directions

In all the experiments reported thus far, the overall best accuracy among each of the datasets is much higher than chance, indicating the effectiveness of pitch-class profile as a feature vector for rāga identification. It is encouraging to find that a simple first-order pitch distribution provides considerable information about the underlying rāga. Including the ornamentation regions in the pitch-class distribution did not help. As mentioned before, the gamakas play an important role in characterizing the rāga 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 seems inadequate for the task.

7.1 Motif identification

Certain predetermined svara sequences characterize a rāga, a fact used by musicians to delineate a rāga in concert. The automatic recognition of such phrases could be based on sequence matching using a lexicon of phrases for each rāga. Challenges are posed by the required segmentation of a continuous melody into phrases and the local variations that occur in performance.

The notion of such melodic phrases correlates with the general concept of motifs in music. Motif is a short fragment of a continuous musical space which has an identity in itself. It is the smallest musically meaningful unit which characterizes an artist and is manifested as a recurring figure. With this idea the concept of melodic motif encapsulates the above-mentioned melodic phrases. An in-depth melodic motivic analysis will help us understand the complexities of the rāga framework by enabling us to understand it at various levels: gamakas, short catch phrases to long characteristic phrases. This analysis will also enable us to establish relationships between the artists based on melodic motives.

As motif identification relies on identifying the characteristic phrases of a rāga, a starting point to gather data would be the rāgas which are often chosen (together) for rāgamālikas[8] in Karṇāṭaka music. In such scenarios, the listener relies more than anything else, on phrases characteristic to the rāga in order to identify the transitions.

7.2 Changing rāgas

In rāgamālika, unlike most compositions in Karṇāṭaka music, multiple rāgas are used one after the other. And in a technique called grahabēdaṁ, making the second svara of a rāga as the new *Sa* makes it a different rāga. There are several classes of such rāgas, which can be derived from each other by grahabēdaṁ. Both these cases pose a difficult challenge where, besides identifying the rāga, the system also needs to identify the time boundaries of the segments in different rāgas.

8. Conclusions

In this article, we have presented the theoretical aspects of rāga of relevance for computational approaches. In a brief survey of current rāga recognition approaches, we have outlined the contributions and limitations of each one. We have evaluated the first-order pitch-distribution based approaches on a larger and comprehensive database. In template matching, we have explored two methods to determine rāga templates from pitch contour. In the approaches based on distributions constrained to stable regions, we have reported our experiments varying the parameters involved in determining stable regions. Further, on unconstrained pitch distributions, the impact of different bin resolutions and the effect of adding tonic information are reported. We have also tested how intonation analysis helps in modelling the differences in varying properties of a svara across rāgas.

The experiments have helped us in understanding the influence of several parameters specific to each approach based on first-order pitch distributions, on their performance in the rāga classification task. The results are encouraging, and we have identified the future directions to further improve them.

8. Acknowledgements

Discussions with musicians, music students and rasikas from Hyderabad and Chennai (India) have been extremely useful in our work, especially for Section 3. This research was partly funded by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583). G.K. Koduri acknowledges the academic facilities provided by International Institute of Information Technology – Hyderabad.

[8]Literally the garland of rāgas. It is a compositional type in which multiple rāgas are used.

References

Bagchee, S. (1998). *NAD Understanding Raga Music*. Mumbai: Business Publications Inc. (India).

Belle, S., Joshi, R., & Rao, P. (2009). Raga identification by using Swara intonation. *Journal of ITC Sangeet Research Academy*, *23*.

Chakravorty, J., Mukherjee, B., & Datta, A.K. (1989). Some studies on machine recognition of ragas in Indian classical music. *Journal of Acoustic Society of India*, *XVII* (3&4), 1–4.

Cheveigné, A. de, & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111* (4), 1917–1930.

Chordia, P., & Rae, A. (2007). Raag recognition using pitch-class and pitch-class dyad distributions. In *Proceedings of the International Society for Music Information Retrieval Conference*, *Vienna*, *Austria* (pp. 431–436). Vienna: Austrian Computer Society (OCG).

Clayton, M. (2000). *Time in Indian Music: Rhythm, Metre and Form in North Indian Rag Performance*. Oxford: Oxford University Press.

Das, D., & Choudhury, M. (2005). Finite state models for generation of Hindustani classical music. In *Proceedings of the International Symposium on Frontiers of Research in Speech and music (FRSM - 2005)*, (pp. 59–64), Bhubaneswar, India.

Datta, A.K.A.K., Sengupta, R., Dey, N., & Nag, D. (2006). *Experimental Analysis of Shrutis from Performances in Hindustani Music*. Kolkata: ITC Sangeet Research Academy.

Gedik, A.C., & Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, *90* (4), 1049–1063.

Janakiraman, S.R. (2008). *Essentials of Musicology in South Indian Music*. Madras: The Indian Music Publishing House.

Koduri, G.K., Serrà, J., & Serra, X. (2012). Characterization of intonation in Carnatic music by parametrizing pitch histograms. In *ISMIR* (pp. 199–204).

Krishna, T.M., & Ishwar, V. (2012). Carnatic music: Svara, gamaka, motif and raga identity. In *2nd CompMusic Workshop* (pp. 12–28).

Krishnaswamy, A. (2004). Melodic atoms for transcribing carnatic music. In *International Society for Music Information Retrieval Conference*, Barcelona, Spain, pp. 1–4.

Levy, M. (1982). *Intonation in North Indian Music*. New Delhi: Biblia Implex Pvt.

Narayanaswami, P.P., & Jayaraman, V. (2012). *Sangita Sampradaya Pradarsini* [Online] (English Edition). Retrieved from: http://ibiblio.org/guruguha/ssp.htm (date of access: 20/05/2012).

Narmada, M. (2001). *Indian Music and Sancharas in Raagas*. Delhi: Somnath Dhall, Sanjay Prakashan.

Pandey, G., Mishra, C., & Ipe, P. (2003). Tansen: A system for automatic raga identification. In *Indian International Conference on Artificial Intelligence*, Hyderabad, India, pp. 1350–1363.

Ranjani, H., Arthi, S., & Sreenivas, T. (2011). 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 (WASPAA)*. New Paltz, NY, pp. 29–32.

Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing 20* (6), 1759–1770.

Salamon, J., Gulati, S., & Serra, X. (2012). A Multipitch Approach to Tonic Identification in Indian Classical Music In *ISMIR* (pp. 499–504).

Sambamoorthy, P. (1998). *South Indian Music*. (6 Volumes). Madras: The Indian Music Publishing House.

Serra, X. (2011). A multicultural approach in music information research. *Proceedings of ISMIR*. Retrieved from: http://ismir2011.ismir.net/papers/OS2-1.eps.

Serra, X. (2012). Data gathering for a culture specific approach in MIR. In *Workshop on Advances in Music Information Research, World Wide Web Conference*, Lyon, France, pp. 867–868.

Shankar, V. (1983). *The Art and Science of Carnatic Music*. Chennai: Music Academy Madras.

Sharma, P., & Vatsayan, K. (1992). *Brihaddeshi of Sri Matanga Muni*. Columbia, MO: South Asian Books.

Shetty, S., & Achary, K.K. (2009). Raga mining of Indian music by extracting Arohana-Avarohana pattern. *International Journal of Recent Trends in Engineering*, *1*.

Sinith, M., & Rajeev, K. (2006). Hidden Markov model based recognition of musical pattern in South Indian classical music. In *IEEE International Conference on Signal and Image Processing*, Hubli, India.

Sridhar, R., & Geetha, T. (2009). Raga identification of carnatic music for music information retrieval. *International Journal of Recent Trends in Engineering*, *1* (1), 1–4.

Subramanian, M. (2007). Carnatic ragam thodi – Pitch analysis of notes and gamakams. *Journal of the Sangeet Natak Akademi*, *XLI* (1), 3–28.

Upadhye, R., & Sahasrabuddhe, H. (1992). On the computational model of raag music of India. In *Workshop on AI and Music: European Conference on AI*, Vienna, Austria.

Viswanathan, T., & Allen, M.H. (2004). *Music in South India*. Oxford: Oxford University Press.