To appear in: Proc. of Int. Soc. of Music Information Retrieval (ISMIR)
Suzhou (China), October 23-27, 2017

TOWARDS COMPUTATIONAL MODELING OF THE UNGRAMMATICAL IN A RAGA PERFORMANCE

Kaustuv Kanti Ganguli **Preeti Rao**
Department of Electrical Engineering,
Indian Institute of Technology Bombay, Mumbai, India.
{kaustuvkanti,prao}@ee.iitb.ac.in

ABSTRACT

Raga performance allows for considerable flexibility in interpretation of the raga grammar in order to incorporate elements of creativity via improvisation. It is therefore of much interest in pedagogy to understand what ungrammaticality might mean in the context of a given raga, and possibly develop means to detect this in an audio recording of the raga performance. One prominent notion is that ungrammaticality is considered to occur only when the performer "treads" on another, possibly allied, raga in a listener's perception. With this view, we consider modeling the technical boundary of a raga as that which separates it from another raga that is closest to it in its distinctive features. We wish to find computational models that can indicate ungrammaticality using a data-driven estimation of the model parameters; i.e. the raga performances of great artists are used to obtain representations that discriminate most between same and different raga performances. We choose a well-known pair of allied ragas (Deshkar and Bhupali in north Indian classical music) for an empirical study of computational representations for the distinctive attributes of tonal hierarchy and melodic shape of a chosen common descending phrase.

1. INTRODUCTION

The melodic framework in Indian art music is governed by the system of ragas. A raga can be viewed as falling somewhere between a scale and a tune in terms of its defining grammar which includes the tonal material, tonal hierarchy, and characteristic melodic phrases [26, 31]. Description of the raga grammar, as found in text resources or as verbalized in pedagogy, typically comprises of a listing of the allowed notes (svara) of the 12-tone scale, ascending and descending svara patterns, the mention of the most important svaras and a list of common phrases (svara sequences). While the texts do not explicitly describe the precise svara intonations or the actual melodic shapes of the phrases in terms of the transitions between svaras, the audio analyses of raga performances has demonstrated what is well-known to practitioners, i.e. the shape of the continuous pitch contour corresponding to the phrase is characteristic of the raga and therefore relatively invariant across performances in a given raga [28, 29].

The raga grammar can be viewed as a set of constraints within which creativity is given a free hand to realise whatever is aesthetically pleasing in the all-important melodic improvisation component, so characteristic of the genre. It is therefore of interest to understand, and possibly model, the technical boundary of a raga in terms of what might constitute ungrammaticalilty in a performance. The technical boundary would be specified in terms of the defining attributes such as tonal material and hierarchy, and phrase shapes. Such an exercise could lead to the development of computational tools for assessing performance accuracy in pedagogy together with the complementary aspect of creative skill. A popular notion of grammaticality in performance occurs around the notion of preserving a raga's essential distinctiveness in terms of the knowledgeable listener's perception [1–3, 5, 9, 24, 27, 30]. Thus, a performance with possibly many creative elements can still be considered not to transgress the raga grammar as long as it does not "tread on another raga" [24, 27, 35]. The technical boundary of a raga should therefore ideally be specified in terms of limits on the defining attributes where it is expected that the limit depends on the proximity of other ragas with respected to the selected attribute.

The computational modeling of the distinctive attributes of a raga has been the subject of previous research motivated by the task of raga recognition from audio given a large training dataset of performances across several ragas. The tonal material has been represented by a variety of first order pitch distributions and experimental outcomes based on recognition performance have been used to comment on the relative superiority of a given representation [4, 6–8, 11, 17, 22]. Motivated by the pitch-continuous nature of the melody, finely-binned histograms of octave-folded instantaneous pitch values have been used as templates in raga recognition tasks [8, 22]. Alternately, 12-bin distributions of pitch values within detected stable svara regions have been used to represent a raga's tonal content [7, 11]. Melodic shape invariance of phrases, on the other hand, has been used in the modeling of similarity measures for the task of melodic motif detection within and across performances of a given raga [12,19–21,28,29]. Pitch contour shape is typically represented by a tonic-normalized time-series or reduced to a symbol string of the raga notes [14,16].

In this work, we consider the computational modeling of tonal hierarchy and phrase shape based on maximizing the discrimination of 'close' ragas with respect to the given attribute. Such an approach has not been used in previous work. The notion of "allied ragas" is helpful here where we consider ragas that share the same grammar in major attributes while differing in a few. For example, the pentatonic ragas Deshkar and Bhupali have the same set of svaras (S, R, G, P, D corresponding to 0, 200, 400, 700, and 900 cents respectively) and common phrases in terms of svara sequences, e.g. the descending phrase GRS. Learners are typically introduced to the two ragas together and warned against confusing them [24, 30, 34]. Recently, subjective experiments on perceived similarity by musicians of synthetically manipulated raga phrases clearly demonstrated the existence of a sharp boundary between valid variants of a given raga phrase from variants of the same phrase (i.e. in terms of svara sequence) from an allied raga [15].

In the present work, we consider the pair of allied ragas, Deshkar and Bhupali, and use the performances of eminent Hindustani vocalists as proxy for creatively expressed, but grammatically accurate, examples of the stated raga. The performances in the allied raga are likewise considered to be examples of the corresponding ungrammatical renderings. We evaluate known, as well as some new, representations in terms of the achieved discrimination on a dataset of performances across the two ragas. Although the scope of the experiments is restricted to the given pair of ragas and chosen attributes, we expect our outcomes to be generalizable.

In the next section, we introduce our dataset along with the necessary musicological background, and describe the audio processing required to obtain the continuous pitch track that forms the basis for the computational representations under study. This is followed by sections that discuss potential representations for tonal hierarchy and phrase shape with associated distance measures. We next present an experimental study of the discrimination performance followed by our conclusions.

2. DATASET AND AUDIO PROCESSING

Table 1 presents a comparison of the melodic attributes corresponding to the grammars of the allied ragas as compiled from musicology texts. These cover the aspects of duration and intonation of the tonal material that includes ascending (*Ar*) and descending (*Av*) scales, dominant (*Vadi*) and subdominant (*Samvadi*), and characteristic phrases. 'Natural shruti' (last row) refers to the Just Intonation tuning, but there is no quantification of the term 'higher'. Also, there is some indication of a duration constraint on R (as a short or passing svara relative to its neighbours) in the form of braces (e.g. G(R)S in raga Deshkar).

2.1 Dataset and Annotation

The audio recordings used in this study are drawn from the Hindustani music corpus from 'Dunya' [1] compiled as a representative set of the vocal performances in the genre [33]. The editorial metadata for each audio recording is publicly available on the metadata repository MusicBrainz [2]. The Dunya corpus for raga Deshkar comprises 5 concerts of which 4 are selected for the current study, omitting the drut (fast tempo) concert due to the distinctly different style of realising phrases associated with such tempi [24]. Similarly, we selected 5 concerts for the Bhupali test set. We augmented the overall dataset, as described in Table 2, by additional concerts from personal collections.

Next, we annotate the occurrences of the chosen phrase, GRS. As discussed earlier, the GRS phrase is common to the two ragas and a frequently used descending motif. The GRS phrases are distributed across three octaves (upper, middle, and lower octaves), although lower octave instances are fewer. The segmentation of the phrases from the alap and vistar (i.e. improvised sections of the concert) is carried out semi-automatically as follows. A musician indicated the coarse location of each instance of the chosen phrase; this was then refined to obtain segmentation boundaries by automatic onset and offset detection methods described later. A count of the phrases used in this study is presented per concert in Table 2. The phrase-level annotation of the remaining concerts is underway for future work.

2.2 Pitch Time-series Extraction from Audio

Predominant-F0 detection is implemented by an algorithm proposed by [32] that exploits the spectral properties of the voice with temporal smoothness constraints on the pitch contour. The pitch is detected at 10 ms intervals with zero pitch assigned to the detected purely instrumental regions. The pitch values in Hz are converted to the cents scale by normalizing with respect the concert tonic determined automatically using a multi-pitch approach [18]. The final pre-processing step is to interpolate short silence regions

[1] https://dunya.compmusic.upf.edu/Hindustani
[2] https://musicbrainz.org/

Deshkar	Bhupali
Tonal material: SRGPD	Tonal material: SRGPD
Ar: SGPD, SPDS *Av*: S, PDGP, DPG(R)S	*Ar*: SRG, PDS *Av*: SDP, GDP, GRS
Vadi: D, *Samvadi*: G	*Vadi*: G, *Samvadi*: D
Phrases: SG, G(P)DPD, P(D)SP, DGP, DPG(R)S	Phrases: RDS, RPG, PDS, SDP, GDP, GRS
Higher shrutis of R, G, D	Natural shrutis of R, G, D

Table 1. Specification of raga grammar for the two allied ragas of the present study [1, 5, 25, 30]

Raga	# Concerts	Duration (hours)	# Artists	# GRS phrases
Deshkar	6	2:16:50	5	52
Bhupali	6	3:22:00	5	107

Table 2. Description of the test dataset.

below a threshold (250 ms which is empirically chosen as proposed by [13]) indicating musically irrelevant breath pauses or unvoiced consonants, by cubic spline interpolation, to ensure the integrity of the melodic shape. A median filtering with a 50 ms window is performed to get rid of spurious pitch excursions. Eventually, we obtain a continuous time-series of pitch values representing the melody line throughout the vocal regions of the concert.

3. MELODIC REPRESENTATIONS

Our goal is to propose computational representations that robustly capture particular melodic characteristics of the raga in a performance while being sensitive enough to the differences between allied ragas. Given that tonal material and hierarchy of svaras are an important component of the raga grammar, we consider representations of tonal hierarchy computable from the melody extracted from the audio recording of the performance. We also consider the representation of the melodic shape of a selected characteristic phrase.

3.1 Representation of Tonal Hierarchy

Tonal hierarchies, manifested in the relative frequencies and durations with which the tones are sounded in a musical piece, have been linked to key identification in Western art music. In a widely known work, Krumhansl [23] used a 12-element vector to code the total duration (in terms of number of beats) of each note of the chromatic scale in a piece and correlated it with each of 24 templates representing the major and minor keys to obtain accurate predictions of key. It is therefore logical to consider the same representation for raga discrimination. However, given the pitch-continuous nature of the music, we are faced with multiple competing options in the definition of a tonal representation. Closest to the tonal hierarchy vector of Krumhansl is the 12-element histogram of the total duration of each of the "stable notes" detected from the melodic contour. Considering the importance of the transitions connecting stable notes as well as microtonal differences in intonation between the same svara in different ragas, a histogram derived from all the pitch values in the melodic contour would seem more suitable. The bin width for such a pitch continuous distribution is also a design choice we must make. Finally, we need a distance measure computable between the histogram representations that correlates well with closeness of the compared performances in terms of raga identity.

3.1.1 Pitch Salience Histogram

The input to the system is tonic normalized pitch contour (cents vs time). The pitch values are octave-folded (0 - 1200 cents) and quantized to p bins of equal width (the bin resolution is $\frac{1200}{p}$). The bin centres are arithmetic mean of the adjacent bin edges. The salience of each bin is proportional to the accumulated duration of the pitch value corresponding to that bin. The normalization is to construct a probability distribution fuction (pdf) where the area under the histogram sums to unity. Given the number of bins, the histogram is computed as:

$$H_k = \sum_{n=1}^{N} \mathbb{1}_{[c_k \leq F(n) \leq c_{k+1}]} \quad (1)$$

where H_k is the salience of the k^{th} bin, $F(n)$ is the array of pitch values, (c_k, c_{k+1}) are the bounds of the k^{th} bin and $\mathbb{1}$ is an indicator random variable [3]. Figure 1 shows the pitch salience histogram for $p = 1200$ (1 cent bin resolution). For a bin resolution of 100 cents, the representation is equivalent to the conventional pitch class distribution (PCD) [7].

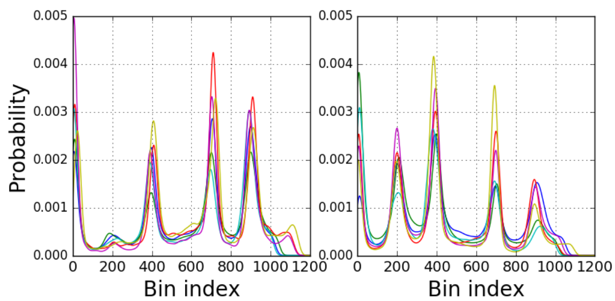

Figure 1. Pitch salience histograms (octave folded, 1 cent bin resolution) of 6 concerts each in ragas Deshkar (left) and Bhupali (right).

3.1.2 Svara Salience Histogram

The svara salience histogram is not equivalent to the PCD. The input to the system is segmented stable svaras which is a subset of the pitch contour. We use a previously proposed algorithm [16] that obtains a simple melodic transcription retaining only the stable svara regions of a pitch contour while discarding the transitory pitch regions. The stable svara regions are segmented by identifying the fragments of pitch contour that are within T_{tol} (35 cents) of the svara frequencies that are located via the peaks of a continuous pitch histogram. Next, the svara fragments that are smaller than T_{dur} (250 ms) in duration are filtered out, as they are too short to be considered as perceptually meaningful held svaras [28]. This leaves a string of fragments each labeled by a svara. Fragments with the same svara value that are separated by gaps less than 100 ms are merged. The svara salience histogram is obtained as:

[3] A random variable that has the value 1 or 0, according to whether a specified event occurs or not is called an indicator random variable for that event.

$$H_k = \sum_{n=1}^{N} \mathbb{1}_{[F(n) \in S_k, k \in (1,2,\cdots,12)]} \tag{2}$$

where H_k is the salience of the k^{th} bin, $F(n)$ is the array of pitch values, and S_k is the k^{th} svara of the octave. H_k is always a 12-element vector. Figure 2 shows the tonal hierarchy in the form of svara salience histogram. One major difference between pitch salience histogram and svara salience histogram is that the precise intonation information is lost in the latter.

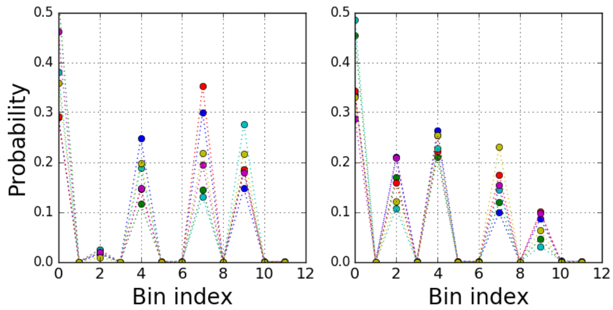

Figure 2. Svara salience histograms (octave folded) of 6 concerts each in ragas Deshkar (left) and Bhupali (right).

3.2 Representation of Phrase Shape

The phrase is a sequence of svara whose melodic realization includes specific intonations and transitions to/from neighboring svaras [24]. While computational models for measuring melodic similarity between phrases have employed distance measures between time-series of pitch values of the phrase segments, we might expect that a more discriminative representation is possible by explicitly incorporating features that contrast the two ragas.

Figure 3 shows a representative GRS phrase from each of the ragas. Distinctive features suggested by the comparison are: (i) durations of each of the stable svara regions, (ii) the durations of the glides connecting the svaras, and (iii) the pitch interval of the svara G. The implementation of these features would involve decisions on segmentation of stable svaras, and determining the pitch interval value from the pitch continuum in the region. Further, it is important to figure out the kind of normalization that is needed to reduce possible variability due to the tempo of the performance.

We describe a phrase as a sequence of melodic 'events' that can each be described by the chosen features. For the GRS phrase in question, we consider the following five events, i.e. svaras $\mathrm{G}, \mathrm{R}, \mathrm{S}$, and the $\mathrm{G}-\mathrm{R}$ and $\mathrm{R}-\mathrm{S}$ transitions. The selected features are: (i) $Start\_time$: onset of an event, (ii) End_time : offset of an event, (iii) $Duration$: difference of the two, (iv) $Intonation$: precise pitch interval location of a stable svara in the octave obtained as the median pitch value over the duration of the svara, and (v) $Slope$: gradient between the mean of last 20% and the first 20% pitch samples of a stable svara segment.

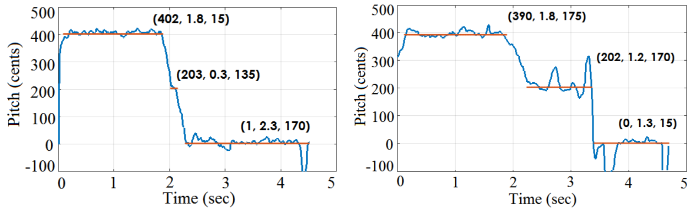

Figure 3. Two representative GRS phrases from ragas Deshkar (left) and Bhupali (right). The tuple corresponding to each svara denotes the extracted features ($Intonation$, $Duration$, $Slope$) for that event.

4. EXPERIMENTAL RESULTS AND DISCUSSION

We present experiments that help us identify the aspects of optimal representation, for each of tonal hierarchy and phrase shape, that discriminate the two ragas maximally based on our labeled dataset of 12 concerts across two ragas. Our common approach for both attributes is to use unsupervised clustering (k-means with k=2) of the feature vectors and optimise the separation between the clusters over the considered choices for feature implementation. Performance in unsupervised clustering can be measured via the 'cluster purity' which is obtained by assigning each obtained cluster to the underlying class that is most frequent in that cluster, and computing the resulting classification accuracy.

We also use the receiver operated characteristic (ROC) curve, with respect to detecting similarity within a raga pair given the feature vectors and distance measure, to evaluate the tuning parameters of tonal hierarchy. An objective function to compare ROCs across configurations is the area under curve (AUC) measure. Closer the AUC value to 1, better is the performance. For the evaluation of phrase shape, we additionally use 'feature selection' to estimate the most significant features.

4.1 Tonal Hierarchy

There are three main aspects of tonal hierarchy that need to be addressed in order to maximize the separation between the two clusters, these are: (i) representation, (ii) distance measure, and (iii) time-scale of analyses. The experiments apply to both pitch salience and svara salience histograms. We describe them in order with discussions on the insights obtained.

4.1.1 Optimal Representation

Our base representation is the octave folded pitch salience histogram normalized to be a pdf. We try different bin resolutions ranging from 1 to 100 cents, with a denser sampling between 20 and 40 cents, motivated from a previous study by Datta et al. [10]. We perform an unsupervised clustering (euclidean distance) to obtain labels for each element in the 2 classes. Figure 4 shows the cluster purity values, where we note that no degradation in the evaluation measure is observed for 1 through 30 cents bin resolution. Each value on the curve is obtained by an average of 5 runs of the clustering algorithm to nullify the effect of any local minima. For the case of svara salience histograms, the average cluster purity value is obtained as 0.96.

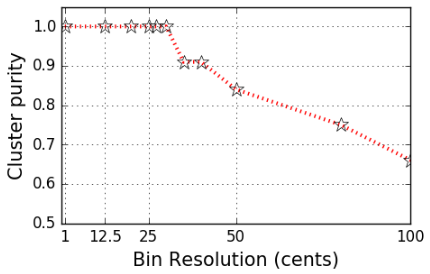

Figure 4. Cluster purity values obtained for different values of bin resolution for the pitch salience histograms.

4.1.2 Comparison of Distance Measures

In order to determine a suitable distance metric between the histograms representing raga tonal hierarchy, we test different metrics on the 25 cent binned pitch salience histogram (given that no degradation was observed at this resolution in the previous experiment) and on the svara salience histogram. As the distribution is a pdf, a natural distance measure suggested in the literature is the (symmetric) KL divergence [22]. We also try Cityblock (L-1 norm), Euclidean (L-2 norm), and Correlation distances. The last one is strongly motivated by the cognitive model of Krumhansl [23]. We present ROC curves (and AUC) for four distance measures in Figure 5. We find that the best performing distance measure is the KL divergence for svara salience histograms. The performance of cityblock distance is observed to be comparable to that of correlation distance for pitch salience histograms. We use the latter in the following experiments since it is more favored in previous work [6, 23].

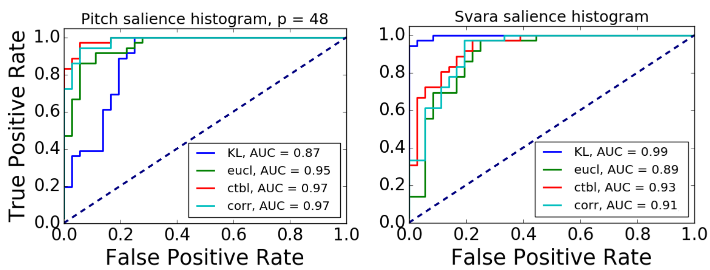

Figure 5. ROCs obtained for four different distance measures from pitch salience (left) and svara salience (right) histograms.

4.1.3 Time-scale for a Stable Tonal Hierarchy

It is of interest to determine at what time-scale, the measured tonal hierarchy qualifies as a stable representation. We therefore carry out the previous discrimination experiments on segmented concerts. We divide each concert to its $\left(\frac{1}{n}\right)^{th}$ ($n = 1, \cdots, 5$) portion and construct a distance matrix with each $\left(\frac{1}{n}\right)^{th}$ part. The goal is to find out the minimum proportion of the full concert that is necessary to robustly discriminate between the two classes. The ROC obtained from $\left(\frac{1}{4}\right)^{th}$ (and below this) of a concert results in an AUC < 0.5 which indicates that the time-scale is too small to constitute a stable tonal hierarchy. We augment our test dataset by considering each half and $\left(\frac{1}{3}\right)^{rd}$ of each concert, making the dataset size 24 and 36 respectively. Figure 6 shows a comparison of best performing systems for the full (n=1) and partial (n=2,3) concerts. We additionally investigate if a finer binned pitch salience histogram shows improvement. A finer bin resolution of 12.5 cents (p=96) is observed to perform better that the 25 cent binned pitch salience histogram (p=48). This suggests that the microtonal differences in intonation become more important when the concert segment duration is not long enough to capture 12-tone hierarchy in a stable manner. While the improvement (in terms of AUC value) might seem rather small, it was observed to be consistent with each distance measure under test. Moreover with full-concerts (n=1) for p=96, the maximum true positive rate for zero false positive rate is higher (≈ 0.95) than that for p=48 (≈ 0.75) indicating an improved performance. We also separately observed that considering only the beginning segments in the $\left(\frac{1}{n}\right)^{th}$ portions shows better performance compared to other locations. The first few minutes of each concert spans the alap and bandish (composition) that play a crucial role in raga delineation thus adhering closely to raga grammar.

4.2 Phrase Shape

We use the svara segmentation method outlined in Section 3.1.2 to obtain the components of the phrase shape corresponding to the sequence of stable svaras as well as the transient regions. In this section we present a statistical description of the features corresponding to the different events. We also compare the discrimination powers of the different features via a clustering experiment.

4.2.1 Distribution of Event Durations

Given the $Duration$ values of each event in the GRS phrases (52 instances in raga Deshkar and 107 instances in raga Bhupali), we present the distributions in Figure 7 of the event $Duration$s in the form of boxplots of the raw measurements in seconds. We observe distinctions between the two ragas in nearly all the duration parameters, most notably in the R $Duration$s. That the R duration is small and constrained in raga Deshkar is supported by the raga grammar specification in Table 1 which indicates R in parantheses, suggesting a "weak note that is never sustained" [30]. Overall, the dispersion in the parameters is smaller in the phrase in raga Deshkar compared with Bhupali, consistent with the fact that it is a grammatically more constrained raga [1, 5, 30]. An exception is the realisation of the S svara with several outlying values of duration due to its location at phrase end, where a number of contextual considerations influence the note offset.

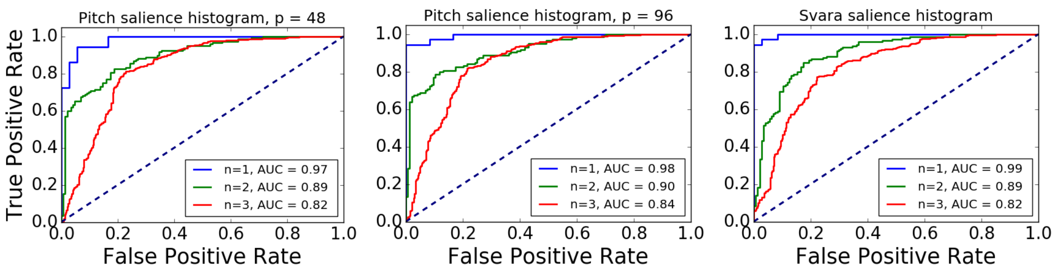

Figure 6. Comparison of ROCs obtained with correlation distance (for pitch salience histograms) and KL divergence (for svara salience histogram) for different time-scales (n=1,2,3) of the concerts.

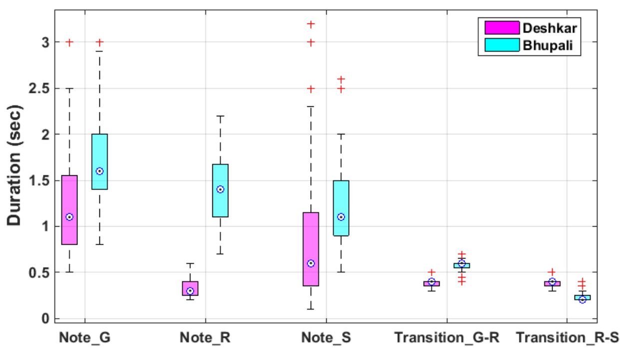

Figure 7. Distributions of event *Durations* across the GRS phrase instances in the two ragas.

4.2.2 *Feature Selection and Evaluation*

To compare the predictive powers of the measured acoustic features, we perform 'Feature Selection' using Weka [4] datamining tool. We use the "InfoGainAttributeEval" as the attribute selector that evaluates the value of an attribute by measuring the information gain with respect to the class, in conjunction with the "Ranker" search method that ranks attributes by their individual performances. We construct a feature vector for each instance of the GRS phrase with 5 *Duration* features, one for each event, and *Intonation* and *Slope* features corresponding to each of the three stable svaras as implemented in Section 3.2. Of these 11 features, we obtained the most significant features in terms of predictive power as the following: (i) R *Duration*, and (ii) G *Intonation*, with the third feature in the list placed considerably lower. This outcome is consistent with the raga grammars where these two aspects are considered distinctive properties of raga Deshkar.

Next, we perform an unsupervized clustering of the 159 phrases into two classes using the Euclidean distance between the 2-element vectors of the two selected features. The achieved cluster purity value is 0.99 (i.e. only 2 instances of the 159 are misclassified). As a next step, we investigate whether duration normalization is helpful. Given that overall phrase duration is correlated with tempo [34], it is natural to expect that the variability of phrase event durations may be reduced by normalization by the overall phrase duration. However, it turned out the the cluster purity with the duration-normalized *Duration* (of R svara) feature coupled with the previous *Intonation* (of G svara) feature reduced to 0.95 (i.e. 8 instances were misclassified). This indicates that musicians interpret the raga grammar in terms of raw durations rather than relative to the tempo.

[4] http://www.cs.waikato.ac.nz/ml/weka/

5. CONCLUSION AND FUTURE WORK

Based on the notion of grammaticality in raga performance, we examined computational representations for some of the key attributes of raga grammar based on discriminating allied ragas. In particular, both the pitch salience histograms and the stable-note based svara salience histograms were considered for tonal hierarchy with a variety of distance measures to derive a combination of histogram parameters and distance metric that best separated same-raga pairs from the allied-raga pairs. It was found that svara salience histograms worked best at the time-scale of full concerts whereas finer bin-widths of pitch salience histograms were superior for segmented concerts. Overall, full concerts with KL divergence as distance measure between 12-tone svara histograms performed best. A phrase level representation that considered only the discriminating elements of the same-phrase variants across the ragas in terms of absolute duration and pitch interval of key events (i.e. for R and G svaras respectively) was able to achieve a high degree of separation between the two allied ragas. Our results suggest that a pedagogy tool that measures ungrammaticality can indeed be designed based on modeling the raga attributes for any raga with the methodology presented here. Future work involves: (i) validation on other allied raga sets, (ii) correlating predicted ungrammaticality with perceived ungrammaticality by expert listeners, and (iii) determining the relative weighting of the different raga attributes for an overall rating, possibly at different time-scales, based on the expert judgments.

6. ACKNOWLEDGEMENT

This work received partial funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 267583 (CompMusic).

7. REFERENCES

[1] Music in Motion: The automated transcription for Indian music (AUTRIM) project by NCPA and UvA. url: https://autrimncpa.wordpress.com/. Last accessed: April 26, 2017.

[2] S. Bagchee. *Nād: Understanding Raga Music*. Business Publications Inc, 1998.

[3] S. Bagchee. *Shruti: A Listener's Guide to Hindustani Music*. Rupa Co, 2006.

[4] S. Belle, R. Joshi, and P. Rao. Raga identification by using swara intonation. *Journal of ITC Sangeet Research Academy*, 23, 2009.

[5] Distinguishing between Similar Ragas. url: http://www.itcsra.org/Distinguishing-between-Similar-Ragas. Last accessed: April 26, 2017.

[6] B. Bozkurt. An automatic pitch analysis method for Turkish Maqam music. *Journal of New Music Research (JNMR)*, 37(1):1–13, 2008.

[7] P. Chordia and A. Rae. Raag recognition using pitch-class and pitch-class dyad distributions. In *Proc. of Int. Soc. for Music Information Retrieval Conf.*, pages 431–436, 2007.

[8] P. Chordia and S. Şentürk. Joint recognition of raag and tonic in north Indian music. *Computer Music Journal*, 37(3):82–98, 2013.

[9] A. Danielou. *The ragas of Northern Indian music*. Munshiram Manoharlal Publishers, 2010.

[10] A. K. Datta, R. Sengupta, N. Dey, and D. Nag. *Experimental analysis of Shrutis from performances in Hindustani music*. Scientific Research Department, ITC Sangeet Research Academy, 2006.

[11] P. Dighe, H. Karnick, and B. Raj. Swara histogram based structural analysis and identification of Indian classical ragas. In *Proc. of Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 35–40, 2013.

[12] S. Dutta, S. PV Krishnaraj, and H. A. Murthy. Raga verification in Carnatic music using longest common segment set. In *Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 605–611, 2015.

[13] K. K. Ganguli, S. Gulati, X. Serra, and P. Rao. Data-driven exploration of melodic structures in Hindustani music. In *Proc. of the International Society for Music Information Retrieval (ISMIR)*, pages 605–611, August 2016. New York, USA.

[14] K. K. Ganguli, A. Lele, S. Pinjani, P. Rao, A. Srinivasamurthy, and S. Gulati. Melodic shape stylization for robust and efficient motif detection in hindustani vocal music. In *National Conference on Communications (NCC)*, 2017.

[15] K. K. Ganguli and P. Rao. Exploring melodic similarity in hindustani classical music through the synthetic manipulation of raga phrases. In *Cognitively-based Music Informatics Research*, 2016.

[16] K. K. Ganguli, A. Rastogi, V. Pandit, P. Kantan, and P. Rao. Efficient melodic query based audio search for Hindustani vocal compositions. In *Proc. of the International Society for Music Information Retrieval (ISMIR)*, pages 591–597, October 2015. Malaga, Spain.

[17] A. C. Gedik and B. Bozkurt. Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, 90(4):1049–1063, 2010.

[18] S. Gulati, A. Bellur, J. Salamon, H. G. Ranjani, V. Ishwar, H. A. Murthy, and X. Serra. Automatic tonic identification in Indian art music: Approaches and Evaluation. *Journal of New Music Research (JNMR)*, 43(1):53–71, 2014.

[19] S. Gulati, J. Serrà, V. Ishwar, S. Şentürk, and X. Serra. Phrase-based rāga recognition using vector space modeling. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 66–70, 2016.

[20] S. Gulati, J. Serrà, V. Ishwar, and X. Serra. Mining melodic patterns in large audio collections of Indian art music. In *Int. Conf. on Signal Image Technology & Internet Based Systems (SITIS-MIRA)*, pages 264–271, 2014.

[21] S. Gulati, J. Serrà, and X. Serra. Improving melodic similarity in Indian art music using culture-specific melodic characteristics. In *Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 680–686, 2015.

[22] G. K. Koduri, S. Gulati, P. Rao, and X. Serra. Rāga recognition based on pitch distribution methods. *Journal of New Music Research*, 41(4):337–350, 2012.

[23] C. L. Krumhansl. *Cognitive Foundations of Musical Pitch*, chapter 4: A key-finding algorithm based on tonal hierarchies, pages 77 – 110. Oxford University Press, New York, 1990.

[24] S. Kulkarni. *Shyamrao Gharana*, volume 1. Prism Books Pvt. Ltd., 2011.

[25] V. Oak. 22 shruti. url: http://22shruti.com/. Lat accessed: April 26, 2017.

[26] H. S. Powers and R. Widdess. *India, subcontinent of*, chapter III: Theory and practice of classical music. New Grove Dictionary of Music. Macmillan, London, 2nd edition, 2001. Contributions to S. Sadie (ed.).

[27] D. Raja. *The Raga-ness of Ragas: Ragas Beyond the Grammar*. D.K. Print World Ltd., 2016.

[28] P. Rao, J. C. Ross, and K. K. Ganguli. Distinguishing raga-specific intonation of phrases with audio analysis. *Ninaad*, 26-27(1):59–68, December 2013.

[29] P. Rao, J. C. Ross, K. K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, and H. A. Murthy. Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research (JNMR)*, 43(1):115–131, April 2014.

[30] S. Rao, J. Bor, W. van der Meer, and J. Harvey. *The Raga Guide: A Survey of 74 Hindustani Ragas*. Nimbus Records with Rotterdam Conservatory of Music, 1999.

[31] S. Rao and P. Rao. An overview of Hindustani music in the context of Computational Musicology. *Journal of New Music Research (JNMR)*, 43(1), April 2014.

[32] V. Rao and P. Rao. Vocal melody extraction in the presence of pitched accompaniment in polyphonic music. *IEEE Trans. on Audio, Speech & Language Processing*, 18(8), 2010.

[33] X. Serra. Creating research corpora for the computational study of music: the case of the Compmusic project. In *Proc. of the 53rd AES Int. Conf. on Semantic Audio*, London, 2014.

[34] W. van der Meer. *Hindustani Music in the 20th Century*. Martinus Nijhoff Publishers, 1980.

[35] K. G. Vijaykrishnan. *The Grammar of Carnatic Music*. De Gruyter Mouton, 2007.