# Classification of Melodic Motifs in Raga Music with Time-series Matching

Preeti Rao*, Joe Cheri Ross*, Kaustuv Kanti Ganguli*, Vedhas Pandit*, Vignesh Ishwar#, Ashwin Bellur#, Hema Murthy#

Indian Institute of Technology Bombay*

Indian Institute of Technology Madras#

This is an Author's Original Manuscript of an Article whose final and definitive form, the Version of Record, has been published in the Journal of New Music Research, Volume 43, Issue 1, 31 Mar 2014, available online at: http://dx.doi.org/10.1080/09298215.2013.831109 [dx.doi.org]

**Abstract**

Ragas are characterized by their melodic motifs or catch phrases that constitute strong cues to the raga identity for both, the performer and the listener, and therefore are of great interest in music retrieval and automatic transcription. While the characteristic phrases, or *pakads*, appear in written notation as a sequence of notes, musicological rules for interpretation of the phrase in performance in a manner that allows considerable creative expression, while not transgressing raga grammar, are not explicitly defined. In this work, machine learning methods are used on labeled databases of Hindustani and Carnatic vocal audio concerts to obtain phrase classification on manually segmented audio. Dynamic time warping and HMM based classification are applied on time series of detected pitch values used for the melodic representation of a phrase. Retrieval experiments on raga-characteristic phrases show promising results while providing interesting insights on the nature of variation in the surface realization of raga-characteristic motifs within and across concerts.

**Keywords:** raga-characteristic phrases, melodic motifs, machine learning

## 1 Introduction

Indian music is essentially melodic and monophonic (or more accurately, heterophonic) in nature. In spite of this apparent simplicity, it is considered a highly evolved and sophisticated tradition. Melodic and rhythmic complexities compensate for the absence of the harmony, texture and dynamics, attributes that play a vital role in the aesthetics of Western music. Ragas form the cornerstones of melody, and correspondingly, *talas* of rhythm. The raga and *tala* framework is common to Hindustani and Carnatic classical music, serving both for composition and improvisation. Ragas are thought to have originated in folk and regional songs, which explains the nature of a raga as lying somewhere between a modal scale and a tune (Powers & Widdess, 2001). That is, the essence of a raga is captured in a set of melodic phrases known as the *pakad* ("catch phrases")

widely used in both compositions in the raga, and in improvisation where the identity of the raga is revealed by the artistethrough the use of the raga's melodic motifs. The set of catch-phrases form the building blocks for melodic improvisation by collectively embodying the raga's melodic grammar and thus defining a raga's "personality" (Raja, 2005).

It was the desire to rationalise musical forms in the 15th century that led to the development of a raga grammar for Carnatic music in terms of specifying explicitly the tones (*svara-s*) forming the scale, their hierarchy (*vadi, samvadi, nyas* etc.), ascending (*aroha*) and descending (*avaroha*) patterns and general progression. A partial adoption of this approach by Bhatkande in the early 20th century gave rise to the ten *that-s* of Hindustani music (Rao & Rao, 2013). The phrases of a raga are viewed as a sequence of expressive *svaras*. The expression (pitch movement) that modulates a *svara* is its*gamaka* in Carnatic music terminology. The breaking down of the melodic phrase into *svaras*and *gamakas* has been part of the exercise of understanding Carnatic music (Krishna & Ishwar, 2012). The*gamaka* have been independently categorized into 15 basic shapes. The melodic context of the *svara*within the phrase, along with the aesthetics of the genre and the raga, dictate the choice of the *gamaka*. The *shruti* or the "microtonal" perceived pitch of a *svara*seems to be related to the pitch inflections (Rao & Rao, 2013). The "phrase intonation" therefore can be considered an elaboration of the prescriptive notation (sequence of *svaras*of the characteristic phrase) by the performer under the constraints of the raga grammar. Thus the repetitions of a raga-characteristic phrase in a concert are strongly recognizable despite a surface variability that makes them interesting.The sequence of melodic phrases comprises a "musical statement" and spans a rhythmic cycle sometimes crossing over to the next. Boundaries between such connected phrase sequences are typically marked by the *sam*of the*tala*cycle. Thus we have various musically meaningful time-scales for analysis as we progress from *svara*to phrase to rhythm cycle durations. While a *svara*, drawn from the permitted notes of a raga, can be considered to have a certain pitch position plus a certain movement within its pitch space through its *gamaka*, it is really always an integral part of the larger phrase. Listeners are known to identify the raga by the occurrence of its permitted melodic phrases or "movements" (*calana*) which can comprise of as few as 2 or 3 notes, but often more. Thus the continuous pitch curve representing the raga-characteristic phrase can be viewed as a fundamental component of raga grammar. Having said this, it must be mentioned that while many ragas are completely recognised by their *pakads* (set of melodic motifs), there are ragas that are characterised by aspects such as the dominant *svaras* or overall progression of the melody over time. This is especially true of the *mela*ragas in Carnatic music (newer ragas that have been defined purely based on their*svaras* and no associated tradition of phraseology (Krishna & Ishwar, 2012).

Given the central role played by raga-characteristic phrases in the performance of both the Indian classical traditions, computational methods to detect specific melodic phrases or motifs in audio recordings 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 characteristic phrases (*pakad*) corresponding to each raga (Chakravorty, Mukherjee, & Datta, 1989). Given that the identity of a raga-characteristic phrase is captured by the *svaras* and *gamakas* that constitute it, a melodic representation of the phrase would be crucial in any automatic classification task. Characteristic-phrase recognition can help in the automatic music transcription of Indian classical music which is notoriously difficult due to its interpretive nature. Such phrase-

level labeling of audio can be valuable in musicological research apart from providing for an enriched listening experience for music students (Rao, Ross & Ganguli, 2013).

Computational approaches to raga identification have so far been limited to the scale aspect of ragas. Pitch class histograms, as well as more fine sub-semitone interval histograms, have been applied previously to Indian classical vocal music audio where the vocal pitch has been tracked continuously in time (see review in Koduri, Gulati, Rao, & Serra, 2012). The first-order distributions have been observed to be dispersed to various extents around prominent peaks facilitating certain interpretations about the underlying tuning of the music as well as cues to raga identity in terms of *svara* locations and dispersion attributed to the *gamaka*. Although characteristic phrases are musicologically germane to raga music, there have been few studies on phrase level characteristics in audio.

This work addresses the automatic detection of raga-characteristic phrases in Hindustani and Carnatic vocal classical music audio. We next reviewthe literature on melodic motivic analysis with an introduction to the specific problem in the context of Indian classical music. We present the audio databases used in the present work which helps provide a concrete framework for the discussion of phrases and their properties in the two classical traditions. The present work is restricted to the classification of segmented raga-characteristic phrases based on supervised training. The goal is to investigatesuitable phrase-level melodic representations in the framework of template and statistical pattern matching methods for phrase recognition. The description of experiments is followed by a discussion of the results and proposals for future work.

## 2 Melodic Motivic Analysis

Given the well-known difficulties with extracting low-level musical attributes such as pitch and onsets from general polyphonic audio recordings, most work in motivic analysis for music has been restricted to symbolic scores. Melodic segmentation as well as motif discovery via string comparisons have been actively researched problems (Juhász, 2007; Cambouropoulos, 2006; Kranenburg, Volk, Wiering, & Veltkamp, 2009). Relative pitch and inter-onset duration intervals derived from the scores constitute the strings. Dannenberg *et al.*(Dannenberg, & Hu, 2003) implemented repeated pattern searching in audio using both note-based and frame-based (segmentation into equal duration time frames) pitch sequences derived from the audio. To appreciate the applicability of these methods to motivic analysis in Indian classical music, we note some aspects of its written notation. Considering the melodic sophistication of the tradition, written notation for Indian classical music is 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. The performer's interpretation of the notation invokes his background knowledge including a complete awareness of raga-specified constraints. The interpretation involves supplying the *gamakas* for the notated *svaras*, and possibly also volume and timbre dynamics. In the present work, we restrict ourselves to a pitchbased description. Fig. 1 shows an extract from a performance in raga Alhaiya-Bilawal by vocalist Ashwini Bhide. The 25 sec duration audio segment has been processed to obtain the vocal melodic pitch versus time. A Hindustani musician familiar with the raga provided the note-level transcription shown in the lower grid where the raga-characteristic phrases are indicated in bold font. We observe that the raga-characteristic phrases are representedsimply as a sequence of svaras in the written notation. The correspondence between the melodic figure representing the phrase in Fig. 1and its notation is not obvious.

This reminds us of a remark by Widdess about his need to collaborate with the performer to achieve the transcription of a recorded concert (Widdess, 1994). What appear, in the melodic contour, to be inflections of other pitches are explicitly notated while other significant pitch deviations are treated as pitch inflections of the notated svara. We see, however, the striking similarity of the continuous pitch curve across repetitionsof the raga-characteristic phrase. Such continuous pitch movements may beviewed as "figures" andare likely to serve cognitively as the best units of melody, as memorized and transmitted through the oral-aural route of Indian classical traditions.That melodic segmentation by listeners depends on both low-level sensory data as well as on learned schemas is borne out by studies on segmentation of Arabic modal improvisations by listeners from different cultures (Lartillot & Ayari, 2008)

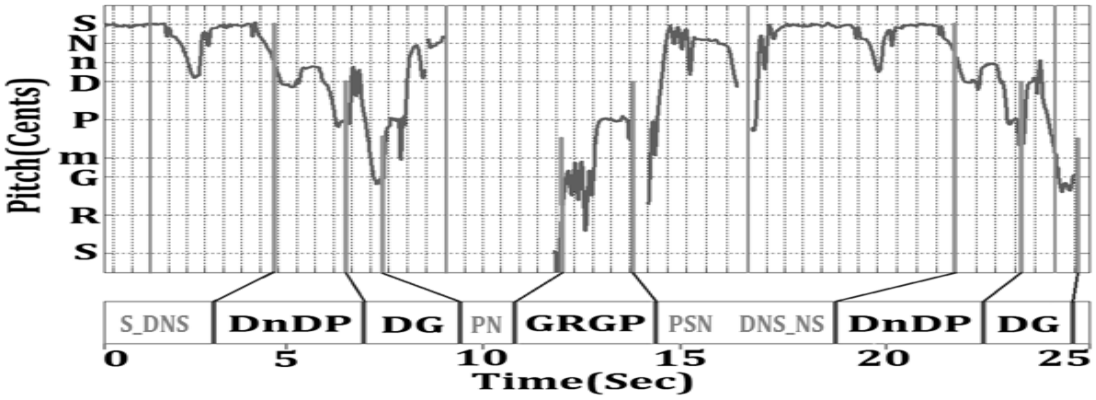



Figure 1. Melodic pitch curve extracted from a concert in raga Alhaiya-Bilawal by a well-known Hindustani vocalist with corresponding written notation in the lower layer. The raga-characteristic phrases are delimited by bold lines and notated in bold font. Horizontal lines mark *svara* positions. Thin vertical dotted lines mark beat instants in the 16-beat cycle. The four *sam* instances are marked by thin solid lines.

There has been some past work on melodic similarity in Hindustani music based on using trained N-grams on note sequences where the detected "notes" are obtained by heuristic methods for automatic segmentation of the pitch contour into notes(Pandey, Mishra, & Ipe, 2003; Chordia & Rae, 2007). Automatic segmentation of the continuous pitch contour of a phrase, extracted from the audio signal, into *svaras* is a challenging and poorly defined task given that it comprises the practically seamless concatenation of *gamakas* as we have seen with reference to Fig. 1.It may be noted that note segmentation based on timbre changes is not applicable here due to the absence of syllables in the often melismatic style of Indian classical vocal music. Further, whereas Carnatic music pedagogy has developed the view of a phrase in terms of a sequence of notes each with its own*gamaka*, the Hindustani perspective of a melodic phrase is that of a gestalt. The various acoustic realizations of a given characteristic phrase form almost a continuum of pitch curves, with overall raga-specific constraints on the timing and nature of the intra-phrase events. There is often only a loose connection between melodic events such as note onsets and the underlying rhythmic structure supplied by the percussion especially in the widely performed *khayal* genre of Hindustani music.

In view of the above observations, the choice of a data representation for the melodic shape of a phrase needs careful consideration. Similarity matching methods on such a data representation must discount the

variabilities in a raga-characteristic phrase intonation that are expected to arise across artistes, concerts and tempo. Variable duration time-series occur in many audio retrieval applications including speech recognition. Classic solutions have involved exemplar-based search using a dynamic time warping (DTW) (Berndt & Clifford, 1994) distance measure, and statistical model based search using a generative model such as the Hidden Markov Model (HMM) (Rabiner & Juang, 1986). Both these are powerful techniques that address the classification of time sequential patterns. HMMs are widely used in speech recognition and can serve the similar role in the present task where the detected pitches constitute observations and the underlying svara the hidden states. The HMM framework can learn models from labeled training data with little prior knowledge provided the training data comprises a sufficiently large number of instances of each class. The flexibility allowed in phrase rendition across tempo and style can potentially be learned from the training data if it incorporates such diversity. In the absence of a large dataset, exemplar based matching using representative templates of each phrase class can work if the similarity measure is suitably implemented. DTW has been applied to melody-based retrieval in a query-by-humming system where segment duration mismatches between time series representing a user-sung query and a stored reference melody are compensated for by DTW alignment (Zhu & Shasha, 2003). However in the present context of raga motif detection, it is not obvious whether this approach is directly applicable given the absence of explicit musical knowledge about the phrase intonation. A small change in a *gamaka* could alter phrase intonation sufficiently to indicate a different raga (Krishna & Ishwar, 2012). Further the variation in phrase intonation with tempo is phrase-dependent and has remained something that is taught only through examples rather than rules in the oral-aural mode of Indian classical music pedagogy (Rao & Rao, 2013). Classic DTW with a fixed global constraint was shown to be effective in within-concert *mukhda* (refrain) detection in *khayal* audio provided the particular rhythmic alignment of the segment was exploited (Ross, Vinutha, & Rao, 2012).

In summary, melodic motif segmentation and labelling in recorded audio signals is an interesting task in the context of Indian classical music with its raga basis. In this work, we restrict ourselves to the problem of classifying pre-segmented phrases in concert audio recordings within a closed set of characteristic phrases of one or more ragas. To help define the problem better, properties of the characteristic phrases are discussed in the context of the datasets selected for this study in the next section. HMM based and template based approaches are investigated for classification depending on the size of the training dataset.

## 3 Database and Annotation

Concert audio recordings were used to put together the raga motifs database. For the Carnatic music database, selected raga-characteristic phrases were segmented from *alap* sections (unmetered improvisation) of vocal recordings where the accompanying instruments were *tanpura* (drone) and violin. In the *alap* section, which typically is the initial part of raga elaboration, the artiste tends to use the key phrases of the raga in their pristine form to convey the raga identity effectively. Further, the absence of percussion (*tabla* or *mridangam*) makes automatic melodic pitch detection easier. Hindustani *khayal* concerts on the other hand are characterised by very short duration (around 1 minute) *alap* except in the Agra *gharana*. The Hindustani dataset is assembled from the *badakhayal* sections where the *vistar* is the improvised segment that occurs within the constraints of the rhythmic cycle framework of the *bandish*. Sequences of raga-characteristic phrases appear between occurrences of the *mukhda* (or refrain) of the composition. Several general properties

of the melodic motifs are illustrated through examples from these datasets. Figure 2 presents the *svara* notation employed for each of the traditions. The *svara* refers to the pitch interval with respect to a chosen tonic (the note C in the case of Figure 2).

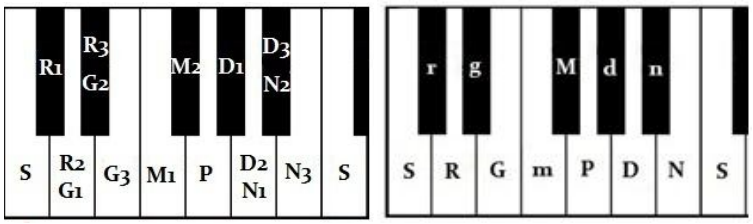



Figure 2: *Svara* names in Carnatic (left) and Hindustani(right) music traditions

## 3.1 Hindustani music

For the motif identification task, we choose the characteristic phrases of the raga Alhaiya-Bilawal, a commonly performed raga of the Bilawal group, which includes ragas based on the major scale (Rao, Bor, Van Der Meer, & Harvey, 1999). It is considered to be complex in its phraseology and is associated with a sombre mood. While its notes include all the notes of the Western major scale, it has additionally the *komal Ni* (n) in the descent (*avaroha*). Further *Ma* is omitted from the ascent. The typical phrases used for raga elaboration in a performance appear in Table 1. A specific phrase may appear in the *bandish* itself or in the *bol-alap* and *bol-taan* (improvised segments). It may be uttered using the words or syllables of the *bandish* or in *aakar* (melismatic singing on the syllable /a/). What is invariant about the *calana* is its melodic form which may be described as a particular-shaped pitch trajectorythrough the nominal notes (*svaras*) in Table 1.Raga Kafi,

| *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)<br>D~ n D \P (DnDP)<br>D \G<br>G m R G P m G | R g R– m m P<br>g- m P m P<br>D m n\P g R<br>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 ("Music in motion," 2013; Rao, Bor, Van Der Meer, & Harvey, 1999 ). 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 parentheses

| Song ID | Artiste | *Tala* | *Laya* | *Bandish* | Tempo (bpm) | Dur. (min) | Phrase Class | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | DnDP | | mnDP | GRGP |
| | | | | | | | Char. | Seq. | | |
| 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 | 10 |
| 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 | --- |
| | | | | **Total no. of phrases** | | | 122 | 52 | 70 | 76 |

Table 2: Description of database with phrase counts in the musician's transcriptionof each concert; all concerts are in raga Alhaiya-Bilawal except the last (AC) in raga Kafi. "Char." = characteristic of the raga; "Seq." = note sequence

whose description also appears in Table 1, is used in this work primarily as an "anti-corpus", i.e. to provide examples of note sequences that match the prescriptive notation of a chosen characteristic phrase of Alhaiya-Bilawal but in a different raga context (and hence are not expected to match the melodic shape, or intonation).

A total of eight selected audio recordings of RagaAlhaiya Bilawal and one recording of RagaKafi, by eminent Hindustani vocalists, from commercial CD and NCPA AUTRIM archive for Music in Motion ("Music in motion," 2013) have been used for the study. The common accompanying instruments were *Tanpura* (drone), *Tabla* and *Harmonium*, except one having *Sarangi* in place of *Harmonium*. The concert sections of *bandish*with its *vistar*(only non-*taan*section) spanning slow (*vilambit*) to medium (*madhya*) tempi. Although the size of the database is limited, it has been designed to present a challenging scenario by including related phrases in raga Alhaiya-Bilawal that share a common *nyas* (focal or ending svara). From Table 1, these are DnDP and GRGP. Additionally, the phrase mnDP which occurs in the *mukhda* in several of the recordings is also used due to its similarity to DnDP in terms of shared *svaras*. Finally, the melodic segments corresponding

to the DnDP sequence from raga Kafi are included due to their shared prescriptive notation with the raga-characteristic phrase in Alhaiya-Bilawal. The phrases were labeled by a musician (and later validated by another) using the PRAAT[1] interface. A steady note or rest generally cues a phrase end. Themusician listened for occurrences of the P-*nyas* ending phrases of Table 1. Every recognized instance was only coarsely delimited to minimize musician effort, and labeled with the corresponding phrase name. The musicians observed that the Alhaiya-Bilawal DnDP phrase was clearly distinguished from the DnDP sequence segments based on phrase intonation and sometimes, preceding context. The actual phrase boundaries were refined via automatic segmentation described later. A count of the phrases of each category appears in Table 2.

### 3.2 Carnatic music

For the Carnatic database, the 5 ragas listed in Table 3 are selected with concerts taken from a personal collection of audio recordings. The unmetered *alap* sections of the concerts were used and a musician labeled the raga-characteristic phrases by listening within the audio interface of Sonic Visualizer[2]. Table 3 showsthe ragas of the songs in the database along with the count of the number of instances in each phrase category. The labeled phrases were validated by two other musicians.

| Raga | Phrase label | Notation | Average duration (sec) | Min/Max duration (sec) | # Phrases labeled |
|---|---|---|---|---|---|
| **Bhairavi** | m4 | R2 G1 M1 P D1,P, | 1.88 | 0.87 / 3.64 | 72 |
| | m10 | R2,P,G1,,,R2,S, | 1.59 | 1.05 / 2.63 | 52 |
| **Shankarabharana** | m6 | S,,,P, | 1.19 | 0.56 / 3.32 | 96 |
| | m2 | S,D2 R2 S N3 D2 P | 2.24 | 0.59 / 4.40 | 80 |
| | m3 | S,D2 N3 S | 1.19 | 1.35 / 7.53 | 52 |
| **Kamboji** | m3 | S,,,N2 D2 P,D2,,,, | 1.32 | 0.84 / 1.87 | 104 |
| | m6 | M1 G2 P D2 S | 1.89 | 0.64 / 4.59 | 48 |
| | m14 | D2 S R2 G2 M1 G2 | 1.55 | 0.89 / 2.31 | 44 |
| **Kalyani** | m5 | N3 R2 S S N3 N3 D2 P M2 | 1.47 | 0.94 / 3.07 | 52 |
| **Varali** | m1 | G1,,R2 S N3 | 0.85 | 0.58 / 1.54 | 52 |

Table 3: Description of ragas, phrases and phrase counts in the Carnatic music database.

## 4 Audio Processing

In this section, we discuss the processing of the audio signal to extract the melodic representation of the phrase. The continuous pitch versus time is a complete representation of the melodic shape of the phrase, assuming that volume and timbre dynamics do not play a role in motif recognition by listeners. Most of the audio processing steps, as outlined in Figure 3, are common to the Hindustani and Carnatic datasets, with differences mentioned where applicable. We present a few sample pitch curves of different raga-characteristic phrases and discuss their observed properties in order to appreciate better the issues that arise in melodic similarity modeling.

[1]Praat: http://www.fon.hum.uva.nl/praat/
[2]Sonic Visualiser: http://www.sonicvisualiser.org/

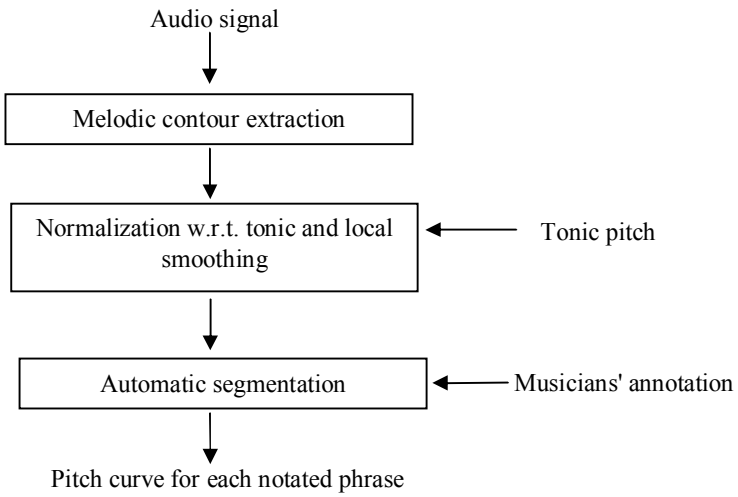



Figure 3: Block diagram for audio processing

## 4.1 Audio processing stages

### 4.1.1 Melodic contour extraction

The singing voice usually dominates over other instruments in a vocal concert performance in terms of its volume and continuity over relatively large temporal extents although the accompaniment of *tabla* and other pitched instruments such as the drone and *harmonium*or violin is always present. Melody extraction is carried out by predominant-F0 detection methods (Rao & Rao, 2010; Salamon &Gómez, 2012). These methods exploit the local salience of the melodic pitch as well as smoothness and continuity over time to provide a pitch estimate and voicing decision every frame. We use frame durations of 10 ms giving us a continuous pitch curve (Hz versus time), sampled at 10 ms intervals, corresponding to the melody.

### 4.1.2 Normalization and local smoothing

The *svara* identity refers to the pitch interval with respect to the artiste-selected tonic. In order to compare the phrase pitch curves across artistes and concerts, it is necessary to normalize the pitches with respect to the chosen tonic of the concert. Thus the pitches are represented in cents with respect to the detected tonic of the performance (Salamon, Gulati, & Serra, 2012).The pitch curve next is subjected to simple 3-point local averaging to eliminate spurious perturbations that may arise from pitch detection errors.

### 4.1.3Segment boundary refinement

Since the scope of the present work is restricted to classification of segmented phrases, we use the musicians' labeling to extract the pitch curve segments corresponding to the phrases of interest. Since the musicians' labeling is carried out relatively coarsely on the waveform in the course of listening, it is necessary to refine the segment boundaries in order to create pitch segments for the training and testing sets of similarity computation especially in the case of exemplar based matching. Thus phrase segmentation is carried out on the

Hindustani audio melodic contours in a semi-automatic manner by detecting the onset and offset of the starting and ending notes respectively. An onset or offset of a *svara* is reliably detected by hysteresis thresholding with thresholds of 50 and 20 cents within the nominal pitch value. Figure 4 shows theDnDP phrase segment where the phrase boundaries are eventually marked at offset of the n (descending from S), and onset of the P-*nyas*.

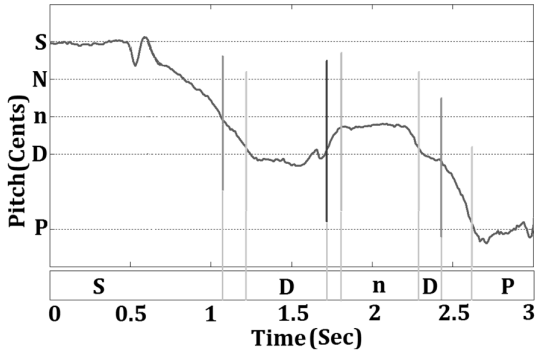



Figure 4: Illustrating *svara* onsets and offsets (vertical bars) in a DnDP phrase pitch curve. Dark bars: exiting a *svara* by descending/ascending; light bars: approaching a *svara* from below/above. Phrase boundaries are selected from these instants based on the starting and ending *svaras* of the phrase.

The output of the audio processing block is the set of segmented pitch curves that correspond to the phrases of interest as shown in Table 2. Thus each phrase is represented by a tonic-normalised cents versus time continuous pitch curve. Figure 5 shows examples of the pitch curves. These are of varying duration depending on the duration of the phrase in the audio.

## 4.2 Phrase-level pitch curve characteristics

Figure 5(Image 1) shows some representative pitch contours for DnDP phrases in various melodic contexts selected from different concerts in our Hindustani database (Table 2). The contexts typically correspond to the two possibilities: approach from higher and approach from lower *svara*. The vertical lines mark the rhythmic beat (*matra*) locations whenever these were found in the time region covered in the figure. We consider the phrase

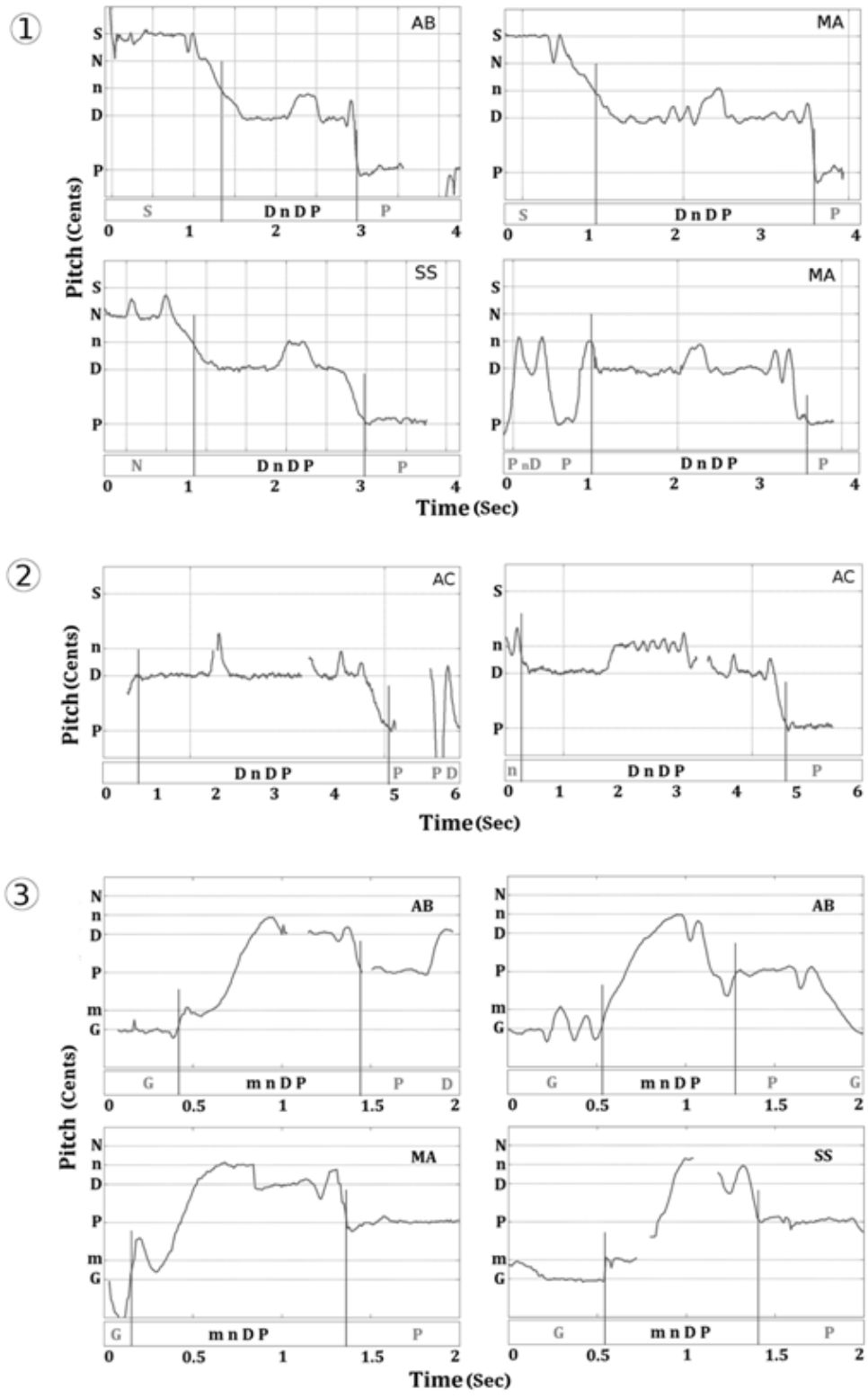



Figure 5:Pitch contours (cents vs time) of different phrases in various melodic contexts by different artistes(adapted from (Rao, Ross & Ganguli, 2013)). Horizontal lines mark *svara* positions. Thin vertical lines mark beat instants. Thick lines mark the phrase boundaries for similarity matching; 1. Alhaiya-Bilawal DnDP, 2.Kafi DnDP, 3.Alhaiya-Bilawal mnDP

duration, indicated by the dark vertical bars, as spanning from D-onset or m-onset (or rather, from the offset of the preceding n through which *svara* the phrase is approached) to P-onset. The final Pis a resting note and therefore of unpredictable and highly varying duration. From the spacing between beat instant markers in Figure 5, 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, tempi span a large range (as seen in Table 2) while the maximum duration of the DnDP phrase in any concert ranges only between 1.1 to 2.8 sec with considerable variation within the concert. Further, any duration variations of sub-segments are not linearly related. For example, 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 Figure 5 (listed in Sec. 2), we note that the raga Kafi phrases (in which raga DnDP is not a characteristic phrase but merely an incidental sequence of notes) display a greater variability in phrase intonation while conforming to the prescriptive notation of DnDP (Rao, Ross & Ganguli, 2013).

We observe the similarity in melodic shape across realizations of a given phrase in the Alhaiya-Bilawal raga. Prominent differences are obvious too, such as the presence or absence of n as a touch note (*kan*) in the final DP transition in DnDP and varying extents of oscillation on the first D. The similar comments apply to the different instances of mnDPshown in Figure 5. Variations within the phrase class may be attributed to the flexibility accorded by the raga grammar in improvisation. Consistent with musicological theory on *khayal* music at slow and medium tempi, (i) there is no observable dependence of phrase duration upon beat duration, (ii) relative note durations are not necessarily maintained across tempi, and (iii) the note onsets do not necessarily align with beat instants except for the *nyas*, considered an important note in the raga.

The Carnatic phrases from the raga Kamboji depicted in Figure 6 display variations relating to both the extent and number of oscillations on a *svara*. Since all the phrases are from the *alap* sections, we cannot comment on dependence of duration or melodic shape on tempo.

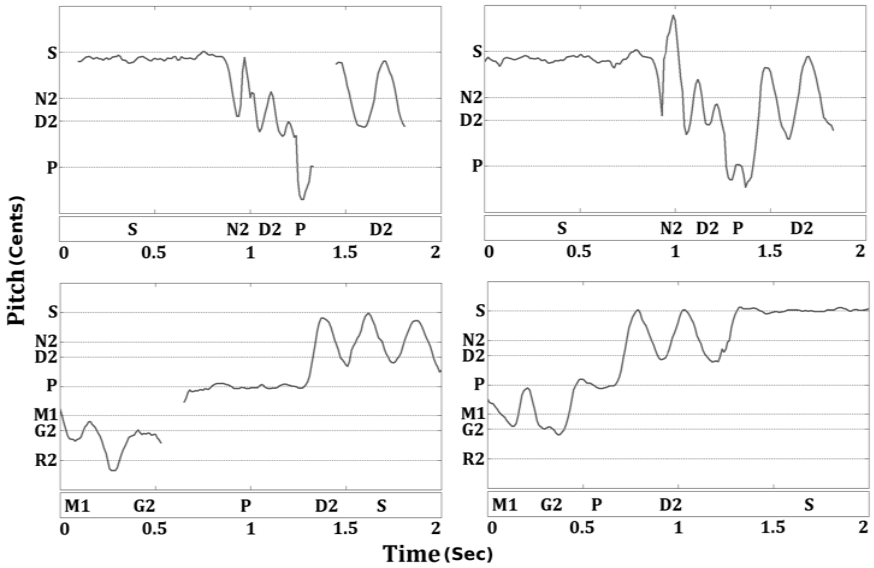



Figure 6: Pitch contours of phrases 'm3'(top row) and 'm6'(bottom row) of Raga Kamboji in Carnatic music

The task of identifying the phrase class from the pitch curve representing the melodic shape can be implemented by either time-series matching or by statistical pattern recognition. The former method is applied to the Hindustani database due to the limited available dataset. On the Carnatic dataset, we use the statistical framework of HMM based classification.

# 5 Similarity Computation

We present two different pattern matching methods to achieve the classification of the test melodic segments, obtained as described in the previous section, into phrase classes. A part of the labeled dataset of phrases is used for training the classifier, and the remaining for testing.

## 5.1 Exemplar-based matching

Reference templates for each phrase class of interest are automatically identified from the training set of phrases. A DTW based distance measure is computed between the test segment and each of the reference templates. The detected phrase class is that of the reference template that achieves the lowest distance, provided the distance is below a pre-decided threshold. DTW distance computation combines a local cost with a transition cost, possibly under certain constraints, both of which must be defined meaningfully in the context of our task of melodic matching. Apart from deriving the reference templates, training can be applied to learning the constraints. The various stages of exemplar-based matching are discussed below.

### 5.1.1 Classification of test segment

The test pitch curve obtained from the audio processing previously described is prepared for DTW distance computation with respect to the similarly processed reference template as shown in the block diagram of Figure 7. The phrases are of varying duration, and a normalization of the distance measure is achieved by interpolating the pitch curves to the fixed duration of the reference template that it is being compared with. The fixed duration of a reference template is the average duration of the phrases that it represents. Before this, the short gaps within the pitch curve arising from unvoiced sounds or singing pauses are linearly interpolated using pitch values that neighbor the silence region up to 30 ms. Zeros are padded at both the ends of the phrases to absorb any boundary frame mismatches which has adverse effect on DTW matching.

To account for the occurrence of octave-transposed versions of phrases with respect to the reference phrase, transposition of +1 and -1 octave creating 3 versions of the test candidate for the similarity matching are computed.We also investigate the quantization of pitch in the melodic representation. 12-semitone quantization to an equitempered scale (with respect to the tonic) and 24 level quantization to quartertones are obtained for evaluation.

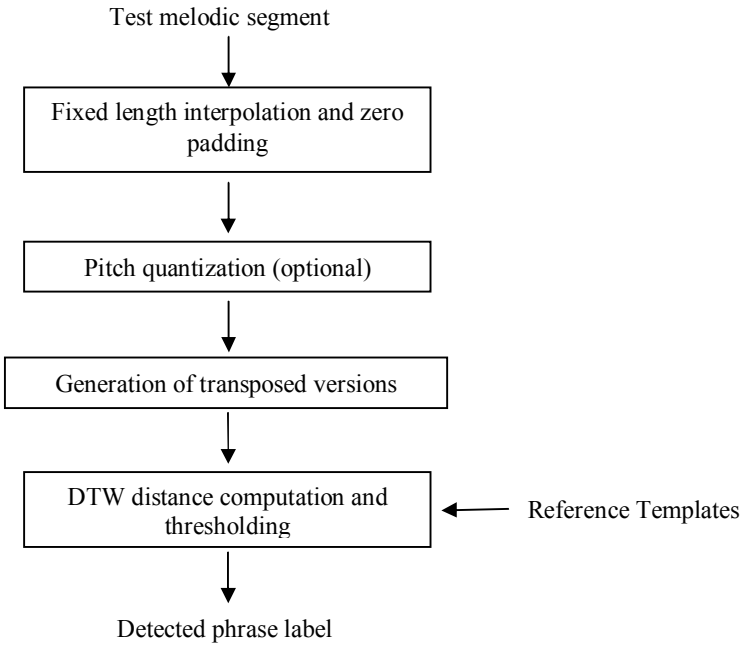



Figure 7: Block diagram for similarity computation for Hindustani music

DTW distance between each of the template phrases and the transposed versions of the test pitch curve is computed. The template phrases comprise the "codebook" for the task and together represent the phrase classes of interest. For example, in our task we are interested in the recognition of DnDP and mnDP of raga Alhaiya-Bilawal. The template phrases are therefore representative pitch curves drawn from each phrase class of interest in the labeled training dataset. Vector quantization (VQ), as presented in the next section on training, is used to obtain the codebook of phrases. The detected phrase class is that of the reference template that achieves the lowest distance, provided the distance is below a pre-decided threshold. From an informal examination of the labeled pitch curves, it was felt that two templates per phrase class would serve well to capture intra-class variabilities. Training is also used to learn DTW path constraints that can potentially improve retrieval accuracy.

### 5.1.2 Vector Quantization based Training

The k-means algorithm is applied separately to each training set phrase class (DnDP and mnDP) with k=2. A DTW distance measure is computed between fixed-length pitch curves. In each iteration of the k-means procedure, the centroid for each cluster is computed asthe mean of corresponding pitches obtained after DTW time-aligning of each cluster member with the previous centroid. Thus we ensure that *corresponding* sub-segments of the phrase are averaged. Figures 8 and 9 show sample pitch curves from the clusters obtained after vector quantization of DnDP and mnDP phrases respectively. We observe that a prominent distinction between members of the DnDP phrase class is the presence or absence of then-*kan* (touch note) just before the P-*nyas*. (Note that the P-*nyas svara* is not shown in the figure since it is not included in the segmented phrase as explained earlier). In the mnDP class, the clusters seem to be separated on the basis of the modulation extent of the initial m*svara*. Visually, the cluster centroids (not shown) are representative of phrase pitch curves of the

corresponding cluster. Thus VQ of pitch curves serves well to capture the variations observed in phrase intonation.

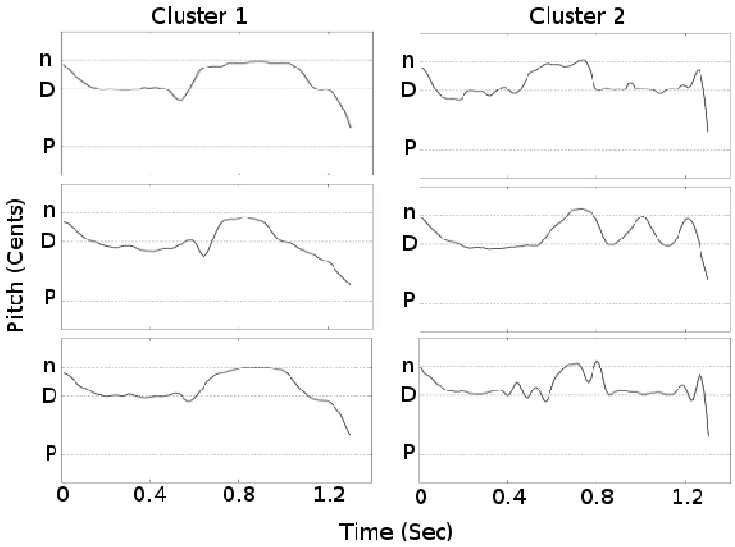



Figure 8: Examples from each of the two VQ clusters obtained for the DnDP instances from the AB and MA concerts (all phrases interpolated to uniform length of 1.3 seconds)

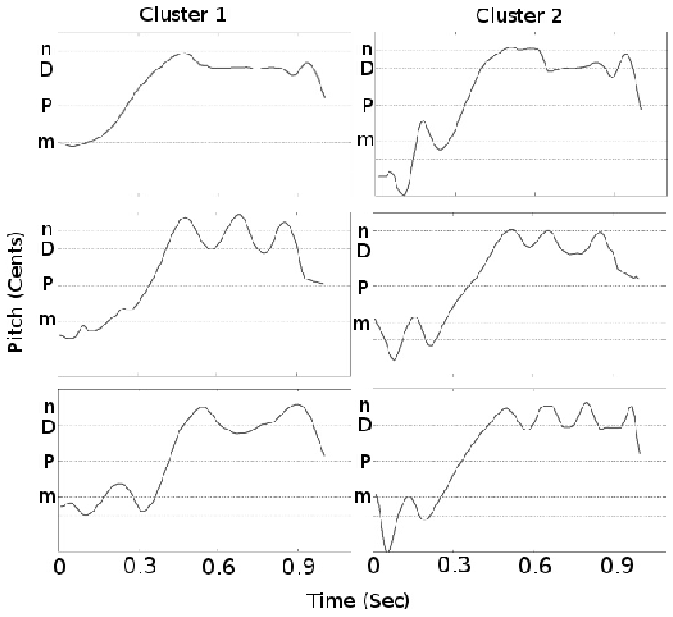



Figure 9: Examples from each of the two VQ clusters obtained for the mnDP instances from the AB and MA concerts (all phrases interpolated to uniform length of 1 second)

### 5.1.3 Constraint Learning

Global path constraints applied in DTW distance computation can restrict unusually low distances arising from pathological warping between unrelated phrases especially those that have one or more *svaras* in common. The global constraint should be wide enough to allow for the flexibility actually observed in the phrase intonation across artistes and concerts. As noted in Section 4, the elongation or compression observed in one instance of a phrase with respect to another is not uniform across the phrase. Certain sub-segments actually remain

relatively constant in the course of phrase-level duration change. Thus it is expected that the ideal global constraint would be phrase dependent and varying in width across the phrase length. Apart from the global path constraint, we are also interested in adjusting the local cost (difference of corresponding pitches of reference and test templates) so that perceptually unimportant pitch differences do not affect the DTW optimal path estimate. Also, we would like the path to be biased towards the diagonal transition if the local distances in all directions are comparable.

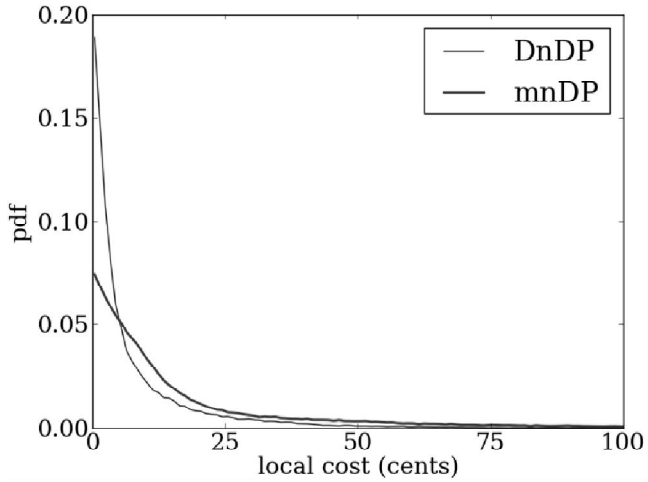



Figure 10: Local error distribution between corresponding pitch values after DTW alignment of every pair of phrases within each cluster across DnDP and mnDP phrase classes from the AB and MA concerts.

We present an iterative procedure to optimize both the above parameters: shape of the global path constraint and a local cost difference lower bound, on the training set of AB and MA concert phrases. For a given phrase class and VQ cluster, we obtain the DTW paths for all possible pairs of member phrases. The differences between all the corresponding pitch values over all the paths are computed to obtain the error distribution plot of Figure 10 for each of the DnDP and mnDP phrase classes. We observe that the error is largest near to 0 and falls off quite rapidly beyond 25 cents. We therefore use this value as a local cost lower bound i.e. pitch differences within 25 cents are ignored in the DTW distance computation. Next, for each phrase class and cluster, we estimate a global path constraint that outer bounds all pair-wise paths corresponding to that cluster. The above two steps are iterated until there is no substantial difference in the obtained global path constraint.

Figure 11 shows the outer bounds as irregular shapes that are obtained by choosing the outermost point across all paths as we move along the diagonal of the DTW path matrix. This "learned" global constraint encompasses all the observed alignments between any two phrases in the same cluster. Some interesting observations emerge from the inspection of the changing width of the global constraint with reference to the representative melodic shapes of each class-cluster in Figures 8 and 9. The relatively narrow regions in the global constraint shapes correspond roughly to the glides (D\P in DnDP-cluster 1, and m/n in both mnDP clusters). This suggests that overall phrase duration variations affect the transitions such as glides less compared to the flat pitch segments within the phrase pitch contour. This has implications for the understanding of characteristic-phrase intonation behavior under different speeds (or tempo) of rendering.

In Figure 11, we also obtain a Sakoe-Chiba constraint, shown by the parallel lines, the area between which just encompasses all the observed paths. The Sakoe-Chiba constraint is widely used in time-series matching to restrict pathological warpings and to improve search efficiency (Sakoe & Chiba, 1978). The learned global constraints are compared with unconstrained-pathDTW and the fixed-width Sakoe-Chiba constraint in the experiments presented later.

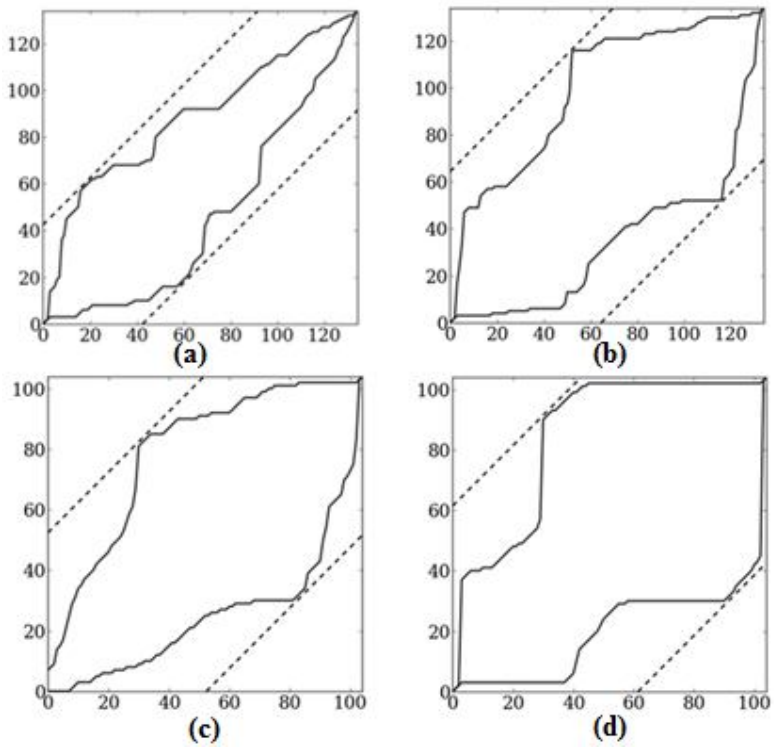



Figure 11:Learned (continuous)and Sakoe-Chiba (dashed) global path constraints obtained by bounding the DTW paths for:(a) DnDP cluster1 (b) DnDP cluster2 (c) mnDP cluster1 (d) mnDP cluster 2

## 5.2 Statistical pattern matching

Due to the availability of a large dataset for Carnatic phrase classification work, we choose to apply the HMM framework. This gives us a powerful learning model with minimal dependence on manual parameter settings. The phrases can be viewed as a sequence of *svaras* and hence transiting through distinct "states". Training the HMM model on the labeled pitch curves enables the learning of the state transition probabilities for each phrase class. Further, the estimated pitch values constitute the observation at each time instant. Given that a *svara*occupies a pitch range as dictated by its *gamaka*, the distribution of observations in each state can be modeled by a 2 mixture Gaussian (with one mode for each of two possible octave-separated regions of the *svara*). Figures 12, 13 show the training and testing procedures. The number of states in the HMM structure was based on the changes that we observed in the pitch contour. A left-right HMM without skips,but with self-loops on all emitting states, was used. The state in an HMM corresponds to an invariant event. The HMM structure is approximately dependent on the number of notes that make up a phrase since this is the invariant throughout the phrases whatever may be the variations rendered by different artists. Figure 14 shows the invariant events in the phrase Kamboji m3(Table 3) across renditions by 4 different artists. One can observe that the state number for the invariant events across the different examples for the same phrase are the same.

Inthe example taken for illustration in Figure 14, the invariant is the sequence of the *svaras* which is S N2 D2 P D2. The regions marked in Figure 14 show these invariant events.

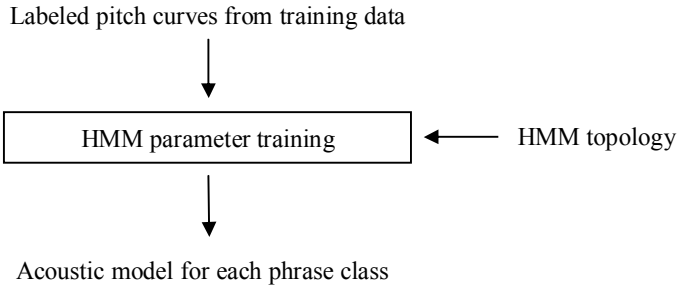



Figure 12**:** Block diagram for acoustic model training for Carnatic music

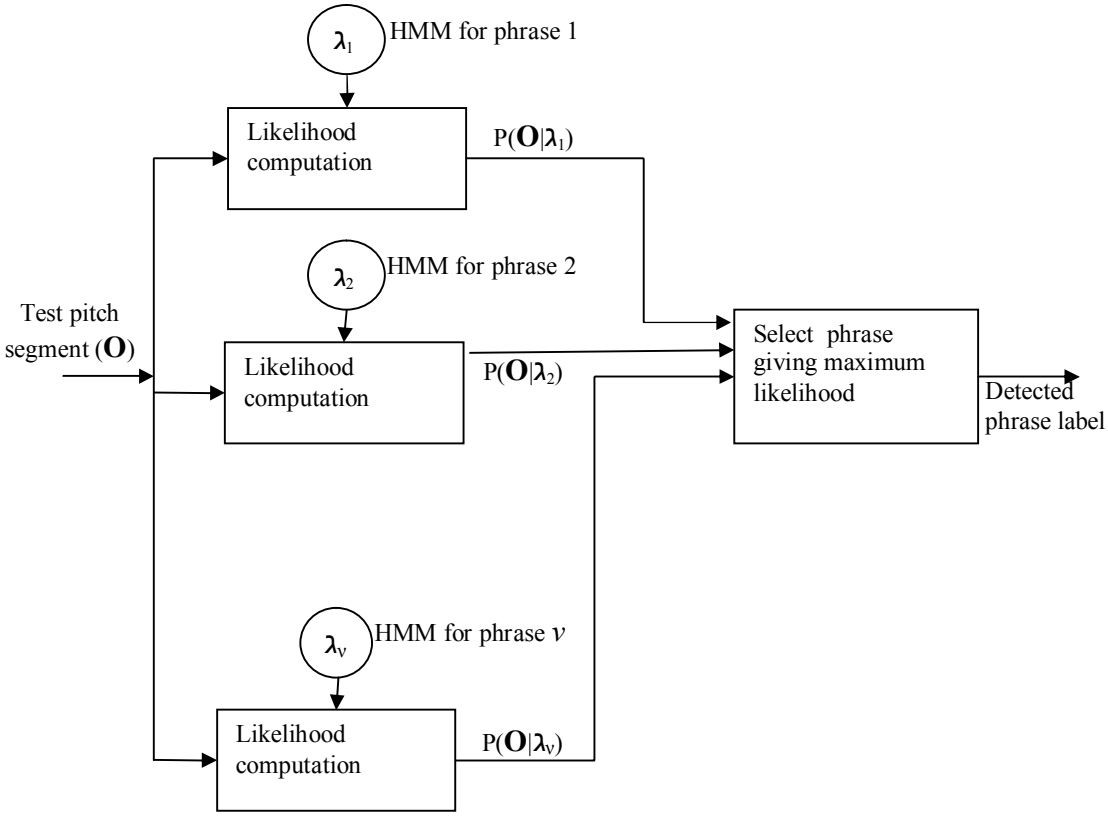



Figure 13: Block diagram for similarity computation for Carnatic music

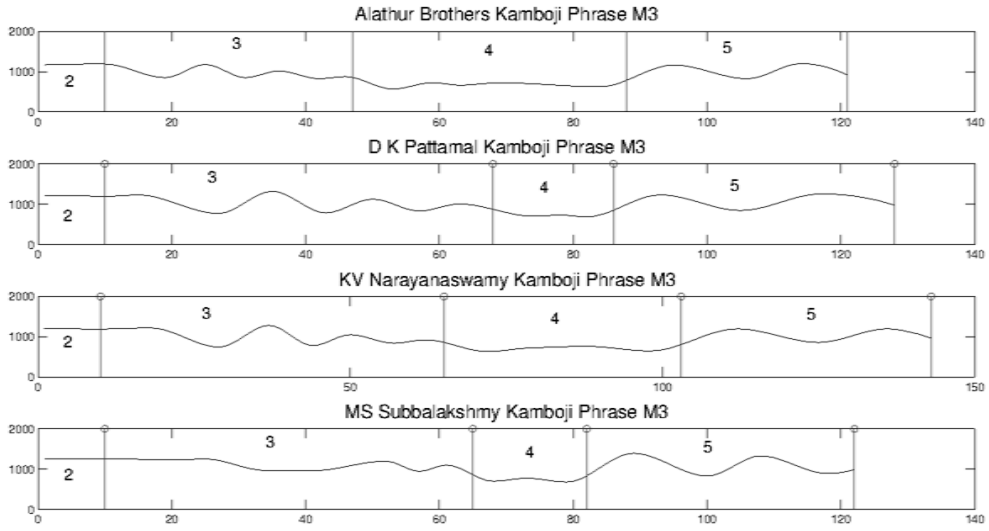



Figure 14: Viterbi alignment for Kamboji phrase 'm3'

# 6 Experiments

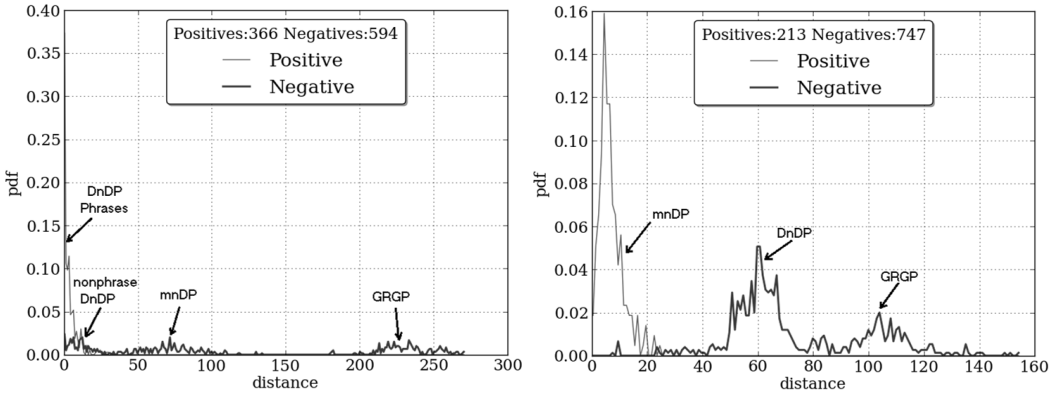
We report the results of motif detection experiments carried out as discussed in the previous sections on Hindustani and Carnatic audio datasets. Evaluation results are presented in terms of retrieval accuracies measured for each of a set of selected phrase classeson the test data comprised of several phrase classes.

## 6.1 Hindustani music

We consider the retrieval of DnDP and mnDP phrases given reference templates obtained by vector quantization on the training set of the same phrase classes from the AB and MA concerts of Table 2. The test set includes all the phrases of the corresponding class drawn from Table 2, plus all phrases not in that class across all the concerts. From Table 2, we note that the Hindustani test dataset comprises phrase classes that share the ending P-*nyas svara*. Further, some phrases have several*svaras* in common e.g. mnDP and DnDP. The test data also includes the DnDP segment from Kafi raga (similarly notated but musicologically different since it is not an Alhaiya-Bilawal raga-characteristic phrase). Thus the design of the Hindustani test data, although small, is challenging. We study the dependence of retrieval accuracy on the choice of the DTW global constraint and on pitch quantization choice.

The retrieval performance for a given phrase class can be measured by the hit rates for a various false alarm rates by sweeping a threshold across the distribution of DTW distances obtained by the test phrases. In the DTW distance distribution, each test phrase contributes a distance value corresponding to the minimum distance achieved between the test pitch curve and the set of reference templates (one per cluster) for the phrase class of interest. Our reference templates are the VQ cluster centroids discussed in Sec. 5.1.2. In order to increase the evaluation data, we create 2 additional sets of reference templates for each phrase class by selecting training set phrases that are close to the VQ obtained centroid of the respective clusters.Figure 15 shows such a distribution of the test phrase distances across the 3 reference template sets for each of the DnDPand mnDPphrase classescomputed without global path constraints.In each case, based on the ground-truth labels of the test phrases, two distributions are plotted, viz. one corresponding to “positive” (true) phrases

and the other to "negative" (i.e. all other phrases including non-raga characteristic). We note that the positive distances are concentrated at low values. The negative distances are widely spread and largely non-overlapping with the positive distances, indicating the effectiveness of our similarity measure. The negative distribution shows clear modes which have been labeled based on the ground truth of the test phrases. In Figure 15(a), as expected, GRGP phrases are most separated while mnDP, with more shared *svaras*, is close to the DnDP distribution. Finally, the non-phrase DnDP overlap most with the positives and are expected to be the cause of false alarms in the raga-characteristic DnDP phrase detection. In Figure 15(b), the positive phrase distances are even more separated from the negative phrase distances due to the more distinct melodic shape differences between the positive and negative phrase classes with their distinct initial svaras. This leads to near perfect retrieval accuracy for the mnDP phrase, at least within the limitations of the current dataset.A hit rate of over 99% (212 phrases detected out of 213) is obtained at a false alarm rate of 1% (8 false detections out of 747 non-mnDP test phrases). We present more detailed results for DnDP phrase detection where the numbers of positive and negative phrases in the dataset are more balanced.




(a) (b)

Figure 15:Distributions of distances of P-*nyas* phrases from (a)DnDP raga-characteristic templates (b)mnDP raga-characteristic templates.

Table 4 shows the hit rate achieved at a range of acceptable false alarm rates for the DnDP phrase class. Since an important application of raga-characteristic phrase detection is the retrieval of music based on raga identity, a low false alarm rate would be desirable. The false alarm rates depicted for DnDP are relatively high compared with that obtained for mnDP due the more challenging test dataset for the former that includes similarly notated test phrases. We compare the DTW similarity measure obtained with various global path constraints versus that without a constraint. Global constraints obtained by learning as presented in the previous section are compared with the similarly derived Sakoe-Chiba constraint with its constant width. In all cases, the 25 cents error threshold was appliedand, for same local cost, the diagonal direction was preferred in the DTW path. We observe from Table 4, that the learned Sakoe-Chiba constraint performs similar to the unconstrained DTW distance while the learned global constraint performs worse. The Sakoe-Chiba path constraint thus helps in terms of more efficient distance computation while retrieval accuracy is uncompromised. The learned global constraint, on the other hand, poses more detailed path shape assumptions on the test data and can be expected to be reliable if derived from a much larger training dataset than we had access to currently. Table 5 compares the retrieval performance with quantized representations of the pitch

curve, i.e. each pitch sample in the curve is independently quantized to either 12 or 24 levels (per octave) based on the best fitting equitempered set of levels. Unconstrained-path DTW is applied. We observe a drop in accuracy with quantization to 12 levels. 24-level quantization does better and performs similar to the unquantized representation. A more complete quantized representation of the pitch curve would be one that considered temporalquantization as well, to obtain a "note" level representation (known as a "score" in the context of Western music). However the lack of any clear alignment between pitch events and beat locations in the course of elaboration of raga-characteristic phrases, as observed in Figure 5, precludes any advantage of beat-based quantization.

| **FA rate** (Total negatives:594) | **Hit rate**(Total positives: 366) | | |
|---|---|---|---|
| | **No constraints** | **Learned constraints** | **Learned Sakoe-Chiba** |
| **6.06%** 36 | 70.22% 257 | 71.04% 260 | 71.86% 263 |
| **11.95%** 71 | 89.62% 328 | 87.16% 319 | 89.34% 327 |
| **18.01%** 107 | 95.63% 350 | 92.35% 338 | 95.36% 349 |
| **24.07%** 143 | 97.81% 358 | 99.18% 363 | 97.54% 357 |

Table 4: Phrase detection accuracies for DnDPcharacteristic phrase under various global constraints

| **FA rate** (Total negatives: 594) | **Hit rate** (Total positives: 366) | | |
|---|---|---|---|
| | **Unquantized** (threshold dtw enabled) | **q12** | **q24** |
| **6.06%** 36 | 70.22% 257 | 67.76% 248 | 69.67% 255 |
| **11.95%** 71 | 89.62% 328 | 86.61% 317 | 89.07% 326 |
| **18.01%** 107 | 95.63% 350 | 93.99% 344 | 95.08% 348 |
| **24.07%** 143 | 97.81% 358 | 97.81% 358 | 98.36% 360 |

Table 5: Phrase detection accuracies for various quantized representations for DnDP with unconstrained paths

## 6.2 Carnatic music

The Carnatic dataset has a relatively large number of annotated phrase classes across ragas that makes it suited to classification experiments. We apply HMM based classification over the closed set of phrases specified in Table 3. Table 6 gives the classifier output in terms of the confusion matrix for the phrases across these ragas. The results observed were as follows.

• Similar motifs of the same ragas are identified correctly.

• Different motifs of the same ragas are distinguished quite accurately.

• Motifs of different ragas are also distinguished quite accurately, except for sk3 (diagonal elements in Table 6).

• The HMM output must be post processed using duration information of every state.

A deeper analysis of the confusion matrix given in Table6 with respect to the notations of the phrases given in Table 3 shows that the confusion makes musical sense. From Table6, it can be seen that the phrase sk3(m6) of the raga Sankarabharana, is confused with ky1(m5) of the raga Kalyani, kb1(m3) of the raga Kamboji and sk1(m2) of Sankarabharana, a phrase in its own ragas. From Table 3, we can see that ky1(m5) and sk3(m6) have two *svaras*viz. S P in common. But the phrase sk3(m6) is so rendered that its notation becomes S (D2) S (D2) P[1]. This shows that the phrases ky1(m5) and sk3(m6) have 3 *svaras*, S D2 and P in common. Also, the macro-movement of both the phrases are also similar, i.e. a descent from the upper octave S towards P for sk3(m6) and from S to M2 for ky1(m5). Similarly, the phrases sk3(m6) and kb1(m3) share 3 *svaras* in common viz. S D2 and P and the macro-movement across the *svaras* of the two phrases, again, is a descent from S to P. This posterior analysis of the phrases with respect to the notation and movement justifies the confusion with the ragas which have similar notes and movements. The other major confusion in the phrase sk3(m6) is with sk1(m2) of the same raga. This is because of the nature of sequencing and movement across *svaras* and the common *svaras* between the two phrases.

| Raga | bh1(m10) | bh2(m4) | ky1(m5) | kb1(m3) | kb2(m6) | kb3(m14) | sk1(m2) | sk2(m3) | sk3(m6) | Val(m1) |
|---|---|---|---|---|---|---|---|---|---|---|
| bh1(m10) | **40** | 2 | 1 | 0 | 0 | 6 | 0 | 0 | 0 | 3 |
| bh2(m4) | 0 | **61** | 4 | 1 | 4 | 1 | 1 | 0 | 0 | 0 |
| ky1(m5) | 0 | 1 | **23** | 10 | 0 | 0 | 11 | 2 | 5 | 0 |
| kb1(m3) | 0 | 0 | 3 | **91** | 0 | 0 | 6 | 3 | 1 | 0 |
| kb2(m6) | 0 | 0 | 1 | 2 | **44** | 0 | 0 | 0 | 1 | 0 |
| kb3(m14) | 0 | 0 | 2 | 1 | 0 | **41** | 0 | 0 | 0 | 0 |
| sk1(m2) | 0 | 2 | 18 | 13 | 3 | 0 | **28** | 7 | 9 | 0 |
| sk2(m3) | 0 | 0 | 7 | 0 | 1 | 0 | 4 | **34** | 2 | 4 |
| sk3(m6) | 0 | 1 | 23 | 25 | 1 | 0 | 29 | 5 | **10** | 2 |
| val(m1) | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | **48** |

Table 6: Classifier output in terms of the confusion matrix for Carnatic phrases

[1] The *svaras* in the brackets are the *svaras* that are not uttered, but present in melody

Table 7 shows the same classification results but now in terms of phrase detection accuracies. The false alarm rates are low across phrases but several of the confusable phrases suffer misdetections as expected.

| Raga | Phrase label | Hit rate | FA rate |
|---|---|---|---|
| **Bhairavi** | m4 | 0.847 | 0.010 |
| | m10 | 0.769 | 0.005 |
| **Shankarabharana** | m6 | 0.104 | 0.032 |
| | m2 | 0.35 | 0.089 |
| | m3 | 0.654 | 0.028 |
| **Kamboji** | m3 | 0.875 | 0.094 |
| | m6 | 0.917 | 0.017 |
| | m14 | 0.932 | 0.011 |
| **Kalyani** | m5 | 0.442 | 0.1 |
| **Varali** | m1 | 0.923 | 0.015 |

Table 7: Results of classification carried out on Carnatic music database

# 7 Conclusions and future work

Raga-characteristic phrases constitute the building blocks of melody in Indian classical music traditions. Different compositions and improvisation in a specific raga are related to each other chiefly through the repetition of the characteristic motifs. This make the representation of melodic phrases and their similarity characterization crucial for music retrieval for Indian classical music. Written notation, as currently available, represents a phrase by a sequence of svara. That this woefully inadequate is seen from the different surface forms of the same notated phrase in different ragas. The continuous-pitch segments obtained by melodic pitch detection, on the other hand, display a strong visual similarity across instances of a raga-characteristic phrase within and across performances and artistes.Simple (non-characteristic) melodic phrases that share the svara sequence can show considerable variability even within a performance, as also previously observed (Rao et al., 2013).

We considered a pitch-continuous representation for melodic phrases segmented from vocal classical music audio in Hindustani and Carnatic styles. Time-series similarity measures were explored to accommodate the variability that characterizes melodic motives in terms of temporal and pitch dynamics. Both DTW based template matching and HMM based statistical modeling provide promising classification performance. The observed variability in within-class phrase intonation is captured by supervised learning of the models from training data. Segmentation of the audio into phrases was not considered here and is an important aspect for future work. Since melodic segmentation itself involves the detection of repetitive structures, the present work can be useful for this task. With Indian classical music concerts covering a very wide performing tempo range, future work should also address the modeling of dependence of melodic shape on tempo such as the nature of detail reduction at higher speeds (Srikimar & Wyse, 2011). Moving ahead from time-series representations, a more event-based representation of the phrase in terms of basic melodic shapes is likely to be less influenced by any “allowed” improvisation in phrase intonation. An example is recent work on Carnatic alapanas which considers salient points on the pitch curve such as saddle points to characterize melodic shape (Vighnesh et al.,

2013).Finally, the potential of including musicological expertisemore explicitly in motivic similarity computation must be exploited by innovative modeling methods.

## 8 Acknowledgement

This work is partly supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

## 9 References

Berndt, D., & Clifford, J. (1994, July). Using dynamic time warping to find patterns in time series. In *KDD workshop* (Vol. 10, No. 16, pp. 359-370).

Cambouropoulos, E. (2006). Musical Parallelism and Melodic Segmentation:: A Computational Approach. *Music Perception*, *23*(3), 249-268.

Chakravorty, J., Mukherjee, B., & Datta, A. K. (1989). Some Studies in Machine Recognition of Ragas in Indian Classical Music. *Journal of the Acoust. Soc. India*, *17*(3&4).

Chordia, P., & Rae, A. (2007, September). Raag recognition using pitch-class and pitch-class dyad distributions. In *Proceedings of International Conference on Music Information Retrieval* (pp. 1-6).

Dannenberg, R. B., & Hu, N. (2003). Pattern discovery techniques for music audio. *Journal of New Music Research*, *32*(2), 153-163.

Ishwar, V., Dutta, S., Bellur, A. and Murthy, H. (2013). Motif spotting in an alapana in Carnatic music. In *Proceedings of the Conference of the International Society for Music Information Retrieval Conference, Curitiba, Brazil.*

Juhász, Z. (2007). Analysis of melody roots in Hungarian folk music using self-organizing maps with adaptively weighted dynamic time warping. *Applied Artificial Intelligence*, *21*(1), 35-55.

Koduri, G. K., Gulati, S., Rao, P., & Serra, X. (2012). Rāga recognition based on pitch distribution methods. *Journal of New Music Research*, *41*(4), 337-350.

Krishna, T. M., & Ishwar, V. (2012). Carnatic music: Svara, gamaka, motif and raga identity. In *Proceedings of the 2nd CompMusic Workshop;* Istanbul, Turkey.

Lartilott, O. & Ayari, M. (2008). Segmenting Arabic modal improvisation:Comparing listeners' responses with computer predictions. In *Proceedings of the 4th Conference on Interdisciplinary Musicology;* Thessaloniki, Greece.

Music in motion. (2013). Retrieved June 3, 2013, from http://autrimncpa.wordpress.com/.

Pandey, G., Mishra, C., & Ipe, P. (2003). Tansen: A system for automatic raga identification. In *Indian International Conference on Artificial Intelligence* (pp. 1350-1363).

Powers, H. S., & Widdess, R. (2001). India, sub-continent of,. in S. Sadie (ed.), *The New Grove Dictionary of Music.* Macmillan, London.

Rabiner, L. , & Juang, B. (1986). An introduction to hidden Markov models. In *IEEE Acoustics, Speech and Signal Processing Magazine.*

Raja, D.(2005). *Hindustani music*. DK Printworld.

Rao, P., Ross, J.C. & Ganguli, K.K. (2013). Distinguishing raga-specific intonation of phrases with audio analysis. *To appear in the Journal of the ITC Sangeet Research Academy*.

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

Rao,S., & Rao,P. (2013). An Overview of Hindustani Music in the Context of Computational Musicology. *Manuscript submitted for publication*.

Rao, V., & Rao, P. (2010). Vocal melody extraction in the presence of pitched accompaniment in polyphonic music. *Audio, Speech, and Language Processing, IEEE Transactions on*, *18*(8), 2145-2154.

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

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43-49.

Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *Audio, Speech, and Language Processing, IEEE Transactions on*, *20*(6), 1759-1770.

Salamon, J., Gulati, S., & Serra, X. (2012). A Multipitch Approach to Tonic Identification in Indian Classical Music. In *Proc. 13th International Conference on Music Information Retrieval (ISMIR)*.

van Kranenburg, P., Volk, A., Wiering, F., & Veltkamp, R. C. (2009). Musical models for folk-song melody alignment. In *Proc. International Conference on Music Information Retrieval (ISMIR)* (pp. 507-512).

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

Zhu, Y., & Shasha, D. (2003, June). Warping indexes with envelope transforms for query by humming. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, 181-192.