

# Automatic Carnatic Raga Classification

June 16, 2009

## Abstract

*Raga* is the central melodic concept of two distinct systems of music in India, Hindustani (North Indian) and Carnatic (South Indian) music. Previous work demonstrated that pitch-class distributions (PCDs) and pitch-class dyad distributions (PCDDs) could be effectively used for *raga* classification in North Indian classical music (NICM). We find that PCDs and PCDDs are also effective for Carnatic *raga* classification, despite substantial differences in the manner of presentation, ornamentation, and melodic types as compared with NICM. In a thirty target experiment a 92.4% classification accuracy was achieved using a Bayesian classifier with both PCDs and PCDDs together. This ability of PCD and PCDD features to generalize to a novel musical context suggests that they are effective for capturing essential melodic characteristics of Indian classical music.

## 1 Introduction

Underlying the two systems of classical music in the Indian subcontinent is a melodic system known as *raga*. Performance in Indian classical music is always within a *raga*, except for solo percussion. *Raga* is a system within which performers improvise and compose. A given *raga* defines what types of phrases that are permissible. Phrases that define that *raga* are often more than just sequences of discrete pitches and include particular ways of handling notes, such as slides and various types of vibrato, as well as points of emphasis, achieved through elongation or position (phrase-ending notes). Pitches in both Hindustani and Carnatic music are almost always selected from the twelve chromatic tones of the just-intoned scale. Microtonal alterations are anchored to these twelve tones.

The *raga* is developed according to the principles of the style and form that is being presented. In some cases, the *raga* is improvised by a soloist with only a drone accompaniment (*alap*, *alapana*), in other cases a through-composed piece is presented that adheres to the *raga* with the accompaniment of percussion (*bandish*, *krithi*). Thus *raga*, while specifying melodic constraints, may be applied to many different musical forms. *Ragas* are often summarized by the notes they use, though many *ragas* in fact share the same notes. This is often presented as a typical ascending and descending scale. It is important to note, however, that the scale is abstracted from the actual phrases and while it is a convenient summary, it is almost never correct to equate a *raga* with a scale. Aesthetically a *raga* is meant to evoke a particular emotional state in the listener, a mood that is characteristic of the *raga*.

*Ragas* have traditionally been taught aurally from teacher to student. Notation, when it is used, is primarily used as a mnemonic device, and is more commonly employed for rhythmic compositions. It can take a performer many years before she is ac-

quainted with a core repertory of several dozen *ragas*. Likewise, it often takes listeners many years before they are able to recognize commonly used *ragas*. *Raga* recognition is a difficult task even for humans.

## 2 Background

The ability to identify *raga* gives essential information about the melodic type of a composition. It has been a central theoretical concern of Indian classical music. A systematic method for *raga* classification was introduced to Carnatic music by Venkatakavhin in 1635 [1]. The system was based on assigning a *raga* to its nearest seven note scale. The seven note scale, as is also the case in Western music, is considered a typical full scale. In Carnatic music, the tonic and fifth are taken as fixed. The second, third, sixth, and seventh scale degrees have three forms, resulting in enharmonic equivalence between certain notes. These forms are similar to the Western flat, natural, and sharp forms, but differ in the following manner. The second scale degree, which is denoted R, has the forms R1, R2, and R3. If we take C as our tonic then R1 is D $\flat$ , R2 D and R3 D $\sharp$ . The third scale degree also has three forms, G1, G2, and G3, however in this case, G1 is E $\flat$ , G2 is E and G3 is E $\sharp$ . A similar convention is used for the sixth and seventh scale degrees. The motivation for this naming system is to avoid conceiving of *ragas* as using two forms of the same scale degree. The fourth scale degree is an exception and only has two forms, F and F sharp. In addition to a scale containing only one form of each scale degree, adjacent tones in the scale must be ascending. In other words, a scale which had R3 (D $\sharp$ ) and G1 (D) would not be permissible. These rules lead to 72 possible scales, known as *melakarthis*. This system gained acceptance largely due to its adoption by three of Carnatic music's greatest composers: Tyagaraja, Dikshitar and Syama Sastry. Interestingly, this theoretical system laid the foundation for the introduction of scale types that were not being used at the time, and many new compositions were written in these new scale types by these three composers.

The above discussion suggests an approach to automatic *raga* classification based on identification of notes used. However, since many *ragas* share the same parent scale, identification of the *melakarta* would be insufficient for classification.

## 3 Related Work

*Raga* classification is formally taught through the identification of *sanchara prayoga*, a technical name given to the *raga*-defining phrases described earlier. In many cases, however, non-technical listeners are able to identify a *raga* simply through its overall gestalt, even when they are unable to name notes.

Early work that described approaches to automatic *raga* classification sought to formalize the idea of identifying key musical phrases. It was hoped that the musical surface could be reduced to a set of discrete pitch symbols. This string of pitch symbols could then be searched for substrings that matched the key phrases. In cases where exact matches were not possible, because of inaccuracies in the transcription or intentional variations by the performer, it was hoped that string comparison techniques such as edit distance would be able to yield near matches [2].

More recently Pandey et al. [3] attempted a simple *raga* classification experiment using a Markov model. A success rate of 77% was reported on thirty-one samples in

a two-target test, although the methodology was not well documented. An additional stage that searched for specific pitch sequences improved performance by 10%.

An initial attempt by Chordia attempted *raga* classification on a small database without explicit pitch-tracking using chroma features, where spectral energy is folded into one octave and divided into semi-tone bins [4]. Perfect results were obtained using a K-NN classifier with 60/40% train/test split. This was further developed in [5] where PCDs and PCDDs were used as features with more sophisticated learning algorithms. In a seventeen-target experiment with one hundred forty two segments, classification accuracy of 94% was attained using 10-fold cross-validation. However, the significance of the results in both cases was limited by the size of the database. Recent work by Chordia and Rae [6] on a large database of North Indian classical music, containing thirty-one *ragas* and twenty hours of audio data, attained 99% accuracy using PCDs and PCDDs with a Support Vector Machines (SVM) classifier. Accuracy for PCDs alone was 78%.

Pitch-class distributions, also referred to as pitch-class profiles, have been widely used for key recognition [7, 8] in tonal Western music. The most commonly used approach has been to calculate empirical PCDs on MIDI or symbolic data for short segments. These were then correlated with average PCDs from a large-sample of music with key information. Later approaches used a more formal machine learning approach in which the PCDs were modeled statistically [9].

The motivation for the use of PCDs in key recognition came from the seminal work of Krumhansl [10], who showed that listeners rate certain tones from the chromatic scale as more fitting than others when presented with a key defining context such as a scale or cadence. This work has been broadly repeated, demonstrating that key-defining contexts elicit stable pitch expectations [11]. It has been shown that tones that were rated as most appropriate for a given key were the most commonly used in that key. For example, the fifth and tonic are the most frequently used scale degrees in nearly every key, and these are the tones that people judge as most fitting the key context. This research led to the observation that PCDs are relatively stable for excerpts from clearly tonal works and can be used as effective features for key recognition. This result is surprising since PCDs do not capture any sequential information, while it is precisely through sequential gestures such as a melody or a cadence that key is typically defined.

Work by Temperley and Marvin [12] has shown that artificial melodies generated according to PCDs for a given key are correctly identified approximately 50% of the time, suggesting that distributional properties alone, while informative, are not sufficient for key characterization. Nevertheless, PCDs have been shown to be psychologically relevant and Huron [13] has shown that such distributional characteristics are unconsciously assimilated by listeners through mere exposure.

In the MIR community, direct calculation of PCDs has been rare because of the difficulty of pitch-tracking in polyphonic contexts. For this reason chroma features that can easily be automatically calculated from the spectrogram of a signal have been used instead with good results [14].

This background suggested that it might be possible to use PCDs as simple, stable representations of *ragas*. Despite that fact that *ragas* are defined by linear pitch sequences, notes are emphasized through repetition and placement, and a tonal hierarchy emerges, with some notes more prominent than others. The results of Chordia [6] demonstrated this.

Additionally, it has been shown that listeners are sensitive to transitions between notes [15]. Using higher order sequences such as bi-grams, and more generally n-

grams, is common in language modeling and serves as a simple way of capturing sequential structure. However, the use of higher-order features requires substantially more training data because the parameter space increases exponentially.

The extension to Carnatic music is an essential test of whether PCDs and PCDDs can be generalized to novel musical contexts where the underlying principles are similar but the musical surfaces significantly different. Successful *raga* classification by similar means would show that in both systems melodies are constructed in ways that lead to well-defined tonal hierarchies that can be effectively modeled by PCDs and PCDDs.

## 4 *Raga* Database

The *raga* database consists of forty-three performances in thirty *ragas*, totaling five hours and thirteen minutes of recorded material. The *ragas* were chosen to represent a selection of the most commonly performed *ragas* in contemporary practice. The scale types for each of these *ragas* is shown in Table 1. In most cases, at least two recordings from different performers for each *raga* were included, with a total of twenty-two performers. The recordings were primarily vocal recordings of many of the most highly regarded performers of Carnatic music, such as GN Balasubramanian, Semmangudi Srinivasa Iyer, DK Pattammal, and KV Narayanaswamy. Instrumental recordings of violin, flute, clarinet, *chitraveena*, and saxophone were included as well. Selections were drawn from the *alapana* and *tanam* forms where there is no *mridangam* (percussion) accompaniment but in most cases have the usual accompaniment of violin and drone (*tambura*). In many cases the level of the violin accompaniment was comparable to the soloist.

Because an emphasis was placed on the authenticity of the recordings, there was no attempt to select for high-fidelity. In many cases substantial interfering noise and artifacts such as ground hum were present. It was thought that since many important recordings have these characteristics, this would be an important test for the algorithm and would be more representative of true performance than simply including clean studio recordings. The database (carnatic2008) thus represents a first attempt at compiling a realistic cross-section of Carnatic music for music information retrieval research.

## 5 Method

### 5.1 Overview

Figure 1 shows a block diagram of the system. Each performance is segmented into sixty seconds frames and pitch-tracked. The continuous pitch-track is converted into scale degrees using a tonic (in Hz) that has been manually annotated for each performance. The scale degree values are then counted by frame to determine the PCD. It can be seen that onset detection is performed on each segment in an attempt to segment the recording into notes. This information is then used to construct the PCDDs, which record how often different two note-sequences (dyads) appear. The PCDs and PCDDs are used as features for the classification stage. We attempt classification using a variety of techniques, such as Bayes rule with a multivariate model and Random Forests. We describe each step in greater detail below.

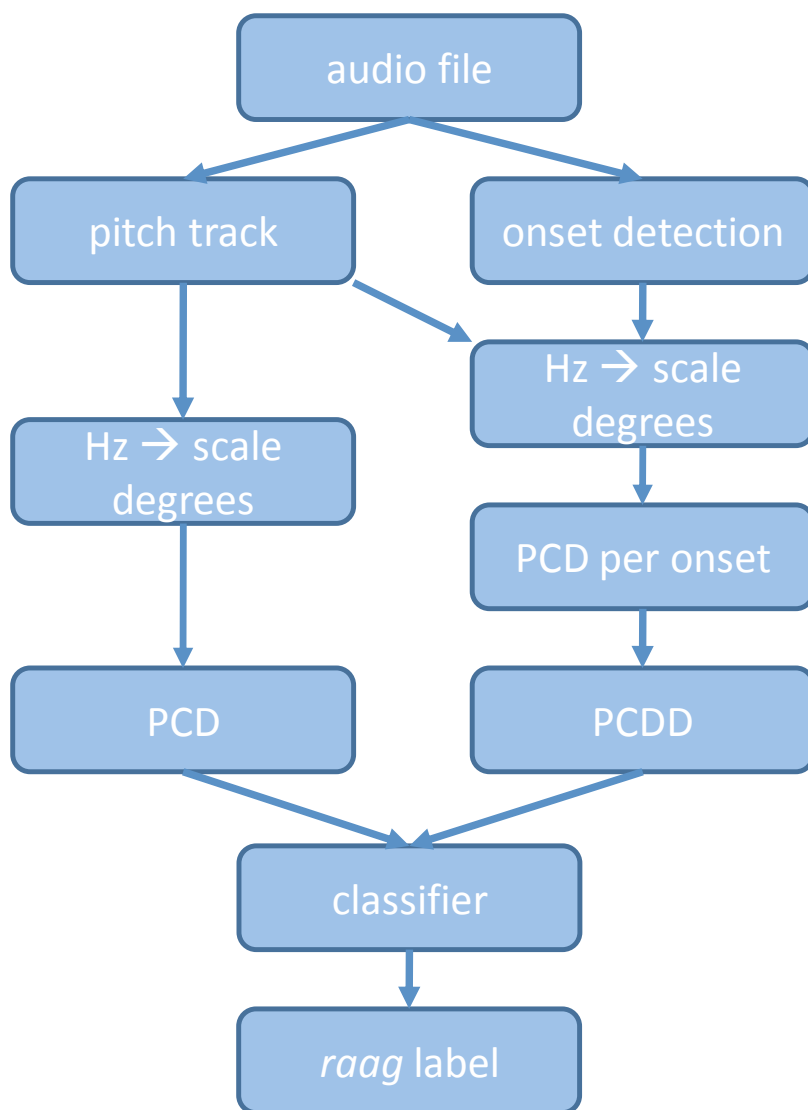


Figure 1: Block diagram of the *raga* classification system.

	C	D $\flat$	D	E $\flat$	E	F	F $\sharp$	G	A $\flat$	A	B $\flat$	B
Abheri	•		•	•		•		•		•	•	
Anandhabhairavi	•		•	•		•		•	•		•	
Atana	•		•		•	•		•		•		•
Bhairavi	•		•	•		•		•	•		•	
Bilahari	•		•		•	•		•		•		•
Dhanyasi	•	•		•		•		•	•		•	
Dwijavanti	•		•	•		•		•		•	•	
Hamsadwani	•		•		•			•				•
Hindolam	•			•		•			•		•	
Kaanada	•		•	•		•		•			•	
Kalyani	•		•		•		•	•		•		•
Kambhoji	•		•		•	•		•		•	•	
Karahapriya	•		•	•		•		•		•	•	
Keeravani	•		•	•		•		•	•			•
Khamas	•		•		•	•		•		•	•	
Madhyamavathi	•		•			•		•			•	
Mayamalavagowlai	•	•			•	•		•	•			•
Mohanam	•		•		•			•		•		
Nalinakanthi	•		•		•	•		•				•
Nattakurinji	•		•		•	•		•		•	•	
Pantuvarali	•	•			•		•	•	•			•
Purvikalyani	•	•			•		•	•		•		•
Reethigowlai	•		•	•		•		•		•	•	
Sahana	•		•		•	•		•		•	•	
Sankarabharnam	•		•		•	•		•		•		•
Shanmugapriya	•		•	•			•	•	•		•	
Simhendramadhyamam	•		•	•			•	•	•			•
Sriranjani	•		•	•		•				•	•	
Todi	•	•		•		•		•	•		•	
Varali	•	•	•				•	•	•			•

Table 1: Summary of scale degrees used by the thirty *ragas* in the database. Notes are listed with C as the tonic.

## 5.2 Pitch-Detection

Pitch detection was done using both autocorrelation-based YIN algorithm [16] as well as an implementation of the Harmonic Product Spectrum (HPS) approach with subharmonic thresholding for comparison [17].

In the YIN algorithm the audio is divided into overlapping frames and the pitch is estimated every millisecond. The core of YIN centers on the squared difference function of a time-domain signal, conceptually similar to the autocorrelation function. Further refinements include normalizations to correct for erroneously picking very high frequencies, increasing precision of the estimates through parabolic interpolation, and a number of other extensions. YIN has been widely used because of its excellent published results on a large database of speech and its speed compared with fully probabilistic approaches. In addition to returning a pitch estimate YIN also returns a measure of aperiodicity associated with the estimate. If the signal is perfectly periodic then sub-

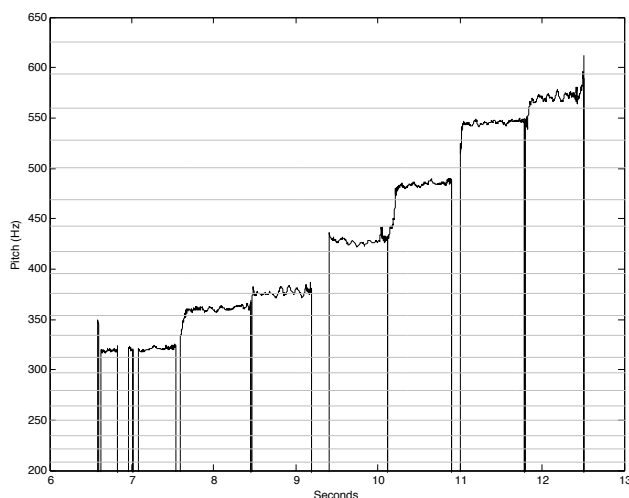


Figure 2: A typical pitch track using the YIN algorithm. Horizontal lines show logarithmic boundaries between scale degrees.

tracting the signals at the period will result in the difference function being zero at that lag. The depth of this dip in the curve can be used as a measure of aperiodicity or periodicity, with a shallow maximum dip corresponding to a relatively aperiodic section. We describe an experiment below testing whether incorporating such pitch salience information improves classification results. A typical pitch track excerpt is shown in Figure 2.

The HPS algorithm, which works in the frequency domain, assumes that an frequency is a good pitch candidate if substantial energy is found at its harmonics. Given the DFT, each possible frequency candidate is evaluated by summing the energy of a specified number of harmonics. The principle drawback to this method is that it is prone to octave errors. This was not problematic in our case, however, because pitch values were all folded into one octave in forming the PCDs. In our implementation each segment was divided into 40ms frames, using a Gaussian window. The frames were overlapped by 75%, leading to a pitch estimate every 10 ms.

### 5.3 Pitch-class Distribution

Typically there is little to no consistency in the tonic frequency between recordings, as an individual performer will sing or play at whatever pitch is comfortable. For this reason the tonic was manually annotated for each recording by tuning an oscillator by ear, and the frequencies of the scale degrees were calculated from this tonic. The pitch estimate at each frame was assigned to the nearest semitone bin, thus converting the pitch track to a sequence of scale degrees. The register information for each scale degree was discarded leaving a sequence of pitch classes. A PCD was then calculated for each segment by taking a histogram of the pitch class values over all the frames. At this stage, no information about note onsets was used. In this procedure, PCDs automatically take into account note durations, since a held note will occur over many frames. PCDs were also calculated using the aperiodicity measure returned by YIN. Two measures of pitch salience, one calculated as 1 minus the aperiodicity and the

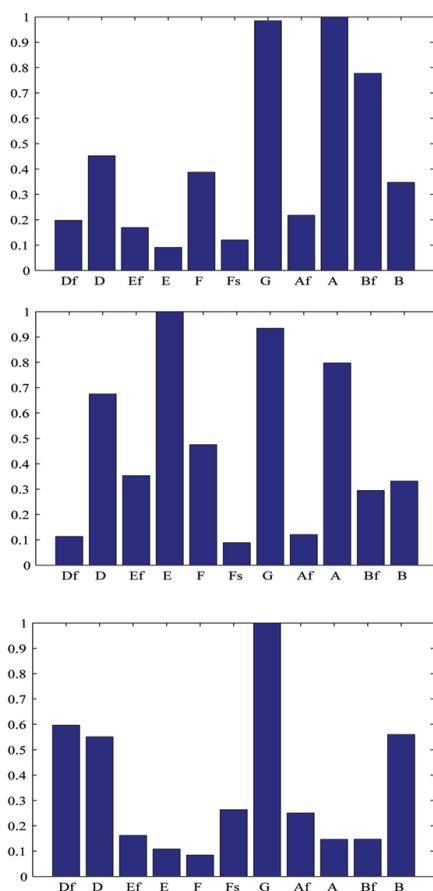


Figure 3: Pitch-class distributions for *ragas Atana*, *Bilahari*, and *Varali*. The tonic is omitted due to its consistent and overwhelming presence. Both *Atana* and *Bilahari* use the notes of the Western major scale.

other as the inverse, were used to weight the pitch estimates, with the effect that less clearly pitched regions of the recordings, often indicating pauses by the soloist, were de-emphasized. PCDs for three *ragas* are shown in Figure 3. *Ragas Atana* and *Bilahari* share the same scale (a Western major scale), while *Varali* uses an entirely different scale; nonetheless, it can be seen that the first two PCDs are distinguishable. Further insight into the discriminative power of PCDs can be seen from the boxplot diagram in Figure 4. Even with a single feature, many of the *ragas* can be told apart.

## 5.4 Onset Detection

In order to determine PCDDs, recordings were segmented by note onsets. We did this using a spectrally based onset detection algorithm described in Duxbury et al. [18]. Here we used a complex detection function that looked at the deviation of each spectral frame from its predicted phase given the previous frame under the assumption

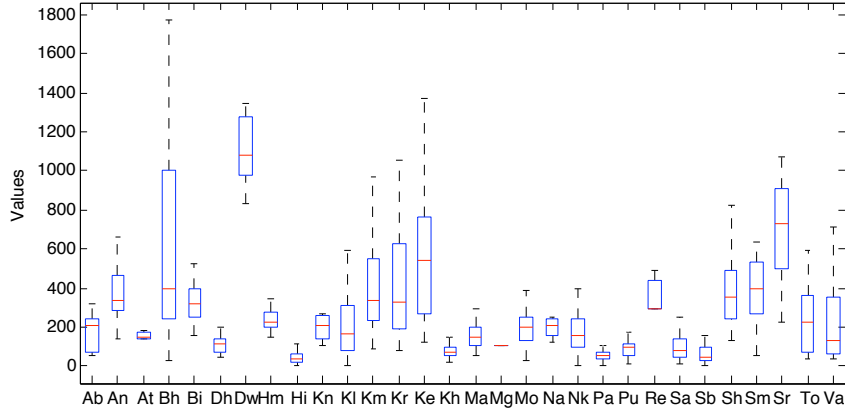


Figure 4: Box plot comparison of scale degree  $D$  across all target *ragas* showing discriminative power of a single feature.

that no transient was present. In such a case we expect a linear phase increase in each frequency band; deviations from this prediction indicate transient regions. The deviation at each time step forms a detection function. Each segment was divided into 128 sample regions, overlapped 50% using a rectangular window. An onset was defined to be when local maxima in the detection function exceeded a constant factor time the sliding median, correcting for many spurious local maxima in the detection function, particularly during low energy sections.

## 5.5 Pitch-class Dyad Distribution

PCDDs were calculated by forcing the continuous pitch track into discrete notes. This was done by using all pitches between two detected onsets to determine the most reasonable static pitch. In all cases the pitch varies due to vibrato, ornaments, slides, performer mistakes and pitch-tracking mistakes. This step can be thought of as pitch quantization. A simple algorithm was employed: a PCD was calculated for the inter-onset region by the above method, with the peak giving the pitch-class estimate for the note. This again reduced the pitch track to a series of pitch-classes, but with only one for each detected onset. The notes were assembled into dyads and counted, yielding a 144 dimensional vector for each segment that counted the frequency of occurrence of each possible dyad. Given the number of notes in a sixty second segment, this matrix was usually quite sparse.

A significant problem with the current approach to calculating PCDDs is the unreliability of the onset detection. In many cases onsets are inserted or missed. Insertions often happen because of noise or accompaniment, while omissions can occur because of slides and tonal onsets that can easily elude the system. This leads to a very noisy PCDD that may contain features that are specific to the recording. Thus they may not generalize well, and if there are insufficient recordings for a given *raga* they may overstate the discriminatory power of PCDDs thus calculated, identifying specific recordings rather than underlying *raga* structure. However it is important to note that they are not in themselves timbral features; two recordings of different *ragas* with nearly identical performer style and recording quality would not likely share similar PCDDs.

Features Used	Classifier Type	
	Bayesian	Random Forest
Yin:		
PCD	44.7	67.1
PCD-Linear	50.9	70.2
PCD-Ratio	65.1	77.4
PCD + PCDD	87.9	74.0
PCD-Lin + PCDD	86.2	73.2
PCD-Ratio + PCDD	90.5	76.5
HPS:		
PCD	50.4	72.2
PCD + PCDD	92.4	86.3

Table 2: Summary of primary classification results using Multivariate Bayesian and Random Forest classifiers for each of the two pitch detection algorithms. “Linear” and “Ratio” refer to the two methods of weighting the PCDs by pitch salience.

## 6 Classification

Soundfiles were segmented using a rectangular window of sixty seconds, each overlapped by 50%, leading to a total of 788 segments that were used for classification. Success rates were calculated using 10-fold cross-validation (CV). In addition to the techniques discussed below several other methods, such as SVM, K-Star, and Tree-based (CART) were tried but performed consistently worse and are not discussed further.

### 6.1 Multivariate Normal (MVN)

The feature vector was modeled using an MVN distribution. The parameters were estimated from the training data using a pooled covariance matrix for each class. The priors were calculated empirically from the training data. The label was selected using a maximum a posteriori (MAP) rule.

### 6.2 Random Forest

Classification was attempted using the Random Forests (RF) method [19] as implemented in WEKA [20]. This somewhat newer algorithm is essentially an aggregate of decision trees, where each is grown by taking a bootstrap sampling of the training set, and each node of a given tree is constructed by randomly choosing some small subset of features and choosing the best split; the trees are not pruned. The resulting set of tree classifiers (forest) outputs a decision by taking a vote over all the individual trees.

## 7 Results and Discussion

Table 2 summarizes the principle results. In the best case 92.4% accuracy was attained using the Bayesian classifier using both the PCD and PCDD features. PCD without pitch salience weighting attained an accuracy of 67.1% with the RF classifier. Notably, the use of pitch salience improved accuracy substantially for PCD features: in the

case of the RF classifier performance improved by 10.3 percentage points when pitch salience was used and 20.4% for the Bayesian classifier. For every condition PCDD features lead to large gains, except for the RF classifier when using YIN along with pitch salience weighting. Interestingly, HPS outperformed YIN without pitch salience weighting.

The results confirm the discriminatory power of PCD and PCDDs for *raga* classification. Given the substantial differences between Carnatic and Hindustani music, this is a substantial result suggesting common structural properties that are effectively captured by these features. However, it is important to reiterate that the PCDD results are almost certainly overly optimistic, a problem due in part to the small number of examples in certain *ragas*. As the database grows, we will be able to more rigorously determine the generalization properties of PCDDs. Nevertheless it is clear that they lead to substantial marginal improvement over PCDs alone. This work also establishes the utility of weighting PCDs by measures of of pitch salience. It is difficult to know exactly why HPS outperformed YIN. Visualization of the pitch tracks show that HPS is substantially noisier, but primarily due to octave errors which as noted in 5.2 are irrelevant. The difference in performance is possibly due to a greater robustness of HPS to simultaneous tones arising from accompaniment such as the omnipresent drone.

## 8 Future work

We hope to continue to build the database so that it includes a large number of examples in each *raga*. Further, we would like to incorporate source-separation so that effects of accompaniment can be minimized. This will also allow us to deal more effectively with a wider range of forms, including rhythmic sections with percussive accompaniment. Also, the current method of mapping the continuous pitch track to a sequence of discrete notes is inadequate and sensitive to errors. By learning more about the relationship between expressive melodic performance and the perception of discrete pitch categories we will be better able to analyze the music.

## References

- [1] Ludwig Pesch, *The Illustrated Companion to South Indian Classical Music*, Oxford University Press, 1999.
- [2] H.V. Sahasrabuddhe and R. Upadhy, “On the computational model of raag music of india,” in *Proc. Indian Music and Computers: Can Mindware and Software Meet?*, 1994.
- [3] Gaurav Pandey, Chaitanya Mishra, and Paul Ipe, “Tansen : A system for automatic raga identification,” in *Proceedings of the 1st Indian International Conference on Artificial Intelligence*, 2003, pp. 1350–1363.
- [4] Parag Chordia, “Automatic rag classification using spectrally derived tone profiles,” in *Proceedings of the International Computer Music Conference*, 2004.
- [5] Parag Chordia, “Automatic raag classification of pitch-tracked performances using pitch-class and pitch-class dyad distributions,” in *Proceedings of International Computer Music Conference*, 2006.

- [6] Parag Chordia and Alex Rae, “Raag recognition using pitch-class and pitch-class dyad distributions,” in *Proceedings of International Conference on Music Information Retrieval*, 2007.
- [7] E. Gomez and P. Herrera, “Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies,” in *Proceedings of International Conference on Music Information Retrieval*, 2004.
- [8] Craig Sapp, “Visual hierarchical key analysis,” *Computers in Entertainment*, vol. 3, no. 4, October 2005.
- [9] David Temperley, *Music and Probability*, MIT Press, 2007.
- [10] C. Krumhansl and R. Shepard, “Quantification of the hierarchy of tonal functions within a diatonic context,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 5, no. 4, pp. 579–594, 1979.
- [11] MA Castellano, JJ Bharucha, and CL Krumhansl, “Tonal hierarchies in the music of north india,” *Journal of Experimental Psychology*, 1984.
- [12] David Temperley and Elizabeth West Marvin, “Pitch-class distribution and the identification of key,” in *Proceedings of the The Society for Music Perception and Cognition*, 2007.
- [13] David Huron, *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press, 2006.
- [14] Steffen Pauws, “Musical key extraction from audio,” in *Proceedings of International Conference on Music Information Retrieval*, 2004.
- [15] B. Aarden, “Expectancy vs. retrospective perception: Reconsidering the effects of schema and continuation judgments on measures of melodic expectancy,” in *Proceedings of the 7th International Conference on Music Perception and Cognition*, 2002, pp. 469–472.
- [16] Alain de Cheveigne and Hideki Kawahara, “Yin, a fundamental frequency estimator for speech and music,” *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917 – 1930, 2002.
- [17] Xuejing Sun, “A pitch determination algorithm based on subharmonic-to-harmonic ratio,” in *In Proc. of International Conference of Speech and Language Processing*, 2000.
- [18] C. Duxbury, J. P. Bello, M. Davies, and M. Sandler, “A combined phase and amplitude based approach to onset detection for audio segmentation,” in *Proc. of the 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-03)*, London, 2003, pp. 275–280.
- [19] Leo Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, 2001.
- [20] Ian H. Witten and Eibe Frank, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2005.