RESEARCH ARTICLE

On the distributional representation of ragas: experiments with allied raga pairs

Kaustuv Kanti Ganguli⁎, Preeti Rao†

Abstract
Raga grammar provides a theoretical framework that supports creativity and flexibility in improvisation while carefully maintaining raga distinctiveness 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. As is well known about 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 interesting insights on how these apparently divergent requirements are ably met in practice. Among the relatively few such studies is one by [Widdess [2011]] who presents the detailed analysis of a sitar performance recording from the point of view of a listener to obtain an understanding of how the raga characteristics are manifested in the 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

⁎Department of Electrical Engineering,
†Indian Institute of Technology Bombay, Mumbai, India.
distinguish different raga performances.

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 the 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 have been found to correspond with the tonal hierarchies in different keys (Smith and Schmuckler 2000; Raman and Dowling 2016; Temperley 1999). Further, the MIR 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 powerfully demonstrated via the raga music transcription visualization available in (Autrim-NCPA 2017). Recognizing the importance of distributional information, we find that this 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 for example. While the raga grammar as specified in music theory texts is not precise enough to help 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 occurs around the notion of 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 can still be 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 the 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; 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 maximizing the recognition of ungrammaticality in terms of the distributional information in a performance of a stated raga (Ganguli and Rao 2017). We do not consider structural information in the form of phrase characteristics in this work. We use audio recordings of performances of eminent Hindustani vocalists as very creatively rendered, but grammatically accurate, references of the stated raga. Performances in the allied raga serve as proxies for the corresponding ungrammatical performances. The obtained model will further be applied to obtain possibly new 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 previous literature on pitch distributions in the MIR context with focus on raga music with a view to motivating 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 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. The only other inscription we see is the parentheses, e.g. (R) that indicates 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 to be 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 infer the technical boundaries of each of the ragas more easily.

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

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; Kulkar, 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) mentions 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 in (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 “fit” 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).

There have been other studies that found 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

| Feature | Deshkar | Bhupali | Puriya | Marwa | Multani | Todi |
|---------|---------|---------|--------|-------|---------|------|
| Tonal material | SRGPD | SRGPD | SrGMDN | SrGMDN | SrgMPdN | SrgMPdN |
| Aroha | SGPD, SPDS | SRG, PDS | N(r)G, MN(r)D | DNr, GMd, NrS | NSrgMP, PNS | SrgMP, MnNS |
| Avroha | S, PDGP, DPG(r)S | SPD, GDP, GRS | NM, DG, MG(r)S | RND, DMGr, DrS | SN(d)p, gmG, (r)NS | SngPd, MgrS |
| Vadi, Samvadi | D, G | G, D | G, N | r, D | S, P | d, g |
| Char. phrases | SGPD, P(D)SP | RDS, RPG, PDS | N(r)S, N(r)G, MDG | DNr, rGMd, DrS | rNS, gMP, PNS | Srg, PMd, dNS |
| Shruti | Higher R, G, D | Natural R, G, D | Lower r | Higher r, D | Higher N | Lower r, g |

Table 1: Specification of raga grammar for three allied raga-pairs (ITC-SRA, 2017; Rao et al., 1999; Autrim-NCPA, 2017; Oak, 2017; Kulkarni, 2011; SwarGanga, 2017; Mahajan, 2010).
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).

In a similar study, Raman and Dowling (2016) used four Carnatic (south Indian classical music repertoire) ragas to come up with a baseline template and then employed two techniques of modulation, namely ghaharbedham (changing the tonic while keeping the tones unchanged) and ragamalika (shifting to any distinctly different raga). By tracking responses of both Western and Indian listeners to the modulations in Carnatic music, they identified the various cues that listeners utilized in order to discern the modulation. The authors reported that the Westerners’ responses matched those of the Indians on ragas with structures similar to Western scales, but differed when ragas were less familiar, and surprisingly, they registered the shifts more strongly than Indian musicians. These findings, according to the authors, converged with previous research in identifying three types of cues: (i) culture-specific cues – schematic and veridical knowledge – employed by Indians, (ii) tone-distribution cues – duration and frequency of note occurrence – employed by both Indians and Westerners, and (iii) transference of schematic knowledge of Western music by Western participants.

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 never 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 the 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 as 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 12/24/36-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 given a training dataset of performances across several ragas (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 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 pitch class distribution (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) melody in the corresponding pitch range. 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 in Koduri et al. (2012, 2013) and Belle et al. (2009, 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 distributions were tested: (i) \( P_{\text{continuous}} \): unconstrained pitch histogram with a choice of bin resolutions (1, 2, \cdots, 100 cents), (ii) \( P_{\text{duration}} \): constrained to stable notes only, weighted by the durations thereof (12-bin), and (iii) \( P_{\text{instance}} \): constrained to stable notes only, weighted by the count of instances (12-bin). The only distance measure tested was the symmetric KL divergence. Overall, \( P_{\text{duration}} \) performed the best. For \( P_{\text{continuous}} \), there was not a 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 dataset 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.
K. K. Ganguli and P. Rao: On the distributional representation of ragas: experiments with allied raga pairs

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

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 segmented note (svara) 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 MusicBrainz². 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 Table 4 in Appendix. As anyone who is familiar with the khayal genre can testify, the selected artists are the stalwarts of Hindustani vocal music spanning the past 7 or 8 decades.

| 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)** |

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

---

¹[https://dunya.compmusic.upf.edu/Hindustani](https://dunya.compmusic.upf.edu/Hindustani)
²[https://musicbrainz.org/](https://musicbrainz.org/)
verse 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 to 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 (en-
ergy 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 percent, 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 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, $T_{tol}$, is empirically chosen to be $\pm 35$ 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 hierarchy of svaras are an important component 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 connect-
ing 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 Salience 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 salience of each bin is proportional to the accumulated duration of the pitches within that bin. A probability distribution function (pdf) is constructed where the area under the histogram sums to unity. This representation is equivalent to \( P_{\text{continuous}} \) as proposed by Koduri et al. (2012). Given the number of bins, the histogram is computed as:

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

where \( H_k \) is the salience of the \( k^{\text{th}} \) bin, \( F \) is the array of pitch values \( F(n) \) of dimension \( N \), and \( c_k \) and \( c_{k+1} \) are the bounds of the \( k^{\text{th}} \) bin and \( \mathbb{1} \) is an indicator random variable.\(^3\) Figure 4 shows the pitch salience 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 200\(^{\text{th}} \) bin) and in the precise location (intonation) of the \( G \) peak (around 400\(^{\text{th}} \) bin) are observed. For a bin resolution of 100 cents, the representation is equivalent to the pitch class distribution (PCD) (Chordia and Rae, 2007).

4.1.2 Svara Salience Histogram

The svara salience 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 \( P_{\text{duration}} \) proposed by Koduri et al. (2012)). The svara salience histogram is obtained as:

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

where \( H_k \) is the salience of the \( k^{\text{th}} \) bin, \( F \) is the array of pitch values \( F(n) \) of dimension \( N \), and \( S_k \) is the \( k^{\text{th}} \) svara of the octave. \( H_k \) is always a 12-element vector. Figure 5 shows the tonal hierarchy in the form of svara salience histogram where different colors indicate different concerts in the corresponding raga. One major difference between pitch salience histogram and svara salience 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 salience.
This is equivalent to the $P_{\text{instance}}$ proposed by Koduri et al. (2012), where salience is proportional to the frequency of occurrence of each svara. The svara count histogram is obtained as:

$$H_{k} = \sum_{j=1}^{J} [S(j)=S_{k} \in \{1,2,\ldots,12\}]$$

(3)

where $H_{k}$ is the salience 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 salience and count histograms.

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

### 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 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:

$$d_{\text{corr}}(P, Q) = \frac{\sum_{i=1}^{N} (p_{i} - \overline{p})(q_{i} - \overline{q})}{\sqrt{\sum_{i=1}^{N} (p_{i} - \overline{p})^2(q_{i} - \overline{q})^2}}$$

(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 $\overline{p}, \overline{q}$ are the means.

We also consider the Euclidean (L-2 norm) and Cityblock (L-1 norm) distances as they have been successfully used for pitch histogram similarity in previous studies. The Euclidean distance between two n-dimensional vectors $P$ and $Q$ in Euclidean $n$-space is the length of the line segment connecting them, given by:

$$d_{\text{eucl}}(P, Q) = \sqrt{\sum_{i=1}^{N} (p_{i} - q_{i})^2}$$

(5)

Gedik and Bozkurt (2010) advocate the Cityblock distance for its superior performance in the shift-and-compare method for automatic tonic detection and they used the same for Makam recognition. The Cityblock (also known as Manhattan) distance between two vectors $P, Q$ in an $n$-dimensional real vector space $\mathbb{R}$, is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes, given by:

$$d_{\text{ctbl}}(P, Q) = \sum_{i=1}^{N} |p_{i} - q_{i}|$$

(6)

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 measures the similarity between the distributions, and is given by:

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

(7)

### 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:

$$\text{CP} = \frac{1}{N} \sum_{i=1}^{k} \max_{j} |c_i \cap t_j|$$  \hspace{1cm} (8)

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 estimating the grammaticality of a given concert with reference to a selected raga represented by a histogram template. Every performance in the Deshkar-Bhupali dataset serves once as the grammar template of the corresponding raga. This template is paired with every other performance in the 2-raga dataset to obtain grammatical and ungrammatical instances of pairings with reference to the chosen template. We term the inverse of the distance computed between the concert histogram and the template histogram in a pair as the grammaticality coefficient. A low value of the coefficient would indicate an ungrammatical performance with reference to the given raga template. 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 ungrammatical instances (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 raga template histogram, the detection of an ungrammatical instance 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. We use the ROC shape to evaluate the different tuning parameters of the tonal hierarchy representation. An objective function to compare across representations and distance measures is the area under curve (AUC) measure. Closer the AUC value to 1, the better is the performance. Additionally, the shape of an ROC curve is characterized by the Equal Error Rate (EER), where the false positive rate equals the false negative rate (i.e. (1 - true positive rate)). Closer the EER value to 0, better is the performance.

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. Finally, we also investigate the advantages, if any, of using the actual full-range distributions versus octave-folded versions.

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. Datta et al. (2006) proposed a number of possibilities for bin-centres for pitch salience histograms – right/left aligned with respect to bin-edges or aligned at centre (arithmetic/geometric/harmonic mean). Our informal observations indicated that bin centre choices did not affect our results. Datta et al. (2006) also reported different bin-edge configurations based on different tuning systems (just intonation and equal temperament) with the aim of uncovering the underlying tuning system.

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 cents bin resolution. Each value on the curve is obtained by an average of 5 runs of the clustering algorithm using different sets of initial values drawn from the dataset to ensure repeatability. For the case of svara salience and count histograms, the average cluster purity value is obtained as 0.96 and 0.84 respectively (note that these values are considerably higher than that corresponding to the pitch salience distribution with 100-cents bin width) 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 same dimension pitch salience distribution of 100-cents bin width. We will see that the similarity in performance between the low-dimensional svara-based histograms and the high-dimensional fine bin-width pitch-salience histograms is reflected also in the outcomes of Experiment 2. This serves to reinforce the current conclusion about the segmented pitch contour corresponding to the svara regions alone containing sufficient discriminatory information in the context of distributions derived from the melodic contour across the concert. 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.

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.

5.2 Experiment 2: Detecting Ungrammaticality

We create $17 \times 17 = 289$ distinct 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 and hence serve as ungrammatical instances. The ROCs of Figure 8 show the true positive rate versus the false positive rate achieved in the detection of ungrammatical instances 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 indicate undegraded clustering up to this bin width, and the 2 svara based histograms. We observe that the different distance measures perform similarly in the case of the pitch salience histogram whereas the svara-based histograms show distinctly superior performance with the Bhattacharyya distance measure. A more extended discussion appears in the next section in the context of the results obtained across the entire dataset of the 3 allied raga pairs.

Given that raga grammar specification can include a mention of the pitch register or region of main activity as remarked in Section 2.1, we also consider histograms without octave folding to capture the differences between ragas with the same scale but possibly different regions of major activity in a 3-octave region (-1200, 2400) cents with 0 cents representing the tonic. Raga Deshkar and Bhupali, for example, have their vadi svaras in different tetrachords of the octave (upper or uttarang-vadi and lower or poorvang-vadi respectively) and therefore might be expected to have the majority of pitch movements in higher and lower regions respectively. Figure 9, the counterpart of Figure 8, but without octave folding in any of the histograms, does not seem to support the theory however. We see the performance worsening without octave folding. An examination of the extracted melodies showed that the region of most activity in a performance was not clear-cut and showed considerable artist- and concert-dependent variation for the given raga. Further, octave folding suffers less from octave errors in pitch tracking and also from degradations due to the sparsity of the histograms.

6. Discussion

6.1 Summary across Allied Raga-pairs

The ROCs for the detection of ungrammatical instances presented in the previous section were obtained from pairing concerts drawn from the Deshkar-Bhupali raga dataset. A robust measure of grammaticality should ideally not need recalibration with a change of raga. In order to test this for the histogram representations and distance measures under consideration, we compute the distance for each pair of concerts drawn from all pairings of concerts within each of the 3 allied raga datasets in Table 2. That is, we have at hand a total of $17 \times 17 + 19 \times 19 + 19 \times 19 = 1011$ distance values, of which 468 instances correspond to the ungrammatical pairings (i.e. a raga and its corresponding allied raga). ROCs are obtained from the computed distances, separately for each combination of histogram representation and distance measure.

Figure 10 presents the ROCs obtained for four different distance measures from pitch salience ($p=48$),
Figure 8: Raga-pair Deshkar-Bhupali (full concerts, octave-folded): ROCs obtained for four different distance measures from pitch salience (left), svara salience (middle), and svara count (right) histograms.

Figure 9: Raga-pair Deshkar-Bhupali (full concerts, not octave-folded): ROCs obtained for four different distance measures from pitch salience (left), svara salience (middle), and svara count (right) histograms.

Table 3: Summary of results: evaluation measures AUC and EER for all combinations of representations and distance measures for all three allied raga-pairs.
svara salience and svara count histograms from the grammaticality coefficient array computed across the full dataset. More specific observations on the relative performances of the different histogram representations and probability distance measures follow from the ROC evaluation measures presented in Table 3 in the form of AUC and EER. We observe that the ROCs corresponding to the full dataset are highly similar to those obtained for the Deshkar-Bhupali dataset in the previous section. Further, the curves are much smoother due to the larger number of instances involved in the ROC computation. The ROC curves being similar across the different raga-pairs supports the important observation that the grammaticality coefficient values are largely independent of the particular allied raga pair used in the computation. This indicates an interpretation of the computed distance measure that is independent of raga. We note the following.

- At the finest bin resolution ($p=96$), the pitch salience histogram representation is largely insensitive to the choice of the distance measure. This histogram representation obtained from the continuous melodic contour is either as good as or, more often, better than any of the svara-based histograms with any distance measure. 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, on the other hand, 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 theoretical resolution of the tonal material in raga music is a semitone (100 cents), 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, if any, in the precise intonation of the svaras. For 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 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) presented an evaluation of different bin resolutions in terms of the ease of visually locating the precise shruti (intonation) positions of different svaras from the computed histogram given theoretical positions. They reported 27 cents as the bin resolution ($p = 44$) optimal for visualization of the clusters. 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 cents 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 totally lost) actually 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 musical interpretations, if any. From the pitch salience histograms ($p=1200$), we observe certain phenomena which are musically interesting and insightful. In Figure 4, there is a peak observed for $N$ svara (1100 cents $\approx 1100^{th}$ bin) for raga Deshkar, which is theoretically a varjit (disallowed) svara. But the salience

---

4Refer to Appendix for the other two allied raga-pairs.
of the N svara is comparable to that of the R svara. Hence the question arises whether the usage of N svara in the raga performance is equivalent to that of R svara. This ambiguity is resolved by the svara salience (and count) histograms representing raga Deshkar in Figures 5 and 6, where 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 in terms of capturing the R svara while rejecting the disallowed (by the grammar) N svara.

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 ($n = 1, 2, \ldots, 5$) and construct the array of grammaticality coefficients across all the pairings associated with the set of $\left(\frac{1}{n}\right)^{th}$ 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. Here, as before, one segment acts as a reference template for its raga grammar; the other member of the pair is a segment drawn from either the same raga (giving us a grammatical instance) or from the allied raga (giving us an ungrammatical instance).

The ROC obtained from $\left(\frac{1}{4}\right)^{th}$ (and smaller) of a concert results in an 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 Deshkar-Bhupali datasets of concert segments of sizes 34 and 51 segments respectively. Figure 11 shows a comparison of ROCs across the full ($n=1$) and partial ($n=2,3$) concerts for the various representations.

The performance of the pitch salience histogram at $p=96$ (not shown in Figure 11) was found to exceed that at $p=48$. This observed superiority of the fine-binned histogram over the svara-based histogram suggests that the micro-tonal 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.

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 11) indicate that beyond the one-third portion, the time-scale is too small to constitute a stable tonal hierarchy, but this does not completely answer the question yet. The ROC evaluation was based on averaging a number of segments drawn from different regions of the concert. Also the concerts in the dataset varied in their total durations. 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 investigate the nature of the distribution computed from the initial segment of the concert versus that from same duration but later occurring segments. 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. Figure 12 shows the ROCs of these annotated segments in comparison with the ROCs computed from the full (n=1) and half (n=2) concerts. We observe that the ROCs of the concert-initial segments are indeed as good as those of the full concerts and much better than those computed from across half-concert segments for all the histogram representations.

6.4.2 Distribution at Cycle Level
One of the smallest recognizable time-scales in the concert is that of the rhythmic cycle. Each cycle which can range in duration from 5 sec to 90 sec (madhya to vilambit laya), depending on the local tempo, contains one or several melodic phrases. A performer typically has a plan for the overall evolution of the concert in terms of the melodic content that is based on individual and stylistic influences (van der Meer, 1980). We explore the application of the histogram representation, but now at the cycle level, to the visualization of local melodic features. This 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 6 concerts in our dataset corresponding to Deshkar raga and 6 concerts corresponding to Bhupali raga. 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 13 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 14 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 a gradual expansion of the melodic range. The nature of the above variation is strikingly similar across the three concerts of a given raga. While in raga Deshkar, the two concerts by the same artist Ajoy Chakrabarty (DK_AC-1 and DK_AC-2) 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.

That the common trend of gradually unfolding a raga in course of vistar is maintained by the artists in both the ragas is brought out very nicely by the time-normalized summaries of the 6 concerts of each raga in Figure 15. A plot of the most salient svara in each cycle versus the cycle index is computed for
Figure 11: Raga-pair Deshkar-Bhupali (partial concerts): Comparison of ROCs obtained with Bhattacharyya distance for different time-scales (n=1,2,3) of the concerts.

Figure 12: Raga-pair Deshkar-Bhupali (initial portion of concerts): Comparison of ROCs obtained with Bhattacharyya distance for annotated initial portion (alap+vistar) of the concerts. We plot the ROCs for full (n=1) and half (n=2) concerts for performance comparison.

Figure 13: Svara salience histograms across tala cycles for DK_AC-1 (left), DK_AC-2 (middle), and DK_KA (right).

Figure 14: Svara salience histograms across tala cycles for BP_AC-1 (left), BP_AC-2 (middle), and BP_RK (right).
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 15. 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 in course of the vistar, as shown in Figure 11 in (Widdess, 2011) 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.

Figure 15: Time-normalized summary of cycle-level salient svara curves over the chosen 6 concerts in raga Deshkar (left) and raga Bhupali (right).

Similar observations were reported by Ganguli et al. (2016) where the authors computed the svara salience histograms at the time scale of the breath phrase 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 15 was observed. The visual similarity among Figure 15 in the current work, Figure 11 in (Widdess, 2011), and Figure 4 in (Ganguli et al., 2016) indicates that performers stick to broad schemata of progressing in the melody from a lower to a higher svara (and swiftly returning to a lower svara to mark the intended ending), at different time-scales (tala cycle, fixed interval of a minute, and breath phrase).

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 the analyses of the performances of the 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 on the practical realization of a raga's distinctiveness in performance.

Tonal hierarchy can be estimated from the detected melody line of a performance recording. 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. The experiment implementation involved maximizing 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 the considered distinct representations of tonal hierarchy. Next, we considered a variety of distance measures in order to derive a combination of histogram parameters and distance metric that best separated same-raga pairs from the allied-raga pairs of concerts.

It was found that svara salience histograms worked best at the time-scale of full concerts whereas the finer bin resolution of 25 cents in the case of the continuous pitch salience histogram served to capture raga grammar better for the segmented shorter portions of concerts. In the latter, the pitch distributions between
the main peaks contributed usefully to the discrimination indicating the importance of continuous melodic phrase shapes. 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.

Interesting insights into the practical realization in performance of the musicological concepts of raga delineation and melodic progression at the concert time scale were obtained by applying the developed computational model to performance data. 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 concerts in allied ragas.

In this work, we gainfully used the allied raga performance as the ungrammatical realization of a given raga. A more direct, but considerably more challenging, validation of the proposed computational model would involve relating the predicted ungrammaticality of a performance to the ungrammaticality actually perceived by expert listeners. Future work must also address the modeling of the structural aspect of raga grammar, corresponding to the phrases, since this is the more easily accessed cue to raga identity by 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.

Acknowledgements
Omitted for blind review…

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., and 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 (JNMR), 37(1):1–13.

Castellano, M. A., Bharucha, J. J., and Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. Journal of Experimental Psychology: General, 113(3):394.

Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. City, 1(2):1.

Chordia, P. and Rae, A. (2007). Raag recognition using pitch-class and pitch-class dyad distributions. In Proc. of the International Society for Music Information Retrieval (ISMIR), pages 431–436.

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

Danielou, A. (2010). The ragas of Northern Indian music. Munshiram Manoharlal Publishers.

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

Dighe, P., Karnick, H., and Raj, B. (2013). Swara histogram based structural analysis and identification of Indian classical ragas. In Proc. of the International Society for Music Information Retrieval (ISMIR), pages 35–40.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8):861–874.

Fujishima, T. (1999). Realtime chord recognition of musical sound: a system using common Lisp music. In Proc. of International Computer Music Conference (ICMC), pages 464–467. Beijing. International Computer Music Association.

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

Ganguli, K. K. and Rao, P. (2017). Towards compu-
tational modeling of the ungrammatical in a raga performance. In Proc. of the International Society for Music Information Retrieval (ISMIR). Suzhou, China.

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

Ghosh, P. K. and Narayanan, S. S. (2009). Pitch contour stylization using an optimal piecewise polynomial approximation. IEEE Signal Processing Letters, 16(9).

Gomez, E. (2006). Tonal description of polyphonic audio for music content processing. INFORMS Journal on Computing, 18(3):294–304.

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

Gulati, S., Serra, J., Ganguli, K. K., Senturk, S., and Serra, X. (2016). Time-delayed melody surfaces for raga recognition. In Proc. of the International Society for Music Information Retrieval (ISMIR), pages 751–757, New York, USA.

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/Distinguishing-between-Similar-Ragas. Last accessed: October 26, 2017.

Koduri, G. K., Gulati, S., and Rao, P. (2011). A survey of raga recognition techniques and improvements to the state-of-the-art. Sound and Music Computing.

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

Koduri, G. K., Ishwar, V., Serra, J., and Serra, X. (2014). Intonation analysis of ragas in Carnatic music. Journal of New Music Research, 43(1):72–93.

Krumhansl, C. L. (1990). Cognitive Foundations of Musical Pitch, pages 77–110. Oxford University Press, New York. Chapter 4: A key-finding algorithm based on tonal hierarchies.

Krumhansl, C. L. and Cuddy, L. L. (2010). A Theory of Tonal Hierarchies in Music, pages 51–87. Springer New York.

Krumhansl, C. L. and 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.

Kulkarni, S. (2011). Shyamrao Gharana, volume 1. Prism Books Pvt. Ltd.

Kumar, V., Pandya, H., and Jawahar, C. V. (2014). Identifying ragas in Indian music. In International Conference on Pattern Recognition, pages 767–772. 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 Proc. of the International Society for Music Information Retrieval (ISMIR), pages 115–120.

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

Raja, D. (2016). The Raga-ness of Ragas: Ragas Beyond the Grammar. D.K. Print World Ltd.

Rao, P., Ross, J. C., and 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., and Murthy, H. A. (2014). Classification of melodic motifs in raga music with time-series matching. Journal of New Music Research (JNMR), 43(1):115–131.

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

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

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

Ross, J. C., Mishra, A., Ganguli, K. K., Bhattacharyya, P., and Rao, P. (2017). Identifying raga similarity through embeddings learned from compositions' notation. In *Proc. of the International Society for Music Information Retrieval (ISMIR)*, Suzhou, China.

Ross, J. C., Vinutha, T. P., and Rao, P. (2012). Detecting melodic motifs from audio for Hindustani classical music. In *Proc. of the International Society for Music Information Retrieval (ISMIR)*, Porto, Portugal.

Serra, X. (2011). A multicultural approach to music information research. In *Proc. of the International Society for Music Information Retrieval (ISMIR)*, pages 151–156.

Serra, X. (2014). Creating research corpora for the computational study of music: the case of the CompMusic project. In *Proc. of the 53rd AES International Conference on Semantic Audio*, London, UK.

Smith, N. A. and Schmuckler, M. A. (2000). Pitch-distributional effects on the perception of tonality. In *Proc. of the International Conference on Music Perception and Cognition (ICMPC)*, pages 5–10.

Smith, N. A. and 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.

Srinivasamurthy, A. and Serra, X. (2014). A supervised approach to hierarchical metrical cycle tracking from audio music recordings. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5217–5221.

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 re-considered. *Music Perception: An Interdisciplinary Journal*, 17(1):65–100.

Temperley, D. and Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception: An Interdisciplinary Journal*, 25(3):193–212.

van der Meer, W. (1980). *Hindustani Music in the 20th Century*. Martinus Nijhoff Publishers.

Vidwans, A., Ganguli, K. K., and Rao, P. (2012). Classification of Indian classical vocal styles from melodic contours. In *Proc. of the 2nd CompMusic Workshop*.

Vijaykrishnan, K. G. (2007). *The Grammar of Carnatic Music*. De Gruyter Mouton.

Widdess, R. (1994). Involving the performers in transcription and analysis: a collaborative approach to Dhrupad. *Ethnomusicology*, 38(1):59–79.

Widdess, R. (2011). Dynamics of melodic discourse in Indian music: Budhadiya mukherjee’s alap in rag Puriya-Kalyan. pages 187–224. Oxford University Press.

Widdess, R. (2013). Schemas and improvisation in Indian music. Kempson, Ruth and Howes, Christine and Orwin, Martin, (eds.), Language, Music and Interaction. London: College Publications.
### Appendix

#### Metadata description of the dataset

| Raga       | Audio ID | Artist       | Tonic in Hz | Tala (Laya)       | Duration in min |
|------------|----------|--------------|-------------|------------------|-----------------|
| Deshkar    | DK_AC-1  | Ajoy Chakrabarty | 146.8       | Ikwai (Madh)     | 21:25           |
|            | DK_AC-2  | Ajoy Chakrabarty | 138.5       | Teental (Madh)   | 16:22           |
|            | DR_KA    | Kishori Amonkar  | 233.0       | Teental (Vil)    | 29:55           |
|            | DR_RK    | Rashid Khan      | 155.5       | Teental (Madh)   | 21:54           |
|            | DK_VK    | Venkatesh Kumar  | 155.5       | Jhaptal (Madh)   | 20:45           |
|            | DK_UK    | Ulhas Kashalkar  | 155.5       | Teental (Madh)   | 16:29           |
| Bhupali    | BP_AC-1  | Ajoy Chakrabarty | 130.7       | Ektal (Vil)      | 33:46           |
|            | BP_AC-2  | Ajoy Chakrabarty | 146.8       | Ektal (Vil)      | 37:57           |
|            | BP_RK    | Rashid Khan      | 155.5       | Teental (Madh)   | 22:21           |
|            | BP_OD    | Omkar Amonkar    | 155.5       | Teental (Vil)    | 55:46           |
|            | BP_DVP   | D V Paluskar     | 158.7       | Ektal (Vil)      | 28:11           |
|            | BP_BGAK  | B Ghulam Ali Khan| 140.7       | Ektal (Vil)      | 23:59           |
|            | BP_ARK   | Abdul Rashid Khan| 137.3       | Teental (Madh)   | 17:30           |
|            | BP_KA-1  | Kishori Amonkar  | 236.9       | Teental (Vil)    | 29:40           |
|            | BP_KA-2  | Kishori Amonkar  | 239.9       | Teental (Vil)    | 21:34           |
|            | BP_KG    | Kumar Gandharv   | 192.2       | Teental (Madh)   | 16:37           |
|            | BP_SK    | Sameehan Kashalkar| 146.8      | Tillwada (Vil)   | 25:02           |
| Puriya     | PR_RSM   | Rajan Sajan Mishra | 155.5      | Ektal (Vil)      | 20:03           |
|            | PR_J     | Jasraj          | 138.5       | Ektal (Vil)      | 15:44           |
|            | PR_BJ    | Bhumse Joshi    | 155.5       | Teental (Madh)   | 18:17           |
|            | PR_DVP   | D V Paluskar    | 155.5       | Ektal (Vil)      | 28:40           |
|            | PR_AK    | Amir Khan       | 146.8       | Ektal (Vil)      | 17:30           |
|            | PR_JMS   | Ratan Mohan Sharma| 138.5      | Ektal (Vil)      | 19:23           |
|            | PR_BMJ   | Brajeswar Mukherjee| 146.8       | Ektal (Vil)      | 30:49           |
| Marwa      | MR_VD    | Vasantrao Deshpande | 155.5    | Ektal (Vil)      | 18:57           |
|            | MR_BGAK  | B Ghulam Ali Khan| 138.5       | Ektal (Vil)      | 16:32           |
|            | MR_AC    | Ajoy Chakrabarty | 146.8       | Teental (Madh)   | 21:31           |
|            | MR_RK-1  | Rashid Khan      | 155.5       | Ektal (Vil)      | 25:52           |
|            | MR_RK-2  | Rashid Khan      | 155.8       | Ektal (Vil)      | 24:06           |
|            | MR_OD    | Omkar Amonkar    | 155.5       | Ektal (Vil)      | 32:17           |
|            | MR_AK    | Amir Khan        | 138.5       | Jhuma (Vil)      | 18:29           |
|            | MR_CRV   | C R Vyas         | 157.9       | Jhaptal (Madh)   | 14:58           |
|            | MR_HB    | Hirabai Barodekar| 210.6      | Teental (Madh)   | 13:32           |
|            | MR_J     | Jasraj          | 139.7       | Ektal (Vil)      | 27:44           |
|            | MR_SJ    | Srinivas Joshi   | 152.2       | Ektal (Madh)     | 17:02           |
|            | MR_SK    | Sameehan Kashalkar| 146.8      | Ektal (Vil)      | 33:59           |
| Multani    | ML_RSM   | Rajan Sajan Mishra | 155.5    | Ektal (Vil)      | 29:56           |
|            | ML_AC    | Ajoy Chakrabarty | 146.8       | Ektal (Vil)      | 21:05           |
|            | ML_KC    | Kaushiki Chakrabarty| 233.0    | Teental (Madh)   | 25:04           |
|            | ML_HB    | Hirabai Barodekar| 207.0      | Ektal (Vil)      | 13:01           |
|            | ML_BJ    | Bhumse Joshi     | 146.8       | Ektal (Vil)      | 19:07           |
|            | ML_DV    | Omkarnath Thakur | 146.8       | Ektal (Vil)      | 14:42           |
|            | ML_DOD   | Omkar Amonkar    | 155.5       | Teental (Vil)    | 27:52           |
| Todi       | TD_OT    | Omkarnath Thakur | 155.5       | Ektal (Vil)      | 16:50           |
|            | TD_HB    | Hirabai Barodekar| 207.0      | Ektal (Vil)      | 13:36           |
|            | TD_VS    | Veena Sahasrabuddhe| 220.0     | Teental (Madh)   | 19:48           |
|            | TD_KA    | Kishori Amonkar  | 207.0       | Ektal (Vil)      | 29:23           |
|            | TD_BJ    | Bhumse Joshi     | 146.8       | Ektal (Vil)      | 19:06           |
|            | TD_J     | Jasraj          | 138.5       | Ektal (Vil)      | 20:05           |
|            | TD_AK    | Amir Khan       | 139.4       | Jhuma (Vil)      | 27:34           |
|            | TD_KG    | Kumar Gandharv  | 186.3       | Jhuma (Vil)      | 32:01           |
|            | TD_KKG   | Kaivalya K Gurav| 174.4       | Ektal (Vil)      | 29:03           |
|            | TD_OD    | Omkar Amonkar    | 153.1       | Teental (Vil)    | 17:36           |
|            | TD_PA    | Prabha Atre     | 237.8       | Ektal (Vil)      | 26:14           |
|            | TP_BRM   | Ritik Rajnish Mishra| 156.0     | Rupak (Madh)     | 28:39           |

Table 4: Metadata associated with each concert in our dataset. We list the artist, tonic, tala, laya, and duration of each concert. The acronym correspondences are: Vil ≡ vilambit, Madh ≡ madhyalaya.

### Raga-pair Puriya-Marwa

We carry out the same set of experiments for representation, distance measures for the Puriya-Marwa allied raga-pair. The correspondence of the figures that describe the tonal hierarchy representation is as follows: Figure 4
From the pitch salience histogram (Figure 16) of raga Puriya (left), we observe the svaras G and N as most prominent, which corresponds to the *vadi* and *samvadi* svaras. However, the peak height for the *vadi* G being less than that for the *samvadi* N is a result of octave-folding. N has more salience (accumulated duration) because of its presence in both lower and middle octaves, whereas G had most of its occurrences in the middle octave only. The r and D svaras have very low salience, which reinforces the grammaticality in the eminent artists’ performance. The r peak is almost merged with the S peak, this is indicative of the fact that r is taken only within the glide (e.g. N(r)S, N(r)G, G(r)S) and is never sustained as an individual svara. In contrast, the r svara has the highest salience for raga Marwa (right), which also happens to be its *vadi* svara. The second most salience peak corresponds to the *samvadi* D svara.

**Figure 16:** Pitch salience histograms (octave folded, 1 cent bin resolution) of 7 concerts in raga Puriya (left) and 12 concerts in raga Marwa (right).

**Figure 17:** Svara salience histograms (octave folded) of 7 concerts in raga Puriya (left) and 12 concerts in raga Marwa (right).

**Figure 18:** Svara count histograms (octave folded) of 7 concerts in raga Puriya (left) and 12 concerts in raga Marwa (right).
The interpretations from the svara salience histograms (Figure 17) are similar. There is a high visual similarity between the pitch salience and svara salience histograms for the corresponding ragas (ignoring the bin mapping due to difference in resolution), this supports the musicological knowledge that major portion of a phrase in course of the raga development is covered by stable svaras (Kulkarni, 2011). However, in comparison, we find an interesting observation on the relative salience of the r svara in the svara count histograms (Figure 18). In raga Puriya (left), the svara count is relatively high though the same has a very low salience in the svara salience histograms (Figure 17). This indicates a large number of detected r svaras that were of short durations, thereby accumulated to a low peak in the svara salience histogram. In contrast, the average duration of the r svaras in raga Marwa (right) are high, thereby accumulating to a high peak in the svara salience histogram for a moderate peak in the svara count.

The correspondence of the figures for the distance measures and time-scale is as follows: Figure 8 ≡ Figure 19. The trend of the shape of the ROCs (and relative differences of the AUC values) for the corresponding distance measures is observed to be same.

The trend of the shape of the ROCs (and relative differences of the AUC values) for the corresponding distance measures is observed to be same.

**Raga pair Multani-Todi**

Similarly, we perform the same experiments for the Multani-Todi allied raga-pair. The correspondence of the figures that describe the tonal hierarchy representation is as follows: Figure 4 ≡ Figure 20, Figure 5 ≡ Figure 21, Figure 6 ≡ Figure 22.

The pitch salience histogram (Figure 20) for raga Multani (left) has the salience of r and d svaras so feeble (and the peaks are merged with that of S and P respectively) that the tonal hierarchy almost resembles a pentatonic scale. This is also indicative of the fact that r and d svaras are taken only within the glide (e.g. N(r)S, N(d)P, M(g(r)S) and is never sustained as an individual svara. The vadi svara S and samvadi P, as hypothesized, are placed at the top two ranks in the tonal hierarchy. In contrast, the svaras in raga Todi (right) are distinctly visible in the tonal hierarchy. While the vadi svara d has a relatively high peak, samvadi g is lower down the order. The salience of r, again, is contributed by its presence in the middle and higher octaves.

The svara salience histogram (Figure 21) for raga Multani (left) also preserves the salience order of the S and P svaras. r and samvadi d svaras, in a similar way, have negligible peak heights. For raga Todi (right) also, visually, correlation (ignoring bin-mapping) between the pitch salience and svara salience histograms are high. The svara count histograms (Figure 22) are highly correlated with the svara salience histograms. For raga Multani (left), the count of the g svara is relatively height as one might expect from its salience in the svara salience histogram (Figure 21). This indicates presence of larger no. of detected g svaras, each of short duration.

The figure correspondence for the distance measures and time-scale is as follows: Figure 8 ≡ Figure 23. Our observation is that the dispersion in the svara counts per bin for these raga pairs are much higher as compared to the Deshkar-Bhupali raga-pair.
Figure 20: Pitch salience histograms (octave folded, 1 cent bin resolution) of 7 concerts in raga Multani (left) and 12 concerts in raga Todi (right).

Figure 21: Svara salience histograms (octave folded) of 7 concerts in raga Multani (left) and 12 concerts in raga Todi (right).

Figure 22: Svara count histograms (octave folded) of 7 concerts in raga Multani (left) and 12 concerts in raga Todi (right).

Figure 23: Raga-pair Multani-Todi (full concerts, octave-folded): ROCs obtained for four different distance measures from pitch salience (left), svara salience (middle), and svara count (right) histograms.