DISTINGUISHING RAGA-SPECIFIC INTONATION OF PHRASES WITH AUDIO ANALYSIS

Preeti Rao*, Joe Cheri Ross[Ŧ] and Kaustuv Kanti Ganguli*

Department of Electrical Engineering* Department of Computer Science and Engineering[Ŧ]
Indian Institute of Technology Bombay,
Mumbai 400076, India
{prao, kaustuvkanti}@ee.iitb.ac.in*
joe@cse.iitb.ac.in[Ŧ]

Abstract

The identity of a raga in a performance is exposed through the occurrence of characteristic melodic phrases or "catch phrases" of the raga. While a phrase is represented as a simple sequence of notes in written notation, the interpretation of the notation by a performer invokes his knowledge of raga grammar and thus the characteristic continuous pitch movement that represents the particular phrase intonation. A performer's skill is closely linked with his ability to render creatively and unambiguously the distinctive aspects of a raga including its phrases or melodic motifs. Similarly, listeners identify the raga by the characteristic melodic shape of the phrase. We present a study to validate implicit musical knowledge that raga-characteristic phrases are relatively invariant across concerts and artistes by means of a time series similarity measure applied to melodic pitch contours extracted from audio recordings of Hindustani classical vocal concerts. The measure, which can discriminate phrases, even with the same svara sequence, corresponding to different ragas, has applications in music retrieval. Further, it has potential in the objective evaluation of raga phrases rendered by a music learner.

1. Introduction

Along with the *svaras* that define the raga in Hindustani and Carnatic music, the identity of a raga in a performance is exposed through the occurrence of characteristic melodic phrases or *pakads* of the raga. For a given raga, there are predefined characteristic phrases which recur in both pre-composed as well as purely improvised sections of the raga performance [1,2]. Indian classical music, as is well known, is transferred through the oral-aural mode. Written notation is typically a sparse form, not unlike basic Western staff notation with pitch class and duration in terms of rhythmic beats specified for each note. When used for transmission, the notation plays a purely "prescriptive" role considering the actual melodic sophistication of the tradition. Thus the audio version, or sound, of the written notation is actually the performer's interpretation, and invokes his background knowledge including a complete awareness of raga-specified constraints.It may be noted that the interpretation involves not only supplying the volume and timbre dynamics but also the pitches "between notes" [3]. An important dimension of a raga's grammar is its phraseology, a set of phrases that provide the building blocks for melodic improvisation by collectively embodying its melodic grammar and thus defining a raga's "personality" [4]. The melodic phrases of a raga are typically associated with prescriptive notation which can comprise of as few as 2 or 3 notes, but often several notes.The recurrences of a raga-characteristic phrase in a concert can therefore be viewed as aesthetically controlledinterpretations marked by an inherent variability that makes them interesting (i.e. not sounding unnecessarily repetitious) but still strongly recognizable.The sequencing of phrases is embedded in the rhythmic accompaniment provided by the percussion but only loosely connected with the underlying beats except for the *tan*(fast tempo) sections in the case of the widely performed Hindustani *khayal* genre [4]. The sequence of phrases comprises a "musical statement" and spans a rhythmic cycle sometimes crossing over to the next. Boundaries between such phrase sequences are typically marked by the *sam*(first beat) of the rhythmic cycle supplied by the *tabla*.

In view of the central role played by raga-characteristic phrases in the performance of both the Indian classical traditions, computational methods to detect specific phrases in audio recordings can have important applications. Raga-based retrieval of music from audio archives can benefit from automatic phrase detection where the phrases are selected from a dictionary of phrases corresponding to each raga [5]. The same methods can potentially be extended to automatic music transcription of Indian classical music, which is a notoriously difficult task due to its interpretive nature. Widdess [6] describes the significant collaborative effort he needed with the performer of a *dhrupad* improvisation to achieve its transcription into Western notation. He mentions how the performer included certain notes in the transcription that were either short or otherwise not explicit because they were important to the raga(presumably belonging in the prescriptive notation of the phrase). The phrase-level labelingof audio, or simply its alignment with available symbolic notation, can be valuable in musicological research apart from providing for an enriched listening experience for music students.Finally, a robust measure for melodic distance can serve in the objective evaluation of music learners' skills.

There is limited previous work onthe audio based detection of melodic phrases in Hindustanimusic.Pitch class distributions and N-gram distributions from the sequence of detected "notes" obtained by heuristic methods of automatic segmentation of the pitch contour have been used in raga recognition (seereview in [7]). While Carnatic music pedagogy has developed the view of a phrase in terms ofa sequence of notes each with its own *gamaka* (movement in its pitch space), the Hindustani perspective of a melodic phrase is more of a gestalt. The various acousticrealizations of a given phrase formpractically a continuum of pitch curves, with overall raga-specific constraints on the timing and nature of the intra-phrase events. This makes explicit note segmentation of debatable value in phrase recognition forIndian classical vocal music.

Recently, the detection of the *mukhda*, a recurring melodic motif in Hindustani performances of *bandish*,was achieved with template matching of continuous pitch curves segmented by exploiting known timing alignment within the metrical cycle [8].However in the general case, there is no fixed correspondence between a characteristic phrase and underlying metrical structure especially in the *vilambit* and *madhya layas* (slow and medium tempos) in Hindustani *khayal* music. That characteristic phrases often terminate on a steady note (the resting note or *nyas svara* of the raga) was exploited to achieve the segmentation of phrase candidates to be used in a similarity based measure with pre-decided templates [9]. In both the above, classic DTW (dynamic time warping) was shown to be effective in measuring inter-phrase similarity via the non-uniform time warping thatrepresents permittedphrase variability across renditions. The present work generalizes thisapproach using a more challenging audio dataset that (1) contains concerts of the selected raga with different *laya* and *tala*across a larger number of artistes, and (2) includes a different ragawhere a certain phrase with the same prescriptive notation as a characteristic phrase of the previous raga happens to occur frequently. The objective is to investigate the possibility of discriminating the chosen characteristic phrase from other phrases within the raga as well as from instances of the identically notated segment from a different raga.The database also allows us to observe thepractically achievedintra-phrase-class variability of characteristic phrases versus that of non-characteristic phrases (note sequences).

In the next section, we review musicological material on raga phraseology and outline the challenges posed to computational methods. The database and evaluation methods are presented next. The audio processing and similarity computation follow. Experiments are presented followed by a discussion of the results and prospects for future work.

2. Raga and Phraseology

The 12 notes of the octave in Hindustani music are denoted as S r R g G m M P d D n N. A given raga is characterized by a set of *svara-s*, ascending and descending scale and its *pakad* (characteristic phrases). We present the relevant musicological concepts with reference to the ragas chosen for the present study, namely Alhaiya-bilawal and Kafi. The former is the most prominent raga of the Bilawal scale (corresponding to the Western major scale). As shown in Table 1, in addition to the 7 notes of the Bilawal scale, raga Alhaiya-Bilawal makes context-dependent use of n (flat-7th). As indicated in the set of characteristic phrases, noccurs during descent towards P, and always between two D [2]. Raga Kafi, whose description also appears in Table 1, is used in this work primarilyas an "anti-corpus", i.e. to provide examples of note sequences that match the prescriptive notation of achosen characteristic phrase of Alhaiya-Bilawal but in a different raga context (and hence are not expected to match in melodic shape, or intonation).

| *Raga* Characteristics	Alhaiya Bilawal	Kafi
Tone Material	S R G m P D n N	S R g m P D n
Characteristic Phrases	G~ R G /P (GRGP) D~ n D \P (DnDP) D \G G m R G P m G	R g R– m m P g- m P m P D m n\P g R S n \P g R
Comments	'n' is used only in the descent, and always in between the two 'D'-s as D n D P	Movements are flexible and allow for melodic elaboration

Table 1. Raga descriptionsadapted from [raga guide, autrim]. The characteristic phrases are provided in the reference in enhanced notation including ornamentation. The prescriptive notation for the phrases used for the present study appears in parantheses.

Apart from the specified sequence of *svara-s*, a characteristic phrase satisfies raga-dependent constraints on the manner of intonation (the timing, duration and linking of the notes). The continuous pitch variation with time, within the phrase, may be viewed as arising from the sequence of *svara-s* (notes) represented by their respective durations and *gamaka-s*, that is, the approach and movement within the "space" of the corresponding *svara*. This "phrase intonation" can be considered as the descriptive representation and an elaboration of the prescriptive notation under the constraints of the raga grammar [10]. The context of the phrase itself (preceding and succeeding phrases) determines how it is extended on either side, and provides a further definition of the phrase in the context of the specific raga. The same phrase intoned in different ways can suggest a different raga [11]. Hence the flexibility available in the interpretation of the phrase varies from raga to raga, and phrase to phrase, driven by the need to retain its distinctiveness and minimize confusion with similar characteristic phrases of other ragas, something that a skilled performer strives to achieve.

Fig. 1 shows some representative pitch contours (computed as described in Sec. 4) for DnDP phrases in various melodic contexts selected from different concerts in our Alhaiya-bilawal database (Table 2). The contexts correspond to the two possibilities: approach from higher *svara* (S or N), and approach from lower svara. In both cases there is a passage through n, consistent with the raga theory. The vertical lines mark the rhythmic beat (*matra*) locations. The phrases are chosen from medium (AB, SS) and slow (MA) tempo performances in the same *tala*. We consider the phrase duration, indicated by the dark bars, as spanning from D-onset (or rather, from the offset of the preceding n) to P-onset (the finalPis a resting note and therefore of unpredictable duration itself). We observe the similarity in melodic shape across realisations. Prominent differences are obvious too such as the presence or absence of n as a touch note (*kan*) in the final DP transition and varying extents of oscillation on the first D. These may be attributed to the flexibility accorded by the raga grammar in improvisation. Consistent with musicological theory on *khayal* music at slow and medium tempos, (i) there is no observable dependence of phrase duration upon beat duration, (ii) relative note durations are not necessarily maintained across tempos, and (iii) the note onsets do not necessarily align with beat instants except for the *nyas svara*, considered an important note in the raga.
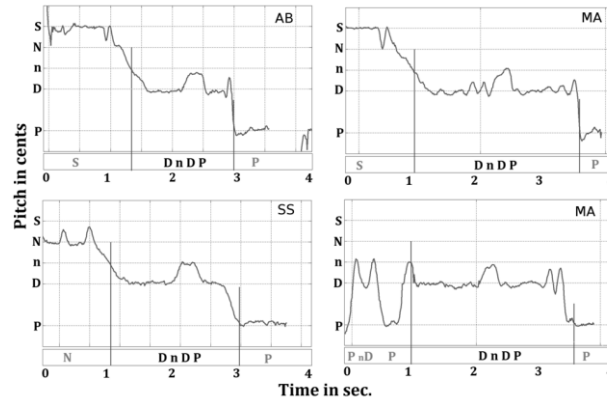
Figure 1. Pitch contours (cents vs time) of Alhaiya-Bilawal DnDP phrases in various melodic contexts by different artistes. Horizontal lines mark svara positions. Thin vertical lines mark beat instants. Thick lines mark the phrase boundaries used in the similarity computation.
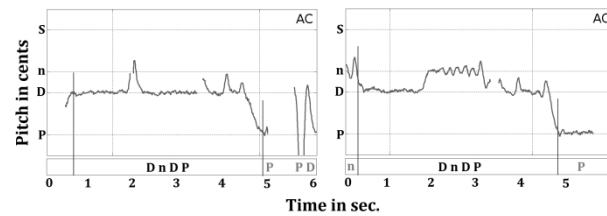
Figure 2. Pitch contours extracted from the Kafi concert described by DnDP notation.

From the computational perspective, we need a melodic representation that captures the essential invariance of the melodic shape while discounting raga-permitted variations. From the previous observations, we note that conventional quantization of pitch and temporal dimensions will not work. Rather the pitch-continuous form of the phrase realization suggests a time-series similarity measure. Considering the non-uniform time-warping that accompanies phrase duration changes, a DTW based distance would be more appropriate. Phrase segmentation too cannot exploit rhythmic structure, and other cues to phrase boundaries linked to the perception of closure that a melodic phrase induces must be sought.

3. Database and Annotation

The concert sections chosen for this study are obtained from one of [1,2] and from commercial CDs. They represent a diversity of well-known contemporary Hindustani *khayal* artistes, compositions, *tala* and *laya* as displayed in Table 2. In all cases, the accompanying instruments are the *tanpura* (drone), *harmonium (sarangi* in one case*)* and *tabla*. The section of each concert corresponding to *bandish* and *vistar* (raga elaboration by improvisation) is extracted.

Song ID	Artiste	*Tala*	*Laya*	*Bandish*	Tempo (bpm)	Dur. (min)	DnDP Char.	DnDP Seq.	mnDP	GRGP
AB	Ashwini Bhide	Tintal	Madhya	Kavana Batariyaa	128	8.85	13	2	31	5
MA	Manjiri Asanare	Tintal	Vilambit	Dainyaa Kaahaan	33	6.9	12	1	13	6
SS	Shruti Sadolikar	Tintal	Madhya	Kavana Batariyaa	150	4.15	3	0	14	3
ARK	Abdul Rashid Khan	Jhaptal	Madhya	Kahe Ko Garabh	87	11.9	44	0	0	14
DV	Dattatreya Velankar	Tintal	Vilambit	Dainyaa Kaahaan	35	18.3	14	4	4	9
JA	Jasraj	Ektal	Vilambit	Dainyaa Kaahaan	13	22.25	19	18	0	29
AK-1	Aslam Khan	Jhumra	Vilambit	Mangta Hoon Tere	19	8.06	10	0	8	6
AK-2	Aslam Khan	Jhaptal	Madhya	E Ha Jashoda	112	5.7	7	0	0	3
AC	Ajoy Chakrabarty	Jhumra	Vilambit	Jago Man Laago	24	30.3	27		0	---

Table 2. Description of database with phrase counts in the musician's transcription; all concerts are in raga Alhaiya-bilawal except the final one (AC) in raga Kafi. "Char." = characteristic of the raga; "Seq." = note sequence.

Manual annotation of selected phrases was carried out by a musician (one of the authors), and validated by a second musician (outside the group). The annotation was based on listening, and marking on the waveform, phrase labels and approximate boundaries using the PRAAT audio interface. The prescriptive notation of each P-*nyas* ending phrase in the audio was provided by the musician based on his knowledge of the raga. The counts per concert of frequently occurring labeled phrases appears in Table 2. GRGP and DnDP are characteristic phrases in raga Alhaiya-Bilawal while mnDP is part of the chosen compositions (*bandish*) and appears frequently as the *mukhda*. Further, the DnDP occurrences in the Alhaiya-Bilawal concerts were separated based on whether they captured the raga identity with only the immediately preceding context. Those that did not were termed non-characteristic phrases or just "note sequences". It was observed that the JA concert had a large number of non-characteristic DnDP owing to influences from a particular singing style. It

was also noted that listeners use some of the preceding context (previous note) to reliably detect raga-characteristic DnDP. The Kafi concert was annotated for all note sequences corresponding to the notation DnDP.

4. Audio Processing and Experiments

Vocal pitch detection is achieved at 10 ms intervals throughout the vocal regions of the audio signalwith a predominant-F0 detection method as described in [8, 12].Any gaps in the pitch contour of a phrase segment due to silence or unvoiced sounds are linearly interpolated. The pitch is normalized with respect to the tonic to obtain the melodic contour.Next phrase segmentation is carried out on the melodic contour in a semi-automatic manner. The coarsely annotated ground-truth phrase segments are searched for exact segment boundaries corresponding to P-onset and D-onset in the region between 1.2 and 3 seconds prior to P-onset(based on the observed range of phrase duration). In the case of DnDP, the phrase boundaries are marked as shown in Fig. 1 i.e. offset of the n, and onset of the P-nyas. An onset or offset is reliably detected by hysteresis thresholding with thresholds of the 50 and 20 cents within the nominal *svara* pitch value. The similar segmentation is applied to the other P-nyas phrases.

Figures 1 and 2 show a few representative pitch contour segments from each raga with the DnDP phrase indicated between thick vertical markers.From the beat instant markers, we note that the MA concert tempo is low relative to the others. However the phrase durations do not appear to scale in the same proportion. It was noted that across the concerts, tempos span a large range (as seen in Table 2) while the maximum duration of the DnDP phrase in any concert varies between 1.1 to 2.8 sec only with considerable variation within the concert. Further, any duration variations are not linearly related. It is observed that the n-duration is practically fixed while duration changes are absorbed by the D svara on either side.There was no observable dependence of phrase intonation on the *tala*.Apart from these and other observations from Fig. 1 (listed in Sec. 2), we note that the Fig. 2 Kafi phrases (in which ragaDnDP is not a characteristic phrase) display a greater variability in phrase intonation while conforming to the prescriptive notation of DnDP.

DTW is used directly on the segmented time-series to account for non-uniform time scaling. The phrase segments are each linearly interpolated to a constant duration of 1.3 sec (a value arbitrarily selected within the observed duration range of the phrase) to compute a duration–normalized DTW distance, after zero-padding at each end to absorb any boundary frame mismatches. Classical DTW is used with its monotonicity and single increment step-size conditions [13]. We introduce a condition to bias the DTW path towards the diagonal by ignoring differences less than a quarter semitone in the local cost function. However the distance computation itself includes the actual differences of aligned pitches. This approach was found to be effective in limiting pathological warps while still accounting for all pitch differences.We present next experiments on phrase similarity on our database.

4.1 Experiment 1: Intra-phrase-class similarity

We wish to compare the pitch curve similarity across DnDP phrases of the raga-characteristic class with that across the non-characteristicDnDP phrase class. The DTW distance measure as described above is applied to all pairs created from two distinct DnDP phrases drawn from the same class. If implemented for each concert audio, this could tell us something about the variability of the DnDP phrase in that concert. To compensate for the shorter durations of some of the concerts however, we group the 8 Alhaiya-Bilawal concerts in Table 2 into 4 sets. The resulting distribution of pairwise phrase distances for each set is shown in Fig. 3(a). The number of pairs is given by N(N-1) where N is the count of DnDP phrases in that set (e.g. N=17 for AK1+AK2). Also shown is the distance distribution created out of DnDP phrase pairs from raga Kafi (where it is not a characteristic phrase).
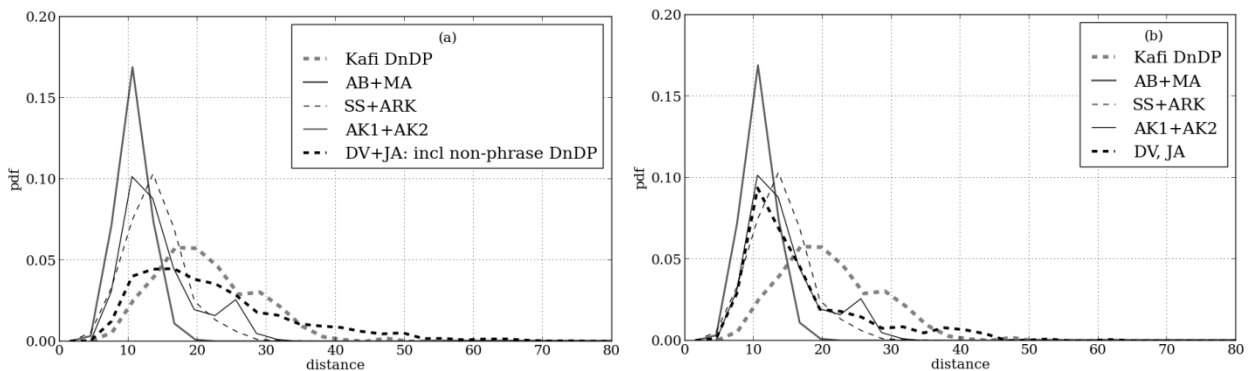
Figure 3. Intra-phrase-classdistance distributions for the different concerts listed in Table 2. (a) All DnDP sequences included (b) Non-characteristic DnDP excluded in the Alhaiya-Bilawal concerts.

4.2 Experiment 2: Phrase detection

In a retrieval scenario, we would like to identify the raga based on the detection of its characteristic phrases in unknown audio based on previously stored templates of the phrase. We simulate this situation by using the first two concert audios to source templates for all other occurrences of the raga-characteristic DnDP in Table 2. Two templates are chosen to capture the variability of the phrase. These are manually identified by visual inspection of the pitch contour as representingdistinct melodic shapes from among the set of 25 phrases from the AB and MA audios. In future, vector quantization can be explored for the template selection task.

In order to obtain sufficient data to measure retrieval accuracy, we carry out the experiment over 3 distinct template sets. Each candidate test phrase is compared with both the reference templates, and the minimum distance is retained as that corresponding to the test phrase. A Sakoe-Chiba constraint is applied in the DTW distance computation to discourage pathological warpings of mismatched phrases. A Sakoe-Chiba width of 25% is chosen based on the observation that this includes most paths obtained in pair-wise distance computation involving the true raga-characteristic DnDP phrases within AB and MA audios.

5. Results and Discussion

Fig. 3(a) shows that the intra-phrase-class distances in all Alhaiya-Bilawal concerts (except DV+JA) are narrowly dispersed about a mean value close to 12.0, indicating the low variability in the intonation of the raga-characteristic phrase across the concert. In contrast, the Kafi raga distribution has a mean near 20.0 and higher standard deviation, implying greater variability in DnDPintonation. These observations are consistent with musicological knowledge about the strictness that applies to the melodic shape of raga-characteristic phrases as opposed to that of non-characteristic phrases. The DV+JA concerts show a relatively greater spread due to the presence of non-characteristic DnDP (as indicated in Table 2). When these phrases are eliminated from the computed distances, we have the more concentrated distribution for DV+JA in Fig 3(b). Thus we see that the mean and spread of the inter-phrase distance distribution clearly captures the raga characteristics with respect to the given phrase.
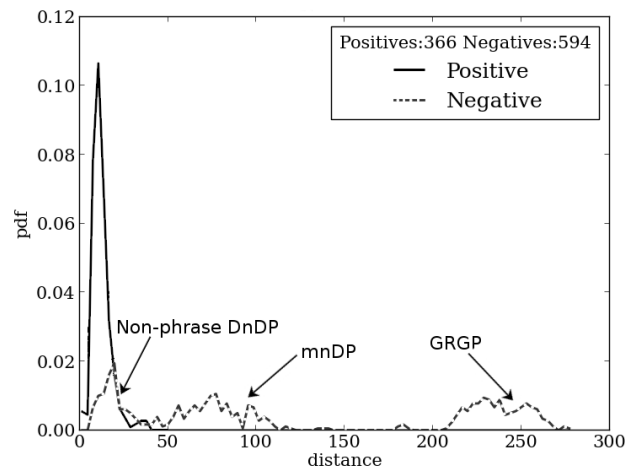
Figure 4. Distributions of distances of P-nyas phrases from the DnDP raga-characteristic templates. Bold: raga-characteristic DnDP; dashed: all other phrases.

The results of Experiment 2 appear in Fig. 4 as the distribution of distances of all candidate phrases (i.e. P-nyas phrases in Table 2) from the selected template DnDP phrases. The ground-truth raga-characteristic DnDP distribution appears in bold line. We see that the spread is very narrow relative to that of the remaining phrases' distance distribution. The latter distribution has 3 distinct modes that were observed to correspond to GRGP (most distant), mnDP and non-characteristic DnDP (near-overlapping). This trend can be explained by the expected greater dissimilarity between the GRGP and DnDP pitch contours where only one svara is common. From Fig. 4, we obtained a hit-rate of **0.85** for the detection of raga-characteristic DnDP at a false-alarm rate of **0.1** with most false alarms coming from the non-characteristic DnDP. The number of raga-characteristic phrase templates was limited to two here. Considering the variety of realizations observed in practice, an improved detection performance is expected with a slightly larger codebook of phrase templates that possibly also represent the preceding context explicitly.The codebook may be systematically derived by using vector quantization methods on a ground-truth characteristic phrases dataset.

6. Conclusion

A raga-characteristic phrase is easily identified by the knowledgeable listener, across concerts of the same raga across artistes and compositions of different *tala* and *laya*. Experiments on a limited corpus showed that the continuous pitch curve can be robustly discriminated, using a suitable time-series measure, from pitch curves of other phrases of the same raga and, more significantly, from those corresponding to the same-notation phrase in a different raga. The distribution of the inter-phrase distances within a selected phrase-class can serve as an indicator of raga identity.

Considering that the flexibility available to the artiste in rendering a raga-characteristic phrase is driven partly by the need to retain its distinctiveness with respect to other ragas, it would be worthwhile to learn the melodic representation by training on a larger corpus of the same raga including an anti-corpus of different ragas where the similar phrase occurs. Such "discriminative training" could help in the robust detection of deviations from raga grammar in the objective evaluation of singing skill. Apart from fully automating phrase segmentation, further work should incorporate volume and timbre dynamics so that the exploration of the constancy of raga-characteristic phrases in view of the flexibility available to performers in the improvisatory framework of Hindustani classical music can be more complete. Such work can be invaluable to music students, listeners and to musicologists with its potential to achieve the empirical testing of the "implicit knowledge of experts" [14].

7. References

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

2. S. Rao and W. van der Meer: "Music in Motion: The Automated Transcription for Indian Music,"[online]. Available: http://autrimncpa.wordpress.com/

3. N. Cook, *Music, A Very Short Introduction*, OUP, Oxford, 1998

4. D. Raja, Hindustani Music, *A Tradition in Transition*, D.K. Printworld (P) Ltd., 2005.

5. J. Chakravorty, B. Mukherjee and A. K. Datta: "Some Studies in Machine Recognition of Ragas in Indian Classical Music," *Journal of the Acoust. Soc. India*, Vol. 17, No.3&4, 1989.

6. R.Widdess, "Involving the Performers in Transcription and Analysis: A Collaborative Approach to Dhrupad,"*Ethnomusicology*, Vol. 38, no. 1, 1994.

7. G. Koduri, S. Gulati, P. Rao and X. Serra, "Raga Recognition based on Pitch Distribution Methods," *Journal of New Music Research*, Vol. 41, No.4, 2012.

8. J. Ross, T.P. Vinutha and P.Rao: "Detecting Melodic Motifs From Audio For Hindustani Classical Music," *Proc. ofInt. Soc. for Music Information Retrieval Conf. (ISMIR),* 2012.

9. J. C. Ross and P. Rao, "Detection of Raga-Characteristic Phrases from Hindustani Classical Music Audio,"*Proc. of 2nd CompMusic Workshop*, Istanbul, 2012.

10. S.K. Subramanian, L. Wyse, K. McGee, "A Two-component Representation for Modeling Gamakas of Carnatic Music," *Proc. of 2nd CompMusic Workshop*, Istanbul, 2012.

11. M. Narmadha, *Indian Music and Sancharas in Ragas*, Vedic Books, 2001.

12. V. Rao and P. Rao: "Vocal Melody Extraction in the Presence of Pitched Accompaniment in Polyphonic Music," *IEEE Trans. Audio Speech and Language Processing*, Vol. 18, No.8, 2010.

13. M. Müller, "Dynamic Time Warping,"*Information Retrieval for Music and Motion,* pp. 69-84,Springer, 2007.

14. A. Volk and A. Honingh, "Mathematical and computational approaches to music: challenges in an interdisciplinary enterprise,"*Journalof Mathematics and Music,* Vol. 6-2, 2012.