

RESEARCH

On the Distributional Representation of Ragas: Experiments with Allied Raga Pairs

Kaustuv Kanti Ganguli and Preeti Rao

Raga grammar provides a theoretical framework that supports creativity and flexibility in improvisation while carefully maintaining the distinctiveness of each raga in the ears of a listener. A computational model for raga grammar can serve as a powerful tool to characterize grammaticality in performance. Like in other forms of tonal music, a distributional representation capturing tonal hierarchy has been found to be useful in characterizing a raga's distinctiveness in performance. In the continuous-pitch melodic tradition, several choices arise for the defining attributes of a histogram representation of pitches. These can be resolved by referring to one of the main functions of the representation, namely to embody the raga grammar and therefore the technical boundary of a raga in performance. Based on the analyses of a representative dataset of audio performances in allied ragas by eminent Hindustani vocalists, we propose a computational representation of distributional information, and further apply it to obtain insights about how this aspect of raga distinctiveness is manifested in practice over different time scales by very creative performers.

Keywords: Indian art music; Allied ragas; Tonal hierarchy; Grammaticality; Improvisation

1. Introduction

The melodic form 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 specifies the tonal material, tonal hierarchy, and characteristic melodic phrases (Powers and Widdess, 2001; Rao and Rao, 2014). The rules, which constitute prototypical stock knowledge also used in pedagogy, are said to contribute to the specific aesthetic personality of the raga. In the improvisational tradition, performance practice is marked by flexibility and creativity that coexist with the strict adherence to the rules of the chosen raga. The empirical analyses of raga performances by eminent artists can lead to insights on how these apparently divergent requirements are met in practice. Among the relatively few such studies, Widdess (2011) presents a detailed analysis of a sitar performance recording from the point of view of a listener to obtain an understanding of how raga characteristics are manifested in a well structured but partly improvised performance. In this work, we consider a computational approach towards using recordings of raga performance to investigate how the tonal hierarchy of a raga, as a prominent aspect of music theory, influences performance practice. A computational model would need to incorporate the

essential characteristics of the genre and be sufficiently descriptive to distinguish performances of different ragas.

In Western music, the psychological perception of “key” has been linked to distributional and structural cues present in the music (Huron, 2006). Krumhansl (1990) derived key profiles from tonality perceived by listeners in key-establishing musical contexts thereby demonstrating that listeners are sensitive to distributional information in music (Smith and Schmuckler, 2000). The distribution of pitch classes in terms of either duration or frequency of occurrence in scores of Western music compositions has been found to correspond with the tonal hierarchies in different keys (Smith and Schmuckler, 2000; Raman and Dowling, 2016; Temperley, 1999). Further, the task of automatic key detection from audio has been achieved by matching pitch chroma (the octave-independent relative strengths of the 12 pitch classes) computed from the audio with template key profiles (Fujishima, 1999; Gomez, 2006; Peeters, 2006). As in other tonal music systems, pitch distributions have been used in characterizing raga melodies. In contrast to the discrete 12-tone pitch intervals of Western music, raga music is marked by pitch varying continuously over a range as demonstrated via a recent raga music transcription visualization (AUTRIMNCPA, 2017). Recognizing the importance of distributional information, we find that the pitch-continuous nature of Indian classical traditions gives rise to several distinct possibilities in the choice of the computational parameters of a histogram representation, such as bin width. While

the raga grammar as specified in music theory texts is not precise enough to resolve the choices, suitable questions posed around observed performance practice can possibly facilitate a data-driven solution.

An important function of the computational model would be to capture the notion of grammaticality in performance. Such an exercise could eventually lead to computational tools for assessing raga performance accuracy in pedagogy together with the complementary aspect of creative skill. A popular notion of grammaticality in performance is preserving a raga's essential distinctiveness in terms of the knowledgeable listener's perception (Bagchee, 1998; Raja, 2016; Danielou, 2010). Thus, a performance with possibly many creative elements is considered not to transgress the raga grammar as long as it does not "tread on another raga" (Vijaykrishnan, 2007; Raja, 2016; Kulkarni, 2011). 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 respect to the selected attribute. We therefore consider deriving a computational representation of distributional information based on maximizing the discrimination of "close" ragas.

The notion of "allied ragas" is helpful here. These are ragas with identical scales but differing in attributes such as the tonal hierarchy and characteristic phrases (Mahajan, 2010), as a result of which they may be associated with different aesthetics. For example, the pentatonic ragas Deshkar and Bhupali have the same set of notes (scale degrees or *svaras*): S;R;G;P;D, corresponding to 0, 200, 400, 700, and 900 cents respectively from the tonic (see solfege in **Figure 1**). Learners are typically introduced to the two ragas together and warned against confusing them (Kulkarni, 2011; Rao et al., 1999; van der Meer, 1980). Thus, a deviation from grammaticality can be induced by any attribute of a raga performance appearing suggestive of an allied, or more generally, different raga (Mahajan, 2010; Raja, 2016). The overall perception of the allied raga itself is conveyed, of course, only when all attributes are rendered according to its own grammar.

A goal of the present work is to develop a distributional representation for raga music that links the mandatory raga grammar with observed performance practice. The parameters of the computational model are derived by optimizing the classification performance when comparing pairs of performances within a set of allied ragas (Ganguli and Rao, 2017). We do not consider structural information in the form of phrase characteristics in this work. We use

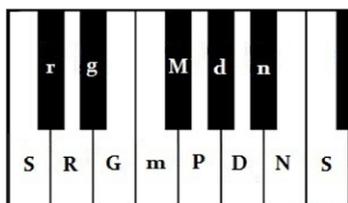


Figure 1: The solfege of Hindustani music shown with an arbitrarily chosen tonic (S) location.

audio recordings of performances of eminent Hindustani vocalists as very creatively rendered, but grammatically accurate, references of the stated raga. The obtained model will further be applied to obtain insights on consistent practices, if any, observed in raga performances apart from what is specified by the grammar.

In the next section, we introduce raga grammar as typically provided in text resources and discuss aspects that a distributional representation should capture. We provide a critical review of tonal hierarchy and pitch distribution representations proposed for different musics including those applied in MIR tasks. Our datasets and audio processing methods are presented next. Performances drawn from selected allied raga pairs comprise the test dataset. Experiments designed to compare the effectiveness of different distributional representations and associated distance measures are presented followed by discussion of the insights obtained in terms of the practical realization of musicological concepts in raga performance.

2. Background and Motivation

We present a brief introduction to raga grammar as commonly found in music theory texts and attempt to relate this with the theory of tonal hierarchy in music. We also review literature on pitch distributions in the MIR context with a focus on raga music in order to motivate our own approach in this work.

2.1. Raga Grammar and Performance

In the oral tradition of raga music, music training is imparted mainly through demonstration. Explicit accounts of the music theory however are available in text resources where the typical presentation of the grammar of a raga appears as shown in any one of the columns of **Table 1**. We observe both distributional and structural attributes in the description. The tonal material specifies the scale while the *vadi* and *samvadi* refer to the tonal hierarchy, indicating the two most dominant or salient *svaras* in order. The remaining notes of the scale are termed *anuvadi* (allowed) *svara*; the omitted notes are termed *vivadi* or *varjit* *svara*. The typical note sequences or melodic motifs are captured in the comma separated phrases in the *aroha-avaroha* (ascent-descent) and the Characteristic Phrases row of **Table 1**. Finally, the *shruti* indicates the intonation of the specified scale degree in the given raga in terms of whether it is higher or lower than its nominal place (i.e. in Just Intonation). The only other inscription we see is the parentheses, e.g. (R) that indicate that the note R is "weak" (*alpatva*) in Deshkar (Rao et al., 1999).

Other descriptors that are often available are the type of emotion evoked by the raga, tempo or pace at which it is performed, and the pitch register of major activity. Finally, most texts discussing a raga's features explicitly mention the corresponding allied raga, if any, which has the same scale but supposedly different "treatment" of notes (Rao et al., 1999). In music training too, the allied raga pairs are taught concurrently so that learners can

Table 1: Specification of raga grammar for three allied raga-pairs (ITC-SRA, 2017; Rao et al., 1999; AUTRIMNCPA, 2017; Oak, 2017; Kulkarni, 2011; SwarGanga, 2017; Mahajan, 2010). Gaps in *Aroha-Avaroha* indicate temporal breaks (held svara or pause), under/overline are lower/higher octave markers.

Feature	Deshkar	Bhupali	Puriya	Marwa	Multani	Todi
Tonal material	SRGPD	SRGPD	SrGMDN	SrGMDN	SrgMPdN	SrgMPdN
<i>Aroha</i> <i>Avaroha</i>	SGPD \overline{S} PD \overline{S} \overline{S} PDGP DPG(R)S	SRG PD \overline{S} \overline{S} DP GDP GRS	\underline{N} (r)G MN(D) \overline{S} NM DG MG(r)S	\underline{D} Nr GMD N \overline{r} \overline{S} \overline{r} ND DMGr \underline{D} rS	N \underline{S} gMP P \overline{N} \overline{S} SN(d)P gMg (r) \underline{N} S	SrgMP Md \overline{N} \overline{S} SNdP MrgrS
<i>Vadi</i> , <i>Samvadi</i>	D, G	G, D	G, N	r, D	S, P	d, g
Char. phrases	SGPD, P(D) \overline{S} P DGP, DPG(R)S	RDS, RPG, PD \overline{S} \overline{S} DP, GDP, GRS	\underline{N} (r)S, \underline{N} (r)G, MDG DNM, DG, MG(r)S	\underline{D} Nr, rGMD, D \overline{r} \overline{S} \overline{r} ND, DMGr, ND \overline{r} S	r \underline{N} S, gMP, P \overline{N} \overline{S} NdP, MgM, g(r)S	Srg, PMd, d \overline{N} \overline{S} NdP, Md, rgrS
Shruti	Higher R, G, D	Natural R, G, D	Lower r	Higher r, D	Higher N	Lower r, g

infer the technical boundaries of each of the ragas more easily.

Table 1 presents a comparison of the melodic attributes corresponding to the grammars of the allied ragas as compiled from musicology texts (ITC-SRA, 2017; Rao et al., 1999; AUTRIM-NCPA, 2017; Oak, 2017; Kulkarni, 2011; SwarGanga, 2017; Mahajan, 2010). It may be noted that the presented information is common across the cited sources. To represent raga diversity, the allied raga pairs are drawn from each of 3 types of scales (pentatonic, hexatonic, and heptatonic). Apart from Deshkar-Bhupali, a well-known allied raga-pair is Puriya-Marwa. In the description of raga Puriya, Rao et al. (1999) mention that if either r or D were emphasized in Puriya, it would immediately create an impression of Marwa; therefore strict care should be taken to highlight only G and N. The complementary warning finds its place in the description of raga Marwa. The third allied raga-pair that we include for the study is Multani-Todi where r and d are weak in the former and omitted from ascending phrases.

In performance, a raga is typically presented in the form of a known composition (*bandish*) set in any appropriate metrical cycle (*tala*) within which framework a significant extent of elaboration or improvisation takes place (Widdess, 1994; Rao et al., 1999). The performer is free to choose the tonic that is then sounded by the drone throughout the performance. A concert begins with slow elaboration of the chosen raga's characteristics in an unmetered section called the *alap*. This is followed by the *vistar* which comprises the chosen composition and elements of improvisation. The progress of the concert is marked by the gradual increase of tempo with the duration of the rhythmic cycles decreasing and slightly more fast paced and ornamented rendering of the melodic phrases (van der Meer, 1980; Widdess, 2013). While the raga grammar specifies a characteristic melodic phrase by a sequence of svara, as in **Table 1**, the pattern serves only as a mnemonic to the trained musician to recall the associated continuous melodic shape or motif. Overall, we note that the raga grammar specification, while skeletal, comprises clear distributional (tonal hierarchy) and structural (melodic phrases) attributes. In this work, we

focus on a computational representation of the former in a manner suited to the empirical analysis of raga performances.

2.2. The Theory of Tonal Hierarchy in Music

Tonal hierarchy, as discussed by Krumhansl and Cuddy (2010) refers to both a fundamental theoretical concept in describing musical structure and a well-studied empirical phenomenon. As a theoretical concept, the essential idea is that a musical context establishes a hierarchy of tones. Krumhansl (1990) proposed a key-finding algorithm that is based on a set of "key-profiles" first proposed by Krumhansl and Kessler (1982), representing the stability or compatibility of each pitch-class relative to each key. The key-profiles are based on experiments in which participants were played a key-establishing musical context such as a cadence or scale, followed by a probe-tone, and were asked to judge how well the probe-tone "fits" given the context (on a scale of 1 to 7, with higher ratings representing better fitness). Given these key-profiles, the algorithm judges the key of a music piece by generating a 12-element input vector corresponding to the total duration of each pitch-class in the piece. The correlation is then calculated between each key-profile vector and the input vector; the key whose profile yields the highest correlation value is the preferred key. In other words, the listener's sense of the fit between a pitch-class and a key is assumed to be highly correlated with the frequency and duration of that pitch-class in pieces in that key (Krumhansl and Cuddy, 2010).

Other studies also found that the tonal hierarchy measured in empirical research mirrors the emphasis given to the tones in compositions by way of frequency of occurrence and duration (Smith and Schmuckler, 2004; Raman and Dowling, 2016; Temperley and Marvin, 2008). This relationship between subjective and objective properties of music provides a strong musical foundation for the psychological construct of the tonal hierarchy. Castellano et al. (1984) used contexts from 10 North Indian ragas. As reviewed earlier, raga music theory describes a hierarchy of the tones in terms of importance. In a probe tone study, ratings of the 12 tones of the Indian

scale largely confirmed these predictions suggesting that distributions of tones can convey the tonal hierarchy to listeners unfamiliar with the style. Both American and Indian groups of listeners gave the highest ratings to the tonic and the fifth degree of the scale. These tones are considered by Indian music theorists to be structurally significant, as they are immovable tones around which the scale system is constructed, and they are sounded continuously in the drone. Relatively high ratings were also given to the *vadi* svara, which is designated for each raga and is considered the dominant note. The ratings of both groups of listeners generally reflected the pattern of tone durations in the musical contexts. This result suggests that the distribution of tones in music is a psychologically effective means of conveying the tonal hierarchy to listeners whether or not they are familiar with the musical tradition. Beyond this, only the Indian listeners were sensitive to the scales (lit. *thaat*) underlying the ragas. For Indian listeners, multidimensional scaling of the correlations between the rating profiles recovered the theoretical representation of scale (in terms of the tonal material) described by theorists of Indian music. Thus, the empirically measured tonal hierarchy induced by the raga contexts generates structure at the level of the underlying scales, but its internalization apparently requires more extensive experience with music based on that scale system than that provided by the experimental context (Castellano et al., 1984).

There has been no notable work on analyzing tonal hierarchy from symbolic notation in raga music. This could be due to the availability of scores, if at all, only for very well-known compositions. Performances are dominated by improvisation and rarely transcribed. A recent work (Ross et al., 2017) analyzed the scores of 3000 compositions across 144 ragas to find that the pitch class distribution served well to cluster ragas with similar scales together.

2.3. First-order Pitch Distributions in Audio MIR

The methods mentioned above for key estimation worked with the symbolic representation of music. A large number of the approaches for key estimation in audio recordings of Western music are essentially motivated by the same methodology. However, the task of key estimation from audio recordings becomes much more challenging due to the difficulty of extraction of a reliable melody representation, such as a music score, from polyphonic music recordings (Gomez, 2006). Based on pitch chroma features computed from the audio signal spectrum, a 12-element pitch-class profile (PCP) vector is estimated (Gomez, 2006; Peeters, 2006). Next the correlation of this estimated tonal profile is implemented with all possible theoretical key-profiles derived in (Krumhansl and Kessler, 1982). The key-profile that results in the maximum correlation is marked as the key of the music piece.

Another music tradition where the concept of tonality has been studied with reference to pitch distributions is Turkish Makam music. Due to the presence of the micro-tonalities and continuous melodic movements between the notes, a fine grained pitch distribution is considered as a feature for modeling tonality Bozkurt (2008); Gedik and Bozkurt (2010) use a pitch distribution with bin-width of $\frac{1}{3}$ Holdrian comma (approximately 7.5 cents). This results in a 159-dimensional pitch-class distribution (PCD) vector that performs significantly better in a Makam recognition task compared to a 12, 24, or 36-dimensional PCP vector often used for tonal analysis of Western music.

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 (Chordia and Senturk, 2013; Koduri et al., 2012; Belle et al., 2009; Chordia and Rae, 2007; Dighe et al., 2013). The tonal material has been represented by a variety of first order pitch distributions as depicted in **Figure 2**. Experimental outcomes based on recognition

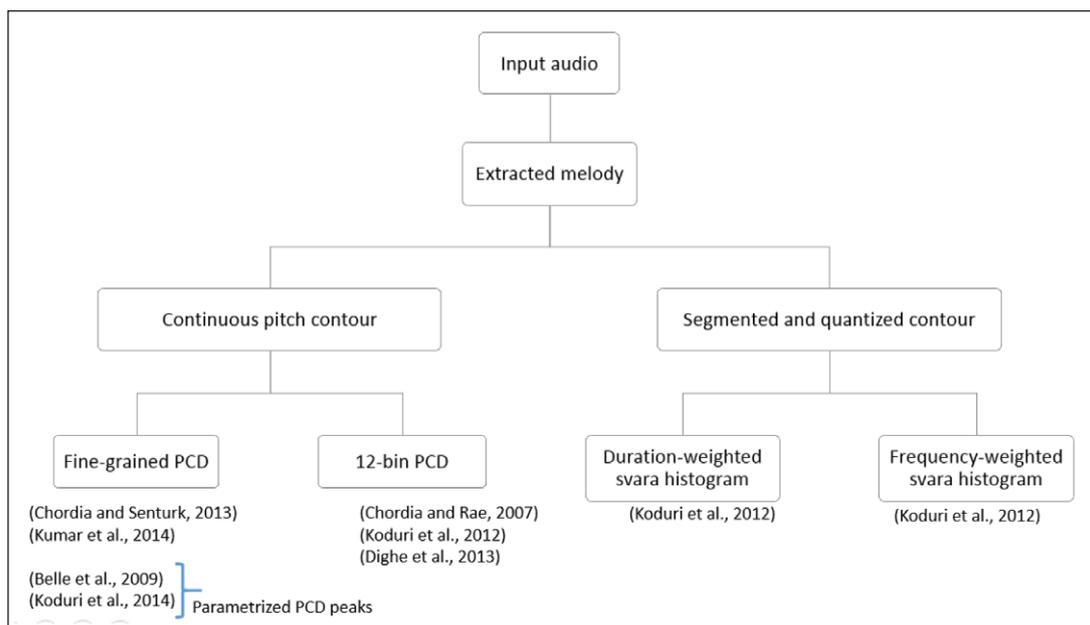


Figure 2: Taxonomy of the previous endeavors in raga recognition from first order pitch distributions.

performance have been used to comment on the relative superiority of a given representation as a feature vector in either a template-based or trained model-based classification context. Motivated by the pitch-continuous nature of the melody, histograms of different bin widths computed from octave-folded instantaneous pitch values have been used as templates in raga recognition tasks. The taxonomy, presented in **Figure 2**, summarizes the distinct first-order pitch distribution representations proposed in the raga recognition literature. The top-level classification is based on whether the continuous melodic contour is used as such, or segmented and quantized prior to histogram computation.

As we have seen in Section 2.1, along with the set of svaras, raga grammar defines the functional roles of these svaras in terms of their saliences. A formal definition of the salience of a svara in a melody does not exist, and therefore, several methods have been proposed to quantify it. Chordia and Rae (2007) and Dighe et al. (2013) represent svara saliences using a 12-bin PCD computed as a histogram of the tonic-normalized and octave-folded pitch sequence. The pitch is detected at uniform intervals (audio frames) across the recording to obtain a time series representing the melodic contour. The salience of a bin in the histogram is therefore related to the total duration of the (octave-folded) pitch in the melody. This global feature is robust to pitch octave errors and is shown to perform well on a sizable dataset. A simple extension to the 12-bin PCD feature mentioned above is to compute the pitch distribution using fine grained bins, e.g. at 1 cent resolution. Such a fine grained PCD is used widely (Chordia and Senturk, 2013; Koduri et al., 2012; Belle et al., 2009; Kumar et al., 2014). These studies report a superior performance in recognizing ragas by using the high resolution PCD as compared to a 12-bin PCD. Belle et al. (2009) and Koduri et al. (2014) proposed a parametrized version of the PCD, wherein the parametrization is performed for the pitch distribution shape across each svara region. Both works exploit the distinctions between ragas in the intonation of shared svaras via peak position, amplitude and shape in the high-resolution PCD.

The above reviewed distributions used the continuous pitch contour (time-series) extracted from the audio. An alternate approach is that of Koduri et al. (2011) who computed the distribution from only the stable-pitch regions of the contour. Based on certain heuristic considerations, such regions were segmented out and used to construct the PCD corresponding to the 12 svaras. Two variants of svara salience estimation were implemented. One of their proposed approaches treats the total duration of a svara as its salience similar to the previously mentioned approaches with the continuous pitch contour (e.g. Chordia and Rae, 2007). The other approach considers the frequency of occurrence (i.e. the count of instances) of the svara (irrespective of the duration of any specific instance) as its salience. Thus, three types of first order pitch-class distributions were tested: (i) *P_continuous*: unconstrained pitch-class histogram with a choice of bin resolutions (1, 2, ..., 100 cents), (ii) *P_duration*: constrained to stable notes only, weighted by the durations thereof

(12-bin), and (iii) *P_instance*: constrained to stable notes only, weighted by the count of instances (12-bin). The only distance measure tested was the symmetric KL divergence. Overall, *P_duration* performed the best. For *P_continuous*, there was no noticeable difference across different choices of bin width below 50 cents.

2.4. Motivation

As shown in this review, several methods based on pitch distribution have been applied to the raga recognition task. Although the outcomes are expected to depend on the design of the dataset, this aspect has received hardly any careful consideration. The test datasets used in previous work typically comprised a number of performance audio recordings arbitrarily chosen for an equally arbitrarily chosen set of ragas. In the face of this diversity of datasets, it is difficult to justify the conclusions or predict how the results generalize to other datasets of performances and ragas. The ragas in the test sets often correspond to different scales; given this distinction in the set of notes, the precise implementation of the first-order distribution is probably not relevant. We propose to develop and tune the parameters of a computational representation for the distribution using a dataset and evaluation methods that are sensitive to changes in the parameters within the reasonable space of parameter choices. This is achieved with the use of allied ragas and a more musicologically meaningful criterion related to the technical boundary of the raga in the distributional feature space.

There has also been a lack of attention, in the literature, to the choice of probability distribution distance measures. This, of course, partly owes itself to the emphasis on classifier-based approaches with input melodic features such as first-order distributions. Here the focus has been on gross performance of the raga recognition system in terms of classification accuracy rather than on obtaining insights into the computational equivalents of musicological concepts.

Koduri et al. (2011, 2012, 2014) proposed a handful of representations (in terms of first-order pitch distributions) but always applied the KL divergence as a distance measure. In contrast Gedik and Bozkurt (2010) used a bunch of distance measures but did not report their sensitivity to the different possible pitch distribution representations. While Datta et al. (2006) have commented on many possible configurations of the bin centres/edges and their precise locations, we do not expect the precise locations of bin centres/edges to affect the performance of a high bin-resolution representation in the raga recognition task.

In the present work, we systematically investigate the choice of bin width and distance measure for continuous pitch distributions computed from performance recordings. We also consider discrete-pitch representations derived from svara-segmented note regions. Given that melody transcription for raga music itself is a challenging (or rather, ill-defined) task (Widdess, 1994), it is necessary to rely on heuristic methods for svara segmentation and transcription as presented in the next section.

3. Dataset and Audio Processing

The music collection used in this study was compiled as a part of the CompMusic project (Serra, 2011). The audio recordings used in this study are drawn from the Hindustani music corpus from “Dunya”¹ as a representative set of the vocal performances in the *khayal* genre (Serra, 2014). The editorial metadata for each audio recording is publicly available on the metadata repository MusicBrainz.² We augmented the dataset derived from the Dunya corpus with concerts from personal collections to obtain the overall test dataset presented in **Table 2**. Additionally, the metadata associated with each concert in our dataset is presented in the Appendix (Table 1). The

Table 2: Description of the dataset in terms of number of artists, concerts, and their durations.

Raga	# Concerts	Duration (hours)	# Artists
Deshkar	6	2:16:50	5
Bhupali	11	5:12:23	9
Puriya	7	2:30:26	7
Marwa	12	4:19:12	12
Multani	7	2:30:57	7
Todi	12	4:39:55	12
# Total	55	21:29:43	52 (26 unique)

artists are selected as the stalwarts of Hindustani vocal music spanning the past 7 or 8 decades.

The selected music material in our collection is diverse in terms of the number of artists (26) and recordings. The chosen allied raga-pairs belong to the set of 7 most popularly performed allied pairs (Mahajan, 2010; Raja, 2016). Our collection includes a total of 55 concerts from the 3 widely performed allied raga-pairs that are diverse both in terms of the number of svaras and their pitch-classes. Note that our dataset comprises diversity in terms of tonal material (pentatonic, hexatonic, and heptatonic scales). All of the allied raga pairs happen to share the phenomenon of a duration constraint on the second scale degree (r/R svara) occurring in one of the ragas of the pair. All the concerts belong to either *vilambit* (slow) or *madhya* (medium) *laya* (tempo). Some concerts consist of a *drut bandish* towards the end for a short duration. Standalone *drut* (fast tempo) concerts are not included due to their typically short durations (less than 12 min) severely limiting both the composition and improvisation sections available for analyses. The accompanying instruments are *tanpura* (drone), *tabla* (percussion), and harmonium or *sarangi* as melodic accompaniment. The pitch range of any performance spans approximately two octaves (octave of the tonic and about half of each of the lower and upper octaves). All of the concerts comprise elaborations based on a selected *bandish*. The number of concerts is unequally distributed across ragas due the greater availability of concerts in the slow and medium tempo ranges in some ragas. We present next the audio processing methods applied to the concert audio to

obtain the two categories of distributions, viz. continuous pitch based and segmented svara based, as depicted in **Figure 3**.

3.1. Pitch Time-series Extraction from Audio

Predominant-F0 detection is implemented by an algorithm from Rao and Rao (2010) that uses harmonicity-based grouping of sinusoidal partials detected in the polyphonic audio to estimate one or more F0 candidates in each frame. Next, the spectral and temporal properties of the singing voice are exploited to discriminate its partials from those of the melodic accompaniment. Further temporal smoothness constraints minimize pitch octave errors. The melodic pitch is detected at 10 ms intervals throughout the audio recording with no F0 values assigned in the detected purely instrumental regions. The algorithm also returns the intensity (energy of the vocal harmonics) at each time step. The predominant-F0 detection algorithm is based on a small set of tunable parameters, which are selected from preset values, for pitch tracking with high accuracy using available knowledge such as singer gender and whether the expected pitch variation is rapid or slow.

As our dataset comprises performances by a number of artists, male and female, the detected vocal melody must be normalized with respect to the tonic pitch. The fundamental frequency (F0) values in Hz are converted to the cents scale by normalizing with respect to the concert tonic determined using a classifier based multi-pitch approach to tonic detection (Gulati et al., 2014). With an accuracy of over 90%, any gross errors are easily corrected based on raga (or rather, allied raga group) information.

The final preprocessing step is to interpolate short silence regions below a threshold (250 ms as proposed by Ganguli et al. (2016)) indicating musically irrelevant breath pauses or unvoiced consonants, by cubic spline interpolation, to ensure the integrity of the melodic contour. Median filtering with a 50 ms window is performed to get rid of irrelevant local pitch fluctuations. Eventually, we obtain a continuous time series of pitch values representing the melody line throughout the vocal regions of the concert.

3.2. Svara Segmentation and Transcription

The stylization of the continuous pitch contour has been of interest in both music and speech. In Western music, piece-wise flat segments are used to model the melody line corresponding to the note values and durations in the underlying score. Speech signals, on the other hand, have smoothly varying pitch which can be stylized, for example, with polynomial fitting (Ghosh and Narayanan, 2009). In Indian art music, we have something in between, with smoothly varying melodic contours but peaky overall pitch distributions coinciding with the discrete svara intervals. The pitch contour of a melodic phrase can thus be viewed as a concatenation of events of two categories: (i) a pseudo-steady segment closely aligned with a raga svara, and (ii) a transitory

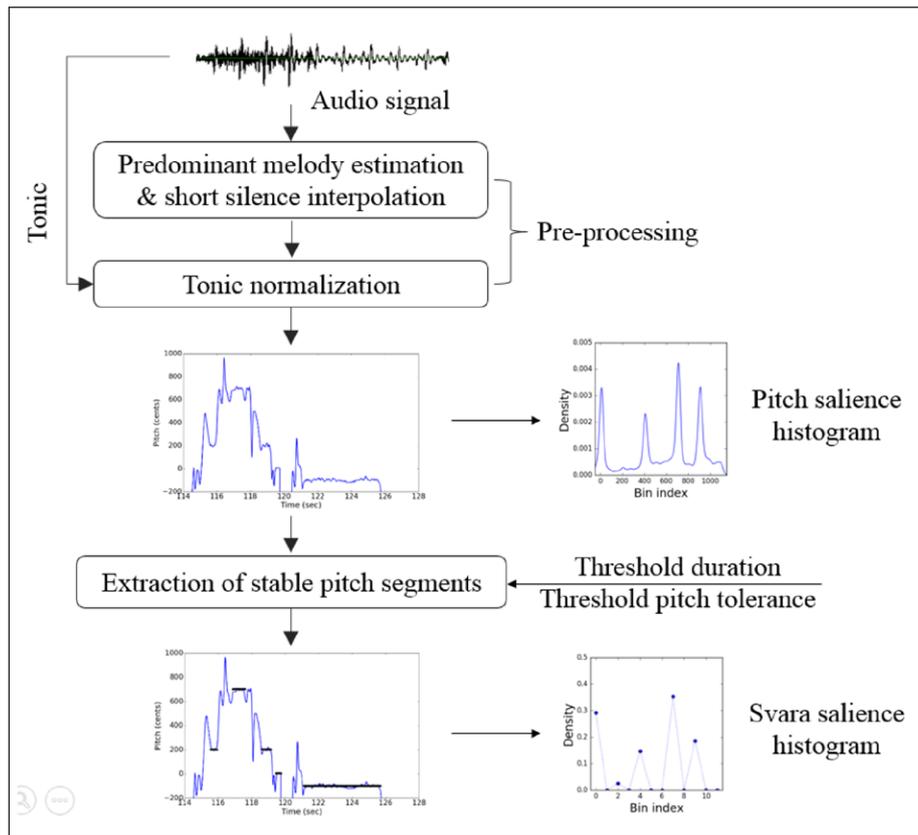


Figure 3: Block diagram of the signal processing chain from audio signal to pitch distributions.

segment which connects two such consecutive steady segments. The latter is often referred to as an *alankar* or an ornament comprising figures such as *meend* (glide), *andolan* (oscillation), *kan* (touch note), etc. A stylization corresponding to a sequence of svara can be achieved by detecting the “stable” segment boundaries and discarding the time segments connecting these.

The underlying scale interval locations or svara are estimated from the prominent peaks of the long-term tonic-normalized pitch histogram across the concert. The allowed pitch deviation about the detected svara location, T_{tot} is empirically chosen to be ± 35 cents. This is based on previous work where this value was found to optimize the recognition of the svara sequence corresponding to a phrase based on the time series representing the melodic shape across many different instances of the same phrase extracted from audio recordings (Rao et al., 2014).

The above steps provide segments of the pitch time-series that approximate the scale notes while omitting the pitch transition regions. Next, a lower threshold duration of T_{dur} is applied to the fragments to discard fragments that are considered too short to be perceptually meaningful as held svaras (Rao et al., 2013). T_{dur} is empirically set to 250 ms, supported by previous subjective listening experiments (Vidwans et al., 2012). This leaves a string of fragments each labeled by the corresponding svara. Fragments with the same note (svara) value that are separated by gaps less than 100 ms are merged. The svara sequence information (i.e. scale degree and absolute

duration of each steady pitch segment) across the concert recording is stored.

4. Distributional Representations

Our goal is to propose computational representations that robustly capture the particular melodic features of the raga in a performance while being sensitive enough to the differences between allied ragas. Given that tonal material and the hierarchy of svaras are important components of the raga grammar, we consider representations of tonal hierarchy computable from the continuous-pitch melody extracted from the audio recording of the performance.

4.1. Representing Tonal Hierarchy

Given the pitch-continuous nature of raga music, we are faced with multiple competing options in the definition of a tonal representation. Closest to the tonal hierarchy vector of Krumhansl (1990) is the 12-bin histogram of the total duration of each of the svara segments detected from the melodic contour as described in Section 3.2. Considering the importance of the transitions connecting stable notes as well as micro-tonal differences in intonation between the same svara in different ragas, a higher dimensional histogram derived from all the pitch values in the continuous 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.

4.1.1. Pitch Saliency Histogram

The input to the system is the tonic normalized pitch contour (cents versus time). The pitch values are octave-folded (0–1200 cents) and quantized into p bins of equal width (i.e. the bin resolution is $\frac{1200}{p}$). The bin centre is the arithmetic mean of the adjacent bin edges. The saliency of each bin is proportional to the accumulated duration of the pitches within that bin. A probability distribution function is constructed where the area under the histogram sums to unity. This representation is equivalent to $P_continuous$ as proposed by Koduri et al. (2012). Given the number of bins, the histogram is computed as:

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

where H_k is the saliency of the k^{th} bin, F is the array of pitch values $F(n)$ of dimension N , (c_k, c_{k+1}) are the bounds of the k^{th} bin and 1 is an indicator random variable. **Figure 4** shows the pitch saliency histogram for $p = 1200$ (1 cent bin resolution) where different colors indicate different concerts in the corresponding raga. Comparing the Deshkar and Bhupali distributions, differences in the heights of the R peak (around bin 200) and in the precise location (intonation) of the G peak (around bin 400) are observed. For a bin resolution of 100 cents, the representation is equivalent to the pitch-class distribution (Chordia and Rae, 2007).

4.1.2. Svara Saliency Histogram

The svara saliency histogram is not equivalent to the PCD. The input to the system is the string of segmented stable svaras extracted from the melodic contour as described in Section 3.2 (and similar to the $P_duration$ proposed by Koduri et al. (2012)). The svara saliency histogram is obtained as:

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

where H_k is the saliency of the k^{th} bin, F is the array of pitch values $F(n)$ of dimension N , and S_k is the k^{th} svara of the octave. H_k is always a 12-element vector. **Figure 5** shows the tonal hierarchy in the form of svara saliency histogram where different colors indicate different concerts in the

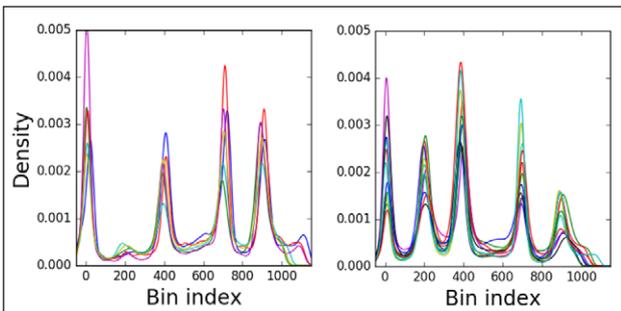


Figure 4: Pitch saliency histograms (octave folded, 1 cent bin resolution) of 6 concerts in raga Deshkar (left) and 11 concerts in raga Bhupali (right).

corresponding raga. One major difference between pitch saliency histogram and svara saliency histogram is that the precise intonation information is lost in the latter.

4.1.3. Svara Count Histogram

The frequency of occurrence of the notes was reported by Smith and Schmuckler (2004) to strongly correlate with the hierarchy of tones; hence we decide to investigate the same as a potential measure of svara saliency. This is equivalent to the $P_instance$ proposed by Koduri et al. (2012), where saliency is proportional to the frequency of occurrence of each svara. The svara count histogram is obtained as:

$$H_k = \sum_{j=1}^J 1_{[S(j)=S_k, k \in \{1, 2, \dots, 12\}]} \quad (3)$$

where H_k is the saliency of the k^{th} bin, S is the array of segmented svaras $S(j)$ of dimension J , and S_k is the k^{th} svara of the octave. H_k is always a 12-element vector. **Figure 6** shows the tonal hierarchy in the form of a svara count histogram where different colors indicate different concerts in the corresponding raga. We observe a high visual similarity between the svara saliency and count histograms.

4.2. Distance Measures

There exist several distinct distance measures between probability distributions with different physical interpretations (Cha, 2007). In the case of first-order pitch distributions, we are looking for a similarity in the tonal

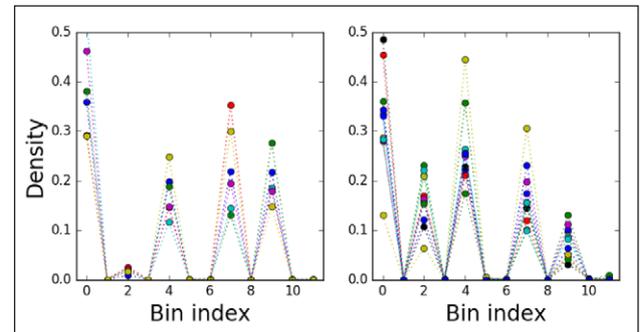


Figure 5: Svara saliency histograms (octave folded) of 6 concerts in raga Deshkar (left) and 11 concerts in raga Bhupali (right).

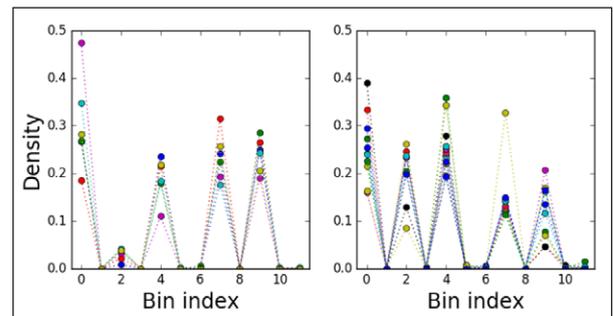


Figure 6: Svara count histograms (octave folded) of 6 concerts in raga Deshkar (left) and 11 concerts in raga Bhupali (right).

hierarchy captured by the distribution. The psychological model of Krumhansl (1990) is the most influential one and presents one of the most frequently applied distance measures in previous studies, the Correlation distance (Gedik and Bozkurt, 2010). This measure is often used in cognitive studies. This measure does not require the compared entities to be probability distributions but rather any two patterns of same dimension. The correlation distance is given by:

$$d_{\text{corr}}(P, Q) = \frac{\sum_{i=1}^N (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^N (p_i - \bar{p})^2 (q_i - \bar{q})^2}} \quad (4)$$

where P, Q refer to the two distributions under test, p_i, q_i are the masses of i^{th} bins of the distributions, and \bar{p}, \bar{q} are the means.

We consider the Euclidean, d_{eucl} (L-2 norm), and City-block, d_{ctbl} (L-1 norm) distances as they have been successfully used for pitch histogram similarity in previous studies. Gedik and Bozkurt (2010) advocate the City-block (Manhattan) distance for its superior performance in the shift-and-compare method for automatic tonic detection and they used the same for Makam recognition.

We also consider the Bhattacharyya distance as a suitable measure for comparing distributions. It is reported to outperform other distance measures with a PCD-based, as well as with higher-order distribution based, features in the raga recognition task (Chordia and Senturk, 2013; Gulati et al., 2016). For two probability distributions P and Q over the same domain, the Bhattacharyya distance is given by:

$$d_{\text{bhat}}(P, Q) = -\log \left(\sum_{i=1}^N \sqrt{p_i \cdot q_i} \right). \quad (5)$$

5. Evaluation Criteria and Experiments

With a view to identifying the choices in bin-width, type of histogram and distance measure between histograms that best serve in the representation of tonal hierarchy for raga music, we present experiments on our dataset of allied raga performances. The evaluation criteria relate to the achieved separation of performances belonging to different ragas in an allied pair. More specifically, we evaluate the performance of unsupervised clustering with k-means ($k = 2$) with its implicit Euclidean distance measure on each set of allied raga performances. The performance in unsupervised clustering can be quantified by the cluster validation measure *cluster purity* (hereafter *CP*) 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:

$$CP = \frac{1}{N} \sum_{i=1}^k \max_j |c_i \cap t_j| \quad (6)$$

where N is the number of data points, k is the number of clusters, c_i are the clusters and t_j are the underlying

classes. In our context ($k = 2$), a *CP* value of 1 indicates perfect clustering, whereas 0.5 implies random clustering.

We also evaluate the effectiveness of the distance between representations in recognizing whether any pair of concerts from a selected pair of allied ragas is from the same raga or not. Likewise for the other allied raga-pair subsets. We term the inverse of the distance computed between the pair of concert histograms as the *grammaticality coefficient* of the corresponding pair. A low value of the coefficient would indicate that the two representations arose from performances of different raga grammars. The receiver operating characteristic (ROC) can serve to evaluate the efficacy of this measure across the entire dataset of performances.

An ROC curve (Fawcett, 2006) provides a visualization of the trade-off between the true positives and false positives in a detection context. We consider our context to be the detection of non-matching raga pairs from the complete set of pairs. An ROC curve is obtained by varying the threshold applied to the obtained array of grammaticality coefficients, and computing the true positives and false positives. Given a histogram of a raga performance, the detection of a non-matching raga pair is considered a true positive (TP) if the performance under test belongs to the allied raga. It is considered a false positive (FP) if it belongs to the same raga as the template histogram. To evaluate the different tuning parameters of the tonal hierarchy representation, we use the area under curve (AUC) measure (higher values are better) and the Equal Error Rate (EER), where the false positive rate equals the false negative rate (lower values are better).

In summary, there are two main features of the tonal hierarchy model under investigation: (i) the histogram representation, (ii) the between-histograms distance measure. Both continuous-pitch (various bin-widths) and segmented-svara options are tested in combination with a number of distance measures between histograms.

5.1. Experiment 1: Unsupervised Clustering

Our base representation is the octave folded pitch salience histogram, normalized so that it is interpreted as a probability distribution. We test with different uniform bin widths, ranging from 1 to 100 cents, with centres coinciding with the tonic and semitone locations and their integer sub-multiples.

Figure 7 shows the cluster purity values at the considered bin widths (namely, 1, 12.5, 20, 25, 27, 30, 35, 40, 50 and 100 cents). We note that no degradation in purity is observed for 1 through 30 cent bin resolution. Each value on the curve is obtained by an average of 5 runs of the clustering algorithm using different initializations. For the cases of svara salience and count histograms, the average cluster purity values obtained are 0.96 and 0.84 respectively, indicating the slight superiority of the higher dimensional continuous-pitch distributions. Note that the svara histogram performances are considerably higher than the performance of the pitch salience distribution of the same dimensions (100-cent bin width). That the observed clustering in all configurations actually captures raga characteristics was confirmed by noting that each

discovered cluster was heterogeneous in performer and performance metadata.

5.2. Experiment 2: Detecting Non-matching Raga Pairs

For the dataset of concerts corresponding to a given allied raga pair, we create pairs of each concert with every other concert in the set. For example, we obtain 17×17 (=289) pairs out of the 17 concerts in the Deshkar-Bhupali allied raga dataset. Of these, $2 \times 6 \times 11$ (=132) pairs contain mismatched ragas. With this logic, we obtain a total of 1011 (= $17 \times 17 + 19 \times 19 + 19 \times 19$) distance values across the dataset of the 3 allied raga pairs with 468 instances corresponding to the non-matching pairs (i.e. a raga and its corresponding allied raga).

The ROCs of **Figure 8** show the true positive rate versus the false positive rate achieved in the detection of non-matching raga pairs for each of 3 histogram representations and all the 4 considered distance measures. The 3 histogram representations are the pitch salience with 25 cent bin width ($p = 48$), given that our previous experiment indicates undegraded clustering up to this bin width, and the 2 svara based histograms. The ROCs for the individual allied raga pairs are provided in the Appendix (Figures I, VI, and IX). The ROC evaluation measures, AUC and EER, for each allied raga-pair subset are presented in **Table 3**. The similarity in ROC shapes across raga-pairs indicates that the computed distance measure is independent of raga, and can serve as a

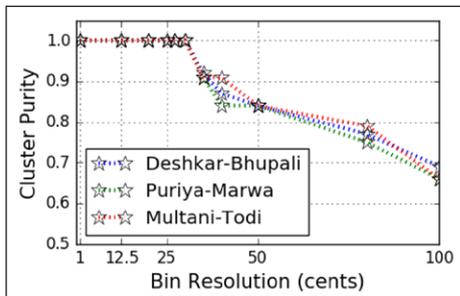


Figure 7: Cluster purity (CP) values obtained for different values of bin resolution for the pitch salience histograms in each of the 3 raga-pairs.

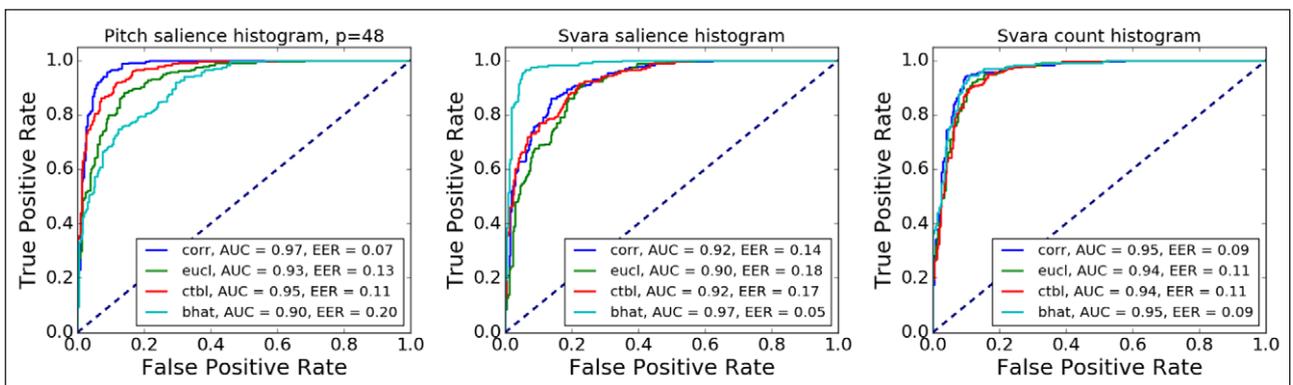


Figure 8: Combination of all three raga-pairs (full concerts): ROCs obtained for four different distance measures from pitch salience (left), svara salience (middle), and svara count (right) histograms from the combined distance vectors for all three raga-pairs.

robust measure of raga discrimination from an allied raga that does not need recalibration with change of raga. Accordingly, a common threshold was applied to each set of 1011 distances computed over the full dataset to obtain the corresponding ROC in **Figure 8**.

6. Discussion

6.1. Performance across Allied Raga-pairs

Salient aspects of the experimental outcomes presented in **Figure 8** and in **Table 3** are summarized next.

- The histogram representation obtained from the continuous melodic contour at the finest bin resolution ($p = 96$) with correlation distance is either as good as or, sometimes, better than any of the svara-based histograms. This indicates that capturing melodic movements such as glides and ornaments in the distributional representation is of value over relying on the stable segments only.
- The pitch salience histogram with $p = 48$ comes close to the performance of $p = 96$ with the correlation distance measure but is overall slightly worse with the other distance measures. As bin width is increased further to obtain the $p = 24$ and $p = 12$ pitch salience histograms, we note a sharp degradation, irrespective of the distance measure, in both the AUC and EER values.
- The svara-based histograms show clearly superior performances relative to pitch salience histograms of comparable dimension ($p = 12$). The Bhattacharyya distance works best for svara based histograms and this performance comes close to that of the $p = 96$ pitch salience histogram.

6.2. Bin Resolution: Coarse or Fine?

Given that the minimum interval between svaras in raga music is a semitone, one may argue that a 12-bin pitch-class distribution should be sufficient to discriminate ragas with different distributional information. However for ragas which share a scale, as with the allied ragas, a finer bin resolution may bring in further value by capturing differences in the precise intonation of the svaras. For

Table 3: Summary of results: evaluation measures AUC and EER for all combinations of representations and distance measures for each of the three allied raga-pairs. In bold font is the highest AUC across distance measures for a given representation and raga-pair.

Allied raga-pair	Distance measure	Evaluation metric	Pitch salience				Svara salience	Svara count
			$p = 96$	$p = 48$	$p = 24$	$p = 12$		
Deshkar-Bhupali	Correlation	AUC	.98	.97	.95	.89	.85	.90
		EER	.04	.06	.08	.19	.19	.15
	Euclidean	AUC	.97	.95	.92	.86	.84	.88
		EER	.08	.09	.15	.20	.18	.18
	City-block	AUC	.98	.98	.94	.90	.85	.88
		EER	.05	.05	.11	.18	.20	.18
Bhattacharyya	AUC	.98	.98	.96	.81	.93	.95	
	EER	.03	.05	.10	.25	.09	.10	
Puriya-Marwa	Correlation	AUC	.99	.95	.90	.84	.93	.94
		EER	.02	.08	.18	.23	.15	.13
	Euclidean	AUC	.92	.90	.82	.80	.94	.95
		EER	.13	.17	.19	.28	.15	.13
	City-block	AUC	.94	.93	.82	.77	.95	.95
		EER	.09	.12	.21	.29	.13	.11
Bhattacharyya	AUC	.95	.87	.79	.72	.98	.94	
	EER	.07	.19	.23	.42	.03	.11	
Multani-Todi	Correlation	AUC	.98	.98	.93	.94	.97	.99
		EER	.04	.05	.12	.15	.06	.02
	Euclidean	AUC	.95	.96	.87	.89	.93	.97
		EER	.10	.08	.18	.16	.12	.06
	City-block	AUC	.96	.97	.92	.92	.96	.98
		EER	.08	.05	.12	.11	.08	.05
Bhattacharyya	AUC	.90	.93	.90	.89	.99	.98	
	EER	.17	.17	.11	.16	.03	.03	

example, this is the case with the Deshkar-Bhupali pair in **Table 1** where at least 3 svaras (R;G;D) have a difference in intonation (*shruti*) for the same scale degree. The question arises about how fine a bin resolution is needed to capture the intonation differences. Datta et al. (2006) reported 27 cents as the optimal bin resolution ($p = 44$) for visually locating the precise shruti positions of different svaras. Our findings, in terms of cluster purity measure in **Figure 7** agree with the observation in that no degradation is observed for 1 through 30 cent bin resolution. However the ROC based evaluation of the grammaticality coefficient showed a slightly improved performance, in terms of AUC measure, for a finer ($p = 96$) bin resolution over a coarser one ($p = 48$). While this has a theoretical justification in terms of the fine intonation differences (e.g. raga Deshkar uses a higher *shruti* of G, by an order of 10 cents), we note that the svara histograms (where such intonation information is lost) perform nearly as well. This indicates that the relative saliences of the svaras (both in terms of

duration and frequency of occurrence), as implemented here, are adequate features as well.

6.3. Validation of Svara Transcription Parameters

It is interesting also to consider the computed histograms in terms of the distributional information provided by music theory. We consider here the information captured by the distribution in terms of musicological interpretations. From the pitch salience histograms ($p = 1200$), we observe certain phenomena which are musicologically interesting. In **Figure 4**, there is a small peak observed for N svara (1100 cents \approx 1100th bin) for raga Deshkar, which is theoretically a *varjit* (disallowed) svara, but its salience is comparable to that of the allowed R svara. In the svara salience (and count) histograms in **Figures 5** and **6**, the peak corresponding to the N svara (11th bin) is insignificant in comparison to that of the R svara. This indicates that the usage of the N svara is different. We confirmed, by interviewing musicians (including a couple

of artists from our dataset), that the N svara is used as a *kan* svara (touch note) in raga Deshkar. This contributed to the pitch salience histogram, but not the svara histograms computed via the stable svara segmentation step. We see that our empirically chosen segmentation parameters ($T_{tol} = 35$ cents, $T_{dur} = 250$ ms) provide a representation that is consistent with the theory.

Further, the chosen svara segmentation parameters ensure that the correlation between the svara salience histogram and the svara count histogram is high. If T_{dur} is set less than 250 ms, the *varjit* (disallowed) svaras would appear in the svara count histograms (the svara salience histogram would not be similarly affected because of the short durations). Additionally, the slow glides would get segmented into svaras and add to the count in the svara count histogram. In contrast, if T_{dur} is set higher, the svaras with *alpatva* (shorter duration, e.g. R svara in raga Deshkar) usage would go undetected and hence vanish from both svara salience and count histograms. This would lead to an inaccurate representation of the raga grammar.

6.4. New Insights on Time Scales

Given that the proposed histogram representations capture the distributional information in the concert, it is of interest to investigate the time scale at which the estimated tonal hierarchy can be considered to be stable and therefore representative of the raga. We carry out the previous allied raga discrimination experiments on segmented concerts. We divide each concert uniformly into n segments ($n = 1, 2, \dots, 5$) and construct the array of grammaticality coefficients across all the pairings associated with the set of $(\frac{1}{n})^{th}$ duration segments. The goal is to determine the smallest fraction of the full concert that is necessary to robustly discriminate between the matched and mismatched raga pairs. In every case, the distance measure chosen is the one with the best performance for the given histogram representation as indicated by **Figure 8**.

ROCs computed for segmentations with $n \geq 4$ were seen to lead to $AUC < 0.5$, which indicates that this time scale is too small to create a stable tonal hierarchy. We therefore consider only the cases of half and one-third segmentation further, giving us two datasets of concert segments of

sizes 110 and 165 segments respectively from the original set of 55 full concerts. **Figure 9** shows a comparison of ROCs across the full ($n = 1$) and partial ($n = 2, 3$) concerts for the various representations where, as in the case of full concerts, the considered pairs for distance computation are constrained to be within allied-raga data subsets.

6.4.1. Raga Delineation in Initial Portion

A good performer is continually engaged in exposing the raga through the improvisation interspersed throughout the concert. This is particularly important in the initial phase of the concert where establishing the raga identity in the mind of the listener is the primary goal. The ROCs (from **Figure 9**) indicate that at lower than the one-half portions, the duration is too small to constitute a stable tonal hierarchy, based on averaging a number of segments drawn from different regions of the concert. On the other hand, the initial phase of the concert comprising the *alap* and the initial part of the *vistar* (e.g. *sthayi* or the chorus line of the composition) is considered by musicians and listeners to fully embody the raga's melodic features. The histogram representation of the initial segment would be expected to be more stable across concerts in a given raga. We consider segments corresponding to the initial slow elaboration (from start of the concert till the end of *vistar* of the first bandish) as annotated by a trained Hindustani musician. The initial portion, so annotated, typically constitutes half the duration of the concert or less than this. We note in **Figure 9** that the ROCs of the concert-initial segments are indeed as good as those of the full concerts and superior to those obtained by considering all segments of similar duration drawn from different locations.

6.4.2. Distribution at Cycle Level

One of the smallest recognizable time-scales in the concert is that of the rhythmic cycle (*tala*). Each cycle, which can range in duration from 5 sec to 90 sec (*madhya* to *vilambit laya*), contains one or several melodic phrases. A performer typically has a plan for the overall evolution of the melodic content, based on individual and stylistic influences (van der Meer, 1980). We explore the application of the histogram representation at the cycle level, to the visualization of local melodic features. This

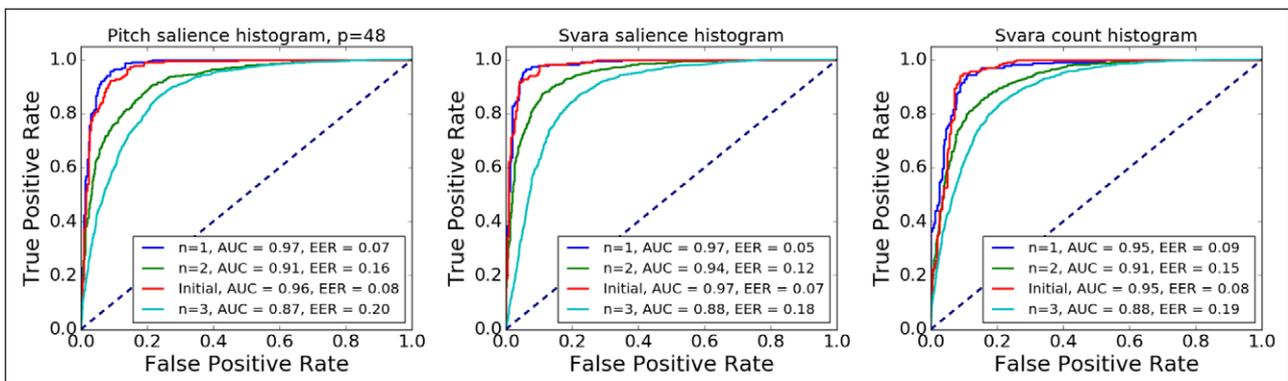


Figure 9: Combination of all three raga-pairs (partial and initial concert segments): Comparison of ROCs obtained with best distance measures (Correlation distance for pitch salience and Bhattacharyya distance for svara histograms) at different time-scales ($n = 1, 2, 3$) and with the initial portion (*alap* + *vistar*).

could be interesting in view of the fact that the *vistar* (lit. *barhat*, meaning expansion) of a raga performance refers to the gradual “unfolding” of a raga by focusing on the different *svaras* in succession on a broad time-scale (Widdess, 2011, 2013; Bagchee, 1998; Kulkarni, 2011). The precise duration spent on each *svara* in the course of this progression is not discussed in the musicology literature.

We select the 6 concerts in our dataset corresponding to Deshkar raga and further choose 6 concerts corresponding to Bhupali raga based on artist diversity while also retaining two same-artist concerts for the comparison of features in this context. The rhythmic cycle boundaries are marked based on the detection of the main accent (*sam*) location (Srinivasamurthy and Serra, 2014; Ross et al., 2012). We compute the *svara* salience histogram corresponding to each *tala* cycle of a concert. **Figure 10** shows the histogram versus cycle index for three concerts in raga Deshkar. The *svara* salience peaks are indicated in the color scale (dark indicates a strong peak). **Figure 11** shows the same representation for three concerts in raga Bhupali. We choose to show two concerts by the same artist (Ajoy Chakrabarty) in both the ragas. As the concert progresses in time, we observe a clear shifting of the location of the most salient *svara* in a cycle as well as variation in the melodic range covered in the cycle. The salient *svara* is seen to move from the lower to higher pitches accompanied by an overall increase in melodic range from the beginning of the concert to the end. The nature of the above variation is similar across the three concerts of each given raga. While in raga Deshkar, the two concerts by the same artist Ajoy Chakrabarty (DK AC-1 and DK AC-2, as presented in Table I in the Appendix) show nearly identical melodic progression

with respect to the *tala* cycles, the two Bhupali concerts by the same artist differ more. This indicates that the grammar of raga Deshkar, being less flexible, causes the artist to be somewhat more constrained during the improvisation.

The common practice of gradually unfolding a raga over the course of the *vistar* is brought out by the time-normalized summaries of the 6 concerts of each raga in **Figure 12**. A plot of the most salient *svara* in each cycle versus the cycle index is computed for each concert. This contour is smoothed using a 7-point median filter. Next the individual concert summaries are each time-normalized and superposed in **Figure 12**. We clearly observe the upward shifting with time of the “focal” note through the allowed *svaras*. The relative duration spent on each *svara* is concert dependent. We note the omission of the R *svara* in Deshkar curves as expected from its grammar. The step-wise movement of the salient *svara* bears a high resemblance to the time evolution of melody over the course of the *vistar*, as shown by Widdess (2011, **Figure 11**) including the sharp fall in pitch of the salient *svara* at the end of many concerts. The latter depicts the descent to a lower *svara* marking the end of the *vistar*.

Similar observations were reported by Ganguli et al. (2016), where the authors computed the *svara* salience histograms at the time scale of the breath phrase³ across 75 concerts drawn from a variety of ragas and performers. By aggregation over several breath phrases, an evolution of the salient *svara* similar to that in **Figure 12** was observed. The visual similarity among **Figure 12** in the current work, **Figure 11** in (Widdess, 2011), and **Figure 4** in (Ganguli et al., 2016) indicates that performers stick to

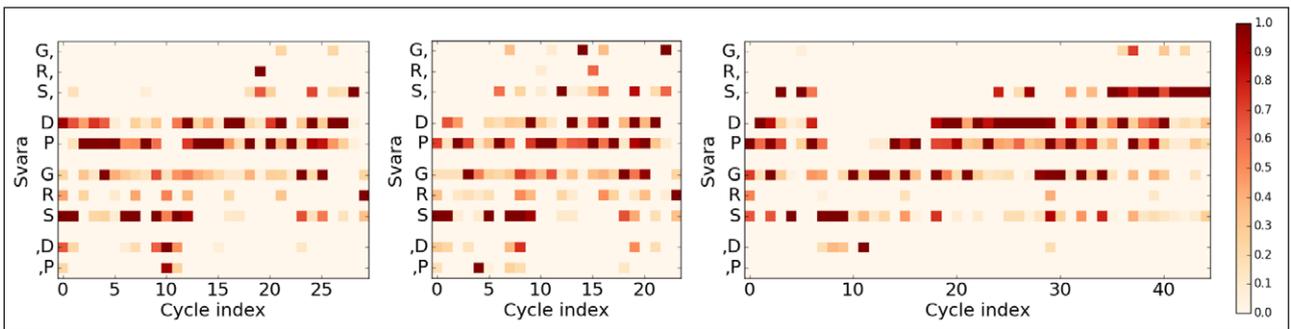


Figure 10: Svara salience histograms across tala cycles for DK AC-1 (left), DK AC-2 (middle), and DK KA (right).

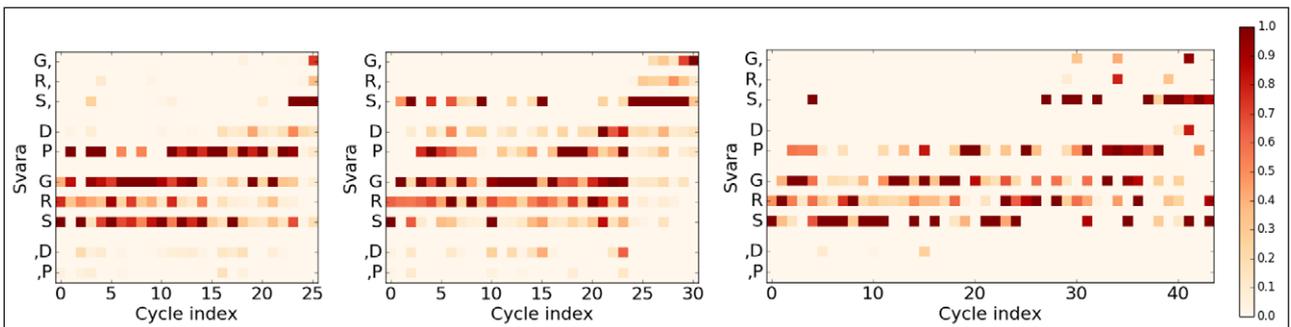


Figure 11: Svara salience histograms across tala cycles for BP AC-1 (left), BP AC-2 (middle), and BP RK (right).

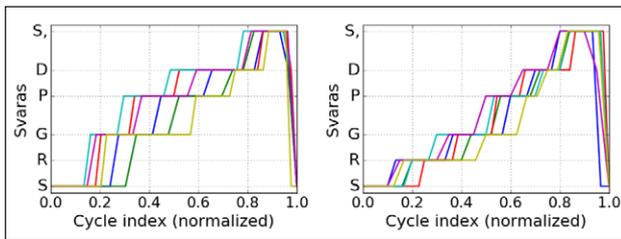


Figure 12: Time-normalized and smoothed cycle-level salient svara curves over the chosen 6 concerts in raga Deshkar (left) and raga Bhupali (right).

a broad schema of progressing in the melody from a lower to a higher svara (and swiftly returning to a lower svara to mark the intended ending).

7. Conclusion

Indian art music is a highly structured improvisational tradition. Based on the melodic framework of raga, performances display much creativity coexisting with adherence to the raga grammar. The performer's mandate is to elaborate upon the raga's essential features without compromising the cues to its identity. Raga grammar in music theory texts comprises only the essential specification of distributional and structural components of the grammar in terms of tonal hierarchies and typical phrases respectively. In the oral tradition, it is expected that there is much to be learnt from analysis of performances of great practitioners. A goal of the present work was to develop a computational representation for distributional information that can eventually be applied in the empirical analyses of audio recordings of performances. This would enable greater insights into the practical realization of a raga's distinctiveness in performance with potential applications to pedagogy.

The tonal hierarchy can be estimated from the detected melody line of a performance. In Western music, pitch-class profiles extracted from music pieces, both written scores and audio, have served well in validating the link between theoretical key profiles and their practical realization. In the pitch-continuous tradition of Indian art music where melodic shapes in terms of the connecting ornaments are at least as important in characterizing a raga as the notes of the scale, it becomes relevant to consider the dimensionality of the first order pitch distribution used to represent distributional information. Music theory however is not precise enough to help resolve the choices of bin width and distance measure between pitch distributions. We use a novel musicological viewpoint, namely a well-accepted notion of grammaticality in performance, to obtain the parameters of the computational representation based on audio performance data, we maximized the discrimination of allied raga performances using well-known evaluation metrics. Pitch salience histograms, as well as the stable segment based svara salience and count histograms, were considered as distinct representations of tonal hierarchy. We considered a variety of distance measures in order to derive a combination of histogram parameters and distance metrics that best separated same-raga pairs from allied-raga pairs.

It was found that svara salience histograms were as good as the high-resolution pitch salience histograms at the time-scale of full concerts, and superior to the svara count histograms. This observation continues to hold for the partial segments of concerts with a few exceptions. For the Deshkar-Bhupali dataset, we found that the continuous pitch salience histograms with the fine bin resolution of 25 cents served to capture raga grammar better for the segmented shorter portions of concerts (see the Appendix: Figure II). That is, the pitch distributions between the main peaks contributed usefully to the discrimination, indicating the importance of continuous melodic phrase shapes in this pentatonic raga-pair where the two most prominent svaras (*Vadi-Samvadi* in **Table 1**) are shared. Overall the best performing distance measures were correlation distance for the continuous pitch histograms and Bhattacharyya distance for discrete svara histograms. The proposed grammaticality coefficient served well to quantify the distributional difference across a pair of performances from same/allied raga independent of the raga.

Insights into the practical realization of the musicological concepts of raga delineation and melodic progression at the concert time scale were obtained. This points to the future possibility of developing the proposed methods for large-scale comparative studies in musicology. Although not the main focus of this work, the obtained outcomes can also be applied to the general raga recognition task, given the performance demonstrated on the relatively challenging sub-problem of discriminating allied ragas.

In this work, we used the discrimination of allied ragas (i.e. ragas which are similar enough in their distributional features so as to be confused easily) to tune the parameters of a computational representation of the raga grammar. A more direct, but considerably more challenging, validation of the proposed computational model would involve predicting deviations from the intended raga grammar as perceived by expert listeners. The use of a corpus, rather than a single concert, in the construction of a histogram template representing the raga grammar, as the basis of comparison, would serve to make this work more practically applicable to detecting ungrammaticality in the context of pedagogy. Future work must also address the modeling of the structural aspect of raga grammar, corresponding to the phrases, since this is a more easily accessed cue to raga identity for listeners (Ganguli and Rao, 2017). Finally, it would be of interest to investigate the relative weighting of the different raga attributes for an overall rating of grammaticality, possibly at the different time scales, based on observed expert judgments.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Discussion on individual allied raga-pairs. DOI: <https://doi.org/10.5334/tismir.11.s1>

Notes

- ¹ <https://dunya.compmusic.upf.edu/hindustani>.
- ² <https://musicbrainz.org/>.

³ The authors defined “breath phrase” as a continuous segment of the melodic contour delimited by musician-intended breath pauses.

Acknowledgements

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).

Competing Interests

The authors have no competing interests to declare.

References

- AUTRIM-NCPA.** (2017). Music in Motion: The automated transcription for Indian music (AUTRIM) project by NCPA and UvA. URL: <https://autrimncpa.wordpress.com/>. Last accessed: October 26, 2017.
- Bagchee, S.** (1998). *Nad: Understanding Raga Music*. Business Publications Inc.
- Belle, S., Joshi, R., & Rao, P.** (2009). Raga identification by using swara intonation. *Journal of ITC Sangeet Research Academy*, 23.
- Bozkurt, B.** (2008). An automatic pitch analysis method for Turkish Maqam music. *Journal of New Music Research*, 37(1), 1–13. DOI: <https://doi.org/10.1080/09298210802259520>
- Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L.** (1984). Tonal hierarchies in the music of north India. *Journal of Experimental Psychology: General*, 113(3), 394–412. DOI: <https://doi.org/10.1037/0096-3445.113.3.394>
- Cha, S. H.** (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4), 300–307.
- Chordia, P., & Rae, A.** (2007). Raag recognition using pitch-class and pitch-class dyad distributions. In: *Proceedings of the International Conference on Music Information Retrieval*, 431–436.
- Chordia, P., & Senturk, S.** (2013). Joint recognition of raag and tonic in north Indian music. *Computer Music Journal*, 37(3), 82–98. DOI: https://doi.org/10.1162/COMJ_a_00194
- Danielou, A.** (2010). *The ragas of Northern Indian music*. Munshiram Manoharlal Publishers.
- Datta, A. K., Sengupta, R., Dey, N., & Nag, D.** (2006). *Experimental analysis of Shrutis from performances in Hindustani music*. Scientific Research Department, ITC Sangeet Research Academy.
- Dighe, P., Karnick, H., & Raj, B.** (2013). Swara histogram based structural analysis and identification of Indian classical ragas. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 35–40.
- Fawcett, T.** (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fujishima, T.** (1999). Realtime chord recognition of musical sound: a system using common Lisp music. In: *Proceedings of the International Computer Music Conference*, 464–467.
- Ganguli, K. K., Gulati, S., Serra, X., & Rao, P.** (2016). Data-driven exploration of melodic structures in Hindustani music. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 605–611.
- Ganguli, K. K., & Rao, P.** (2017). Towards computational modeling of the ungrammatical in a raga performance. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 39–45.
- Gedik, A. C., & Bozkurt, B.** (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, 90(4), 1049–1063. DOI: <https://doi.org/10.1016/j.sigpro.2009.06.017>
- Ghosh, P. K., & Narayanan, S. S.** (2009). Pitch contour stylization using an optimal piecewise polynomial approximation. *IEEE Signal Processing Letters*, 16(9), 810–813. DOI: <https://doi.org/10.1109/LSP.2009.2025824>
- Gomez, E.** (2006). Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, 18(3), 294–304. DOI: <https://doi.org/10.1287/ijoc.1040.0126>
- Gulati, S., Bellur, A., Salamon, J., Ranjani, H. G., Ishwar, V., Murthy, H. A., & Serra, X.** (2014). Automatic tonic identification in Indian art music: Approaches and Evaluation. *Journal of New Music Research*, 43(1): 53–71. DOI: <https://doi.org/10.1080/09298215.2013.875042>
- Gulati, S., Serra, J., Ganguli, K. K., Senturk, S., & Serra, X.** (2016). Time-delayed melody surfaces for raga recognition. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 751–757.
- Huron, D.** (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.
- ITC-SRA.** (2017). Distinguishing between similar ragas. URL: <http://www.itcsra.org/Distinguishingbetween-Similar-Ragas>. Last accessed: October 26, 2017.
- Koduri, G. K., Gulati, S., & Rao, P.** (2011). A survey of raaga recognition techniques and improvements to the state-of-the-art. In: *Proceedings of the 8th Sound and Music Computing Conference*, 39–46.
- Koduri, G. K., Gulati, S., Rao, P., & Serra, X.** (2012). Raga recognition based on pitch distribution methods. *Journal of New Music Research*, 41(4), 337–350. DOI: <https://doi.org/10.1080/09298215.2012.735246>
- Koduri, G. K., Ishwar, V., Serra, J., & Serra, X.** (2014). Intonation analysis of ragas in Carnatic music. *Journal of New Music Research*, 43(1), 72–93. DOI: <https://doi.org/10.1080/09298215.2013.866145>
- Krumhansl, C. L.** (1990). *Cognitive Foundations of Musical Pitch*, 77–110. Oxford University Press, New York. Chapter 4: A key-finding algorithm based on tonal hierarchies.
- Krumhansl, C. L., & Cuddy, L. L.** (2010). *A Theory of Tonal Hierarchies in Music*, 51–87. Springer: New York.
- Krumhansl, C. L., & Kessler, E. J.** (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys.

- Psychological review*, 89(4), 334–368. DOI: <https://doi.org/10.1037/0033-295X.89.4.334>
- Kulkarni, S.** (2011). *Shyamrao Gharana, 1*. Prism Books Pvt. Ltd.
- Kumar, V., Pandya, H., & Jawahar, C. V.** (2014). Identifying ragas in Indian music. In: *International Conference on Pattern Recognition*, 767–772. DOI: <https://doi.org/10.1109/ICPR.2014.142>
- Mahajan, A.** (2010). *Ragas in Hindustani Music: Conceptual Aspects*. Gyan Publishing House.
- Oak, V.** (2017). 22 shruti. URL: <http://22shruti.com/>. Last accessed: October 26, 2017.
- Peeters, G.** (2006). Chroma-based estimation of musical key from audio-signal analysis. In: *Proceedings of the International Conference on Music Information Retrieval*, 115–120.
- Powers, H. S., & Widdess, R.** (2001). *India, subcontinent of*, chapter III: Theory and practice of classical music, 170–210. New Grove Dictionary of Music. Macmillan, London, 2nd edition.
- Raja, D.** (2016). *The Raga-ness of Ragas: Ragas Beyond the Grammar*. D.K. Print World Ltd.
- Raman, R., & Dowling, W. J.** (2016). Real-time probing of modulations in south Indian classical (Carnatic) music by Indian and Western musicians. *Music Perception*, 33(3), 367–393. DOI: <https://doi.org/10.1525/mp.2016.33.3.367>
- Rao, P., Ross, J. C., & Ganguli, K. K.** (2013). Distinguishing raga-specific intonation of phrases with audio analysis. *Ninaad*, 26–27(1), 59–68.
- Rao, P., Ross, J. C., Ganguli, K. K., Pandit, V., Ishwar, V., Bellur, A., & Murthy, H. A.** (2014). Classification of melodic motifs in raga music with timeseries matching. *Journal of New Music Research*, 43(1), 115–131. DOI: <https://doi.org/10.1080/09298215.2013.873470>
- Rao, S., Bor, J., van der Meer, W., & Harvey, J.** (1999). *The Raga Guide: A Survey of 74 Hindustani Ragas*. Nimbus Records with Rotterdam Conservatory of Music.
- Rao, S., & Rao, P.** (2014). An overview of Hindustani music in the context of Computational Musicology. *Journal of New Music Research*, 43(1), 24–33. DOI: <https://doi.org/10.1080/09298215.2013.831109>
- Rao, V., & Rao, P.** (2010). Vocal melody extraction in the presence of pitched accompaniment in polyphonic music. *IEEE Transactions on Audio, Speech & Language Processing*, 18(8), 2145–2154. DOI: <https://doi.org/10.1109/TASL.2010.2042124>
- Ross, J. C., Mishra, A., Ganguli, K. K., Bhattacharyya, P., & Rao, P.** (2017). Identifying raga similarity through embeddings learned from compositions' notation. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 515–522.
- Ross, J. C., Vinutha, T. P., & Rao, P.** (2012). Detecting melodic motifs from audio for Hindustani classical music. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 193–198.
- Serra, X.** (2011). A multicultural approach to music information research. In: *Proceedings of the International Society for Music Information Retrieval Conference*, 151–156.
- Serra, X.** (2014). Creating research corpora for the computational study of music: the case of the Comp-Music project. In: *Proceedings of the 53rd AES International Conference on Semantic Audio*, 1–9.
- Smith, N. A., & Schmuckler, M. A.** (2000). Pitchdistributional effects on the perception of tonality. In: *Proceedings of the International Conference on Music Perception and Cognition*, 5–10.
- Smith, N. A., & Schmuckler, M. A.** (2004). The perception of tonal structure through the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 268–286. DOI: <https://doi.org/10.1037/0096-1523.30.2.268>
- Srinivasamurthy, A., & Serra, X.** (2014). A supervised approach to hierarchical metrical cycle tracking from audio music recordings. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 5217–5221. DOI: <https://doi.org/10.1109/ICASSP.2014.6854598>
- SwarGanga.** (2017). Swarganga music foundation. URL: <https://www.swarganga.org/>. Last accessed: October 26, 2017.
- Temperley, D.** (1999). What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, 17(1), 65–100. DOI: <https://doi.org/10.2307/40285812>
- Temperley, D., & Marvin, E. W.** (2008). Pitch-class distribution and the identification of key. *Music Perception*, 25(3), 193–212. DOI: <https://doi.org/10.1525/mp.2008.25.3.193>
- van der Meer, W.** (1980). *Hindustani Music in the 20th Century*. Martinus Nijhoff Publishers. DOI: <https://doi.org/10.1007/978-94-009-8777-7>
- Vidwans, A., Ganguli, K. K., & Rao, P.** (2012). Classification of Indian classical vocal styles from melodic contours. In: *Proceedings of the 2nd CompMusic Workshop*, 139–146.
- Vijaykrishnan, K. G.** (2007). *The Grammar of Carnatic Music*. De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110198881>
- Widdess, R.** (1994). Involving the performers in transcription and analysis: a collaborative approach to Dhrupad. *Ethnomusicology*, 38(1), 59–79. DOI: <https://doi.org/10.2307/852268>
- Widdess, R.** (2011). Dynamics of melodic discourse in Indian music: Budhaditya Mukherjee's alap in rag Puriya-Kalyan, 187–224. Oxford University Press.
- Widdess, R.** (2013). Schemas and improvisation in Indian music. In: Kempson, R., Howes, C., & Orwin, M. (Eds.), *Language, Music and Interaction*, 197–209. College Publications: London.

How to cite this article: Ganguli, K. K., & Rao, P. (2018). On the Distributional Representation of Ragas: Experiments with Allied Raga Pairs. *Transactions of the International Society for Music Information Retrieval*, 1(1), pp. 79–95. DOI: <https://doi.org/10.5334/tismir.11>

Submitted: 20 January 2018

Accepted: 03 June 2018

Published: 19 December 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[

Transactions of the International Society for Music Information Retrieval is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 