MOTIF SPOTTING IN AN *ALAPANA* IN CARNATIC MUSIC

Vignesh Ishwar
Dept. of
Computer Sci. & Engg.
IIT Madras, India
ishwar@cse.iitm.ac.in

Shrey Dutta
Dept. of
Computer Sci. & Engg.
IIT Madras, India
shrey@cse.iitm.ac.in

Ashwin Bellur
Dept. of
Electrical Engg.
IIT Madras, India
ashwin@ee.iitm.ac.in

Hema A Murthy
Dept. of
Computer Sci. & Engg.
IIT Madras, India
hema@cse.iitm.ac.in

ABSTRACT

This work addresses the problem of melodic motif spotting, given a query, in Carnatic music. Melody in Carnatic music is based on the concept of *raga*. Melodic motifs are signature phrases which give a *raga* its identity. They are also the fundamental units that enable extempore elaborations of a *raga*. In this paper, an attempt is made to spot typical melodic motifs of a *raga* queried in a musical piece using a two pass dynamic programming approach, with pitch as the basic feature. In the first pass, the rough longest common subsequence (RLCS) matching is performed between the saddle points of the pitch contours of the reference motif and the musical piece. These saddle points corresponding to quasi-stationary points of the motifs, are relevant entities of the *raga*. Multiple sequences are identified in this step, not all of which correspond to the the motif that is queried. To reduce the false alarms, in the second pass a fine search using RLCS is performed between the continuous pitch contours of the reference motif and the subsequences obtained in the first pass. The proposed methodology is validated by testing on Alapanas of 20 different musicians.

1. INTRODUCTION

Carnatic music is a sub-genre of Indian classical music prominent in south India. It is a heterophonic musical form which involves multiple instruments performing at the same time along with the voice. This form of music is highly melody centric and thrives on the melodic concept of *ragas*. The *ragas* in Carnatic music consist of a set of inflected musical notes called *svaras*. *Svaras* are the Indian classical equivalent to the solfege and are annotated as *Sa Ri Ga Ma Pa Da Ni Sa*. Theoretically frequencies of *svaras* follow more or less the just intonation ratios unlike the notes in western classical music which follow the equi-temperament scale [16]. In Carnatic music, the *svaras* are seldom rendered as discrete notes. They are rendered as a seamless meandering across the notes using *gamakas* [1] [7, 9, 17]. Figure 1 shows a pitch histogram of two *ragas* namely, *Kalyani* and *Sankarabharana*. Observe that there is a significant band of frequencies around every peak that are also frequent. This band of frequencies correspond to a single *svara* and is referred to as the intonation of that *svara*. This characteristic is observed due to the inflected nature of the *svaras*. It has been conjectured by musicians that a quantization of the pitch on the basis of the absolute frequencies of the *svaras* is erroneous [7].

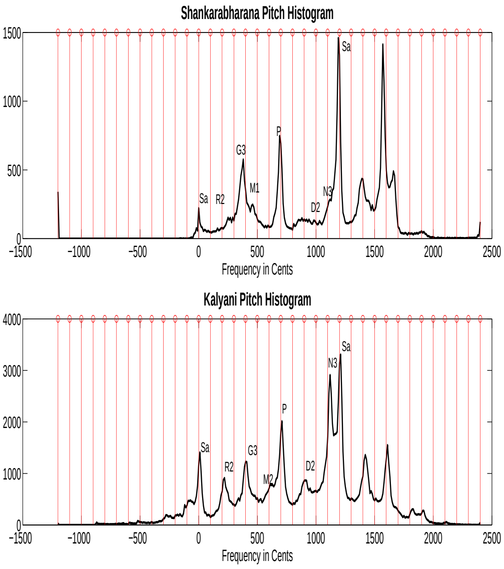

Figure 1. Pitch Histograms of Ragas Kalyani and Shankarabharana

A *raga* in Carnatic music can be characterised by a set of distinctive motifs. Distinctive motifs can be characterised by the trajectory of inflected *svaras* over time. These motifs are of utmost aesthetic importance to the *raga*. Carnatic music is a genre abundant with compositions. These compositions are replete with many distinctive motifs. These motifs are used as building blocks for extempore improvisational pieces in Carnatic music. These motifs can also be used for distinguishing between two *ragas*, and also for archival and learning purposes. The objective of this paper is to spot the location of the

[1] *Gamaka* is a meandering of a note encompassing other permissible frequencies around it.

distinctive motifs in an extempore enunciation of a *raga* called the *Alapana*. For more details on the concepts of *svara, gamaka, phraseology* pertaining to Carnatic music, the reader is advised to refer to [7].

In this paper, pitch is used as the main feature for the task of motif spotting. Substantial research exists on analysing different aspects of Carnatic music computationally, using pitch as a feature. Krishnaswamy et al [8], characterize and analyse *gamakas* using pitch contours. Serra et al [16] study the tuning of Indian classical music using pitch histograms. M. Subramaniam [17] has extensively studied the motifs in the raga Thodi using pitch histograms and pitch contours. All of the above prove the relevance and importance of pitch as a feature for computational analysis of Carnatic music.

In the previous work, the uniqueness of the characteristic motifs was established using a closed set motif recognition experiment using HMMs. As a continuation, in this work an attempt is made to spot motifs given a long *Alapana* interspersed with motifs. Time series motif recognition has been attempted for Hindustani music. J.C. Ross et. al. [13] use the onset point of the rhythmic cycle [2] emphasized by the beat of the tabla (an Indian percussion instrument) as a cue for potential motif regions. In another work, J.C.Ross et. al. [14] attempt motif spotting in a *Bandish* (a type of composition in Hindustani music) using elongated notes(*nyaas svara*).

Spotting motifs in a *raga alapana* is equivalent to finding a subsequence in trajectory space. Interestingly, the duration of these motifs may vary, but the relative duration of the *svaras* is preserved across the motif. The attempt in this work is to use pitch contours as a time series and employ time series pattern capturing techniques to identify the motif. The techniques are customized using the properties of the music. There has been work done on time series motif recognition in fields other than music. Pranav et. al. in their work [11] define a time series motif and attempt motif discovery using the EMMA algorithm. In [3, 18], time series motifs are discovered adapting the random projection algorithm by Buhler and Tompa to time series data. In [2], a new warping distance called Spatial Assembling distance is defined and used for pattern matching in streaming data. In the work of Hwei-Jen Lin et. al. [10], music matching is attempted using a variant of the Longest Common Subsequence (LCS) algorithm called Rough Longest Common Subsequence. This paper attempts similar time series motif matching for Carnatic Music. Searching for a 2-3 second motif (in terms of a pitch contour) in a 10 min *Alapana* (also represented as a pitch contour) can be erroneous, owing pitch estimation errors. To address this issue, the pitch contour of the *Alapana* is first quantized to a sequence of quasistationary points which are meaningful in the context of a *raga*. A two-pass search is performed to determine the location of the motif. In the first pass, a Rough Longest Common Subsequence approach is used to find the region corresponding to the location of the motif. Once the region is located, another pass is made using a fine-grained RLCS algorithm using the raw pitch contour.

[2] The rhythmic cycle in Indian Music is called Tala and the onset point is called the Sam.

The paper is organised as follows. Section 2 discusses the approach employed to extract the stationary points in a *raga*. Section 3 discusses the algorithm to perform the two-level RLCS approach to spot the motif. In Section 3.1 the Rough Longest Common Subsequence(RLCS) algorithm is discussed. In Section 4, the database used in the study is discussed. Section 5 discusses the results. Finally, conclusions are presented in Section 6.

2. SADDLE POINTS

2.1 Saddle Points: Reducing the Search Space

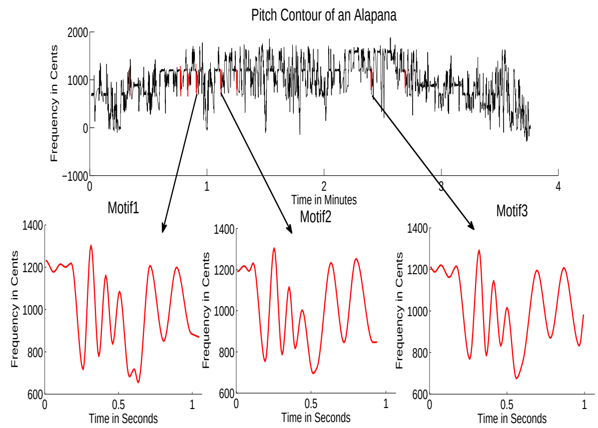

Figure 2. a) Motifs Interspersed in an Alapana ; b) Magnified Motif

The task of this paper is to attempt automatic spotting of a motif that is queried. The motif is queried against a set of *Alapanas* of a particular *raga* to obtain locations of the occurrences of the motif. The task is non-trivial since in *Alapanas*, rhythm is not maintained by a percussion instrument. Figure 2 (a) shows repetitive occurrences of motifs in a piece of music. An enlarged view of the motif is given in Figure 2(b). Since the *Alapana* is much longer than the motif, searching for a motif in an *Alapana* is like searching for a needle in a haystack. After an analysis of the pitch contours and discussions with professional musicians, it was conjectured that the pitch contour can be quantized at saddle points. Figure 3 shows an example phrase of the *raga* Kamboji with the saddle points highlighted.

Musically, the saddle points are a measure of the extent to which a particular *svara* is intoned. In Carnatic music since *svaras* are rendered with *gamakas*, there is a difference between the notation and the actual rendition of the phrase. However, there is a one to one correspondence with the saddle point frequencies and what is actually rendered by the musician (Figure 3). Figure 4 shows the pitch histogram and the saddle point histogram of an *Alapana* of the *raga Kamboji*. The similarity between the two pitch histograms indicates our conjecture that saddle points are important.

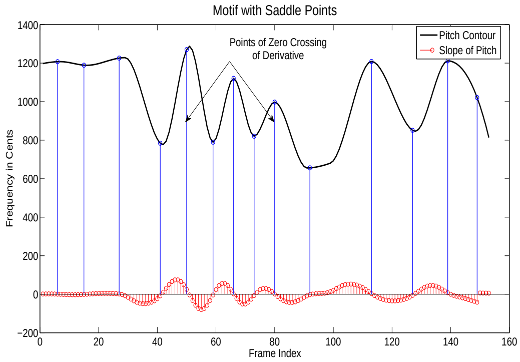

Figure 3. A Phrase with Saddle Points

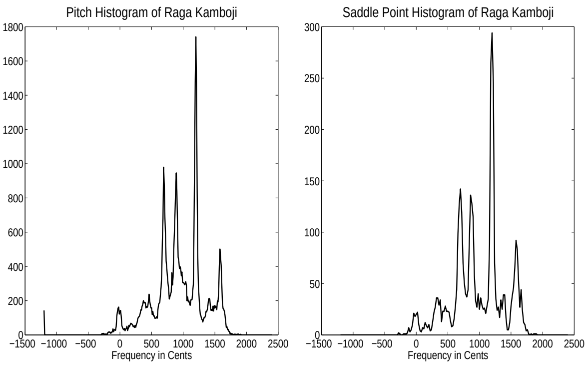

Figure 4. The Pitch and Saddle Point Histograms of the raga Kamboji

2.2 Method of obtaining Saddle Points

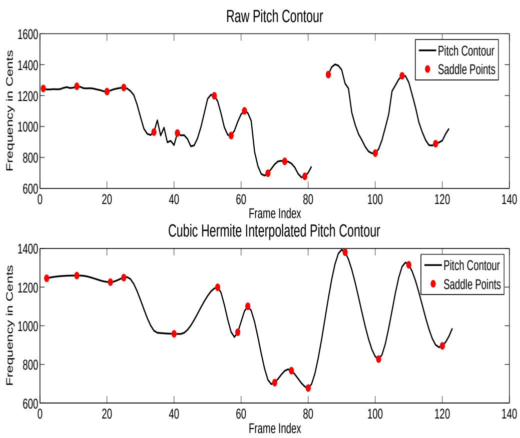

Figure 5. Distorted and Cubic Interpolated pitch contours

Carnatic music is a heterophonic musical form. In a Carnatic music concert, a minimum of two accompanying instruments play simultaneously along with the lead artist. These are the violin and the *mridangam* (a percussion instrument in Carnatic music). Carnatic music is performed at a constant tonic [1] to which all instruments are tuned. This tonic is chosen by the lead artist and is provided by an instrument called the *Tambura*. Thus, the simultaneous performance of many instruments in addition to the voice renders pitch extraction of the predominant voice a tough task. This leads to octave errors and other erroneous pitch values. For this task it is necessary that pitch be continuous. After experimenting with various pitch algorithms, it was observed that the Melodia-Pitch Extraction algorithm [15] produced the fewest errors. This was verified after re-synthesis using the pitch contours. In case of an octave error or any other such pitch related anomaly, the algorithm replaces the erroneous pitch values with zeros. The saddle points are obtained by processing the pitch contour extracted from the waveform. The pitch extracted is converted to the cent scale using Equation 1 to normalise with respect to the tonic of different musicians.

$$centFrequency = 1200 \cdot \log_2 \left(\frac{f}{tonic} \right) \qquad (1)$$

Least squares fit (LSF) [12] was used to compute the slope of the pitch extracted. The zero crossings of the slope correspond to the saddle points (Figure 3). A Cubic Hermite interpolation [4] was then performed with the initial estimation of saddle points to get a continuous curve (Figure 5). The saddle points are then obtained by sampling the interpolated spline using LSF. The interpolated pitch contours were validated by re-synthesizing and listening tests [3] .

3. A TWO PASS DYNAMIC PROGRAMMING SEARCH

In Section 2 it is illustrated that the sequence of saddle points are crucial for a motif. Therefore, RLCS is used to query for the saddle points of the given motif in the *Alapana*.

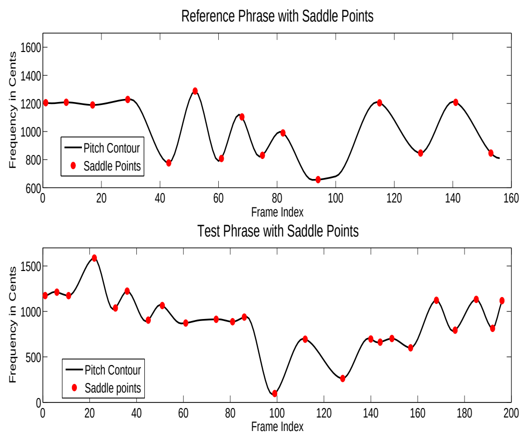

Figure 6. Similar Saddle Points, different contours

Music matching using LCS methods for western music is performed on symbolic music data [5]. The musical

[3] Original and re-synthesized waveforms are available at http://lantana.tenet.res.in/motif_analysis.html

notes in this context are the symbols. However, in the context of Carnatic music, a one to one correspondence between the notation and sung melody does not exist. Hence, in this paper, saddle points are used instead of a symbolic notation. One must keep in mind that saddle points are not symbols but are continuous pitch values (Figure 6). In order to match such pitch values, a rough match instead of an exact match is required. A variant of the LCS known as the Rough Longest Common Subsequence [10] allows such a rough match.

In this paper, a two pass RLCS matching is performed. In the first pass, the saddle points of the reference and query are matched to obtain the candidate motif regions. Nevertheless, given two saddle points, the pitch contour between two saddle points can be significantly different for different phrases (Figure 6). This leads to many false alarms. A second pass of RLCS is then performed on the regions obtained from the first pass to filter out the false alarms from the true motifs.

3.1 Algorithm for the Rough Longest Common Subsequence

RLCS is a variant of LCS which performs an approximate matching between a query and a reference retaining local similarity. Here, a cubic distance measure is used to determine the similarity between two saddle points. If the distance between the two saddle points in the query and reference is less than a threshold, *Td*, they are said to be roughly similar. In the RLCS algorithm, unlike LCS, the similarity measure (cost) is incremented by a weighted quantity (between 0 and 1) depending on the proximity of the points. To account for the local similarity, the width across query (WAQ) and width across reference (WAR) are incorporated. WAR and WAQ are the lengths of the shortest sub-strings, in the reference and query respectively, which contain the longest common subsequence. These measures convey the density of the match given a query and reference. Thus, lesser the WAQ and WAR, denser is the distribution of the RLCS and better is the alignment. The RLCS algorithm gives five matrices, namely, cost matrix, WAQ matrix, WAR matrix, score matrix and direction matrix (for tracing back the common subsequences).

3.2 First Pass: Determining Candidate Motif Regions using RLCS

The RLCS algorithm used in this paper is illustrated in this section. The *Alapana* is first windowed and then processed with the RLCS algorithm. The window size chosen for this task is 1.5 times the length of the motif queried for. The matrices obtained from the RLCS are then processed as follows:

- From the cells of the score matrix with values greater than a threshold, *seqFilterTd*, sequences are obtained by tracing the direction matrix backwards.
- The duplicate sequences which may be acquired are neglected, preserving unique sequences of length greater than a percentage, ρ, of the length of reference. These are then added to a sequence buffer.
- This process is repeated for every window. The window is shifted by a hop of one saddle point.
- The sequences obtained thus are grouped.
- Each group, taken from the first element of the first member to the last element of the last member, represents a potential motif region.

3.3 Second Pass: Determining Motifs from the Groups

In the first pass a matching of only the saddle points is performed. As mentioned above, eventhough the saddle points are matched it is not necessary that the trajectory between them match. This leads to a large number of false alarms. Now that the search space is reduced, the RLCS is performed between the entire pitch contour of the potential motif region obtained in the first pass and the motif queried. The entire pitch contour is used in order to account for the trajectory information contained in the phrases (Figure 7). The threshold *Td* used for the first pass is tightened in this iteration for better precision while matching the entire feature vector. In this iteration, the cell of the score matrix having the maximum value is chosen and the sequence is traced back using the direction matrix from this cell. This sequence is hypothesized to be the motif. The database and experimentation are detailed in the following sections.

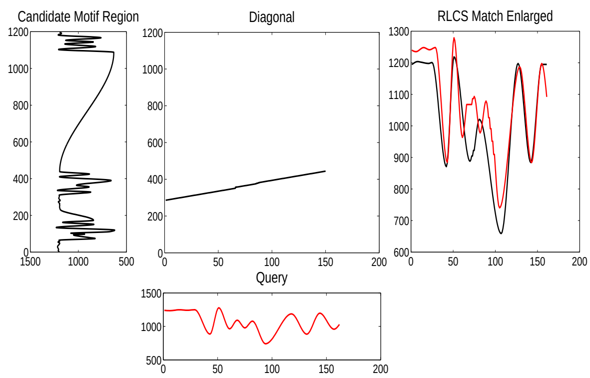

Figure 7. RLCS Matching

4. DATABASE

Table 1 gives the details of the database used in this work. As mentioned above, this task will be performed on *Alapanas*. The reason for using only *Alapana*s and not compositions for spotting is due to the limitations of the pitch extraction algorithms. The presence of multiple instruments and percussion along with the voice makes pitch extraction a non trivial task due to which pitch obtained for compositions is distorted. Pitch extracted for *Alapanas* have less distortion as compared to that of compositions.

Table 1. Database of *Alapanas*; N-Al- No. of *Alapanas*, N-Art- No. of Artists, Avg-Dur- Average Duration, Tot-Dur -Total Duration

Rāga Name	N-Al	N-Art	Avg-Dur (mins)	Tot-Dur (mins)
Kamboji	27	12	9.73	262.91
Bhairavi	21	15	10.30	216.49

Table 2. Phrases Queried

Raga Name	Phrase Notation	Average Duration(seconds)
Kamboji	S..N2 D2 P D2...	1.8837
Bhairavi	R2 G2 M1 P D1 P..	1.3213

The details of motifs of the ragas queried are given in Table 2. The average duration is obtained from the ground truth labeled in the previous work [6].

5. EXPERIMENTS AND RESULTS

RLCS was performed on the database *Alapana*s querying for the motifs of the ragas mentioned in Table 1. The distance function used for RLCS is cubic in nature with the equation given below.

$$\boxed{\delta(i,j) = \begin{cases} \frac{|i-j|^3}{300^3} & ; if\ |i-j| \leq 300 \\ 1 & ;\ otherwise \end{cases}} \quad (2)$$

Here, i and j correspond to the i^{th} saddle point in the *alapana* that matches the j^{th} saddle point in the motif. Due to different styles of various musicians, an exact match between the saddle points of the motif, and the *Alapana* cannot be expected. Hence in this paper a leeway of *extrTd = 300 cents* is allowed between two saddle points. Musically two points 300 cents (3 semitones) apart cannot be called similar, but in this case, due to the different styles of singing and artefacts introduced due to pitch extraction, the saddle points for the same phrase sung by different artists does not match exactly but is generally within a limited range of frequencies. The upper limit of this range is set to 300 cents empirically.

In this work, the phrases sung across octaves are ignored. For this experiment the parameters set were as follows: *Td* = 0.45 ; ρ = 0.8; *seqFilterTd* = 0.45. The parameter $0 < \rho < 1$ is a user defined parameter that ensures that $\rho\times$ *length of the query motif* is matched with that of the *Alapana*. As a sanity check, the regions obtained from the RLCS were verified with the ground truth motifs labelled for the experiment in the previous work [6]. The parameters were tuned to retrieve as many labelled ground truth motifs as possible for the *raga Kamboji*'s phrase. A high percentage of regions coinciding with the ground truth labelled, were retrieved by the RLCS [4] .

The regions obtained from the RLCS were then subjected to a listening test performed by three professional musicians. The regions which contained the phrase queried were marked as true, and those which did not as false. It is observed that the correlation with respect to the carefully marked ground truth by one musician with verification by two other musicians is an average of 0.833. The details of the number of ground truth motifs retrieved and the total number of trues retrieved after verification by the listening test are given in Table 3. The number of false positives, retrieved are however substantial. This is affordable since the objective in the first pass is to obtain the maximum number of the regions similar to the motif. The second iteration of RLCS is performed to filter out the false positives. Now that the candidate motif regions are known, the second pass of RLCS is conducted wherein the same motif from four different artists are queried in the regions retrieved by the first pass. The entire pitch contour of the query and reference are used for this task in order to account for the information of trajectory of pitches between the saddle points. Since the task is to locate the motif in a smaller continuous search space, the threshold *extrTd* was tightened and the allowable leeway for the distance function was reduced to 200 cents (See Eq 2). The parameter *seqFilterTd* is not used in the second pass since the best match with the candidate region is sought (See Section 3.2). The RLCS is repeated for four examples of the same motif by different musicians and the scores are obtained.

Table 3. Retrieved regions First Pass: TR-Total Retrieved, LGT-Labelled Ground Truth, TrR-True Retrieved, PR-Percentage Retrieved

Rāga Name	No of Apalanas	TR	LGT	TrR	PR
Kamboji	27	719	70	58	82.86%
Bhairavi	20	474	103	91	88.35%

5.1 Evaluation and Discussion

This work illustrates the method of spotting a query motif in an improvisational form of Carnatic music called an *alapana*. The requirement of a query motif for such a search is due to the scarce rendition of certain characteristic phrases in an *alapana*. The spotting of such phrases proves to be useful to musicians and students for analysis purposes. To quantify this and evaluate the performance of the algorithm in this context, it was decided to compute the recall, precision and F2-Measure for the motifs retrieved. The F2-Measure was chosen in order to give a higher weightage to the recall. The objective of this work being music exploration through motif spotting, the recall of the motif queried is of greater importance.

The hits obtained in the second pass are sorted according to the RLCS scores. The precision, recall and F2-Measure per *Alapana* are calculated and the average is computed across all *Alapanas*. The motifs are not exact since they correspond to the extempore enunciation by an artist. The results are reported per query, per *alapana*. The hits of the RLCS in an *alapana* are sorted according to their scores and the precision, recall and F2-Measure are calculated for the top 10 sorted hits. The relevant motifs

[4] In the previous experiment not all occurrences of the phrase were labelled. The instrumental phrases were ignored and some were missed due to manual error.

are all the motifs which were marked as true in that *alapana*. The results are illustrated in Table 4. The experiments on the phrase of the *raga Kamboji* were treated as the development set, with parameters optimised to maximise the match. To verify that this works in general, another *raga Bhairavi* was taken up for study. The same parameters used for the *raga Kamboji* were used for the *raga Bhairavi*. The results are described in the Tables 3 and 4. From the results obtained above, it is clear that eventhough the precision is low, the recall is high in most of the cases. Certain partial matches are also obtained where either the first part of the query is matched or the end of the query is matched. These are movements similar to those of the phrases and are interesting for a listener, learner, or researcher. High scores were obtained for certain false alarms. This is primarily due to some significant similarity between the false alarm and the original phrase.

Table 4. Results: Average Precision - Pr%, Recall - Rec%, F2-Measure - F2M% across *alapanas*, Average of Query Scores - AQS

Queries	*Kamboji*-27 *Alapanas*			*Bhairavi*-20 *Alapanas*		
	Pr	Rec	F2M	Pr	Rec	F2M
Query 1	43.18	81.52	64.19	43.13	94.80	71.94
Query 2	33.33	63.318	50.50	38.15	85.10	64.33
Query 3	43.63	83.16	65.28	36.87	81.04	61.04
Query 4	25.90	58.62	42.85	38.12	83.33	63.07
AQS	40.45	76.00	60.00	41.25	91.04	68.91

6. CONCLUSION

In this work, RLCS is used for motif discovery in *alapanas* in Carnatic music. It is illustrated that the saddle points of the pitch contour of a musical piece hold significant music information. It is then shown that quantizing the pitch contour of the *alapana* at the saddle points leads to no loss of information while it results in a significant reduction in the search space. The RLCS method is shown to give a high recall for the motif queried. Given that the objective is to explore the musical traits of a *raga* by spotting interesting melodic motifs rendered by various artists, the recall of the motif queried is of higher importance than the precision. The future work would involve spotting of motifs occurring across multiple octaves. It would also be interesting see if similarity measures can be obtained by this approach across ragas.

7. ACKNOWLEDGEMENTS

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

8. REFERENCES

[1] Ashwin Bellur and Hema A Murthy. A cepstrum based approach for identifying tonic pitch in indian classical music. In *National Conference on Communications*, 2013.

[2] Yueguo Chen, Mario A. Nascimento, Beng Chin, Ooi Anthony, and K. H. Tung. Spade: On shape-based pattern detection in streaming time series. *in ICDE, 2007*, pages 786–795, 2007.

[3] Bill Chiu, Eamonn Keogh, and Stefano Lonardi. *Probabilistic discovery of time series motifs*. 2003.

[4] F. N. Fritsch and R. E. Carlson. *Monotone Piecewise Cubic Interpolation*. SIAM Journal on Numerical Analysis, Vol. 17, No. 2., 1980.

[5] F Scholer I S H Suyoto, A L Uitdenbogerd. Searching musical audio using symbolic queries audio, speech, and language processing. *IEEE Transactions on In Audio, Speech, and Language Processing, IEEE Transactions on, Vol. 16, No. 2.*, pages 372–381, 2008.

[6] Vignesh Ishwar, Ashwin Bellur, and Hema A Murthy. Motivic analysis and its relevance to raga identification in carnatic music. In *Workshop on Computer Music*, Instanbul, Turkey, July 2012.

[7] T M Krishna and Vignesh Ishwar. Svaras, gamaka, motif and raga identity. In *Workshop on Computer Music*, Instanbul, Turkey, July 2012.

[8] A Krishnaswamy. Application of pitch tracking to south indian classical music. *In Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 557–560, 2003.

[9] A Krishnaswamy. Inflexions and microtonality in south indian classical music. *Frontiers of Research on Speech and Music*, 2004.

[10] HWEI-JEN LIN, HUNG-HSUAN WU, and CHUN-WEI WANG. Music matching based on rough longest common subsequence. *Journal of Information Science and Engineering*, pages 27, 95–110., 2011.

[11] Pranav Patel, Eamonn Keogh, Jessica Lin, and Stefano Lonardi. Mining motifs in massive time series databases. *In Proceedings of IEEE International Conference on Data Mining (ICDM)*, pages 370–377, 2002.

[12] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing. Second Edition*. Oxford University Press, 1992.

[13] Joe Cheri Ross and Preeti Rao. Detecting melodic motifs from audio for hindustani classic music. In *ISMIR*, Portugal, October 2012.

[14] Joe Cheri Ross and Preeti Rao. Detection of raga-characteristic phrases from hindustani classical music audio. *Proc. of the 2nd CompMusic Workshop*, July 12-13, 2012.

[15] J. Salamon and E. Gomez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech and Language Processing*, pages 20(6):1759–1770, Aug. 2012.

[16] J Serra, G K Koduri, M Miron, and X Serra. Tuning of sung indian classical music. *In Proc. of ISMIR*, pages 157–162, 2011.

[17] M Subramanian. Carnatic ragam thodi - pitch analysis of notes and gamakams. *in Journal of the Sangeet Natak Akademi*, pages 3–28, 2007.

[18] Dragomir Yankov, Eamonn Keogh, Jose Medina, Bill Chiu, and Victor Zordan. Detecting time series motifs under uniform scaling. *In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, August 12-15, 2007*, 2007.