RAAG RECOGNITION USING PITCH-CLASS AND PITCH-CLASS DYAD DISTRIBUTIONS

Parag Chordia and Alex Rae
Georgia Institute of Technology, Department of Music
840 Mc Millan St., Atlanta GA 30332
{ppc,arae3}@gatech.edu

ABSTRACT

We describe the results of the first large-scale *raag* recognition experiment. *Raags* are the central structure of Indian classical music, each consisting of a unique set of complex melodic gestures. We construct a system to recognize *raags* based on pitch-class distributions (PCDs) and pitch-class dyad distributions (PCDDs) calculated directly from the audio signal. A large, diverse database consisting of 20 hours of recorded performances in 31 different *raags* by 19 different performers was assembled to train and test the system. Classification was performed using support vector machines, maximum a posteriori (MAP) rule using a multivariate likelihood model (MVN), and Random Forests. When classification was done on 60s segments, a maximum classification accuracy of 99.0% was attained in a cross-validation experiment. In a more difficult unseen generalization experiment, accuracy was 75%. The current work clearly demonstrates the effectiveness of PCDs and PCDDs in discriminating *raags*, even when musical differences are subtle.

1 BACKGROUND

1.1 *Raag* in Indian Classical Music

Raag is a melodic abstraction around which almost all Indian classical music is organized. A *raag* is most easily explained as a collection of melodic gestures and a technique for developing them. The gestures are sequences of notes that are often inflected with various micro-pitch alterations and articulated with an expressive sense of timing. Longer phrases are built by joining these melodic atoms together.

By building phrases in this way, a tonal hierarchy is created. Some tones appear more often in the basic phrases, or are sustained longer. Indian music theory has a rich vocabulary for describing the function of notes in this framework. The most stressed note is called the *vadi* and the second most stressed, traditionally a fifth or fourth away, is called the *samvadi*. There are also less commonly used terms for tones on which phrases begin and end. A typical summary of a *raag* includes its scale type (*that*), *vadi* and *samvadi*. A pitch-class distribution (PCD), which gives the relative frequency of each scale degree, neatly summarizes this information.

© 2007 Austrian Computer Society (OCG).

Indian classical music (ICM) uses approximately one hundred *raags*, of which fifty are common. Despite microtonal variation, the notes in any given *raag* conform to one of the twelve chromatic pitches of a standard just-intoned scale. There are theoretically thousands of scale types; in practice, however, *raags* conform to a much smaller set of scales, and many of the most common *raags* share the same set of notes.

The performance context of *raag* music is essentially monophonic, although vocalists will usually be shadowed by an accompanying melody instrument. The rhythmic accompaniment of the *tabla* is also present in metered sections. There is usually an accompanying drone that sounds the tonic and fifth using a harmonically rich timbre.

2 RELATED WORK

2.1 Western Tonality

Krumhansl and Shephard [11] as well as Castellano et al. [3] have shown that stable pitch distributions give rise to mental schemas that structure expectations and facilitate the processing of musical information. Using the now famous probe-tone method, Krumhansl [12] showed that listeners' ratings of the appropriateness of a test tone in relation to a tonal context is directly related to the relative prevalence of that pitch-class in a given key. Huron [10] has shown that emotional adjectives used to describe a tone are highly correlated with that tone's frequency in a relevant corpus of music. Further, certain qualities seemed to be due to higher-order statistics, such as note-to-note transition probabilities. These experiments show that listeners are sensitive to PCDs and internalize them in ways that affect their experience of music.

The demonstration that PCDs are relatively stable in large corpora of tonal Western music led to the development of key- and mode-finding algorithms based on correlating PCDs of a given excerpt, with empirical PCDs calculated on a large sample of related music [6, 14, 9].

2.2 *Raag* Classification

Raag classification has been a central topic in Indian music theory for centuries, inspiring rich debate on the essential characteristics of *raags* and the features that make two *raags* similar or dissimilar [1].

Pandey [13] developed a system to automatically recognize *raags Yaman* and *Bhupali* 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 to 87%.

In an exploratory step, Chordia [4] classified one hundred thirty segments of sixty seconds each, from thirteen *raags*. The feature vector was the Harmonic pitch class profile (HPCP) for each segment. 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 17 target experiment with 142 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.

3 MOTIVATION

Raag is the most important concept in Indian music, making accurate recognition a prerequisite to almost all musical analysis. Further, because a *raag* defines the underlying emotional character of the music, correct classification captures qualities essential to the subjective experience. Practically, automatic *raag* recognition thus has tremendous potential use in music discovery and automatic playlist generation for ICM. It is also useful for interactive work featuring ICM.

An additional motivation is to examine whether conceptions of tonality appropriate to Western tonal music are applicable cross-culturally. If PCDs could be used to identify *raags*, this would mean that they are important to establishing a fundamental tonal context in very different musical traditions; showing this common underlying mechanism would be an important discovery.

4 *RAAG* DATABASE

For this study, a substantial database was assembled from a variety of sources. The samples were chosen to include considerable diversity across several dimensions: in *raag*, musician, instrument, playing style, presence or absence of accompaniment, and recording quality. Commercial recordings were included along with close to twenty hours of unaccompanied *raags* recorded specifically for this study. The performances can be heard at `http://paragchordia.com/research/`.

A total of thirty one *raags* were represented in the database, comprising a significant fraction of the commonly played corpus of ICM. Table 1 summarizes the database by *raag* and pitch content. Most performances used had both a very slow, unmetered section (*alap*), and a faster, rhythmic section as well (*bandish* or *gat*). Nineteen musicians' playing was included; six were instrumentalists playing either the plucked string instruments *sarod* or *sitar*, or the blown instruments *shenai* or flute, and the remaining thirteen were vocalists, both male and female. Recordings made expressly for this project were unaccompanied by either drone or *tabla*, providing a clean and isolated signal, while the commercial recordings contained a full range of accompaniment, and sometimes were of a significantly degraded sound quality.

	C	D♭	D	E♭	E	F	F♯	G	A♭	A	B♭	B
YamanK.	•		•		•	•	•	•		•		•
Yaman	•		•		•		•	•		•		•
MaruBihag	•		•		•	•		•		•		•
GaudSarang	•		•		•	•		•		•		•
Hameer	•		•		•	•		•		•		•
Desh	•		•		•	•		•		•		•
TilakKamod	•		•		•	•		•		•		•
GaudMalhar	•		•		•	•		•		•	•	•
Jaijaiwante	•		•		•	•		•		•	•	•
Khamaj	•		•		•	•		•		•	•	•
Bihag	•		•		•	•	•	•		•		•
Kedar	•		•		•	•	•	•		•	•	•
Rageshri	•		•		•	•				•	•	
Bageshri	•		•	•		•		•		•	•	
Bhimpalasi	•		•	•		•		•		•	•	
A.Bhairav	•		•		•	•	•	•		•		•
Darbari	•		•	•		•		•	•		•	
Jaunpuri	•		•	•		•		•	•		•	
K.Kanhra	•		•	•		•		•	•		•	
Malkauns	•			•		•			•		•	
Multani	•	•		•			•	•	•			•
Shree	•	•			•		•	•	•			•
P.Dhanashri	•	•			•		•	•	•			•
K.R.Asaveri	•	•		•		•		•	•		•	
Todi	•	•		•			•	•	•			•
Bhairavi	•	•		•		•	•	•	•		•	
Ka.Bhairavi	•	•	•	•		•	•	•	•		•	
B.Todi	•	•		•		•			•	•	•	
Bhairav	•	•			•	•		•	•			•
Marwa	•	•			•		•			•		•
Bhatiyar	•	•			•	•	•	•		•		•

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

Individual recordings were between three and sixty minutes in length (with no more than seven minutes taken from any single one), and were segmented into 30s and 60s chunks. There were in total 20 hours of segmented material, forming the largest *raag* classification database to date.

5 FEATURE EXTRACTION

5.1 Annotation

In order to facilitate the creation of pitch profiles relevant to the particular tuning of the performances, each *raag* sample was labeled with the frequency value for the tonic

of that recording. This was done manually, by tuning an oscillator and noting the value in Hz.

5.2 Pitch Detection

Pitch detection was done using a version of the Harmonic Product Spectrum (HPS) algorithm [7, 15]. Each segment was divided into 40ms frames, using a Gaussian window. The frames were overlapped by 75%, so that the pitch was estimated every 10 ms. Visualization of the resulting pitch-tracks showed the estimates to be quite robust, despite occasional octave errors and confusion with accompaniment.

5.3 Onset Detection

Note onsets in each segment were found by thresholding a complex detection function (DF) [8]. The segment was divided into 128 sample regions, overlapped 50% using a rectangular window, and the DFT of each region was computed and used to construct the DF. Adaptive thresholding, proportional to the median over a sliding window, was used to choose the peaks to be labeled as onsets.

5.4 Pitch-Class Distributions

The PCDs were calculated without reference to the detected onsets, by simply taking histograms of the pitch-tracks. The bins corresponded to each note of five octaves of a chromatic scale centered about the tonic for that segment. Specifically, the ratios of the just-intoned scale and the tonic frequency were used to calculate the center of each bin, and the edges were determined as the log mean. The five octaves were then folded into one and the values normalized to create a pitch-class distribution. This nulled any significance of octave errors in the HPS algorithm.

Being frame-based, the PCDs were not vulnerable to errors in onset detection, and were able as well to capture information from lengthy held notes, a feature especially important in the slow *alap*, where a full minute might only include eight to ten discrete notes. On the other hand, they were also guaranteed to include noise, as they added every erroneous pitch estimate to the histogram.

Figures 1 and 2 show an illustrative subset of this data; they demonstrate the discriminative potential of PCDs. Figure 1 shows the distributions for the flat second scale degree (D♭) for each *raag*. The tonic (C) is omitted from Figure 2 due to its ubiquity in all *raags*.

Additionally, harmonic pitch class profiles (HPCP) were calculated from the spectra using the same frequency binning and octave-folding method [6], in order to make a comparison with the PCDs.

5.5 Pitch-class Dyad Distributions

To determine the PCDDs, the detected onsets were used to segment the pitch-tracks into notes. Each note was then assigned a pitch-class label: first the raw pitch estimates were discretized by assigning to each the center

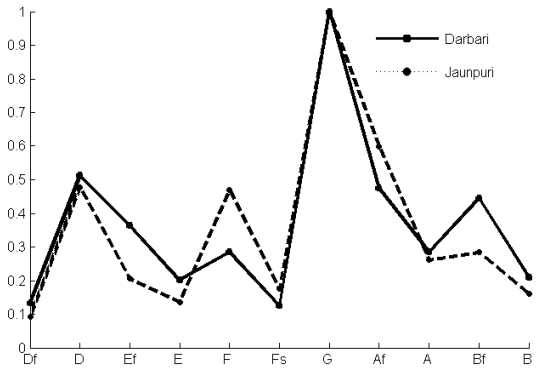

Figure 2. Pitch-class distribution for *raag Darbari* and *Jaunpuri*

value of the bins defined for the pitch histogram, and then the mode was calculated for each note. The label of the corresponding chromatic pitch was assigned to that note. This process dealt quite effectively with variations due to micro-pitch structure, attacks, and errors by the detection algorithm. The octaves were folded into one as with the PCDs.

The pitch-classes were then arranged in groups of two (bi-grams), or in musical terms, dyads. Tables 4 and 5 show two examples, taken from *raags* that share the same set of notes.

A significant complication of calculating PCDDs in this musical context is the occurrence of notes that are played by sliding up or down to that pitch from a previous note, without any clear onset. When pitches are histogrammed for each time-frame, as in PCDs, this poses no problem. However, in PCDDs this characteristic poses difficulties, and ideally the algorithm would not rely on explicit onsets. In the current work, this problem was not solved, and so the PCDDs entail a certain level of abstraction, as some of the values recorded in them are in actuality the bi-grams for the closest pairs of clearly articulated notes, rather than simply bi-grams of adjacent notes. This also makes PCCDs substantially more vulnerable than PCDs to variations in recording quality, accompaniment, and instrumentation.

6 CLASSIFICATION

Soundfiles were segmented using a rectangular window. One set used segments of 30s and the other of 60s, each overlapped by 50%, leading to a total of 4676 thirty second segments and 2248 sixty second segments. Success rates were calculated using 10-fold cross-validation (CV). To further test generalization, classification was attempted using ”unseen” cases, in which *raag* excerpts in the training set came from different performances than those in the training set. In addition to those listed below, classification was also attempted with a number of other methods, specifically feed-forward neural networks, K-star, and a

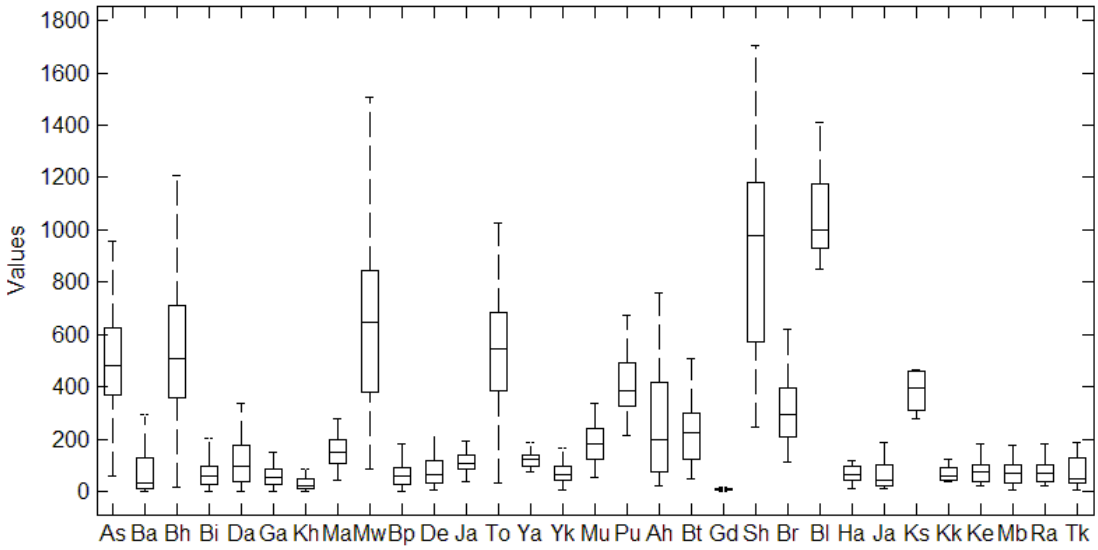

Figure 1. Boxplot comparison of use of 'D♭' in target *raags*

tree-based classifier (CART). The results, however, were significantly worse than the three most effective (typically around 70%), so we exclude them from our discussion.

The feature vector was modeled using an MVN distribution. The parameters were estimated from the training data using a common 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.

Classification was attempted using the Random Forests method [2]. 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.

Support vector machines were learned using the sequential minimal optimization algorithm as implemented in WEKA [16]. A series of binary classifiers, one for each pair of *raags*, were trained and used to make the multi-category decision.

7 RESULTS AND DISCUSSION

Table 2 shows the primary results. In the CV experiment nearly perfect results (99.0%) were attained using PCD and PCDD features with a SVM classifier. In the more difficult unseen case, accuracy was 75%. To determine the contribution of PCD and PCDD separately, we ran the SVM algorithm using each feature set alone, yielding success rates of 78.0% and 97.1% respectively for the CV case, and 75.3% and 57.1% for the unseen case. The

	Experiment Type	
Feature Used	CV	Unseen
PCD	78.0	75.2
PCDD	97.1	57.1
Both	99.0	73.7

Table 2. Summary of primary classification results using for SVM classifier using 60s segments.

SVM classifier easily outperformed MVN (96.1%) and RF (96.1%) in the CV experiment. A 1-3% reduction in performance was observed using 30s segments. This was somewhat surprising considering the paucity of notes in certain slow sections. We were interested to see if similar results to PCD alone could be obtained without explicit pitch tracking, using HPCP features. The success rate was 64.4%, demonstrating the utility of pitch tracking. PCA was attempted but did not improve results in any of the experiments.

In the unseen case, while the PCD results are on par with those from the CV case, PCDDs seem to degrade performance, despite their success in the the CV case. We hypothesize that errors in onset detection lead to distinctive dyads that are dependent on both performance and *raag*. This effect is discussed in Section 5.5. More work needs to be done, however, to show that PCDDs are in general not performance-specific.

It is important to note that informal listening shows that melodic variation from segment to segment within a performance is as great as the variation between performances. Since purely melodic information is being used, without any timbral information, the CV classification results are impressive and indicate that PCDD is capturing something important. Refinement of the PCDD calcula-

	As	Ba	Bh	Bi	Da	Ga	Kh	Ma	Mw	Bp	De	Ja	To	Ya	Yk	Mu	Pu	Ah	Bt	Gd	Sh	Br	Bl	Ha	Ja	Ks	Kk	Ke	Mb	Ra	Tk
As	80	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ba	.	88	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bh	.	.	180	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bi	.	.	.	137	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
Da	.	.	.	.	264	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ga	.	.	.	.	.	72	2	.	.	.	4	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
Kh	.	.	.	1	.	1	148	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ma	.	.	.	.	.	.	.	39	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mw	.	.	.	.	.	.	.	.	68	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bp	.	.	.	.	1	.	.	.	.	176	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
De	.	.	.	.	1	4	.	.	.	.	176	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
Ja	.	.	.	.	.	.	.	.	.	.	.	27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
To	.	.	.	.	.	.	.	.	.	.	.	.	41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ya	.	.	.	.	.	.	.	.	.	.	.	.	.	19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yk	.	.	.	.	.	.	.	.	.	.	.	.	.	.	126	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mu	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Pu	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	22	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ah	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	56	.	.	.	.	.	.	.	.	.	.	.	.	.
Bt	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	31	.	.	.	.	.	.	.	.	.	.	.	.
Gd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	33	.	.	.	.	.	.	.	.	.	.	.
Sh	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	67	.	.	.	.	.	.	.	.	.	.
Br	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	33	.	.	.	.	.	.	.	.	.
Bl	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	.	.	.	.	.	.	.	.
Ha	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	55	1	.	.	.	.	.	.
Ja	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	38	.	.	.	.	.	.
Ks	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.	.	.	.	.
Kk	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	.	.	.	.
Ke	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	20	.	.	.
Mb	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	84	.	.
Ra	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	82	.
Tk	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	19

Table 3. Confusion matrix using 60s segments with SVM classifier.

	C	D	E♭	F	G	A♭	B♭
C	19.97	**1.1**	0.58	0.97	**4.23**	1.2	1.36
D	1.22	2.5	**0.62**	**0.05**	0.29	0.1	0.22
E♭	**0.93**	**0.37**	1.58	0.93	0.39	0.22	0.16
F	0.75	0.21	0.2	1.14	**1.06**	0.14	**0.53**
G	**4.23**	0.32	0.61	0.8	9	1.11	0.64
A♭	0.94	0.06	0.15	0.26	**1**	2.68	**0.83**
B♭	1.26	0.5	0.2	0.3	**0.71**	0.53	2.31

Table 4. Pitch-class dyad distribution for *raag Darbari*. Transitions are from row to column. Numbers given are percentages and sum to 100% for the matrix.

	C	D	E♭	F	G	A♭	B♭
C	17.96	1.59	**0.13**	0.64	**1.13**	1.32	1.72
D	1.23	3.67	0.23	**1.25**	0.49	0.4	0.31
E♭	**0.19**	**0.89**	1.46	0.95	0.06	0.08	0
F	0.74	0.35	0.34	1.77	**2.66**	0.42	**0.1**
G	**1.75**	0.82	0.36	1.31	7.38	1.12	0.56
A♭	1.05	0.22	0	0.43	**2.56**	3.77	**0.34**
B♭	**1.41**	0.09	0	0.02	**0.13**	0.55	0.43

Table 5. Pitch-class dyad distribution for *raag Jaunpuri*. Transitions are from row to column. Numbers given are percentages and sum to 100% for the matrix.

tion, by improving onset detection, will be a major focus of future work and will likely lead to much better generalization.

Table 3 shows the confusion matrix for the SVM classifier on 60s segments. Of a total of twenty one errors, eighteen were made amongst similar *raags* that shared the same scale tones and had phrases in common. The other classifiers made similar types of errors.

Figure 2 shows the average PCD of *raags Darbari* and *Jaunpuri*, two *raags* with the same scale. In *Darbari*, F is often used in passing, and is rarely held. On the other hand, many phrases in *Darbari* linger on E♭, a note used more in passing in *Jaunpuri*. Both these characteristics are clearly visible in the PCDs.

Tables 4 and 5 show the dyad probabilities for *Darbari* and *Jaunpuri*. Significant differences are bolded. Here, more subtle distinctions become apparent. For example, the tendency of *Jaunpuri* to skip E♭ in ascending phrases, a rare event in *Darbari*, can be seen from comparing the probabilities of the dyad 'D F' (1.25% vs .05%). Likewise, many essential phrases in *Darbari* center around the descending transition from B♭ to G (skipping A♭), whereas in *Jaunpuri* the descent is usually taken sequentially through all three notes. The B♭ to G dyad probabilities reveal this (.71% vs. .13%).

8 CONCLUSIONS AND FUTURE WORK

The 99% (CV) and 75% (unseen) success rates clearly demonstrate the efficacy of PCDs and PCDDs for *raag* classification, even when many *raags* used an identical set of notes (Table 1). This suggests that essential melodic characteristics beyond simple scale type are captured, allowing the system to recognize stylistic distinctions which for a human require extensive immersion in the genre to learn.

Future work will focus on more difficult cases, such as relatively loud and complex accompaniment and low SNR conditions. As even larger databases are assembled, it will be possible to make comparisons between instrument type

and performance style. It will also be possible to begin to model sequential structure beyond dyads. As discussed, more robust methods for analyzing melodies with gliding tones will need to be developed.

9 REFERENCES

[1] V.N. Bhatkande. *Hindusthani Sangeet Paddhati*. Sangeet Karyalaya, 1934.

[2] Leo Breiman. Random forests. *Machine Learning*, 45(1), 2001.

[3] MA Castellano, JJ Bharucha, and CL Krumhansl. Tonal hierarchies in the music of north india. *Journal of Experimental Psychology*, 1984.

[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] Ching-Hua Chuan and Elaine Chew. Audio key-finding using the spiral array ceg algorithm. In *Proceedings of International Conference on Multimedia and Expo*, 2005.

[7] Patricio de la Cuadra, Aaron Master, and Craig Sapp. Efficient pitch detection techniques for interactive music. In *Proceedings of the International Computer Music Conference*, pages 403–406, 2001.

[8] 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)*, pages 275–280, London, 2003.

[9] 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.

[10] David Huron. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, 2006.

[11] C. Krumhansl and R. Shepard. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5(4):579–594, 1979.

[12] Carol Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, 1990.

[13] 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*, pages 1350–1363, 2003.

[14] Craig Sapp. Visual hierarchical key analysis. *Computers in Entertainment*, 3(4), October 2005.

[15] Xuejing Sun. A pitch determination algorithm based on subharmonic-to-harmonic ratio. In *In Proc. of International Conference of Speech and Language Processing*, 2000.

[16] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.