CHARACTERIZATION OF INTONATION IN CARNATIC MUSIC BY PARAMETRIZING PITCH HISTOGRAMS

Gopala K. 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 an important concept in Carnatic music that is characteristic of a raaga, and intrinsic to the musical expression of a performer. In this paper we approach the description of intonation from a computational perspective, obtaining a compact representation of the pitch track of a recording. First, we extract pitch contours from automatically selected voice segments. Then, we obtain a a pitch histogram of its full pitch-range, normalized by the tonic frequency, from which each prominent peak is automatically labelled and parametrized. We validate such parametrization by considering an explorative classification task: three raagas are disambiguated using the characterization of a single peak (a task that would seriously challenge a more naïve parametrization). Results show consistent improvements for this particular task. Furthermore, we perform a qualitative assessment on a larger collection of raagas, showing the discriminative power of the entire representation. The proposed generic parametrization of the intonation histogram should be useful for musically relevant tasks such as performer and instrument characterization.

1. INTRODUCTION

Carnatic music is the south Indian art music tradition and Raaga is the melodic framework on which Indian art music thrives. The intonation of a single swara [1] can be different due to the melodic context established by different raagas. Therefore, to understand and model Carnatic music computationally, intonation analysis becomes a fundamental step.

We define intonation as the pitches used by a performer in a given musical piece. From this definition our approach will consider a performance of a piece as our intonation unit. In Carnatic music practice, it is known that the intonation of a given swara can vary significantly depending on the artist and the raaga [8, 15]. The study of intonation differs from that of tuning in its fundamental emphasis. When we talk about tuning we refer to the discrete frequencies with which we tune an instrument, thus it is more of a theoretical concept than the one of intonation, with which we focus in the pitches used during a performance. The two concepts are basically the same when we study instruments that can only produce a fixed set of discrete frequencies, like the piano.

Given than in Indian music there is basically no instrument with fixed frequencies (the harmonium is an important exception), in practice tuning and intonation can be considered the same. Here we will maintain the terms, tuning or intonation, used by the different studies.

Krishnaswamy [7] discusses various tuning studies in the context of Carnatic music, proposing a hybrid tuning scheme based simple frequency ratios plus various tuning systems, specially equal temperament. His work also points out the lack of empirical evidence thus far. Recently, Serrà et al. [12] have shown important quantitative differences between the tuning systems in modern Carnatic and Hindustani [2] musics. In particular, they show that Carnatic music follows a tuning system which is very close to just-intonation, whereas Hindustani music follows a tuning system which tends to be more equi-tempered. Although there are several studies on intonation and tuning in Indian music, the emphasis so far has been in the interpretation of ancient texts rather than on analysing real musical practise(see [7, 8, 12] and references therein).

In a study that was conducted with Hindustani music performances [8], pitch consistency is shown to be highly dependent on the nature of gamaka [3] usage. The swaras sung with gamakas were often found to have a greater variance within and across the performances, and across different performers. Furthermore, the less dissonant swaras were also found to have greater variance. However, it was noted that across the performances of the same raaga by a given performer, this variance in intonation was minor. The same work concludes that the swaras used in the analyzed performances do not strictly adhere to either just-intonation or equal-tempered tuning systems. Belle *et al* [1] use the intonation information of swaras to classify Hindustani raagas. Another recent experiment conducted with

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

[2] The north Indian art music tradition.

[3] Gamakas are a class of short melodic movements sung around and between swaras.

Carnatic music performances draws similar conclusions about the variance in intonation [15]. However, the methodology employed in these experiments cannot easily be scaled to a larger set of recordings due to the human involvement at several phases of the study, primarily in cleaning the data and the pitch tracks, and also in interpreting of the observations made.

Approaches to tuning analysis of real musical practise usually follow a so-called 'stable region' approach, in which only stable frequency regions are considered for the analysis (cf. [12]). However, it is known [14] that the most portion of the performance in Carnatic music is gamaka-embellished. Since gamakas are characteristic to a given raaga, such an approach is not suitable to understand the crucial information provided by them. So far, the tuning analysis was approached to explain the interval positions of Carnatic music with one of the known tuning methods like just-intonation or equal-temperament. But considering that these intervals are prone to be influenced by factors like raaga, performer [8] and instrument [7], computational analysis of swara intonation for different raagas, artists and instruments has much more relevance to the Carnatic music tradition.

2. HISTOGRAM PEAK PARAMETRIZATION

In this contribution we propose a methodology based on histogram peak parametrization that helps to describe the intonation of a given recording by characterizing the distribution of pitch values around each swara. From the observations made by Krishnaswamy [7] and Subramanian [14], it is apparent that steady swaras only tell us part of the story that goes with a given Carnatic music performance. The gamaka-embellished swaras pose a difficult challenge for automatic swara identification. Therefore, alternative means of deriving meaningful information about the intonation of swaras becomes important. The gamakas and the role of a swara are prone to influence the aggregate distribution of a swara. We believe that this information can be derived by parametrizing the distribution around each swara.

Our intonation description method 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. In the third step, using all the performances of each raaga, a pitch histogram for every raaga is computed and its prominent peaks detected (we will refer to them as reference peaks). 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 raaga. 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 distributions are extracted.

2.1 Segmentation

Cleaning the data is a crucial pre-processing step for our experiments. All the Carnatic vocal performances are accompanied by a violin and one or more percussion instruments. We just use the sections in which the voice is alone or very prominent. In order to do this automatically, we train an SVM model [5] on 300 minutes of audio data, equally split between vocal, violin and percussion sections of 10 seconds each. The features extracted from the audio and used in the classification task are [4]: Mel-frequency cepstral coefficients, pitch confidence, spectral flatness, spectral flux, spectral rms, spectral rolloff, spectral strong-peak, zero crossing rate and tristimulus. This method scores an accuracy of 96% in a 10-fold cross validation test.

2.2 F0 Analysis

With the segmentation module in place, we minimize to a large extent pitch errors due to the interfering accompanying instruments. However, there is a significant number of the obtained voice segments in which the violinist fills short pauses or in which the violin is present in the background, mimicing the vocalist very closely with a small time lag. This is one of the main problems we encountered when using pitch tracking algorithms like YIN [3], since the violin was also being tracked in quite a number of portions. The solution has been to extract the predominant melody [10] using a multi-pitch analysis approach. Then, given that the pitch accuracy of YIN is better, we compare the pitch obtained from the multi-pitch analysis with YIN at each time frame, and we only keep the pitch from those frames where both methods agree within a threshold. Though it is a computationally intensive step, this helps in obtaining clean pitch tracks, free of f0-estimation and octave errors. The frequencies are then converted to cents and normalized with the tonic frequency obtained using [11]. The octave information is retained.

2.3 Histogram Computation

As Bozkurt et al. [2] point out, there is a trade-off in choosing the bin resolution of a pitch histogram. A good bin resolution keeps the precision high, but significantly affects the peak detection accuracy. However, unlike Turkish-maqam music where the octave is divided into 53 Holdrian commas, Carnatic music uses roughly 12 swaras [13]. Hence, in this context, choosing a finer bin width is not as much a problem as it is in Turkish-maqam music. In order to retain the preciseness in estimating the parameters for such distribution, we keep the bin resolution at one cent. We then compute the histogram H by placing the pitch values into their corresponding bins:

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

where H_k is the k-th bin count, N is the number of pitch values, $m_k = 1$ if $c_k \leq P(n) \leq c_{k+1}$ and $m_k = 0$ otherwise, P is the array of pitch values and (c_k, c_{k+1}) are the bounds on k-th bin.

Features/Classifier	Naive Bayes	1-Nearest Neigh.	SVM	Logistic Regression	Random Forest
Mean and Height	63.43%	56.67%	61.81%	56.33%	64.62%
All parameters combined	63.76%	68.90%	65.19%	68.86%	70.71%

Table 1. Results of an exploration raaga classification test with 42 recordings in 3 raagas using different classifiers. The random baseline accuracy in this case is 28.57%.

Features/Classifier	Naive Bayes	1-Nearest Neigh.	SVM	Logistic Regression	Random Forest
Mean and Height	39.6%	39.85%	41.25%	43.65%	48.85%
All parameters combined	58.05%	67.6%	74.25%	77.45%	74.45%

Table 2. Results of an exploration raaga classification test with 26 recordings in 2 raagas using different classifiers. The random baseline accuracy is 20% in this case.

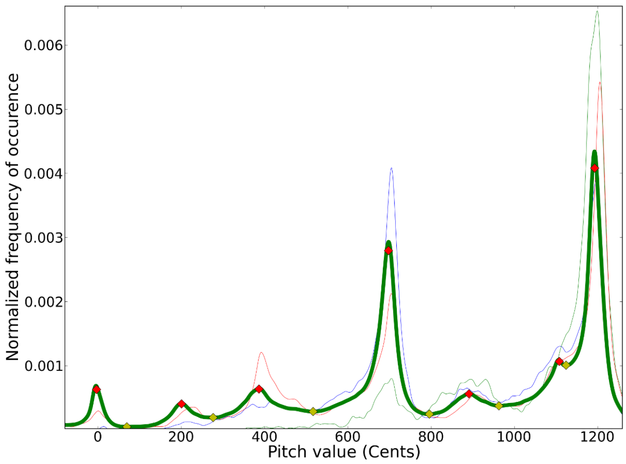

Figure 1. Histograms corresponding to the recordings of Kalyani raaga shown in thin lines of different colors, together with the average histogram labelled with peaks and valleys, shown in thick green line. Only the middle octave, which contains the most information, is shown.

In the histograms of a few performances, we observe that a number of pitch distributions for specific swaras are very narrow. However we can observe that the distributions still play a role in characterizing the performance. To validate this observation, an average histogram is computed for each raaga with all the performances in the raaga. This histogram has a clearer distribution for each swara in the raaga and it serves as a reliable reference to verify the peaks identified in individual performances (Fig. 1).

2.4 Peak Detection and parametrization

A given histogram is convolved with a Gaussian kernel using a standard deviation of five cents. This step is necessary for the peak detection algorithm to avoid identifying the spurious peaks. The peaks are identified using a depth parameter (D_p) and with an empirically set lookahead parameter (L_p). A local maxima is labelled as a peak only if it has valleys deeper than D_p on either side, and is also the maxima at least within L_p values ahead. In the case of H_{avg}, D_p and L_p are set high (in our experiment, they correspond to $2.5 \cdot 10^{-5}$ and 20 respectively), which result in fewer, but reliable peaks. In addition, for each peak in H_{avg}, an upper and lower octave peak is added if there does not already exist a peak in a given proximity. We call these *extended reference peaks*. To compute the histogram of a given performance, D_p and L_p are set to lower values (in our experiment, they correspond to $2 \cdot 10^{-5}$ and 15 respectively), which result in more peaks which include several unwanted ones. However, only those peaks which have a corresponding match in *extended reference peaks* of H_{avg} (within a given proximity) are retained. This step not only helps to identify all the possible peaks for a given performance, but also compensates the choice of higher bin resolution, which otherwise generally results in some unwanted peaks.

In order to parametrize a given peak in the performance, it needs to be a bounded distribution. Generally, we observe that two adjacent peaks are at least 80 cents apart. The valley point between the peaks becomes a reasonable bound if the next peak is close by. But in cases where they are not, we have used a 50 cent bound to limit the distribution. The peak is then characterized by five parameters: peak location, mean, variance, skew and kurtosis.

3. RESULTS & DISCUSSION

One of the crucial factors that influence intonation of a given swara is raaga. This can be attributed to the factors like the role the swara plays in the raaga, the gamakas used with it, and the characteristics of neighbouring swaras. For a given swara, all of them change with the raaga [13]. This affects the distribution of pitches around it. Therefore, we choose to evaluate our approach by using intonation of swaras to characterize raagas. Hence, we focused on a data set which is representative enough of the variations allowed on swaras: 170 performances in 16 raagas each with at least 5 recordings per raaga are selected. These performances feature 35 vocal artists in total.

We evaluate our approach using three self-contained tasks. The first one is an explorative raaga disambiguation task in which the proposed parameters of the peak are shown to consistently increase the accuracy of the system. The second task is a qualitative study showing the use-

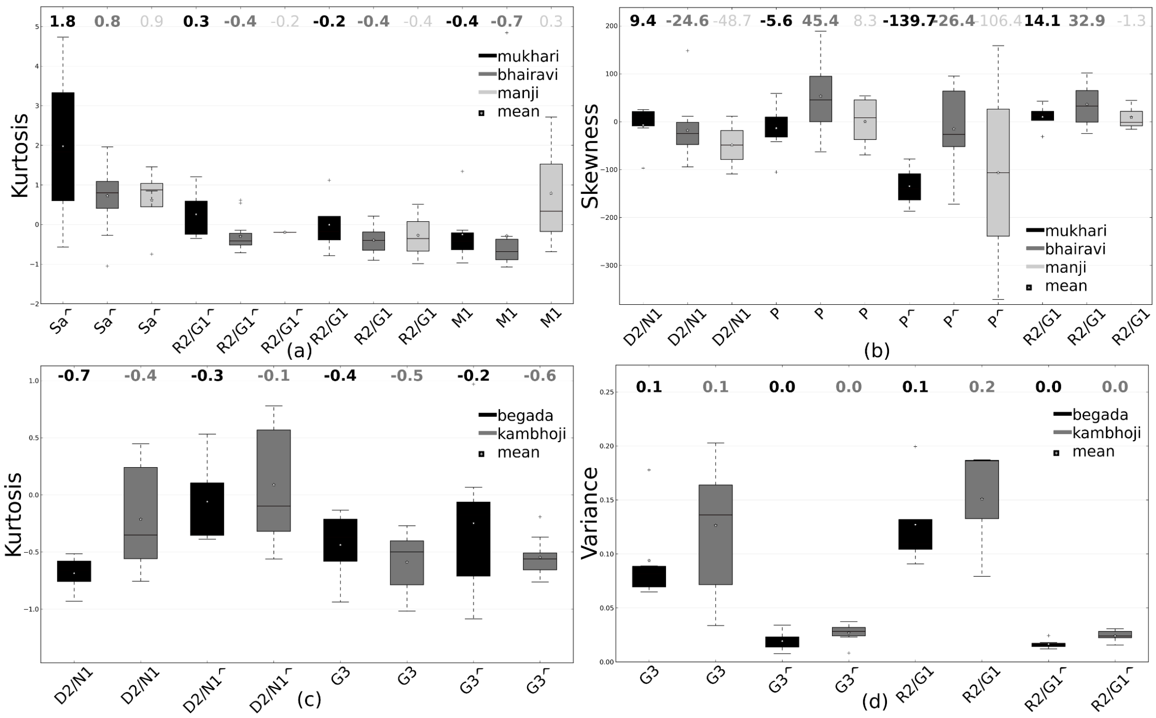

Figure 2. (a). Kurtosis values for swaras of Bhairavi, Mukhari and Manji raagas. (b). Pearson's first skewness coefficient values for swaras of Bhairavi, Mukhari and Manji raagas. (c). Kurtosis values for swaras of Begada and Kambhoji raagas. (d). Variance values for swaras of Begada and Kambhoji raagas. The x-axis corresponds to swara names in all the subplots.

fulness of peak parameters in discriminating raagas which share the same set of swaras. In the third task, the peak positions are used in deriving a general template for preferred mean values of swaras. Further, this general template is juxtaposed against different raagas showing notable deviations.

3.1 Raaga classification task

Previous raaga classification techniques employ only two parameters extracted by histogram analysis: peak position and height [6]. However, in using just these two parameters for classifying raagas which share the same set of swaras, there is a high chance of error. In order to assert the usefulness of our approach to describe intonation, we take 42 recordings in three raagas (Bhairavi, Thodi and Hindolam) which share five common swaras. Due to the limitation on the number of available recordings, choosing many swaras in our task will make it difficult to assess the complementarity of the new parameters, and could also potentially result in over-fitting (more features than instances). Therefore, we chose one swara to perform the raaga classification task.

The task is performed using two feature sets, both having four features. One set consists of just the position and height of the swara in the middle and upper octaves (common swara parametrization, hence used as a baseline). To ensure fairness, we have used two feature selection methods and different classifiers [5, 9] [4] over sub-

[4] The implementations provided in Weka were used with default parameters.

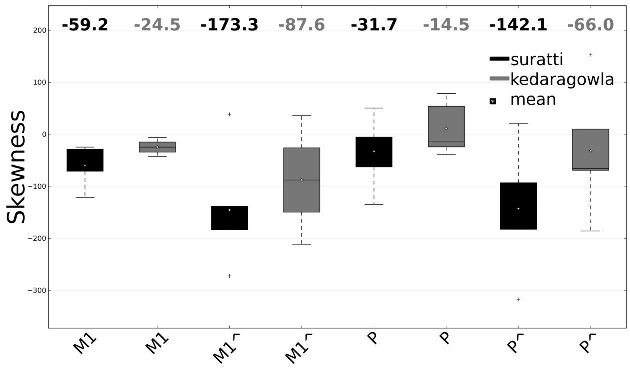

Figure 3. Skewness values for swaras of Surati and Kedaragowla raagas.

sampled data sets in a 10-fold cross validation test. The other feature set is obtained by using information gain-based and correlation-based feature selection methods [9] on a combination of position, height, skewness, kurtosis, variance and the mean of the distribution. Both feature selection methods select new parameters different from the position and the height.

Table 1 shows the averaged results obtained. Though the accuracy increments are not statistically significant, they are indicative of the worthiness of additional parameters. However, if we just consider the classification of two raagas with 26 recordings and perform the same test, the results show a significant increase in the overall accuracy (Table 2). These two explorative tasks show a consistent increase in the accuracy of the system, which indicates the usefulness and complementarity of the proposed peak characteristics.

3.2 Allied Raagas

In the experiment described in the previous section the raagas share a subset of the swaras. However, there is a class of raagas which share exactly the same set of swaras, but have different characteristics, called allied raagas. Since the swaras are common for the raagas, the discriminative capability of the peak position and/or mean will be considerably low. Therefore, these raagas constitute a good repertoire to test our approach. We consider three sets of allied raagas which together have 60 recordings in 7 raagas. The first set has three raagas, and the second and the third sets have two raagas each.

Fig. 2 (a) shows the kurtosis values for swaras of the first set of allied raagas (Bhairavi, Mukhari and Manji). In all the figures that follow, we have only shown the values for the most relevant swaras due to space constraints. $R_2/G_1^\wedge$ [5] and M_1 swaras can be seen to play a notable role in discriminating the raagas in this set. Fig. 2 (b) shows the skewness values for swaras of the same set of allied raagas. The distinction is observed even better through the skewness values of D_2/N_1, P and $P^\wedge$ swaras. Generally, in a given raaga, the melodic context of a swara is kept consistent across octaves, which means that the intonation characteristics of a swara across octaves should be consistent. The skewness values of P and $P^\wedge$ swaras assert this.

Figs. 2 (c) & (d) show the kurtosis and the variance values, respectively, for swaras of the second set of allied raagas (Begada and Kambhoji). All the swaras can be observed to play an equal role in distinguishing raagas of this set. The consistency of the variances of G_3, and R_2/G_1, and kurtosis of G_3 and D_2/N_1 across octaves is quite evident. Fig. 3 shows the skewness values for swaras of the third set of allied raagas (Surati and Kedaragowla). The observations from this set further reinforce the usefulness of the peak parametrization approach to describe intonation.

3.3 Analysis of peak positions

Table 3 shows the average of peak positions of each swara across all the available recordings (we now use the full data set), and the absolute sum of the differences of all observations from the corresponding equi-tempered and just-intonation intervals. The swaras which are observed in less than 20 recordings are not shown. There is a general tendency to just-intonation intervals compared to equi-tempered, which is in agreement with the results obtained by Serrà et al. [12]. However, this tendency is not very evident, supporting the claims in [7]. What interests us more here is the relevance of the values shown in the table for understanding the intonation of swaras in different raagas. For that we consider the set of the average values (Table 3) for each swara to be a general template. A similar template is obtained for each raaga, and the differences between the

[5] $\wedge$ denotes the swara in upper octave.

Swara	Mean	D_E	D_J	Recordings
Sa	1.92	6.43	6.43	142
$R2/G1$	200.93	10.93	11.62	68
$G3$	384.67	16.34	10.51	56
$M1$	495.05	9.94	9.56	123
P	700.18	6.01	6.21	164
$D2/N1$	889.58	13.99	11.06	87
$D3/N2$	987.87	16.0	14.11	56
$Sa^\wedge$	1196.62	6.35	6.35	174
$R2/G1^\wedge$	1401.37	8.86	8.97	96
$G3^\wedge$	1583.63	17.21	10.04	61
$M1^\wedge$	1693.05	12.71	12.17	91
$P^\wedge$	1897.86	7.45	8.08	118
$D2/N1^\wedge$	2095.6	8.78	13.15	54
$D3/N2^\wedge$	2192.15	14.17	13.33	28

Table 3. Mean of peak positions of each swara, and the differences from corresponding Equi-tempered (D_E) and Just-Intonation (D_J) intervals.

two templates are analysed and interpreted to check if they are musically meaningful.

Figs. 4 (a), (b) and (c) show the boxplots for positions of D_2, N_2 and M_1 respectively, in various raagas. They were examined by a trained musician who interpreted them, and asserted that they made sense in the context of today's practice of the raagas. For instance, it is said that D_2 in Khamas raaga is sung without any gamaka, whereas the same swara in Kalyani raaga is sung with a particular gamaka that might have been responsible for the observed phenomenon. This explains the observations made from Fig. 4 (d). Furthermore, as a sanity test, we have plotted the positions of the swara P in various raagas. This swara is normally expected to be sung without any gamaka. Hence, we expect the peak positions corresponding to P of all the raagas to be centered around the mean position observed in the general template (Table 3). Fig. 4 asserts this, except for minor deviations.

4. CONCLUSIONS

We have proposed a peak parametrization approach to describe intonation in Carnatic music and evaluated it qualitatively using three tasks. All the tasks discriminate raagas with the obtained information of swara intonation. However there are a few challenges in this approach. Few swaras, by the nature of the role they play, will not be manifested as peaks at all. Rather, they will appear as a slide that cannot be identified by a peak detection algorithm. Characterizing pitch distributions near all the theoretical intervals or from the general template shown in Sec. 3.3, irrespective of whether it is identified as a peak or not, is one possible way to address this issue. However, identifying few heuristics that will help in locating such slides can be a good substitute, since it falls in line with our methodology in not assuming any particular tuning. The future direction of this work is to extend it to the Hindustani music tradi-

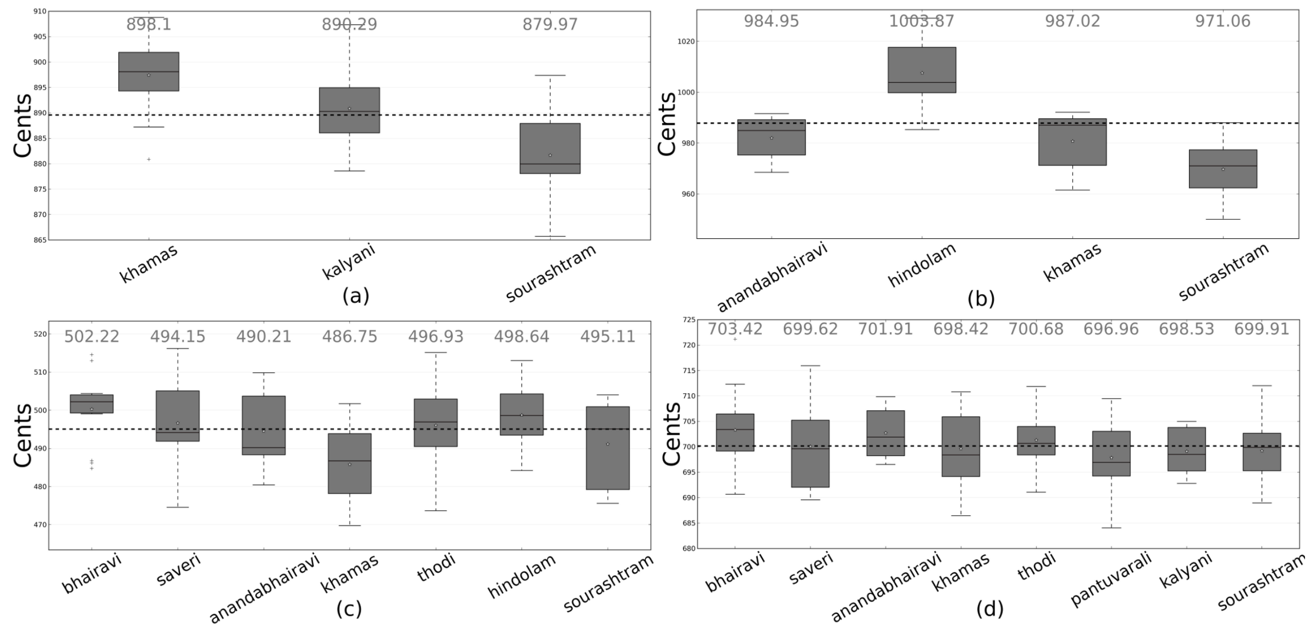

Figure 4. (a). Peak positions of D_2 for recordings in Khamas, Kalyani and Sourashtram raagas. (b). Peak positions of N_2 for recordings in Ananda Bhairavi, Hindoam, Khamas, and Sourashtram raagas. (c). Peak positions of M_1 for recordings in several raagas. (d). Peak positions of P for recordings in several raagas. The dashed line shows the mean of the corresponding swara obtained from the general template.

tion, and also to characterize performers and instruments by their preferred intonation.

5. 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, 2009-SGR-1434 from Generalitat de Catalunya, TIN2009-13692-C03-01 from the Spanish Government, and EU Feder Funds.

6. REFERENCES

[1] Shreyas Belle, Rushikesh Joshi, and Preeti Rao. Raga Identification by using Swara Intonation. *Journal of ITC Sangeet Research Academy*, 23, 2009.

[2] B. Bozkurt, O. Yarman, M. K. Karaosmanolu, and C. Akkoç. Weighing Diverse Theoretical Models on Turkish Maqam Music Against Pitch Measurements: A Comparison of Peaks Automatically Derived from Frequency Histograms with Proposed Scale Tones. *Journal of New Music Research*, 38(1):45–70, 2009.

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

[4] J. D. Deng, C. Simmermacher, and S. Cranefield. A Study on Feature Analysis for Musical Instrument Classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 38(2):429–438, 2008.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, 2009.

[6] G. K. Koduri, P. Rao, and S. Gulati. A Survey Of Raaga Recognition Techniques And Improvements To The State-Of-The-Art. In *Sound and Music Computing*, 2011.

[7] A. Krishnaswamy. On the twelve basic intervals in South Indian classical music. *Audio Engineering Society Convention*, page 5903, 2003.

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

[9] 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*, 36(5):51–52, 2011.

[10] Justin Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *Audio, Speech, and Language Processing, IEEE Transactions on*, (99):1–1, 2012.

[11] Justin Salamon, Sankalp Gulati, and Xavier Serra. A Multipitch Approach to Tonic Identification in Indian Classical Music. In *ISMIR*, page In press, 2012.

[12] J. Serrà, G. K. Koduri, M. Miron, and X. Serra. Assessing the tuning of sung indian classical music. In *ISMIR*, pages 157–162, 2011.

[13] V. Shankar. *The art and science of Carnatic music*. Music Academy Madras, Chennai, 1983.

[14] M. Subramanian. Carnatic Ragam Thodi Pitch Analysis of Notes and Gamakams. *Journal of the Sangeet Natak Akademi*, XLI(1):3–28, 2007.

[15] D. Swathi. *Analysis of Carnatic Music : A Signal Processing Perspective*. Masters thesis, IIT Madras, 2009.