

Automatic rag classification using spectrally derived tone profiles

(Proceedings of the 2004 International Computer Music Conference, *Miami: University of Florida.*)

Parag Chordia

CCRMA, Stanford University
pchordia@ccrma.stanford.edu

Abstract

Rag is the central melodic concept of Indian art music. The current system classifies segments from rag performances beginning at the signal level based on a spectrally derived tone profile. A tone profile is essentially a histogram of note values weighted by duration. The method is shown to be highly accurate (100% accuracy) even when rags share the same scalar material and when segments are drawn from different instruments and different points in the performance. Rag identification has important applications in automatic transcription, genre classification and musical database queries.

1 Background

Rag is the central melodic concept in Indian art music. Roughly, it is a melodic structure that is somewhere between a scale, an abstract class that specifies what notes are used, and through-composed melodies in which notes and durations are fixed. A rag is most easily defined as a collection of basic phrases and a set of transformations that operate on these phrases. Basic phrases are sequences of notes that may contain ornaments and articulation information. Transformations are operations which are used to construct the musical surface from the basic phrases. For example, a larger phrase can be constructed by the operation of concatenating two basic phrases. In this case, concatenation is the operation. Stretching and compression, in which the total duration of the phrase is increased or decreased, are examples of other transformations. Rag music is essentially monophonic, with percussive accompaniment in certain sections, and the constant presence of a drone that gives a rich texture emphasizing the ground note and the fifth or fourth above it.

The use of basic phrases in constructing musical gestures leads naturally to note hierarchies. This is because certain notes appear more often in phrases, or are agogically stressed because they appear at phrase endings. In this way, the basic phrases give rise to a tonal hierarchy. Indian music theory has a rich vocabulary for describing this hierarchy. The most stressed note is called the *vadi* and the second most stressed note *samvadi*. Notes on which phrases

typically begin or terminate are also named. A typical summary of a rag includes its scale type (*that*), *vadi* and *samvadi*. A tone profile, which gives the weighted pitch distribution, can be seen as a natural generalization of this type of description.

Krumhansl (1990) as well as Castellano and Bharucha (1984) have shown that stable tone profiles give rise to a mental schemas that structure expectations and facilitate the processing of musical information. This was demonstrated by Krumhansl (1979) who showed, using the probe tone method, that listeners ratings of the appropriateness of a test tone in relation to a key defining context is directly related to the relative prevalence of that pitch chroma. Recent experiments by Aarden (2002) have given further evidence of the psychological reality of these schemas. Aarden showed that decisions which required listeners to compare a test tone with a previous tone, taken from an established tonal context, were significantly faster and more accurate when the test tone belonged to the same tonal context. In other words, listeners are quite sensitive to pitch distributions and their expectations and reactions are structured by this context. Taken together, these findings provide evidence for the possible utility of the tone profile as a classification tool.

2 Motivations

The current work has several motivations. The first is to determine to what extent *rags* can be discriminated using the non-sequential information present in a tone profile. Even in *rags* with the same set of notes, the basic phrases often differ enough that we see recognizable differences in their tone profiles. For the method to be successful, segments drawn from a given *rag* must have pitch distributions which are more similar than segments drawn from different *rags*. Success of this method would demonstrate that different *rags* give rise to different pitch distributions that are relatively stable when viewed in windows shorter than the total duration of the performance.

Secondly, accurate *rag* classification would give us prior information that would be useful for automatic transcription of Indian art music. Specifically, knowing the *rag* would

allow us to specify an appropriate hidden Markov model that could be used for accurate note transcription.

Finally, automatic *rag* identification would be a useful tool in musical retrieval and classification tasks. For example, because there is a relatively well developed theory linking *rags* to emotional states, *rag* identification would be a useful component in a system that tried to retrieve music based on user's emotional state preferences.

3 Related Work

Computer-based analysis of Indian music has been relatively rare, and up to this point few attempts have been made to automatically classify *rags*. On the other hand, *rag* classification has been a central topic in Indian music theory and has led to a rich debate on the essential characteristics of *rags* and the features that make two *rags* similar or dissimilar (Bhatkande 1934).

Recently Pandey et al. (2003) developed TANSEN, a system to automatically recognize *rags* using a Markov model. Transitions between notes are modeled as a first-order Markov process. For each *rag*, a transition table is learned that gives the probability of proceeding to other notes based on the current note. Note that this will vary by *rag*, giving a different underlying model for each. The classification task then asks, given a certain sequence of notes, which model, i.e. *rag*, is it most likely generated by.

Before proceeding to this stage, pitch information is estimated using a spectrally based fundamental frequency estimation algorithm. This continuous frequency vs. time information is discretized into notes by using a heuristic that looks for note boundaries at local maxima in the time vs. frequency plot.

Although this approach has the same end goal, its methodology does not allow it to address two aims of the current project, namely determining the discriminatory power of the tone profile, and a means of identifying the *rag* without note information to help in the automatic transcription task.

Additionally, they report testing the system on only two *rags*, *Bhupali* and *Yaman Kalyan* which do not share the same set of notes¹, making the classification problem relatively straightforward. Finally, the source and length of the samples are not specified making it difficult to evaluate the results. Nevertheless, their approach is methodologically clear choice when note information is available.

Other related work is that of Sahasrabudde (1992, 1994) who attempted to model *rags* as finite automata by codifying *rag* principles found in standard texts. The goal of this work was to generate note sequences that were grammatically correct. Although an early example of computer modeling of *rags*, it did not address the problem a

the signal level and was concerned more with synthesis of strings than classification.

4 Method

The algorithm proceeded in two steps, the generation of the tone profiles, and classification using a simple k nearest neighbor algorithm.

A central challenge was developing a way of approximating the tone profile without using note information. In the ideal case, if the sounds are perfectly harmonic and the fundamental and first harmonic are relatively strong, then the tone profile can be approximated by taking the DFT of the segment and summing energy in bins surrounding the note centers. In this paper, I will refer to this as the spectral profile.

Even if the notes contained substantial energy in the higher partials, if the timbral structure of the different instruments are close enough, we would expect segments which share the same distribution of notes to have similar spectral profiles. However, in this case the spectral profile would not necessarily resemble the tone profile. For the purpose of classification however, it is sufficient if the spectral profiles of segments drawn from the same pitch distribution are more similar than profiles derived from different pitch distributions.

Although it was expected that pitch distributions would be relatively stable for large segments, such as entire pieces or movements, it was not clear at the beginning how stable they would be for shorter segments. Indeed, because a performer may emphasize certain phrases in certain sections we might expect the tone profile to depend on the section from which the segment was drawn.

<i>Rag Name</i>	<i>Notes used (Given as scale degrees. # = sharp, \$ = flat)</i>
<i>Bageshri</i>	1, 2, 3, 4, 5, 6, \$7
<i>Bahar</i>	1, 2, \$3, 4, 5, 6, \$7
<i>Chayanat</i>	1, 2, 3, 4, 5, 6, \$7, 7
<i>Darbari</i>	1, 2, \$3, 4, 5, \$6, \$7
<i>Jaunpuri</i>	1, 2, \$3, 4, 5, \$6, \$7
<i>Jog</i>	1, 2, \$3, 3, 4, 5, 6, \$7
<i>Kamod</i>	1, 2, 3, 4, #4, 5, 6, 7
<i>Kedar</i>	1, 2, 3, 4, #4, 5, 6, 7
<i>Kirvani</i>	1, 2, \$3, 4, 5, \$6, 7
<i>Lalit</i>	1, \$2, \$3, 4, #4, 5, \$6, 7
<i>Marva</i>	1, \$2, 3, #4, 6, 7
<i>Miyan ki Malhar</i>	1, 2, \$3, 4, 5, 6, \$7, 7
<i>Miyan ki Todi</i>	1,\$2, \$3, 4, #4, 5, \$6, 7

Table 1: Scale degrees of each *rag* in corpus

Ten segments were drawn from thirteen different *rag* performances, for a total of one hundred and thirty segments. Each segment was one minute long. This duration was chosen because it was the minimum duration that we

¹*Yaman Kalyan* has the following note set: {1, 2, 3, 4, #4, 5, 6, 7}. *Bhupali*: {1, 2, 3, 5, 6} Numbers denote the scale degrees.

expected to be able to see a stable tone profile. The *rags* were chosen to represent a diverse set of scale types as well as containing *rags* with the same scale set (Table 1). For

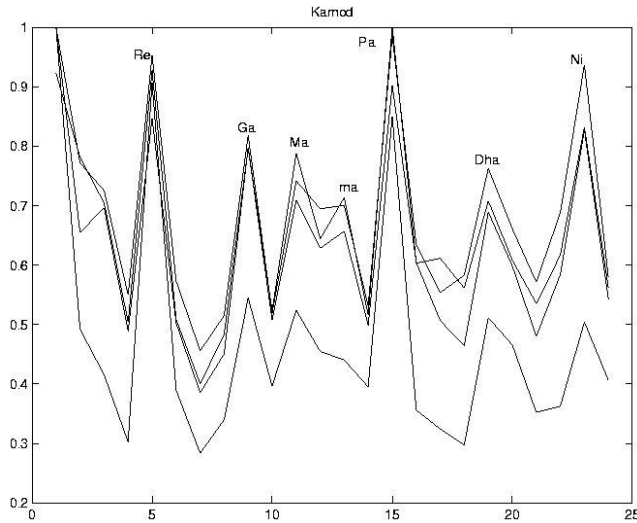


Figure 1. Spectral Profiles of Rag Kamod. (x-axis gives bin number, y-axis relative prevalence)

example, *rags Bageshri* and *Bahar* have notes drawn from the Dorian mode. The only constraint in choosing the examples was that they were excerpted from relatively short performances, that is performances whose total duration was less than twenty minutes. This constraint was introduced because in long renditions of a *rag* a performer might focus on one or two phrases for a minute or longer. In this case the local tone profile will not reflect the average tone profile for the *rag*. On the other hand, in longer renditions it is possible to use longer segments for identification.

The segments were were sung, or were played on the flute or *sarod*. This was important in determining the spectral profile's robustness to different instruments. Further, segments were drawn from different parts of the performance. Some were taken from the *alap*, the unmetered solo section, and some from the portion with percussive *tabla* accompaniment. It was desired that the method generalize to different forms.

Each segment was normalized and the spectral profile calculated. Because only the distribution over pitch chromas was desired, frequency information was folded into one octave before summing. Profiles were calculated by summing the total energy in quarter tone bins surrounding each note. Semitone bins were also tried but it was found that the choice of bin did not effect the classification performance.

Because different segments had different absolute values for the tonic, the tonic had to first be calculated. The presence of the drone, which constantly sounds the tonic, made this straightforward; the largest peak in the DFT invariably corresponded to the tonic.

The segments were then split into training (60%) and test sets (40%). The relative balance of segments from

different *rags* was preserved in each set. Classification was done using a simple k nearest neighbor algorithm. In the nearest neighbor algorithm, the segment to be classified is

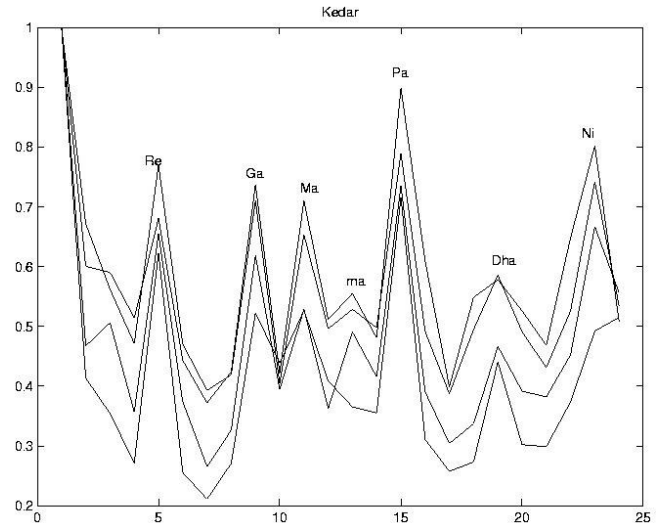


Figure 2. Spectral Profiles of Rag Kedar.

compared with all the segments in the training set. It is assigned the label of the segment most closely resembles it. In other words, the distance of a tone profile from each tone profile in the training set is calculated using a euclidean metric. The k nearest neighbor algorithm generalizes this to using the labels of the k nearest segments and assigning the one that appears most frequently amongst them.

5 Results and Discussion

One hundred percent of the segments were accurately labeled. It was particularly encouraging that *rags* which used the same notes were accurately distinguished. Figures 1 and 2 show *rags Kamod* and *Kedar* respectively. Each curve is the spectral profile of a segment. These two *rags* use notes from the major scale, with the addition of the augmented fourth (see Table 1). *Kamod* and *Kedar* differ, however, in their phraseology, and this is clearly reflected in the spectral profile. The figures show the spectral profiles of a number of segments from a given *rag*.

For example in *Rag Kamod*, the tonic, the fifth (labeled *Pa*) and the second (*Re*) were the most prevalent. This corresponds to the *vadi* and *samvadi* ascribed to this *rag*. In *Rag Kedar* we see that the second (*Re*) is less prominent. This follows from the basic phrases; the second scale degrees appears more frequently and prominently in *Rag Kamod*.

The examples from both these *rags* were played on the *sarod* confirming that the difference between the spectral profiles were due to differences in the underlying note distribution rather than instrument differences. In the case of *Rags Kamod* and *Kedar*, the spectral profile resembled the expected tone profile. In other *rags*, spectral profiles were also quite consistent between samples, suggesting that pitch

distributions are in fact quite stable, even for relatively small segments. However other spectral profiles, although displaying characteristic shapes for each *rag* (Figures 3 and 4), did not correspond closely to expected tone profiles. In

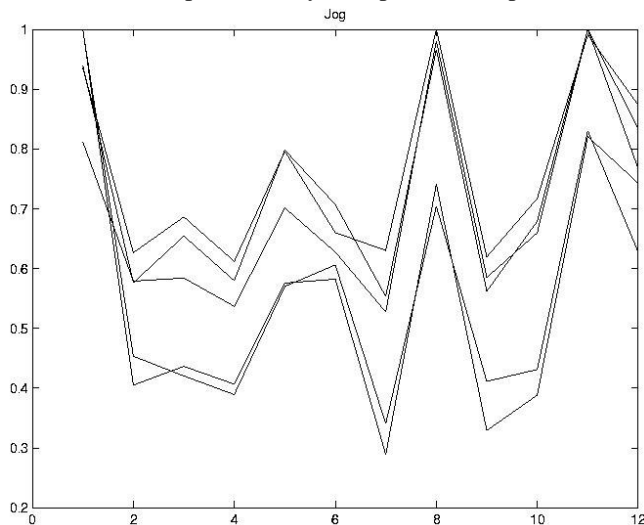


Figure 3. Spectral Profiles of Rag Jog.

other words the peaks did not necessarily correspond to the most used notes.

6 Conclusions and Future Work

Although the data set is still too small to make sweeping generalizations, it appears that spectral profiles in a given *rag* have a characteristic shape independent of the instrument and section of the performance they are drawn from. The current work is promising and suggests that spectral profiles have substantial discriminatory power in automatic *rag* classification.

Future work will address the limits of this approach. For example, as we increase the diversity of our sample set by increasing the number of closely allied *rag*s, and the diversity of instruments we expect that performance will deteriorate. The next step will be to find the performance bounds of this approach. In any case, the algorithm will undoubtedly be useful in *rag* classification when note information is not readily available making it suitable for defining a tonal context within which an automatic transcription system can work.

References

Aarden, B. (2002). "Expectancy vs. retrospective perception: Reconsidering the effects of schema and continuation judgments on measures of melodic expectancy." In: C. Stevens, D. Burnham, G. McPherson, E. Schubert & J. Renwick (Eds.) Proceedings of the 7th International Conference on Music Perception and Cognition. Adelaide: Causal Productions, pp.469-472.

Bhatkande, V.N. (1934). Hindusthani Sangeet Paddhati (6 vols.). Bombay, 1934.

Castellano et al. (1984). "Tonal Hierarchies in the Music of North India" *Journal of Experimental Psychology: General* 113, pp. 394-412.

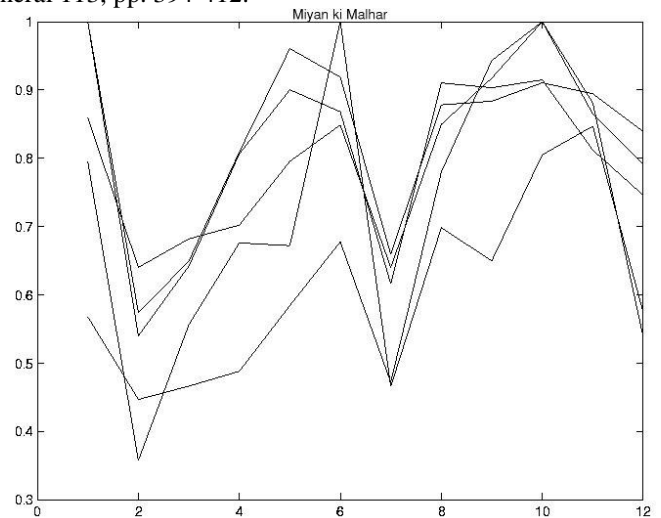


Figure 4. Spectral Profiles of Rag Miyan ki Malhar.

Krumhansl, C., & Shepard, R.N. (1979). "Quantification of the hierarchy of tonal functions within a diatonic context." *Journal of Experimental Psychology: Human Perception & Performance*, Vol. 5, No. 4, pp. 579-594.

Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Pandey, Gaurav, Mishra, Chaitanya and Paul Ipe (2003). "TANSEN : A System for Automatic Raga Identification." Proceedings of the 1st Indian International Conference on Artificial Intelligence, pp. 1350-1363, Hyderabad, India.

Sahasrabudde, H.V., and R. Upadhye (1992). "On the Computational Model of Raag Music of India." Workshop on AI and Music: 10th European Conference on AI, Vienna.

Sahasrabudde, H.V. (1994). "Searching for a Common Language of Ragas." Proc. Indian Music and Computers: Can Mindware and Software Meet?