CHARACTERIZATION OF INTONATION IN KARṆĀṬAKA MUSIC BY PARAMETRIZING CONTEXT-BASED SVARA DISTRIBUTIONS

Gopala Krishna Koduri[1], Joan Serrà[2], and Xavier Serra[1]
[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.
[2] Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain.
gopala.koduri@upf.edu, jserra@iiia.csic.es, xavier.serra@upf.edu

ABSTRACT

Intonation is a fundamental music concept that has a special relevance in Indian art music. It is characteristic of the rāga and intrinsic to the musical expression of the performer. Describing intonation is of importance to several information retrieval tasks like the development of rāga and artist similarity measures. In our previous work, we proposed a compact representation of intonation based on the parametrization of the pitch histogram of a performance and demonstrated the usefulness of this representation through an explorative rāga recognition task in which we classified 42 vocal performances belonging to 3 rāgas using parameters of a single svara. In this paper, we extend this representation to employ context-based svara distributions, which are obtained with a different approach to find the pitches belonging to each svara. We quantitatively compare this method to our previous one, discuss the advantages, and the necessary melodic analysis to be carried out in future.

1. INTRODUCTION

Indian art music has two main branches: Karṇāṭaka and Hindustānī music, the former more prevalent in the Indian peninsular, the latter more prevalent in northern regions of the Indian subcontinent. Rāga is the melodic framework on which Indian art music relies. A given rāga is described by a set of properties: A set of svaras [1], their progressions (ārohaṇa/avarōhaṇa), their intonation involving various movements (gamakas), and their strength, duration, and positions relative to each other (functionality of svaras) [1].

In the literature, it is shown that the intonation of a given svara can vary significantly depending on the artist and the rāga [2–4]. Therefore, obtaining a representation of intonation for computational purposes is a necessary step to characterize rāgas and artists. In our previous work [4], we obtained a compact representation of intonation by parametrizing pitch histograms, of which in this paper we present an extension. In the following sections, we briefly summerize our previous work, present the changes to that approach and compare the performance of both the methods in a rāga classification task.

2. HISTOGRAM PARAMETRIZATION

We hypothesize that the intonation of a svara is manifest in a pitch histogram, specifically in the shape of the distribution of pitches around the svara positions. Therefore, the goal of our intonation description approach is to obtain the parameters that describe the shape of the distribution around each svara in a given histogram. We detailed this parametrization, denoted as M_h, in [4].

The parametrization of the svaras can be broadly divided into six steps. In the first step, the prominently vocal segments of each performance are extracted using a trained support vector machine (SVM) model. In the second step, the pitch corresponding to the voice is extracted using multipitch analysis [5]. In the third step, using all the performances of each rāga, an average pitch histogram for every rāga is computed and its prominent peaks detected. In the fourth step, we compute the pitch histogram for each single performance, detecting the relevant peaks and valleys using information from the overall histogram of the corresponding rāga. In the fifth step, each peak is characterized by using the valley points and an empirical threshold. Finally, in the sixth step, the parameters that characterize each of the peak distributions are extracted: mean, variance, position, amplitude, skew and kurtosis.

In [4], we evaluated M_h using an explorative rāga classification task in which three rāgas were classified based on the parameters of just a single svara. The results showed the usefulness of the approach as it outperformed the baseline approach which consists of just the position and amplitude parameters. However, this approach completely discards the contextual information of pitches: mainly the melodic & temporal contexts. The melodic context of a pitch instance refers to the larger melodic movement of which a given pitch is part of. The temporal context refers to the properties of the modulation: a fast intra-svara movement, a slower inter-svara movement, a striding glide from one svara to another, etc. The histogram analysis is an aggregation-based approach and it is thus not feasible to employ such contextual information.

In M_h, a pitch value gets the same treatment irrespective of where it occurs in pitch contour. Consider the following two scenarios: (i) a given svara being sung steadily for

[1] A svara-sthana is a frequency region which indicates the note and its allowed intonation in different melodic contexts.

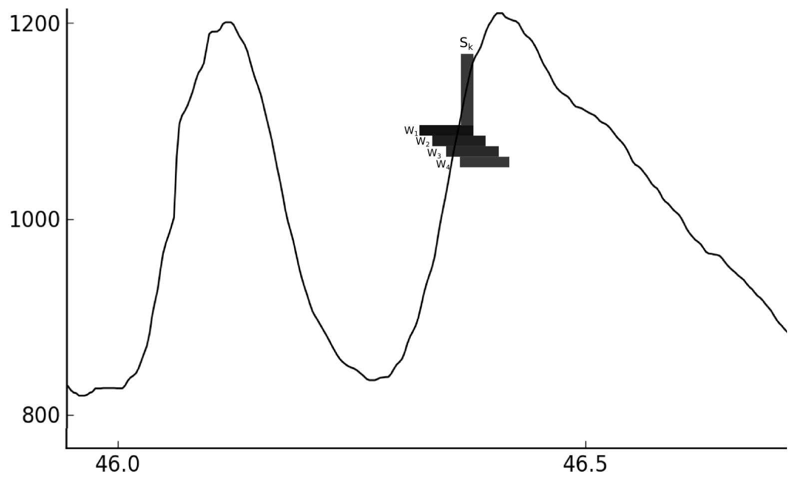


Figure 1: The movement of windows shown for a given segment S_k, which spans t_h milliseconds. In this case, $t_w = t_h*4$.

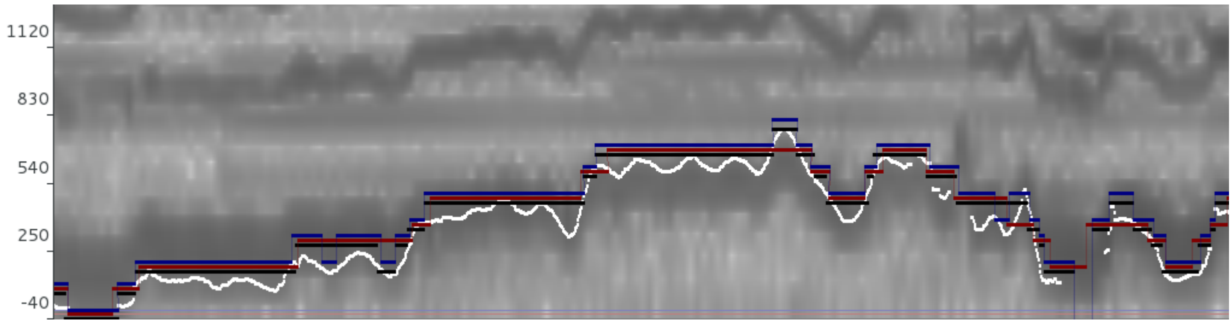


Figure 2: The pitch contour (white) is shown on top of the spectrogram of a short segment from a Karṇāṭaka vocal recording. The red (t_w = 150ms, t_h = 30ms), black (t_w = 100ms, t_h = 20ms) and blue (t_w = 90ms, t_h = 10ms) contours show the svara to which the corresponding pitches are binned to. The red and blue contours are shifted few cents up the y-axis for legibility.

some time duration, and (ii) the same svara appearing in a quick transition between two neighboring svaras. In M_h, it is not possible to handle them differently. But in reality, the first occurrence should be part of the given svara's distribution, and the second occurrence should belong to either of the neighboring svaras depending on which is more emphasized. The objective of the new method we propose, M_c, is to handle such cases by incorporating the local melodic and temporal context of the given pitch value.

3. CONTEXT-BASED SVARA DISTRIBUTIONS

In the context-based parametrization we propose, M_c, the pitches corresponding to each svara distribution are found from the pitch contour itself, taking into account the modulations in the pitch contour surrounding a given pitch instance. The pitch contour is viewed as a collection of small segments. For each segment, we consider the mean values of a few windows containing the segment. The windows are positioned in time such that the segment moves from the end of the first window to the beginning of the last window. The mean values provide us with the necessary contextual information. Figure. 1 shows the movement of windows for a given segment S_k. The f0 samples of the segment belong then to the svara distribution which is nearest to the median of the mean values.

To achieve this, we define a moving window with window size set to t_w milliseconds and hop size set to t_h milliseconds. For a k-th hop on pitch contour P, k=0,1,...$\frac{N}{t_h}$, where N is the total number of samples of the pitch contour, we define segment (S_k), mean (μ_k), number of windows (ϵ) and median ($\bar{m}_k$) as:

$$S_k = P[t_w + (k-1)t_h : t_w + kt_h] \tag{1}$$

$$\mu_k = \frac{1}{t_w} \sum_{i=kt_h}^{t_w+kt_h} P[i] \tag{2}$$

$$\epsilon = \frac{t_w}{t_h} \tag{3}$$

$$\bar{m}_k = \text{median}(\mu_k, \mu_{k+1}, \mu_{k+2} \ \dots \ \mu_{k+\epsilon-1}) \tag{4}$$

S_k therefore, is a subset of pitch values of P as given by Eq. 1. μ_k is the mean of the given window (which contains S_k). ϵ is the total number of windows a given segment S_k can be part of, and is constrained by t_w and t_h. $\bar{m}_k$ is the median of the set of μ_k values of the ϵ windows.

Given the Eqs. 1-4, a pitch-distribution of a svara I in Γ, a predefined array of just-intonation intervals corresponding to four octaves, is obtained as:

$$\mathbb{D}_I = \{S_k \mid argmin_i \ \delta(\Gamma_i, \bar{m}_k) = I\} \tag{5}$$

Rāga	Recordings	Duration (minutes)	Performers
Bēgaḍa	8	60	8
Bhairavi	16	332	14
Hindōlaṁ	7	107	5
Kāṁbhōji	11	254	10
Mukhāri	8	81	8
Tōḍi	19	652	15
Total	**69**	**1487**	**29**

Table 1: Karṇāṭaka music dataset used for the evaluation of the two methods.

Method/Classifier	3-Nearest Neigh.	Naive Bayes	Multilayer Perceptron	SVM
M_h	0.37	0.54	0.43	0.48
M_c	0.50	0.51	0.54	0.55

Table 2: F-measures shown for the two methods in a rāga classification task. The baseline calculated using zeroR classifier lies at 0.15.

where δ is a function that gives difference between the arguments.

The crucial step in this method is to determine the window and hop sizes. A large window has an advantage that it incorporates larger temporal context, but it is bad for inter-svara modulations as they will be averaged to a mean. On the other hand a small window might defeat the whole purpose, by including no meaningful context at all. Eqs. 1 and 3 require the hop size to be reasonably smaller than the window size. This is crucial since for any segment S_k, a larger number of μ_k values carry a larger temporal context. This helps to take a better decision in determining the svara distribution to which S_k belongs to. Also keeping in mind the computational limitations, we found empirically that t_w = 100ms and t_h = 20ms perform considerably good for most performances. Figure 2 shows the results for (t_w = 150ms, t_h = 30ms), (t_w = 100ms, t_h = 20ms) and (t_w = 90ms, t_h = 10ms).

In the figure, the intra-svara movements tend to be associated with the corresponding svara whereas the inter-svara movements are segmented and distributed appropriately. In effect, this method alleviates the need for peak detection and finding the distribution bounds as we obtain each svara distribution independently. These two steps which are part of M_h have their own limitations. The peak detection algorithm is prone to pick erroneous peaks and leave out few relevant ones. The latter is common phenomena when the peak at a svara appears as a bump on the peak of neighboring svara. On the other hand, in order to estimate the parameters it is necessary to determine the bandwidth of peaks from the histogram. In the cases where the valley points of a peak are not so evident and the peak distribution overlapped with that of a neighboring svara, we chose a hard bound of 50 cents on either side of the peak. This affects the parameters computed for the distribution. Such issues does not arise with M_c as it does not require these two steps.

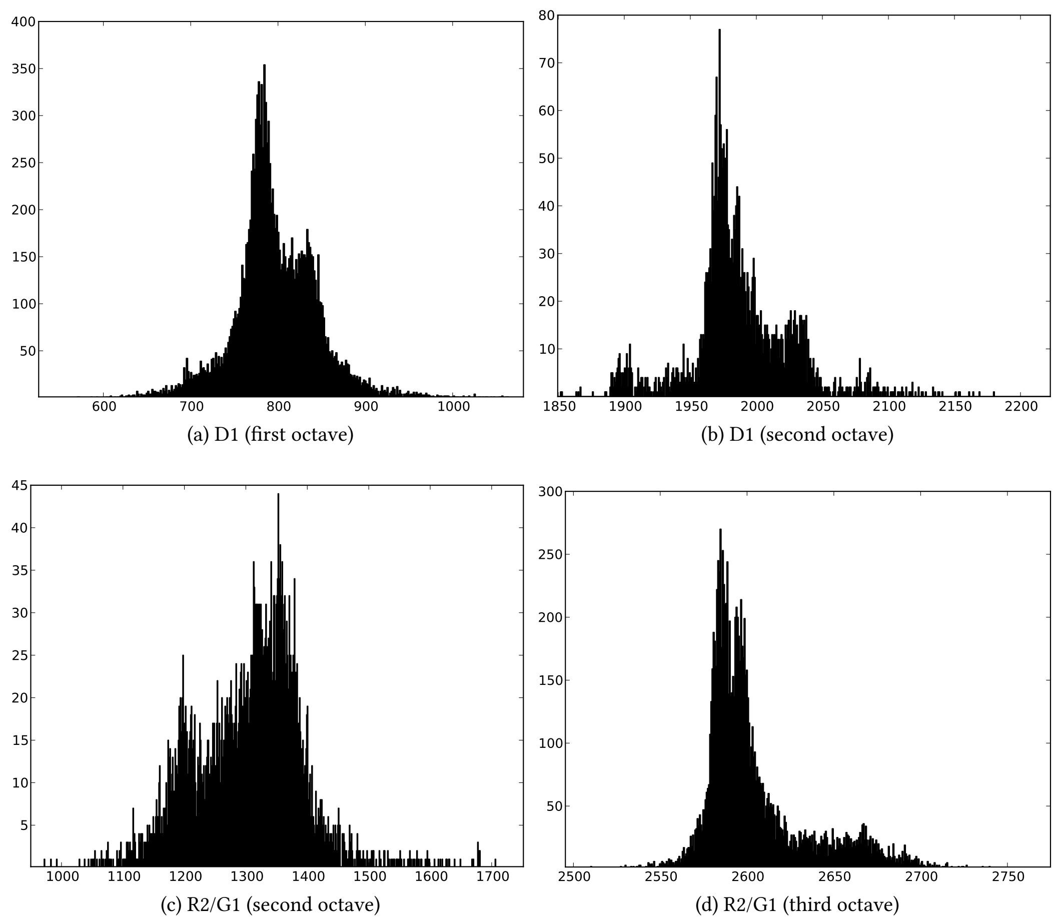
4. EVALUATION & RESULTS

We evaluated both methods using a rāga classification task. For this, we used a Karṇāṭaka music dataset of 69 recordings in six rāgas. Table 1 shows the details of the dataset. The same svara parameters for all the recordings are obtained using both methods: position, amplitude, mean, variance, skew and kurtosis. The position and amplitude parameters are deliberately removed as their values obtained using the two methods are more or less the same. From among the remaining parameters, we got 192 features (4 parameters x 12 svaras x 4 octaves). To reduce this number for classification, we used information-gain feature selection algorithm [6, 7] to select the 10 best parameters for rāga classification. To ensure fairness the dataset is subsampled 10 times with 7 recordings per rāga. Each subsample is split with 2:1 train/test ratio. This is done 10 times randomizing the order of recordings in the sample (10 subsampled datasets x 10 randomizations = 100 times). We chose k-nearest neighbour, naive bayes, multilayer perceptron and SVM classifiers for the experiment [2]. This way, by testing consistent differences in accuracy across several classifiers we can be more confident of the improvements of our newly proposed approach. Table 2 shows the results averaged over each classifier.

The results are indicative of the slightly better performance of M_c over M_h. The only exception occurs with naive bayes, where the difference between their performances is not large.

There is plenty of scope to further improve M_c. At the moment, the overall distribution parameters are the only source of information which we have taken into account. Consider the distributions obtained by M_c shown in Figure 3. Though each of them has one dominant peak, it is also characterized by one or more other minor peaks. These peaks as such may or may not correspond to another svara, but they do signify a melodic context in which

[2] The implementations provided with Weka were used with the default parameters.



(a) D1 (first octave)

(b) D1 (second octave)

(c) R2/G1 (second octave)

(d) R2/G1 (third octave)

Figure 3: Various svara distributions from a Karṇāṭaka vocal recording. Y-axis is in cents. Besides the dominant peak, such distributions are also characterized by the presence of other minor peaks, which can be important in characterizing the intonation of the svara.

the svara occurs: a frequently co-occuring svara, a melodic movement over this svara which involves another svara or viceversa, etc. Though this information is partly contained by a few of the existing parameters such as skew and kurtosis, it might help greatly to consider it as an additional parameter.

5. CONCLUSIONS & FUTURE WORK

In this paper, we have extended our previous work on characterizing intonation in Karṇāṭaka music by incorporating local melodic context in obtaining svara distributions. The results from the evaluation task seem to indicate that our new approach performs better compared to the earlier one. We now discuss the future direction of this work to include the ascending and descending patterns of svaras, and improve the current method to gather more contextual information about the pitches.

5.1 Ārōhaṇa & Avarōhaṇa

Rāga is usually represented as a set of ascending and descending svaras. These more or less define the possible melodic movements, or rather prevent a few movements deeming them inappropriate for the melodic context of that particular rāga (see [8] for more details and an example). This is bound to impact the way svaras are sung: with or without gamakas, the extent of gamakas, svaras sung in a given gamaka and so on. Therefore, obtaining separate distributions for each svara in its ascending and descending contexts might be helpful.

5.2 Better contextual information

M_c derives the context of a pitch value in the melodic contour, which is constrained by t_w and t_h. Another possible method that would alleviate such constraints is to derive the context based on the characteristics of melodic movement which the given pitch instance is located in. We locate the nearest peak and valley on either side of it and calculate the slope with the two points. The slope, and the absolute pitch difference (in cents) between the peak and valley, together will set the melodic context of the given pitch instance. Essentially these two parameters indicate the nature of the modulation. This information over a range of pitches can be used to group them into a meaningful unit (such as a particular gamaka) based on their

Collection	Recordings	Performers (min. 5 recs/all)	Rāgas (min. 5 recs/all)	Compositions (min. 2 recs/all)
Karṇāṭaka	1001	56/60	62/184	39/303
Hindustānī	495	33/64	24/159	10/218

Table 3: F-measures shown for the two methods in a rāga classification task. The baseline calculated using zeroR classifier lies at 0.15.

patterns. For instance, a large kampita [3] will result in a pattern of slopes, pitch diferences which is different from that of a single glide surrounded by other modulations.

The role of amplitude is another direction which deserves attention. It helps in understanding the perceptual importance of a given pitch instance. One possible modification of obtaining parameters would be a simple weighting of pitches with their amplitude.

5.3 Dataset

In our work, our major concern till now was to obtain a meaningful description of intonation for computational purpose. Therefore, we have come up with small rāga classification tasks that helped us to cross check whether the description is useful. Now that we have a reasonably meaningful and useful description, we intend to use it in several tasks: primarily rāga and artist characterization. A possible application out of this is browsing the audio collections with similarity measures based on rāgas and artists. The Karṇāṭaka and Hindustānī collections put together as part of CompMusic project provide a suitable ground for implementing such an application. Table. 3 shows the relevant statistics of the collections in their current form.

Acknowledgments

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). JS acknowledges JAEDOC069/2010 from Consejo Superior de Investigaciones Científicas and 2009-SGR-1434 from Generalitat de Catalunya.

6. REFERENCES

[1] T. Viswanathan and M. H. Allen, *Music in South India.* Oxford University Press, 2004.

[2] D. Swathi, "Analysis of Carnatic Music : A Signal Processing Perspective," Masters Thesis, IIT Madras, 2009.

[3] M. Levy, *Intonation in North Indian Music.* New Delhi: Biblia Implex Pvt. Ltd, 1982.

[4] G. K. Koduri, J. Serrà, and X. Serra, "Characterization of Intonation in Carnatic Music by Parametrizing Pitch Histograms," in *ISMIR*, 2012, p. In press.

[5] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *Audio, Speech, and Language Processing, IEEE Transactions on*, no. 99, pp. 1–1, 2012.

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.

[7] T. Ngo, "Data mining: practical machine learning tools and technique, third edition by Ian H. Witten, Eibe Frank, Mark A. Hell," *SIGSOFT Softw. Eng. Notes*, vol. 36, no. 5, pp. 51–52, 2011.

[8] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Rāga Recognition based on Pitch Distribution Methods," *Journal of New Music Research*, p. In press, 2012.

[3] Kampita is one of the gamakas which means oscillation. It can mean anything between a vibrato on a svara and larger modulations involving different svaras.