In Proc. of the National Conference for Communications (NCC)
March 2017, IIT Madras, India.

Melodic Shape Stylization for Robust and Efficient Motif Detection in Hindustani Vocal Music

Kaustuv Kanti Ganguli, Ashwin Lele, Saurabh Pinjani, Preeti Rao
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai, India.
kaustuvkanti@ee.iitb.ac.in

Ajay Srinivasamurthy, Sankalp Gulati
Music Technology Group
Universitat Pompeu Fabra, Barcelona, Spain.

***Abstract*—In Hindustani classical music, melodic phrases are identified not only by the stable notes at precise pitch intervals but also by the shapes of the continuous transient pitch segments connecting these. Time-series matching via subsequence dynamic time warping (DTW) facilitates the equal contribution of stable notes and transients to the computation of similarity between pitch contour segments corresponding to melodic phrases. In the interest of reducing computational complexity it is advantageous to replace time-series DTW with low-dimensional string matching provided a principled approach to the time-series to symbolic string conversion is available. While the stable notes easily lend themselves to quantization, we address the compact representation of the transient pitch segments in this work. We analyze the design considerations at each stage: pitch curve fitting, normalization (with respect to pitch interval and duration), shape dictionary generation, inter-symbol proximity measure and string matching cost functions. A combination of domain knowledge- and data-driven optimization on a database of raga music is exploited to design the melodic representation of a raga phrase that enables a performance comparable to the time series based matching in an audio search by query task at significantly lower computational cost.**

I. Introduction

Computational models for melodic similarity at the level of a musical phrase require the definition of a representation and a corresponding distance measure. In the case of Western music, the representation draws upon the well-established written score where pitch intervals and note timing are unambiguously captured. The distance measure is then typically cast as a string matching problem where musically informed costs are associated with the string edit distance formulation [1], [2]. Research on melodic similarity measures, of course, also extends to more cognitive modeling based approaches using higher-level features extracted from melodic contours rather than simply surface features [3]. Non-western and folk musics do not lend themselves well to the Western score based transcription [4], [5]. Having originated and evolved as oral traditions, they typically do not have a well-developed symbolic representation thus posing challenges for pedagogy and also for computational tasks such as music retrieval.

Relevant past work on the computational analyses of non-Western music includes the representation of flamenco singing with its highly melismatic form with smooth transitions between notes where onsets are not clearly specified [5]. A symbolic transcription based on sequences of several short notes was fitted to the continuous time-varying pitch contour using dynamic programming based optimization using probability functions for pitch, energy, and duration. In the case of Indian art music forms, melodic phrases are identified not only by the stable notes at precise pitch intervals but also by the shapes of the continuous transient pitch segments connecting them. The continuous pitch curves have certain dynamics that are not accurately represented by a succession of short notes. Some musicological literature proposes a set of 15 basic melodic shapes (alankar) [6]. However there is no general acceptance of this in contemporary Hindustani classical music nor has there been any computational research that exploits this.

In this work, we consider building a robust melodic representation for raga phrases of Hindustani classical vocal music that suits various melodic query based retrieval tasks. Hindustani vocal music is characterized by precise tuning and timing as can be judged from the raga grammar and from the rhythmic framework which invites a clear rigidity in the time placement of certain melodic events with respect to the rhythmic cycle supplied by the tabla accompanist. Our previous work [7] demonstrated that using the continuous pitch curves, as time series, with a suitably constrained DTW based distance measure provides for reasonable retrieval accuracy but at very high computational cost. A grossly simplified representation that discarded the pitch transitions and obtained a string of symbols from the stable notes obtained significant computation reduction at the cost of retrieval accuracy. In the present work, we investigate the symbolic representation of continuous pitch transitions that, when combined with stable notes, leads to the more complete representation of a raga-phrase. In our earlier work [8], [9], a data-driven approach to identifying canonical forms for raga-characteristic phrases together with a DTW distance measure was used to discriminate phrases that shared the same stable notes sequence. We use and extend this approach for the retrieval task by filling in the critical gap in the time-series to symbol conversion stage required for efficient string-matching methods to become applicable. Our approach is to stylize the continuous pitch curve with a low-order representation and obtain an inter-symbol distance measure for the ensuing codebook of shapes.

Another important predictor in melodic similarity paradigm is ‘duration’. E.g. the same sequence of notes with different

relative durations should be tractable and be penalized by the algorithm, more so because Hindustani music contains many such examples where melodic phrases from two different ragas bear the same note sequence but with different relative durations (refer to [10] for further discussion). Our recent work [11], which included duration information in the string based representation, was observed to provide an improvement in retrieval performance over using pitch information alone.

The rest of the paper is organized as follows. The next section presents the methodology to obtain the proposed melodic representation from audio. The training and test datasets are described followed by an overview of the different retrieval schemes under test. The experimental evaluation of the systems is followed by a discussion providing some insights into the observed comparative behaviors.

II. Methodology

The methodology for the representation stage has the following steps as shown in Figure 1. The main contribution of the current work lies in the "Polynomial fitting, normalization & VQ" block, our primary focus is on optimizing the tuning parameters for a reasonably 'good' representation.

A. Time series extraction from audio

Predominant-F0 detection is implemented by an algorithm proposed by [12] that exploits the spectral properties of the voice with temporal smoothness constraints on the pitch. The pitch is detected at 20 ms intervals with zero pitch assigned to the detected purely instrumental regions. Next, the pitch values in Hz are converted to the cents scale by normalizing with respect the concert tonic determined by automatic tonic detection [13]. The final pre-processing step is to interpolate short silence regions below a threshold (250 ms which is empirically chosen based on our previous study [14]) indicating musically irrelevant breath pauses or unvoiced consonants, by cubic spline interpolation, to ensure the integrity of the melodic shape.

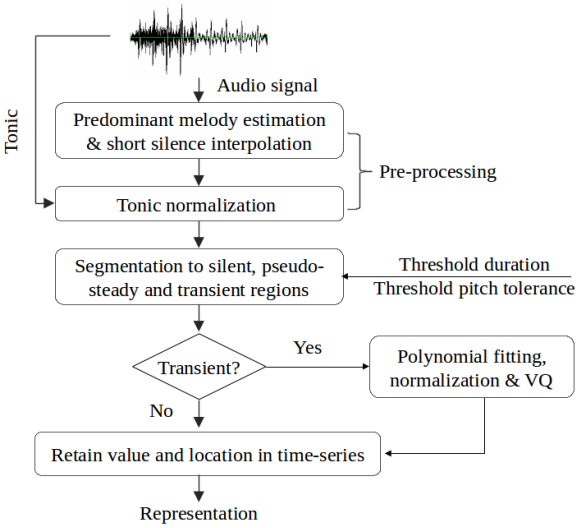

Fig. 1. Block diagram showing the steps from the input audio to the proposed melodic representation.

B. Segmentation

The pitch contour of a melodic phrase can be thought of as a chain of three possible states: (i) stable note segment, (ii) transitory segment which joins two stable segments, and (iii) breath pauses. Segmenting (iii) is trivial and follows from F0 extraction, (i) is segmented by the algorithm proposed in [7] that employs a heuristic thresholding on contiguous segments ($\geq$ 250 ms) of stable pitch regions within a tolerance ($\leq$ 35 cents). The resultant note sequence largely approximates the textbook notation. Hence (ii) is obtained as a residual of the preceding segmentation steps.

C. Pitch contour stylization

Different approaches for representing the pitch contour with several innovative strategies are found in literature, such as polynomial fit [15], SAX [16] and its variations (we proposed a modified version of the SAX [7]), melody transcription [17], melodic shape assignment [18], [19], landmark detection [20]. While many of these approaches addressed the task of melodic representation from a purely retrieval viewpoint, others proposed different approaches from a musicology and pedagogy perspective. Datta et. al. [21], [22] had used 2^{nd} degree polynomial to automatically extract (and hence classify) 'meend' from the performances in Hindustani vocal music. Gupta et. al. [23] had reported superiority of a 3^{rd} degree polynomial over a second-degree (i.e. parabolic) contour for the task of objective assessment of ornamentation in Indian classical singing.

The pseudo-note system [7] representation discarded all melodic transients and only preserved the sequence of stable note segments. We propose a way to consider the transient regions in the modelled contour, with an additional step of quantizing them to a set of codebook vectors. We first normalize each transient segment to lie within 0 to 1 range. A 3^{rd} degree polynomial is fitted and we generate a candidate shape by constructing a unit length (100 samples) contour from the polynomial coefficients. The K-means clustering algorithm with Euclidean distance measure is used to generate a codebook of distinct representative transient shapes (refer to Figure 2). The quantization of a test transient segment is achieved through a nearest neighbor classifier (on the fitted and normalized 3^{rd} degree polynomial) with the same Euclidean distance measure as used during training. If the achieved representation corresponds to some invariant skeleton of the melodic shape of the phrase via a low-degree polynomial, we would anticipate obtaining better matches across variations of the melodic phrase. We address the question of how to bring domain knowledge into this transformation.

Figure 3 shows the pitch contours of an example mukhda phrase: before and after pitch contour stylization. The reconstruction steps are discussed in Appendix A. An informal perceptual experiment was carried out to verify that the simplification has retained the essential characteristics of the test phrase.

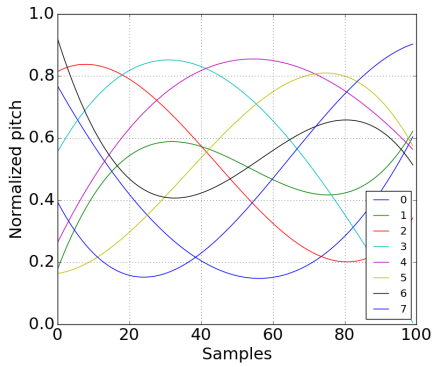

Fig. 2. 8 centroids obtained corresponding to each cluster index from the codebook. Each vector is normalized between [0,1] and contain 100 samples.

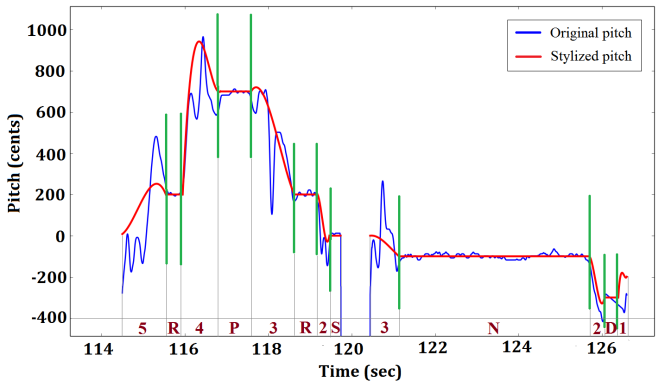

Fig. 3. Original pitch contour segment (blue) with superimposed stylized contour (red). We are able to capture the overall pitch movements after discarding all local pitch excursions.

III. Dataset and Annotation

A. Training corpus

The training corpus comprises 30 songs from 30 different ragas (22,092 transient segments) from the 'Raga Dataset' of 300 songs [24] from the CompMusic [25] collection. From each raga, one song is randomly chosen to constitute the training corpus. We tried several such 30-song sets but the training seems to be no different.

B. Test set

The test set is an extension of the dataset reported in [7], totalling a 75-song set which is disjoint from the training set. The musician's annotation involved in the test set is to mark the start and end boundaries of the mukhda (melodic motif used as the song's refrain) phrases. The description of the test dataset is given in Table I.

TABLE I
Description of the test dataset.

# Song	Dur (hrs)	# GT	Dur (hrs)	Ratio	# Unique	
					Raga	Artist
75	22:18	1754	2:49	13%	55	23

IV. Melodic Phrase Retrieval Systems

In this section, we consider various approaches towards our end goal which involves searching the entire vocal pitch track extracted from the archived audio recording to identify pitch contour sub-segments that match the melodic shape of the query. We discuss the different schemes for scoring strategy, incorporating the inter-symbol similarity score for the transient segments in the Smith-Waterman string alignment algorithm. The inter-symbol distance is derived from the distance matrix (Euclidean distance) of the codebook centroids as shown in Figure 4. The rank ordered codebook indices with respect to increasing distance is quantized into 3 bins of size 1, 3, 4 codevectors respectively. We call these bins (refer to Figure 6) as 'Same', 'Close', and 'Far'. Note that the distance matrix is not symmetric, because each row is min-max normalized (between 0 and 1) in order to get an evenly distributed rank ordered codebook index with respect to each codebook vector as a query.

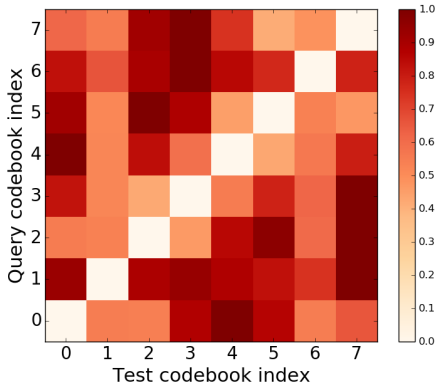

Fig. 4. Distance matrix (Euclidean distance) between the 8 codebook vectors. For each codebook vector, we arrange the other vectors in ascending order of distance and quantize them to 3 bins of of size 1, 3, 4 respectively.

A. Pseudo-note system

Our baseline method is the pseudo-note system as illustrated in our previous work [7]. We take the version C (the then best performing system) that used query dependent preset parameters. For completeness we recall two relevant parameters: (i) substitution score and (ii) gap penalty. In its standard form, the Smith-Waterman algorithm uses a fixed positive score for an exact match and a fixed negative cost for symbol mismatch. In the context of musical pitch intervals, we penalize small differences less than large differences. The two parameter presets are (i) fast varying: substitution score of +1 to symbols differing by upto 2 semitones ('Close'), gap penalty is affine with parameters $m = 1$, $c = 0.5$; and (ii) slowly varying: substitution score of -0.5 to symbols differing by upto 3 semitones ('Close'), gap penalty is affine with parameters $m = 0.5$, $c = 1.5$.

B. Proposed schemes

Our proposed system partially uses the pseudo-note system with the introduction of symbols for the modelled transients.

TRANSIENT \ DURATION	WITHOUT	WITH
WITHOUT	[R,P,R,S], [N,D]	[(R,390),(P,720),(R,550), (S,270)], [(N,4550),(D,280)]
WITH	[5,R,4,P,3,R,2,S], [3,N,2,D,1]	[(5,1070),(R,390),(4,910),(P,720), (3,1040),(R,550),(2,280),(S,270)], [(3,690),(N,4550),(2,440), (D,280),(1,270)]

Fig. 5. Proposed schemes of melodic phrase representation (symbols and duration (ms) information) applied to the pitch contour of Figure 3.

TRANSIENT \ DURATION	WITHOUT	WITH
WITHOUT	1 Same: +3 Close: +1 Far: -1 -m1*X1 + c1	Same: +3/Z 2 Close: +1 Far: -1*Z -m2*X1*Y1 + c2
WITH	Same: +3, +1 Close: +1, +0.33 Far: -1, -0.5 -m1*X2 + c1 3	Same: +3/Z, +1/Z Close: +1, +0.33/Z Far: -1*Z, -0.5*Z -m2*X2*Y2 + c2 4

Fig. 6. Proposed schemes of melodic phrase retrieval systems. The scheme indices are marked in red. The parameters and corresponding values for substitution score and gap penalty are presented.

The main contribution, here, lies in the modified scoring scheme which is discussed next. The notations of interest are as follows. $Z = \max(\frac{T_{Query}}{T_{Candidate}}, \frac{T_{Candidate}}{T_{Query}})$, where T_{Query} and $T_{Candidate}$ are the durations of the note in the query and candidate respectively (refer to the parameter 'Fraction' [11]). $X1$: length of gap (no. of notes), $X2$: length of gap (no. of notes/transients), $Y1$: duration of gap (sec), $Y2$: weighted duration of gap (sec). The optimal values for the slopes and intercepts, as obtained from an empirical observation [11], are: $m1 = 3$, $m2 = 0.02$, $c1 = 1.5$, $c2 = 1.2$. There are four schemes at play, Figure 5 shows the representation of the pitch contour given in Figure 3.

Scoring: The modification of the scoring strategy involves two new parameters. The first, incorporating a parameter Z into the substitution score (schemes 2 and 4) that attenuates the positive score (cases: Same and Close) and amplifies the negative score (case: Far). This is because, by definition $Z \geq 1$ (we assign a lower bound of 1.25 chosen empirically) and a higher value of Z suggests a high mismatch between the query and candidate durations which should be compensated in the substitution score. Secondly, we incorporate a factor called 'weighted duration' $Y2$ for the transient segments in order to balance their importance with respect to the pseudo-notes. The weight is a linear combination of pre- and post-context of transient durations (with an empirically chosen weighting factor of 0.8) for each pseudo-note duration. The proportion of duration contributed by pseudo-notes and transients is very high which is compensated by introduction of the $Y2$ parameter. Figure 6 shows the parameters involved. Schemes 3 and 4 has two values for substitution scores (comma separated) which stand for pseudo-notes and transients respectively.

V. Experiments and Evaluation

The experiments allow us to compare the performance of the different schemes on the task at hand, i.e., detecting occurrences of the mukhda in the audio concert given an audio query corresponding to one instance of the mukhda phrase. In our earlier work [7], we had the assumption that early mukhda repetitions tend to be of the canonical form that a musician might generate to describe the bandish; here instead we consider all annotated mukhdas as queries. We process the database to convert each concert audio to the pitch time series and then to the corresponding stylized representation. Next, the query is converted to the corresponding representation and the search is executed. The detections with time-stamps are listed in the order of increasing distance with the query as computed by the corresponding search distance measure. We disallow the list to grow further twice the number of ground truth mukhdas since this would correspond well to the maximum number of mukhdas expected in the concert given its duration. A detection is considered a true positive if the time series of the detection overlaps at least 50% of that of one of the ground-truth labeled mukhdas in the song. A receiver operating characteristic curve (precision vs recall) is obtained for each query by sweeping a threshold across the obtained distances. The performance for each song is summarized by the value corresponding to the percentage of queries (i.e. mukhdas) that result in at least $n\%$ recall with respect to all the remaining mukhdas in the song. We term this "goodness %" of the song. We report performance for different choices of n.

TABLE II
Evaluation metrics in terms of best [Precision, Recall] pairs.

[Precision, Recall] pairs for best Precision, Recall			
Scheme 1	Scheme 2	Scheme 3	Scheme 4
[1.00,0.57]	[1.00,0.91]	[1.00,0.60]	[1.00,0.60]
[1.00,0.57]	[1.00,0.91]	[0.58,0.96]	[0.61,0.96]

VI. Results and Discussion

From Table II we see that there is improvement of recall (with fixed precision) after including the duration information. Figure 7 shows an improvement of recall in each column from

TABLE III
Average "goodness %" across song (# songs with at least one good query) for different thresholds for the 4 schemes.

Threshold (n)	Average "goodness %" (# songs)			
	Scheme 1	Scheme 2	Scheme 3	Scheme 4
10%	0.56 (63)	0.64 (71)	0.68 (71)	0.77 (71)
30%	0.35 (21)	0.38 (40)	0.39 (44)	0.42 (58)
50%	0.38 (2)	0.35 (9)	0.38 (12)	0.24 (25)

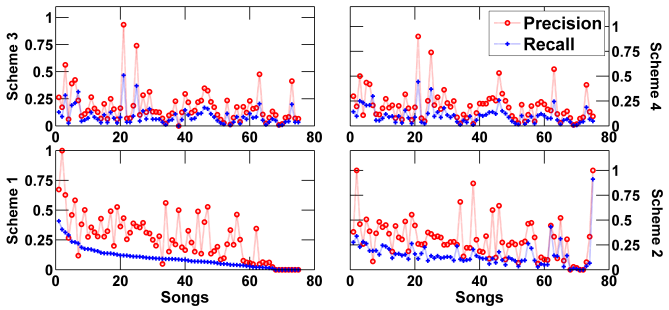

Fig. 7. Comparison of evaluation metrics Precision and Recall for all 4 schemes. The song indices for Scheme 1 are obtained from descending order of recall, the same indices carry over to the other schemes.

bottom to the top row. Given that we attain consistent improvement in recall (while precision is marginally improving too), the proposed measure fraction of "good query" shows an improvement in Schemes 1 through 4. Table III shows the average fraction of good query for different threshold values (at least 10%, 30% and 50% recall). The number of songs where at least one 'good query' is found, improves in Schemes 1 through 4, irrespective of the threshold. On a rigorous error analysis, we find that the rootcause of the songs not having a single good query either belongs to slow tempo (vilambit) compositions where mukhdas are reasonably long or the mukhdas are performed with heavy embellishments that resulted in a long string sequence. Long queries might have resulted in amplification of negative score, a length normalization scheme could be useful to compensate this which is posed as a future work.

We highlight the main contributions and summarize our work in terms of the design considerations at each stage:

(i) We use a representative corpus (diverse in terms of artists and ragas) to train a codebook of melodic transient shapes. The test set for the evaluation task is completely disjoint from the training corpus.

(ii) The choice of codebook size is iterated from a very small (2) to a large (100) value. It is observed that beyond a certain reasonable size (8) redundancy is introduced into the codebook vectors. This indicates about a possible universality in the basic melodic movements between notes, though the same transient shape between the same pair of notes in two different ragas sounds perceivably different just due to the time-scaling (slowness) of the transient.

(iii) The k of k-means is chosen from the hierarchical clustering view. We carried out statistical methods (finding the elbow of the mean squared error curve iterated over k, gap statistics with varying k, observing the density of obtained clusters after truncating the dendrogram at different levels etc.) to conclude that $k = 7$ to 9 is reasonable. Hence we empirically chose $k = 8$.

(iv) There may be a criticism of the fact that the redundancy for a large codebook size might be resulting from a low (3^{rd}) degree polynomial. But observations confirm that a 3^{rd} degree polynomial is good enough to capture the trend (Figure 3). The residue is suggestive of a possibly superposed vibrato-type oscillation ('gamak'). A higher degree polynomial would have the danger of overfitting the transient segments.

Our previous works [7], [11] had shown to have improved retrieval accuracy by incorporating query-dependent preset parameter settings. In line with the same philosophy, we plan to propose task-dependent preset parameter settings for the same retrieval framework in the symbolic measure paradigm. The stylization, per se, smoothes the contour by discarding perceptually irrelevant fast pitch oscillations and also disregards measurement errors in F0 extraction (e.g., refer to Figure 3 at time-stamp 118 sec where a dip in the melodic contour resulted from an unvoiced consonant). In the current experiments we needed only the string sequence (and not the continuous contour), but the stylized contour could find its use in other music information retrieval (MIR) application such as synthesis or perception experiments [26].

VII. Conclusion

While addressing the problem of query based retrieval of raga phrases, we investigated the potential of a time-series to symbol string conversion where both stable notes and smooth pitch curves that serve to connect these are adequately represented. A 3^{rd} degree polynomial model for the pitch transitions provides a pitch contour stylization that preserves the essential form of the melodic shape of a phrase while being relatively insensitive to artist- and context-dependent expressive pitch variations. We plan to add the segment duration as a feature; this may be useful in terms of matching test segment with codebook entries that are similar in duration, overcoming any deficiencies of the normalization methodology. Further, clustering of melodic shapes computed on training data drawn from diverse melodic material has shown to have resulted in a compact and generic dictionary of normalized pitch curves. Tweaking of the same string matching algorithm on the proposed representation has shown improvement in retrieval performance over the previously reported setup used as a baseline. As a future work we plan to closely investigate the residual signal, towards modelling melodic embellishments, e.g. stylistic observations on artist- and context-based expressive singing.

Appendix A
Reconstruction from Quantization

The steps involved to reconstruct the stylized contour is to denormalize the codebook vector to the original time-scale and pitch range. We take the exact reverse steps of the encoding procedure to decode the transient segments (refer to Figure 8). Given a time-series pitch sequence p, where $a = min(p)$ and $b = max(p)$, normalized (within range [0,1]) contour q is obtained as $q = \frac{p-a}{b-a}$. For denormalizing $q = [q[1], q[2], ..., q[L]]$ ensuring end-point matches between two pitch points c and d, we arrive at the equation:

$$q' = \frac{q - \min(q[1], q[L])}{\max(q[1], q[L]) - \min(q[1], q[L])} * \\ [\max(c, d) - \min(c, d)] \quad + \quad \min(c, d) \tag{1}$$

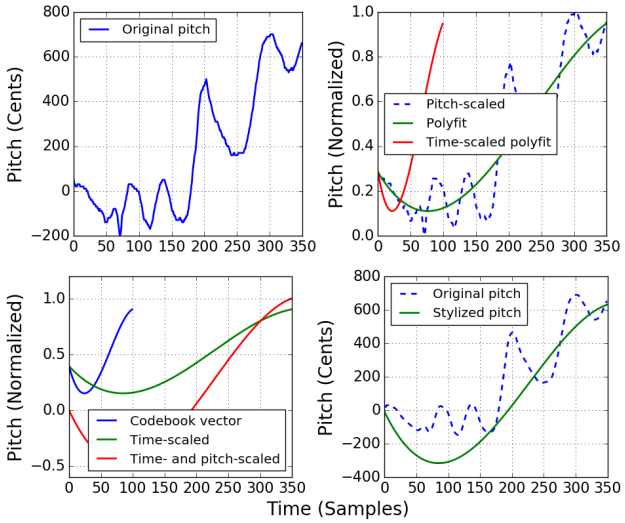

Fig. 8. Steps of (de)normalization of a transient segment to the corresponding stylized contour. The corresponding codebook vector (bottom left) is of index 7. Bottom right shows the (superimposed) original and the stylized contours. The time unit is shown in samples on purpose, refer to Section II-C.

We observe from Figure 8 that the pitch oscillations are neglected and a lowpass trend is obtained as a stylized transient segment. This phenomenon is favorable for retrieval applications, because the oscillations (lit. gamak) is not an essential, but occasional, part of a melodic phrase. Musicians choose to either apply or omit a gamak on based on local context. Hence we are at a better chance of retrieving a phrase independent of the presence of gamak. We have not added smoothness constraints at end-points, a derivative-based approach to ensure smooth transition between stable notes and transients is posed as a future work.

Acknowledgment

This work received partial funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 267583 (CompMusic). Authors Ashwin Lele and Saurabh Pinjani equally contributed to this paper.

References

[1] P. van Kranenburg, A. Volk, and F. Wiering, "A comparison between global and local features for computational classification of folk song melodies," *Journal of New Music Research (JNMR)*, vol. 42, no. 1, pp. 1–18, 2013.

[2] D. Mullensiefen and K. Frieler, "Optimizing measures of melodic similarity for the exploration of a large folk song database," in *Proc. of Int. Soc. for Music Information Retrieval (ISMIR)*, 2004.

[3] J. B. Prince, "Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 6, 2014.

[4] A. Vidwans, K. K. Ganguli, and P. Rao, "Classification of indian classical vocal styles from melodic contours," in *Proc. of the 2nd CompMusic Workshop*, July 2012.

[5] J. J. Cabrera, J. M. Diaz-Banez, F. J. E. Borrego, E. Gomez, F. G. Martin, and J. Mora, "Comparative melodic analysis of a cappella flamenco cantes," in *Fourth Conference on Interdisciplinary Musicology (CIM)*, 2008, thessaloniki, Greece.

[6] S. Bagchee, *Nād: Understanding Raga Music*. Business Publications Inc, 1998.

[7] K. K. Ganguli, A. Rastogi, V. Pandit, P. Kantan, and P. Rao, "Efficient melodic query based audio search for Hindustani vocal compositions," in *Proc. of the International Society for Music Information Retrieval (ISMIR)*, Oct. 2015, pp. 591–597, Malaga, Spain.

[8] 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 (JNMR)*, vol. 43, no. 1, pp. 115–131, Apr. 2014.

[9] P. Rao, J. C. Ross, and K. K. Ganguli, "Distinguishing raga-specific intonation of phrases with audio analysis," *Ninaad*, vol. 26-27, no. 1, pp. 59–68, Dec. 2013.

[10] K. K. Ganguli, "How do we 'See' & 'Say' a raga: A Perspective Canvas," *Samakalika Sangeetham*, vol. 4, no. 2, pp. 112–119, Oct. 2013.

[11] A. Lele, S. Pinjani, K. K. Ganguli, and P. Rao, "Improved melodic sequence matching for query based searching in Indian classical music," in *Proc. of Frontiers of Research on Speech and Music (FRSM)*, Nov. 2016, Baripada, India.

[12] V. Rao and P. Rao, "Vocal melody extraction in the presence of pitched accompaniment in polyphonic music," *IEEE Trans. on Audio, Speech & Language Processing*, vol. 18, no. 8, 2010.

[13] 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 (JNMR)*, vol. 43, no. 1, pp. 53–71, 2014.

[14] K. K. Ganguli, S. Gulati, X. Serra, and P. Rao, "Data-driven exploration of melodic structures in Hindustani music," in *Proc. of the International Society for Music Information Retrieval (ISMIR)*, Aug. 2016, pp. 605–611, New York, USA.

[15] P. K. Ghosh and S. S. Narayanan, "Pitch contour stylization using an optimal piecewise polynomial approximation," *IEEE Signal Processing Letters*, vol. 16, no. 9, Sep. 2009.

[16] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time-series with implications for streaming algorithms," in *Proc. of ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 2003.

[17] T. Eerola and P. Toiviainen, "MIR in Matlab: The MIDI Toolbox," in *Proc. of Int. Soc. for Music Information Retrieval (ISMIR)*, 2004.

[18] C. R. Adams, "Melodic contour typology," *Ethnomusicology*, vol. 20, no. 2, pp. 179–215, May 1976.

[19] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, 2006.

[20] S. Gulati, J. Serra, K. K. Ganguli, and X. Serra, "Landmark detection in Hindustani music melodies," in *Proc. of Int. Computer Music, Sound and Music Computing*, 2014, pp. 1062–1068, Athens, Greece.

[21] A. K. Datta, R. Sengupta, N. Dey, and D. Nag, "A methodology for automatic extraction of 'meend' from the performances in Hindustani vocal music," *Journal of ITC Sangeet Research Academy*, vol. 21, pp. 24–31, 2007.

[22] ——, "Automatic classification of 'meend' extracted from the performances in Hindustani vocal music," in *Proc. of Frontiers of Research on Speech and Music (FRSM)*, 2008.

[23] C. Gupta and P. Rao, *Speech, Sound and Music Processing: Embracing Research in India; CMMR 2011-FRSM 2011*. Springer Berlin Heidelberg, 2012, ch. Objective Assessment of Ornamentation in Indian Classical Singing, pp. 1–25.

[24] S. Gulati, J. Serra, K. K. Ganguli, S. Senturk, and X. Serra, "Time-delayed melody surfaces for rāga recognition," in *Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, Aug. 2016, pp. 751–757, New York, USA.

[25] X. Serra, "Creating research corpora for the computational study of music: the case of the Compmusic project," in *Proc. of the 53rd AES Int. Conf. on Semantic Audio*, London, 2014.

[26] K. K. Ganguli and P. Rao, "Perceptual anchor or attractor: How do musicians perceive raga phrases?" in *Proc. of Frontiers of Research on Speech and Music (FRSM)*, Nov. 2016, Baripada, India.