DETECTION OF *RAGA*-CHARACTERISTIC PHRASES FROM HINDUSTANI CLASSICAL MUSIC AUDIO

Joe Cheri Ross* and Preeti Rao[Ŧ]
Department of Computer Science and Engineering* Department of Electrical Engineering[Ŧ]
Indian Institute of Technology Bombay,
Mumbai 400076, India
joe@cse.iitb.ac.in, prao@ee.iitb.ac.in

ABSTRACT

Melodic motifs form essential building blocks in Indian Classical music. The motifs, or key phrases, provide strong cues to the identity of the underlying *raga* in both Hindustani and Carnatic styles of Indian music. Automatic identification and clustering of similar motifs is relevant in this context. The inherent variations in various instances of a characteristic phrase in a *bandish* (composition) performance make it challenging to identify similar phrases in a performance. A *nyas svara* (long note) marks the ending of these phrases. The proposed method does segmentation of phrases through identification of *nyas* and computes similarity with the reference characteristic phrase.

1. INTRODUCTION

Hindustani classical music is based on the framework of *raga* and *tala*. The *raga* or melodic base is described by the permitted intervals (*svara*) with respect to the tonic and the characteristic phrases of the *raga*. The classical music performance, although largely improvised, is actually an extensive elaboration of the *raga* where the characteristic *svara* and phrases recur throughout thus reinforcing the mood and character of the *raga*. Considering the availability of vast audio archives but typically limited metadata and practically non-existent symbolic scores, it is of interest to develop automatic systems to provide rich transcriptions of concert recordings. *Raga*-characteristic phrases would be an important component of a concert transcription. The automatic detection of phrases can serve well for music retrieval by providing inputs for higher level music attributes such as *raga* or *bandish*.[1]. In this work, we consider the problem of segmenting and clustering melodic phrases or motifs from recorded performances of Hindustani vocal music.

While the problem of melodic phrase detection has been attempted for Western music, Hindustani music presents the challenges of a pitch-continuous tradition where symbolic notation is inadequately developed. Thus phrase detection must rely on the segmentation of the continuous pitch contour followed by identifying the phrases in terms of the sequence of *svara* as well as their particular manifestation in the phrase context. The improvisatory nature of the tradition leads to variability in the actual rendering of the phrase in a manner that adds to the overall expressiveness while still retaining its easily recognizable identity as the particular characteristic phrase of the *raga*.

In the present work, we use a predominant pitch detection algorithm to extract the melodic pitch contour from vocal concert recordings [2]. Selected *raga*-specific phrases are segmented from the contour based on the proposed phrase-ending cues and clustered with an inter-phrase similarity metric. Experimental results are presented on a database of audio concert recordings of a selected *raga* by two vocalists.

2. MUSICOLOGICAL BACKGROUND

The character of the *raga* is manifested not only in the set of permitted notes (*svaras*) but also in the commonly occurring note sequences or phrases. Listeners are known to identify the *raga* by the occurrence of its main phrases (*calana*). Alhaiya-bilawal is the most commonly performed *raga* of the Bilawal group, which mainly includes *raga*s based on the major scale [3]. It is considered to be complex in its phraseology and is associated with a somber mood. While its notes include all the notes of the Western major scale, it has additionally the *komal Ni* (flat *Ni*) 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 trajectory through the nominal notes (*svara*s) in Table 1. The manual transcription of the melodic contour involves the detection of phrase boundaries and the labeling of the characteristic phrases and the other interwoven *svaras*. The commonly used annotation is simply the temporal sequence of *svara*s. However the actual interpretation of the phrase by a performer would involve the background knowledge of the performer in achieving the *raga*-specified intonation of the *svaras* and their transitions within the phrase.

Raga	Characteristic Phrases(*Pakads*)	Scale of *Raga*
Alhaiya-bilawal	mnDP, RGPmG, NDNS, DnDP, GRGP, DG	S R G m P D n N S

Table 1. *Raga*-phrase information

The detection of phrases can be aided by the availability of cues to phrase boundaries. Melodic phrase boundaries can be associated with bounding pauses such as due to a gap or a rest in the notation, or with certain metrical positions in the metrical cycle of the music. In a previous work on motif detection in Hindustani vocal music, the main melodic motif, or *mukhda*, was segmented based on its fixed position in the metrical cycle (coincidence of final note onset with the *sam*) [4]. However this relationship of inter-phrase events to metrical events is weak in the case of non-*mukhda* phrases. While the relative timing of notes is preserved within the phrase, the phrase beginning does not occur on a specific beat necessarily. In Hindustani music, the concept of *nyas svara* can be useful. *Nyas svara*, literally the "resting note", refers to a *svara* that acts as a phrase ending. It is relatively long and stable and likely to be followed by a pause.

As for the duration of a specific phrase, it may vary slightly from instance to instance in a single concert section but it may also change drastically should the *laya* (tempo) of the vocals alter during the concert section. This could happen, for instance, when the vocalist changes from *alap*-style singing during a *bandish* to *taan*-style singing with its faster syllable rate.

In the present study, we choose the set of characteristic phrases of Alhaiya-bilawal *raga* that end on the *nyas svara Pa*. These fall in the two broad categories of ascending (GRGP) and descending (mnDP, DnDP), depending on how the final note is approached. In *raga* Alhaiya-bilawal, GRGP which is rendered for a longer duration due to the presence of 2 'G's, which is a *nyas* for the *raga,* while DnDP has presence of one *nyas* 'P. We restrict ourselves to the descending phrases (mnDP, DnDP) in this study.

3. DATABASE

We selected audio concerts of the *raga* Alhaiya-bilawal performed by well-known Hindustani *khyal* vocalists Ashwini Bhide and Manjari Asnare available at the NCPA AUTRIM archive for Music in Motion [5]. In all cases, the accompanying instruments are the *tanpura* (drone), *harmonium* and *tabla*. The section of each concert corresponding to *bandish* and *vistar* is extracted for this study. Table 2 shows the song, artiste name with other relevant details.

For further processing, the audio is converted to 16 kHz mono at 16 bits/sample. All the phrases of interest, as well as *nyas svaras* of interest, were labeled throughout the audio by a musician. Figure 1 shows a few of the extracted and manually annotated pitch contour segments for the characteristic phrases of Table 2. We observe the complexity of phrase intonation, and variability in the melodic contour of the phrase even within the concert.

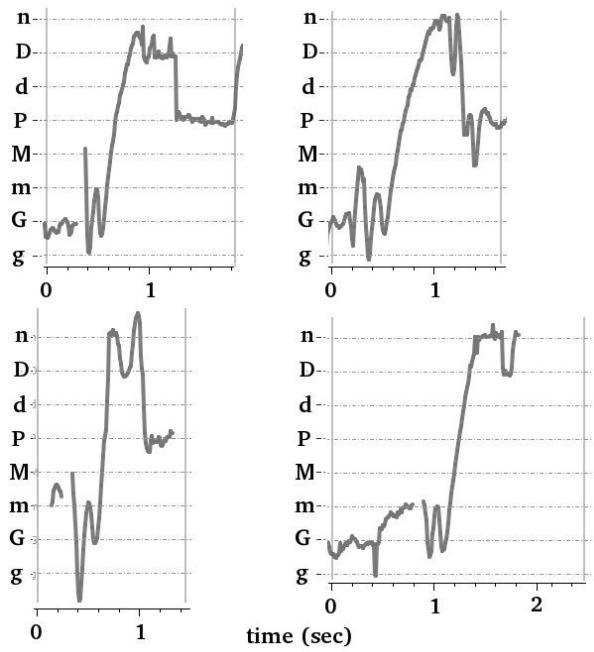

Figure 1. mnDP phrases in *Kavana Batariyaa* by Ashwini Bhide

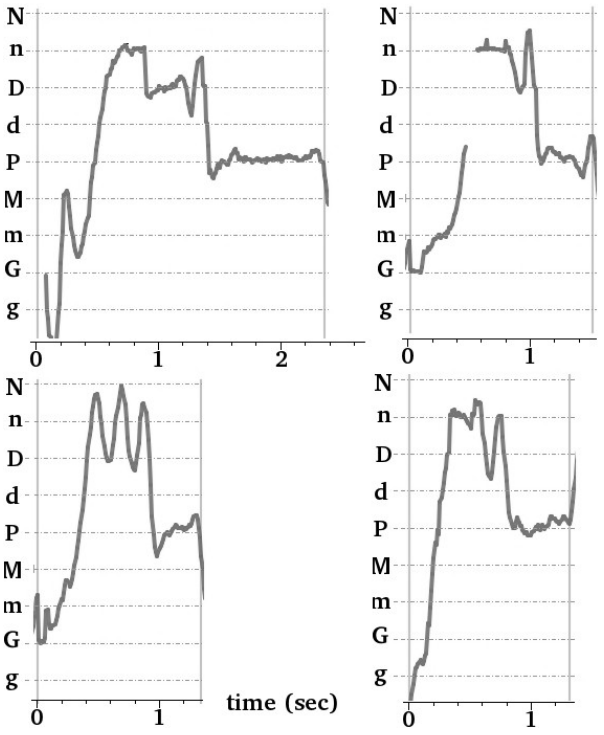

Figure 2. mnDP phrases in *Dainyaa Kaahaan* by Manjiri Asanare

Song ID	Artiste	*Raga*	*Tala*	*Bandish*	Tempo (bpm)	Dur. (min)	#Phrases		#Candidate Phrases
							DnDP	*mnDP*	
AB	Ashwini Bhide	Alhaiya-bilawal	Tintal	Kavana Batariyaa	128	8.85	15	31	67
MA	Manjiri Asanare	Alhaiya-bilawal	Tintal	Dainyaa Kaahaan	33	6.9	12	11	40

Table 2. Description of database

4. AUTOMATIC PHRASE DETECTION

The melodic contour is extracted by applying predominant F0 detection over the entire audio track of interest. Next, we propose to detect and label the characteristic phrases of the raga by computations on the extracted pitch track. Our approach to phrase detection assumes the availability of one or more reference templates in terms of the segmented pitch contour for each raga-characteristic phrase of interest.

4.1 Vocal pitch detection

In Hindustani classical vocal music, the accompanying instruments include the drone (*tanpura*), *tabla*, and often, the *harmonium* as well. The singing voice is usually dominant and the melody can be extracted from the detected pitch of the predominant source in the polyphonic mix. Melody detection involves identifying the vocal segments and tracking the pitch of the vocalist. The drone and *harmonium* are strongly pitched instruments. We therefore employ a predominant-F0 extraction algorithm designed for robustness in the presence of pitched accompaniment. This method is based on the detection of spectral harmonics helping to identify multiple pitch candidates in each 10 ms interval of the audio [6]. Finally, the predominant F0 is selected based on a combination of temporal and spectral constraints.

4.2 *Nyas svara* detection

Figure 3 shows some examples of manually annotated *nyas-svara* superposed on the extracted pitch contour. We observe the longer duration as well as the lower intra-note pitch fluctuations in the *nyas-svara* relative to the other notes in the phrase. These properties can be exploited for automatic detection as follows.

Since we are focused on P-*nyas* phrases, we scan the pitch contour for segments over which there is a less than 50 cents deviation from the P *svara* value (or the fifth with respect to the tonic) over at least a 100 ms duration. The 150 ms following this segment are checked for the same constraint except that now excursions outside the 50 cents range but limited to within 20 ms are permitted. The latter takes care of occasional pitch tracking errors. Any gaps (silences) are included within the 150 ms. A segment that satisfies these criteria is labeled P-*nyas svara*.

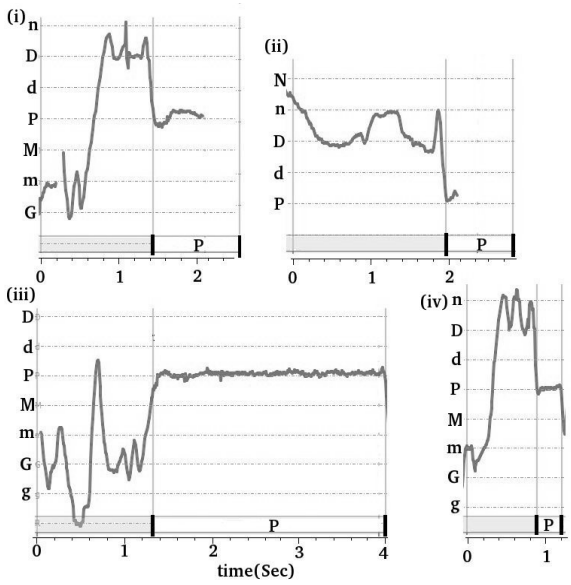

Figure 3. P *nyas-svara* at ending of different phrases. (i) mnDP in AB (ii) DnDP in AB (iii) GRGP in MA (iv) mnDP in MA

4.3 Phrase identification

The *nyas-svara* detected as above help to locate the ending boundaries of candidate phrases for the next step of phrase identification. The next task is to classify each candidate phrase into one or none of the characteristic phrases ending with the specific *nyas-svara*. For instance, P-*nyas* ending phrases could be one of GRGP, mnDP or DnDP in *raga* Alhaiya-bilawal as seen from Table 3.

We assume that we have reference templates available in the form of pitch contour segments extracted from a set of manually annotated phrases. Phrase identification then would essentially involve the computation of a similarity metric between a reference template and each of the candidate pitch segments. We retain the full pitch contour computed by the pitch detection step at 10 ms intervals throughout the vocal region corresponding to a candidate phrase location. While the phrase-ending is reliably located by the *nyas-svara* detection, the phrase beginning may or may not be linked to particular cues. In some cases, the phrase may begin from silence; however it could also begin in continuation of the end of a previously sung phrase or other melodic entity. Due to the expected variability in the phrase duration across the con-

Song	# 'P' *nyas* detected	# 'P' *nyas* associated with characteristic phrases	# Characteristic phrases Ending with 'P'(but not with 'P' *nyas*)
AB	67	44	7
MA	40	27	2

Table 3. Performance of *nyas svara* detection

cert section, the actual duration of the reference template is not useful in phrase segmentation. Further, the phrase similarity measure should take time-warping into account.

In view of the above, a set of candidate phrases is generated from each detected phrase ending by back tracing the phrase beginning to various instances at distances of less than the reference template to twice the duration of the reference template. A constrained dynamic time-warping (DTW) [7] based similarity measure is computed between the reference template and each member of the set. The minimum distance obtained serves as the estimated distance between the reference template and the candidate phrase. Applying a threshold to the estimated distance provides for the decision on phrase detection.

5. EXPERIMENTAL EVALUATION

The evaluation of methods and similarity measure was done with database described in Table 2. We separately evaluate the *nyas svara* detection algorithm followed by the similarity matching of the reference phrase with the detected candidate phrases. The similarity matching is tested on DnDP and mnDP phrases in the songs. The number of manually annotated instances of these phrases is mentioned in Table 2, which serve as the template phrases.

Nyas svara identification is the key to identify right candidates. Table 3 summarizes the results of *nyas svara* detection and details of phrases which ends at 'P' *nyas* and which do not. The performances by Ashwini Bhide and Manjiri Asanare in *raga* Alhaiya-bilawal which we have taken for the experiments are used for evaluation of *nyas svara* identification also. As the phrases considered ends with 'P' *nyas,* the *nyas* identification is evaluated on 'P' *nyas* in the songs mentioned. Identification task is able identify all the *nyas svaras* which appear at the end of the characteristic phrases in the song. Table 3 columns describe the number of 'P' *nyas* identified in the song, the phrases in the song which end with a 'P' *nyas* and the phrases which end with short duration 'P' *svara* respectively.

The candidate phrases are extracted with reference to the location of *nyas svaras* identified. As the characteristic phrases of interest to the experiments end with P *nyas svara*, candidate phrases are extracted from the locations of P *nyas.* As discussed in Section 4.3 candidate phrases with variable duration are generated from each *nyas* identified. The selected durations span the range from 1 second to 3 seconds with step-size of 0.1 second.

From each set of candidate phrases associated with a *nyas svara*, the candidate phrase having the least distance with the template phrase is considered for further processing. This facilitates finding the best candidate phrase associated with a *nyas svara* considering the fact that similar phrases may vary with respect to duration also. While evaluating with a specific characteristic phrase in the song, the positive phrases from candidate phrases are identified from the ground truth annotation of the phrase. The rest of the candidate phrases are identified as negative candidates for the experiment. Distance between all the available instances of a characteristic phrase in the annotation and the candidate phrases are computed. The number of positive candidate phrases may be less than the number of annotated positive phrases when certain positive phrases do not end with a *nyas svara*. For the phrase DnDP in performance by Ashwini Bhide, phrase identification is evaluated on 15x11=165 postive pairs and 15x55=825 negative pairs (i.e., each positive with all negatives).

Table 4 summarizes these experiments with the songs and phrases described in Table 2 along with information on positive and negative pairs. All the experiments evaluate within-concert phrase identification of instances of characteristic phrases given a reference phrase. The first experiment performs identification of candidate phrases similar to DnDP phrase in the performance by Ashwini Bhide. The false alarm rate computed for all experiments is for a fixed hit rate. DTW measure computes the similarity between the template and candidate phrases which are represented as continuous pitch values without applying any pitch quantization. Sakoe-chiba constraint [8] applied to DTW forbids pathological warpings. This helps to reduce matching between very temporally separated pitch instants in phrases.

In the distance computation, differences in pitch values less than 25 cents (quarter tone) are considered to be 0. We also tried to quantize the pitch to 12 semitones / octave before distance computation. However, as can be expected from the example contours seen in Figures 1 and 2, the results are poor due to the loss of details related to the transitory nature of *svara* realization within the phrase.

Song	Phrase	#Phrases		HR	FA
		POS	NEG		
AB	DnDP	165	825	0.95	0.21
	mnDP	868	1178	0.95	0.09
MA	DnDP	132	336	0.95	0.12
	mnDP	110	319	0.95	0.09

Table 4. Performance of motif detection for different phrases. HR= hit rate; FA= false alarm rate

Figure 4 shows distance distribution of mnDP phrase and Figure 5 shows distance distribution in DnDP phrase in performance by Ashwini Bhide. Distances between the positive phrases are clustered better in mnDP distribution than in the DnDP distribution. Also the negative distribution overlaps more with the positive distribution in DnDP distribution. The variation within the phrases of mnDP is less when compared to DnDP in this performance. mnDP being a part of *mukhda* phrase in this *bandish* is expected to be rendered with less amount of variations. The ROC curve in the Figure 6 shows variation of hit rate and false alarm rate for different decision thresholds for the experiment with mnDP phrase in performance by Ashwini Bhide.

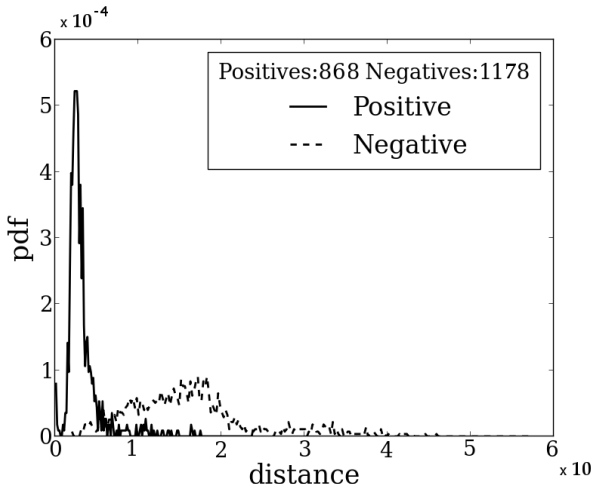

Figure 4. DTW distance distribution for mnDP phrase in AB

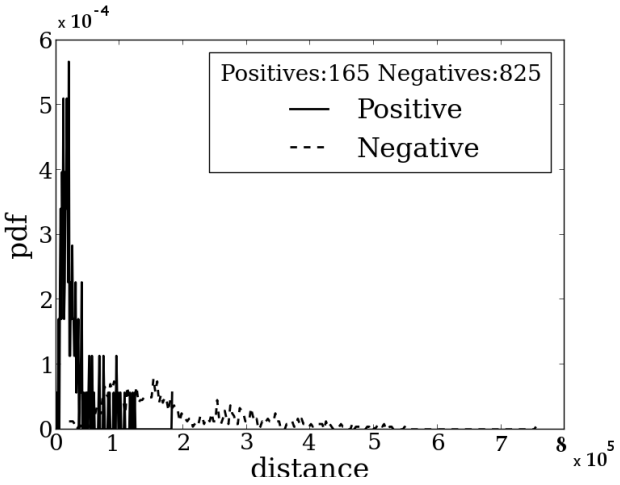

Figure 5. DTW distance distribution for DnDP phrase in AB

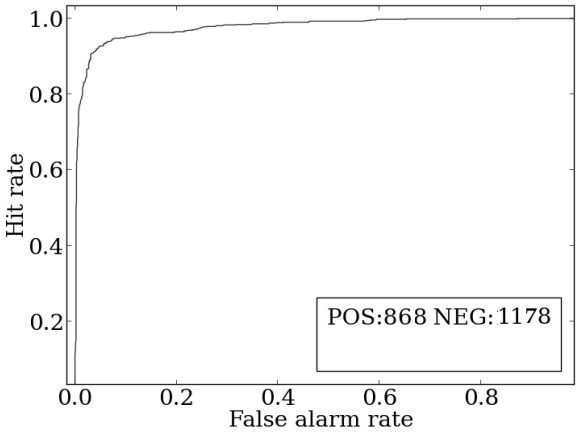

Figure 6. ROC curve for mnDP phrase(AB) distance distribution

6. FUTURE WORK

Variations with respect to time and pitch between the phrases of the same kind make the phrase identification task challenging. The current methods need to be tested on the other characteristic phrases of the *raga* esp. the longer (and visibly more variable) GRGP phrase. Further, the current method checks for a matching phrase at a *nyas* location by extracting candidate phrases of variable length around a *nyas*. A sub-sequence search in a single wider window around the identified *nyas* should find the exact matching phrase more efficiently. A sub-sequence search giving more weights to the invariant segments within the characteristic phrase could lead to better clustering of similar phrases. Finally, an attribute-based matching (rather than the direct matching of pitch values) could serve to achieve robustness to variations due to improvisation, as long as the invariant attributes can be identified e.g. specific *svara* intonations, oscillations or glide transitions.

7. REFERENCES

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

[2] 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.

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

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

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

[6] V. Rao, P. Gaddipati and P. Rao: "Signal-driven Window-length Adaptation for Sinusoid Detection in Polyphonic Music," *IEEE Trans. Audio, Speech, and Language Processing,* Vol. 20, No.1, 2012.

[7] D. Berndt and J. Clifford: "Using Dynamic Time Warping to Find Patterns in Time Series," *AAAI-94 Workshop on Knowledge Discovery in Databases,* 1994.

[8] H. Sakoe and S. Chiba: "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *Acoustics, Speech and Signal Processing, 19 IEEE Transactions on*, Vol.26, No.1, 1978.