IMPROVING MELODIC SIMILARITY IN INDIAN ART MUSIC USING CULTURE-SPECIFIC MELODIC CHARACTERISTICS

Sankalp Gulati[†], Joan Serrà[⋆] and Xavier Serra[†]
[†]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[⋆]Telefonica Research, Barcelona, Spain
`sankalp.gulati@upf.edu, joan.serra@telefonica.com, xavier.serra@upf.edu`

ABSTRACT

Detecting the occurrences of rāgs' characteristic melodic phrases from polyphonic audio recordings is a fundamental task for the analysis and retrieval of Indian art music. We propose an abstraction process and a complexity weighting scheme which improve melodic similarity by exploiting specific melodic characteristics in this music. In addition, we propose a tetrachord normalization to handle transposed phrase occurrences. The melodic abstraction is based on the partial transcription of the steady regions in the melody, followed by a duration truncation step. The proposed complexity weighting accounts for the differences in the melodic complexities of the phrases, a crucial aspect known to distinguish phrases in Carnatic music. For evaluation we use over 5 hours of audio data comprising 625 annotated melodic phrases belonging to 10 different phrase categories. Results show that the proposed melodic abstraction and complexity weighting schemes significantly improve the phrase detection accuracy, and that tetrachord normalization is a successful strategy for dealing with transposed phrase occurrences in Carnatic music. In the future, it would be worthwhile to explore the applicability of the proposed approach to other melody dominant music traditions such as Flamenco, Beijing opera and Turkish Makam music.

1. INTRODUCTION

The automatic assessment of melodic similarity is one of the most researched topics in music information research (MIR) [3,14,30]. Melodic similarity models may vary considerably depending on the type of music material (sheet music or polyphonic audio recordings) [4, 8, 22] and the music tradition [5, 18]. Results until now indicate that the important characteristics of several melody-dominant music traditions of the world such as Flamenco and Indian art music (IAM) need dedicated research efforts to devise specific approaches for computing melodic similarity [23,24]. These music traditions have large audio music repertoires but comparatively very fewer number of descriptive scores (they follow an oral transmission), the automatic detection of the occurrences of a melodic phrase in audio recordings is therefore a task of primary importance. In this article, we focus on this task for IAM.

Hindustani music (also referred to as north Indian art music) and Carnatic music (also referred to as south Indian art music) are the two art music traditions of India [6, 31]. Both are heterophonic in nature, with melody as the dominant aspect of the music. A typical piece has a main melody being sung or played by the lead artist and a melodic accompaniment with the tonic pitch as the base reference frequency [9]. *Rāg* is the melodic framework and *tāl* is the rhythm framework in both music traditions. Rāgs are characterized by their constituent *svars* (roughly speaking, notes), by the *āroh-avroh* (the ascending and descending melodic progression) and, most importantly, by a set of characteristic melodic or 'catch' phrases. These phrases are the prominent cues for rāg identification used by the performer, to establish the identity of a rāg, and also the listener, to recognize the rāg.

The characteristic melodic phrases of a rāg act as the basis for the artists to improvise, providing them with a medium to express creativity during a rāg rendition. Hence, the surface representation of these melodic phrases can vary a lot across their occurrences. This high degree of variability in terms of the duration of a phrase, non-linear time warpings and the added melodic ornaments together pose a big challenge for melodic similarity computation in IAM. In Figure 1 we illustrate this variability by showing the pitch contours of the different occurrences of three characteristic melodic phrases of the rāg Alaiya Bilawal. We can clearly see that the duration of a phrase across its occurrences varies a lot and the steady melodic regions are highly varied in terms of the duration and the presence of melodic ornaments. Because of these and other factors, detecting the occurrences of characteristic melodic phrases becomes a challenging task. Ideally, the melodic similarity measure should be robust to a high degree of variation and, at the same time, it should be able to discriminate between different phrase categories and irrelevant melodic fragments (noise candidates).

For melodic similarity computation, the string matching-based and the set point-based approaches are extensively used for both musical scores and audio recordings [30]. However, compared to the former, the set point-based approaches are yet to be fully exploited for polyphonic audio music because of the challenges involved in melody extrac-

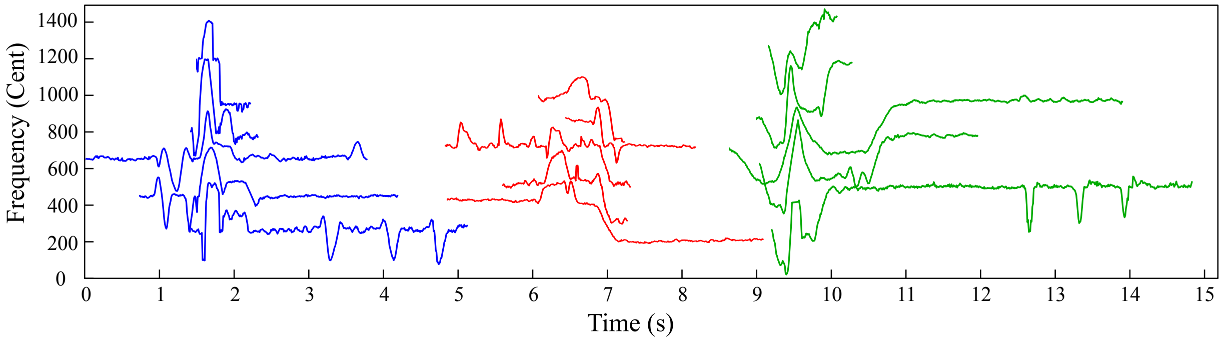

Figure 1. Pitch contours of occurrences of three different characteristic melodic phrases in Hindustani music. Contours are frequency transposed and time shifted for a better visualization.

tion and transcription [4]. A reliable melody transcription algorithm is argued to be the key to bridge the gap between audio and symbolic music, leading to the full exploitation of the potential of the set point-based approaches for audio music. However, for several music traditions such as Hindustani and Carnatic music, automatic melody transcription is a challenging and a rather ill-defined task [25].

In recent years, several methods for retrieving different types of melodic phrases have been proposed for IAM, following both supervised and unsupervised strategies [7, 12, 13, 16, 17, 24, 26, 27]. Ross et al. [27] detect the occurrences of the title phrases of a composition within a concert recording of Hindustani music. The authors explored a SAX-based representation [20] along with several pitch quantizations of the melody and showed that a dissimilarity measure based on dynamic time warping (DTW) is preferred over the Euclidean distance. Noticeably, in that work, the underlying rhythm structure was exploited to reduce the search space for detecting pattern occurrences. An extension of that approach [26] pruned the search space by employing a melodic landmark called nyās svar [11].

Rao et al. [24] address the challenge of a large within-class variability in the occurrences of the characteristic phrases. They propose to use exemplar-based matching after vector quantization-based training to obtain multiple templates for a given phrase category. In addition, the authors propose to learn the optimal DTW constraints in a previous step for each phrase category in order to exploit the possible patterns in the duration variability. For Carnatic music, Ishwar et al. [17] propose a two-stage approach for spotting the characteristic melodic phrases. The authors exploit specific melodic characteristics (saddle points) to reduce the target search space and use a distance measure based on rough longest common subsequence [19].

On the other hand, there are studies that follow an unsupervised approach for discovering melodic patterns in Carnatic music [7, 12]. Since the evaluation of melodic similarity measures is a much more challenging task in an unsupervised framework, results obtained from an exhaustive grid search of optimal distance measures and parameter values within a supervised framework are valuable [13].

In this study, we present two approaches that utilize specific melodic characteristics in IAM to improve melodic similarity. We propose a melodic abstraction process based on the partial transcription of the melodies to handle large timing variations in the occurrences of a melodic phrase. For Carnatic music we also propose a complexity weighting scheme that accounts for the differences in the melodic complexities of the phrases, a crucial aspect for melodic similarity in this music tradition. In addition, we come up with a tetrachord normalization strategy to handle the transposed occurrences of the phrases. The dataset used for the evaluation is a superset of the dataset used in a recent study [13] and contains nearly 30% more number of annotated phrases.

2. METHOD

Before we present our approach we first discuss the motivation and rationale behind it. A close examination of the occurrences of the characteristic melodic phrases in our dataset reveals that there is a pattern in the non-linear timing variations [24]. In Figure 1 we show a few occurrences of three such melodic phrases. In particular, we see that the transient regions of a melodic phrase tend to span nearly the same time duration across different occurrences, whereas the stationary regions vary a lot in terms of the duration. In Figure 2 we further illustrate this by showing two occurrences of a melodic phrase (P_{1a} and P_{2a}). The stationary svar regions are highlighted. We clearly see that the duration variation is prominent in the highlighted regions. To handle such large non-linear timing variations typically a non-constrained DTW distance measure is employed [13]. However, such a DTW variant is prone to noisy matches. Moreover, the absence of a band constraint renders it inefficient for computationally complex tasks such as motif discovery [12].

We put forward an approach that abstracts the melodic representation and reduces the extent of duration and pitch variations across the occurrences of a melodic phrase. Our approach is based on the partial transcription of the melodies. As mentioned earlier, melodic transcription in IAM is a challenging task. The main challenges arise due to the presence of non-discrete pitch movements such as smooth glides and *gamakas* [1]. However, since the duration variation exists mainly during the steady svar regions, transcribing only the stable melodic regions might be suffi-

[1] Rapid oscillatory melodic movement around a svar

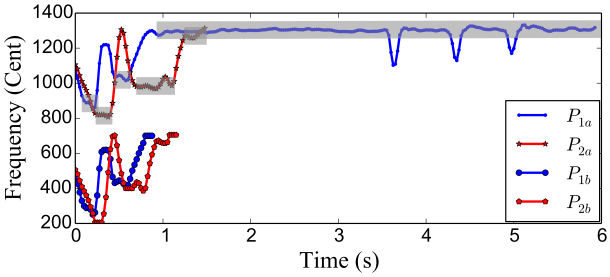

Figure 2. Original pitch contours (P_{1a}, P_{2a}) and duration truncated pitch contours (P_{1b}, P_{2b}) of two occurrences of a characteristic phrase of rāg Alhaiya Bilawal. The contours are transposed for a good visualization.

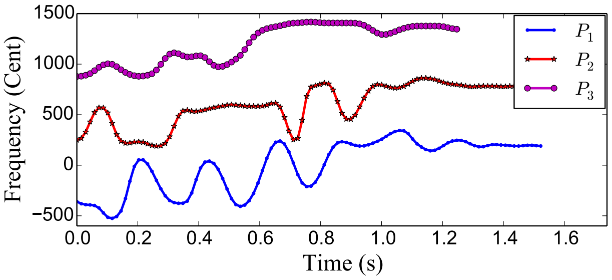

Figure 3. Pitch contours of three melodic phrases (p_1, p_2, p_3). p_1 and p_2 are the occurrences of the same characteristic phrase and both are musically dissimilar to p_3.

cient. Once transcribed, we can then truncate the duration of these steady melodic regions and hence effectively reduce the amount of timing variations across the occurrences of a melodic phrase. Additionally, since the duration truncation also reduces the overall length of a pattern, the computational time for melodic similarity computation is also reduced substantially.

The rapid oscillatory pitch movements (gamakas) in Carnatic music bring up another set of challenges for the melodic similarity computation. Very often, two musically dissimilar melodic phrases obtain a high similarity score owing to a similar pitch contour at a macro level. However, they differ significantly at a micro level. In Figure 3 we illustrate such a case where we show the pitch contours of three melodic phrases P_1, P_2 and P_3, where P_1 and P_2 are the occurrences of the same melodic phrase and both are musically dissimilar to P_3. Using the best performing variant of the similarity measure in [13] we obtain a higher similarity score between the pairs (P_1, P_3) and (P_2, P_3) compared to the score between the pair (P_1, P_2). This tendency of a high complexity time-series (higher degree of micro level variations) obtaining a high similarity score with another low complexity time-series is discussed in [1]. We follow their approach and apply a complexity weighting to account for the differences in the melodic complexities between phrases in the computation of melodic similarity.

In the subsequent sections we present our proposed approach. As a baseline in this study we consider the method that was reported as the best performing method in a recent study for the same task on a subset of the dataset [13]. We denote this baseline method by M_B.

2.1 Melody Estimation and post-processing

We represent melody of an audio signal by the pitch of the predominant melodic source. For predominant pitch estimation in Carnatic music, we use the method proposed by Salamon and Gómez [29]. This method performed favourably in MIREX 2011 (an international MIR evaluation campaign) on a variety of music genres, including IAM, and has been used in several other studies for a similar task [7, 12, 13]. An implementation of this algorithm available in Essentia [2] is used in this study. Essentia is an open-source C++ library for audio analysis and content-based MIR. We use the default values of the parameters for pitch estimation except the frame size and the hop size, which are set to 46 ms and 2.9 ms, respectively. For Hindustani music, we use the pitch tracks corresponding to the predominant melody that are used in several other studies on a similar topic [24, 27] and are made available to us by the authors. These pitch tracks are obtained using a semi-automatic system for predominant melody estimation. This allows us to compare results across studies and avoid the effects of pitch errors on the computation of melodic similarity. After estimating the predominant pitch we convert it from Hertz to Cent scale for the melody representation to be musically relevant.

We proceed to post-process the pitch contours and remove the spurious pitch jumps lasting over a few frames as well as smooth the pitch contours. We first apply a median filter over a window size of 50 ms, followed by a low-pass filter using a Gaussian window. The window size and the standard deviation of the Gaussian window is set to 50 ms and 10 ms, respectively. The pitch contours are finally down-sampled to 100 Hz, which was found to be an optimal sampling rate in [13].

2.2 Transposition Invariance

The base frequency chosen for a melody in IAM is the tonic pitch of the lead artist [10]. Therefore, for a meaningful comparison of the melodic phrases across the recordings of different artists, a melody representation should be normalized by the tonic pitch of the lead artist. We perform this tonic normalization (N_{tonic}) by considering the tonic of the lead artist as the reference frequency during the Hertz to Cent conversion. The tonic pitch is automatically identified using a multi-pitch approach proposed by Gulati et al. [10]. This approach was shown to obtain more than 90% tonic identification accuracy and has been used in several studies in the past.

Tonic normalization does not account for the pitch of the octave transposed occurrences of a melodic phrase within a recording. In addition, estimated tonic pitch sometimes might be incorrect and a typical error is an offset of fifth scale degree. To handle such cases, we propose a novel tetrachord normalization (N_{tetra}). For this we analyse the difference (Δ) in the mean frequency values of the two tonic normalized melodic phrases (p_1, p_2). We offset the pitch values of the phrase p_1 by the frequency in the set {- 1200, - 700, - 500, 0, 500, 700, 1200, 1700, 1900} that is closest to Δ within a vicinity of 100 Cents. In addition to tetrachord normalization, we also experiment with mean normalization (N_{mean}), which was reported to improve the performance in the case of Carnatic music [13].

2.3 Partial Transcription

We perform a partial melody transcription to automatically segment and identify the steady svar regions in the melody. Note that even the partial transcription of the melodies is a non-trivial task, since we desire a segmentation that is robust to different melodic ornaments added to a svar where the pitch deviation from the mean svar frequency can be up to 200 Cents. In Figure 2 we show such an example of a steady svar region (P_{1a} from 3-6 s) where the pitch deviation from the mean svar frequency is high due to added melodic ornaments. Ideally, the melodic region between 1 and 6 s should be detected as a single svar segment.

We segment the steady svar regions using a method described in [11], which addresses the aforementioned challenges. A segmented svar region is then assigned a frequency value corresponding to the peak in an aggregated pitch histogram closest to the mean svar frequency. The pitch histogram is constructed for the entire recording and smoothened using a Gaussian window with a variance of 15 cents. As peaks of the normalized pitch histogram, we select all the local maximas where at least one peak-to-valley ratio is greater than 0.01. For a detailed description of this method we refer to [11].

2.4 Svar Duration Truncation

After segmenting the steady svar regions in the melody we proceed to truncate the duration of these regions. We hypothesize that, beyond a certain value δ, the duration of these steady svar regions do not change the identity of a melodic phrase (i.e. the phrase category). We experiment with 7 different truncation durations $\delta = \{$ 0.1 s, 0.3 s, 0.5 s, 0.75 s, 1 s, 1.5 s, 2 s$\}$ and select the one that results in the best performance. In Figure 2 we show an example of the occurrences of a melodic phrase both before (P_{1a}, P_{2a}) and after (P_{1b}, P_{2b}) the svar duration truncation using $\delta = 0.1$ s. This example clearly illustrates that the occurrences of a melodic phrase after duration truncation exhibit lower degree of non-linear timing variations. We denote this method by M_{DT}.

2.5 Similarity Computation

To measure the similarity between two melodic fragments we consider a DTW-based approach. Since the phrase segmentation is known beforehand, we use a whole sequence matching DTW variant. We consider the best performing DTW variant and the related parameter values for each music tradition as reported in [13]. These variants were chosen based on an exhaustive grid search across all possible combinations and hence can be considered as optimal for this dataset. For Carnatic music we use a DTW step size condition $\{(2,1),(1,1),(1,2)\}$ and for Hindustani music a step size condition $\{(1,0),(1,1),(0,1)\}$. We use Sakoe-Chiba global band constraint [28] with the width of the band as ±10% of the phrase length. Note that before computing the DTW distance we uniformly time-scale the two melodic fragments to the same length, which is the maximum of the lengths of the phrases.

Dataset	Rec.	PC	Rāgs	Artists	Duration (hr)
CMD	23	5	5	14	3.82
HMD	9	5	1	7	1.76

Table 1. Details of the datasets in terms of the total number of recordings (Rec.), number of annotated phrase categories (PC), number of rāgs, unique number of artists and total duration of the dataset.

2.6 Complexity Weighting

The complexity weighting that we apply here to overcome the shortcoming of the distance measure in distinguishing two time series with different complexities is discussed in [1]. We apply a complexity weighting (α) to the DTW-based distance (D_{DTW}) in order to compute the final similarity score $D_f = \alpha D_{DTW}$. We compute α as:

$$\alpha = \frac{\max(C_i, C_j)}{\min(C_i, C_j)}; \qquad C_i = \sqrt[2]{\sum_{i=1}^{N-1}(p_i - p_{i+1})^2} \qquad (1)$$

where, C_i is the complexity estimate of a melodic phrase of length N samples and p_i is the pitch value of the i^{th} sample. We explore two variants of this complexity estimate. One of these variants is already proposed in [1] and is described in equation 1. We denote this method variant by M_{CW1}. We propose another variant that utilizes melodic characteristics of Carnatic music. This variant takes the number of saddle points in the melodic phrase as the complexity estimate [17]. This method variant is denoted by M_{CW2}. As saddle points we consider all the local minimas and the local maximas in the pitch contour which have at least one minima to maxima distance of half a semitone. Since such melodic characteristics are predominantly present in Carnatic music, the complexity weighting is not applicable for computing melodic similarity in Hindustani music.

3. EVALUATION

3.1 Dataset and Annotations

For a better comparison of the results, for our evaluations we use a music collection that has been used in several other studies for a similar task [13, 24, 27]. However, we have extended the dataset by adding 30% more number of annotations of the melodic phrases, which we make available at http://compmusic.upf.edu/node/269. The music collection comprises vocal recordings of renowned artists in both Hindustani and Carnatic music. We use two separate datasets for the evaluation, Carnatic music dataset (CMD) and Hindustani music dataset (HMD) as done in [13]. The melodic phrases are annotated by two professional musicians who have received over 15 years of formal music training. All the annotated phrases are the characteristic

CMD				HMD			
PC	#Occ	L_{mean}	L_{std}	PC	#Occ	L_{mean}	L_{std}
C_1	39	1.38	0.25	H_1	62	1.93	0.98
C_2	46	1.25	0.21	H_2	154	1.40	0.79
C_3	38	1.23	0.24	H_3	47	1.30	0.78
C_4	31	1.11	0.17	H_4	76	2.38	1.33
C_5	45	0.76	0.08	H_5	87	1.17	0.36
Total	199	1.13	0.29		426	1.59	0.99

Table 2. Details of the 625 annotated melodic phrases. PC: pattern category, #Occ: number of annotated occurrences, and L_{mean}, L_{std} are the mean, standard deviation of the lengths of the patterns of a PC in seconds.

phrases of a rāg. In Table 1 we summarize the relevant dataset details. Table 2 summarizes the details of the annotated phrases in terms of their number of instances and basic statistics of the length of the phrases.

3.2 Setup, Measures and Statistical Significance

We consider each annotated melodic phrase as a query and perform a search across all the annotated phrases in the dataset (referred to as target search space). In addition to the annotated phrases, we add randomly sampled melodic segments (referred to as noise candidates) in the target space to simulate a real world scenario. We generate the starting time stamps of the noise candidates by randomly sampling a uniform distribution. The length of the noise candidates are generated by sampling the distribution of the duration values of the annotated phrases. The number of noise candidates added are 100 times the number of total annotations in the entire music collection. For every query we consider the top 1000 nearest neighbours in the search results ordered by the similarity value. A retrieved melodic phrase is considered as a true hit only if it belongs to the same phrase category as the query.

To assess the performance of the proposed approach and the baseline method we use mean average precision (MAP), a common measure in information retrieval [21]. To assess if the difference in the performance of any two methods is statistically significant we use the Wilcoxon signed rank-test [32] with $p < 0.01$. To compensate for multiple comparisons, we apply the Holm-Bonferroni method [15].

4. RESULTS AND DISCUSSION

In Table 3 we summarize the MAP scores and the standard deviation of the average precision values obtained using the baseline method (M_B), the method that uses duration truncation (M_{DT}) and the ones using the complexity weighting (M_{CW1}, M_{CW2}), for both the CMD and the HMD. Note that M_{CW1} and M_{CW2} are only applicable to the CMD (Sec. 2).

We first analyse the results for the HMD. From Table 3 (upper half), we see that the proposed method variant that applies a duration truncation performs better than the baseline method for all the normalization techniques. Moreover, this difference is found to be statistically significant in each case. The results for the HMD in this table correspond to δ =500 ms, for which we obtain the highest accuracy compared to the other δ values as shown in Figure 4. Furthermore, we see that N_{tonic} results in the best accuracy for the HMD for all the method variants and the difference is found to be statistically significant in each case. In Figure 5 we show a boxplot of average precision values for each phrase category and for both M_B and M_{DT} to get a better understanding of the results. We observe that with an exception of the phrase category H_2, M_{DT} consistently performs better than M_B for all the other phrase categories. A close examination of this exception reveals that the error often is in the segmentation of the steady svar regions of the melodic phrases corresponding to H_2. This can be attributed to a specific subtle melodic movement in H_2 that is confused by the segmentation method as a melodic ornament instead of a svar transition, leading to a segmentation error.

We now analyse the results for the CMD. From Table 3 (lower half), we see that using the method variants M_{DT}, M_{CW1} and M_{CW2} we obtain reasonably higher MAP scores compared to the baseline method M_B and the difference is found to be statistically significant for each method variant across all normalization techniques. This MAP score for M_{DT} corresponds to δ =1 s, which is considerably higher than the MAP scores for other δ values as shown in Figure 4. We also see that M_{CW2} performs slightly better than M_{CW1} and the difference is found to

HMD				
Norm	M_B	M_{DT}	M_{CW1}	M_{CW2}
N_{tonic}	**0.45 (0.25)**	**0.52 (0.24)**	-	-
N_{mean}	0.25 (0.20)	0.31 (0.23)	-	-
N_{tetra}	0.40 (0.23)	0.47 (0.23)	-	-
CMD				
Norm	M_B	M_{DT}	M_{CW1}	M_{CW2}
N_{tonic}	0.39 (0.29)	0.42 (0.29)	0.41 (0.28)	0.41 (0.29)
N_{mean}	0.39 (0.26)	0.45 (0.28)	0.43 (0.27)	0.45 (0.27)
N_{tetra}	**0.45 (0.26)**	**0.50 (0.27)**	**0.49 (0.28)**	**0.51 (0.27)**

Table 3. MAP scores for the two datasets HMD and CMD for the four method variants M_B, M_{DT}, M_{CW1} and M_{CW2} and for different normalization techniques. Standard deviation of average precision is reported within round brackets.

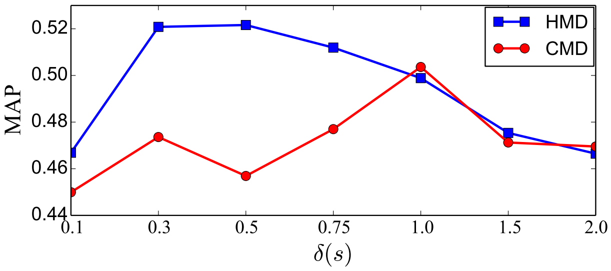

Figure 4. MAP scores for different duration truncation values (δ) for the HMD and the CMD.

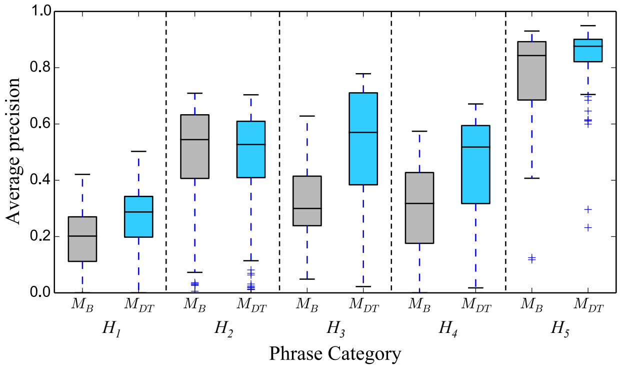

Figure 5. Boxplot of average precision values obtained using M_B and M_{DT} for each melodic phrase category for the HMD. These values correspond to N_{tonic}.

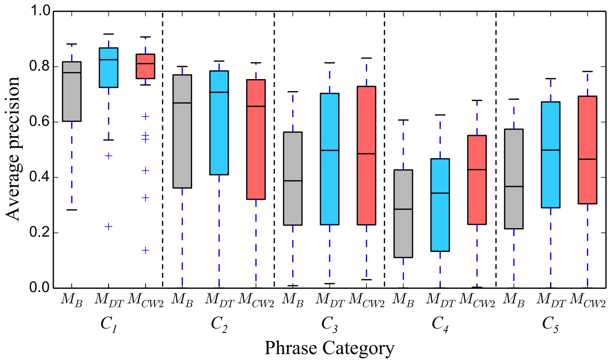

Figure 6. Boxplot of average precision values obtained using methods M_B, M_{DT} and M_{CW} for each melodic phrase category for the CMD. These values correspond to N_{tetra}.

be statistically significant only in the case of N_{tetra}. We do not find any statistically significant difference in the performance of methods M_{DT} and M_{CW2}. Unlike the HMD, for the CMD N_{tetra} results in the best performance with a statistically significant difference compared to the other normalization techniques across all method variants. We now analyse the average precision values for every phrase category for M_B, M_{DT} and M_{CW2}. Since M_{CW2} performs slightly better than M_{CW1} we only consider M_{CW2} for this analysis. In Figure 6 we see that M_{DT} performs better than M_B for all phrase categories. We also observe that M_{CW2} consistently performs better than M_B with the sole exception of C_2. This exception occurs because M_{CW2} presumes a consistency in terms of the number of saddle points across the occurrences of a melodic phrase, which does not hold true for C_2. This is because phrases corresponding to C_2 are rendered very fast and the subtle pitch movements are not the characteristic aspect of such melodic phrases. Hence, the artists often take the liberty of changing the number of saddle points.

Overall we see that duration truncation of steady melodic regions improves the performance in both the HMD and the CMD. This reinforces our hypothesis that elongation of steady svar regions in the melodies of IAM in the context of the characteristic melodic phrase does not change the musical identity of the phrase. This correlates with the concept of nyās svar (nyās literally means home), where the artist has the flexibility to stay and elongate a single svar. A similar observation was reported in [24], where the authors proposed to learn the optimal global DTW constraints a priori for each pattern category. However, their proposed solution could not improve the performance. Further comparing the results for the HMD and the CMD we notice that N_{tonic} results in the best performance for the HMD and N_{tetra} for the CMD. This can be attributed to the fact that the number of the pitch-transposed occurrences of a melodic phrase is significantly higher in the CMD compared to the HMD [13]. Also, since the non-linear timing variability in the HMD is very high, any normalization (N_{mean} or N_{tetra}) that involves a decision based on the mean frequency of the phrase is more likely to fail.

5. CONCLUSIONS

In this paper we briefly presented an overview of the approaches for detecting the occurrences of the characteristic melodic phrases in audio recordings of Indian art music. We highlighted the major challenges involved in this task and focused on two specific issues that arise due to large non-linear timing variations and rapid melodic movements. We proposed simple and easy to implement solutions based on partial transcription and complexity weighting to address these challenges. We also put forward a new dataset by appending 30% more number of melodic phrase annotations to those used in previous studies. We showed that duration truncation of the steady svar regions in the melodic phrases results in a statistically significant improvement in the computation of melodic similarity. This confirms our hypothesis that the elongation of steady svar regions beyond a certain duration does not affect the perception of the melodic similarity in the context of the characteristic melodic phrases. Furthermore, we showed that complexity weighting significantly improves the melodic similarity in Carnatic music. This suggests that the extent and the number of saddle points is an important characteristic of a melodic phrase and is crucial to melodic similarity in Carnatic music.

In the future, we plan to improve the method used for segmenting the steady svar regions so that it can differentiate melodic ornaments from subtle svar transitions. In addition, we see a vast scope in further refining the complexity estimate of a melodic phrase to improve the complexity weighting. It would also be worthwhile to explore the applicability of this approach to music traditions such as Flamenco, Beijing opera and Turkish Makam music.

6. ACKNOWLEDGMENTS

This work is partly supported by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583). We thank Kaustuv K. Ganguli and Vignesh Ishwar for the annotations and valuable discussions and, Ajay Srinivasamurthy for the proof-reading.

7. REFERENCES

[1] G. E. Batista, X. Wang, and E. J Keogh. A complexity-invariant distance measure for time series. In *SDM*, volume 11, pages 699–710, 2011.

[2] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. Essentia: an audio analysis library for music information retrieval. In *Proc. of Int. Society for Music Information Retrieval Conf. (ISMIR)*, pages 493–498, 2013.

[3] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: Current directions and future challenges. *Proc. of the IEEE*, 96(4):668–696, 2008.

[4] T. Collins, S. Böck, F. Krebs, and G. Widmer. Bridging the audio-symbolic gap: The discovery of repeated note content directly from polyphonic music audio. In *Audio Engineering Society's 53rd Int. Conf. on Semantic Audio*, 2014.

[5] D. Conklin and C. Anagnostopoulou. Comparative Pattern Analysis of Cretan Folk Songs. *Journal of New Music Research*, 40(2):119–125, 2010.

[6] A. Danielou. *The ragas of Northern Indian music*. Munshiram Manoharlal Publishers, New Delhi, 2010.

[7] S. Dutta and H. A. Murthy. Discovering typical motifs of a raga from one-liners of songs in Carnatic music. In *Int. Society for Music Information Retrieval*, pages 397–402, 2014.

[8] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming: musical information retrieval in an audio database. In *Proc. of the third ACM Int. Conf. on Multimedia*, pages 231–236. ACM, 1995.

[9] S. Gulati. A tonic identification approach for Indian art music. Master's thesis, Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain, 2012.

[10] S. Gulati, A. Bellur, J. Salamon, H. G. Ranjani, V. Ishwar, H. A. Murthy, and X. Serra. Automatic tonic identification in Indian art music: approaches and evaluation. *Journal of New Music Research*, 43(1):55–73, 2014.

[11] S. Gulati, J. Serrà, K. K. Ganguli, and X. Serra. Landmark detection in hindustani music melodies. In *Int. Computer Music Conf., Sound and Music Computing Conf.*, pages 1062–1068, 2014.

[12] S. Gulati, J. Serrà, V. Ishwar, and X. Serra. Mining melodic patterns in large audio collections of indian art music. In *Int. Conf. on Signal Image Technology & Internet Based Systems (SITIS-MIRA)*, pages 264–271, 2014.

[13] S. Gulati, J. Serrà, and X. Serra. An evaluation of methodologies for melodic similarity in audio recordings of indian art music. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 678–682, 2015.

[14] W. B. Hewlett and E. Selfridge-Field. *Melodic similarity: Concepts, procedures, and applications*, volume 11. The MIT Press, 1998.

[15] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2):65–70, 1979.

[16] V. Ishwar, A. Bellur, and H. A. Murthy. Motivic analysis and its relevance to raga identification in carnatic music. In *Proceedings of the 2nd CompMusic Workshop*, pages 153–157, 2012.

[17] V. Ishwar, S. Dutta, A. Bellur, and H. Murthy. Motif spotting in an Alapana in Carnatic music. In *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, pages 499–504, 2013.

[18] Z. Juhász. Motive identification in 22 folksong corpora using dynamic time warping and self organizing maps. In *Int. Society for Music Information Retrieval Conf.*, pages 171–176, 2009.

[19] H. J Lin, H. H. Wu, and C. W. Wang. Music matching based on rough longest common subsequence. *J. Inf. Sci. Eng.*, 27(1):95–110, 2011.

[20] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proc. of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pages 2–11, 2003.

[21] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.

[22] A. Marsden. Interrogating Melodic Similarity: A Definitive Phenomenon or the Product of Interpretation? *Journal of New Music Research*, 41(4):323–335, 2012.

[23] A. Pikrakis, J. Mora, F. Escobar, and S. Oramas. Tracking melodic patterns in Flamenco singing by analyzing polyphonic music recordings. In *Proc. of Int. Society for Music Information Retrieval Conf. (ISMIR)*, pages 421–426, 2012.

[24] P. Rao, J. C. Ross, K. K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, and H. A. Murthy. Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research*, 43(1):115–131, 2014.

[25] S. Rao. Culture Specific Music Information Processing : A Perspective From Hindustani Music. In *2nd CompMusic Workshop*, pages 5–11, 2012.

[26] J. C. Ross and P. Rao. Detection of raga-characteristic phrases from Hindustani classical music audio. In *Proc. of 2nd CompMusic Workshop*, pages 133–138, 2012.

[27] J. C. Ross, T. P. Vinutha, and P. Rao. Detecting melodic motifs from audio for Hindustani classical music. In *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, pages 193–198, 2012.

[28] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoustics, Speech, and Language Processing*, 26(1):43–50, 1978.

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

[30] Rainer Typke. *Music retrieval based on melodic similarity*. 2007.

[31] T. Viswanathan and M. H. Allen. *Music in South India*. Oxford University Press, 2004.

[32] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, pages 80–83, 1945.