

STRUCTURAL SEGMENTATION AND VISUALIZATION OF SITAR AND SAROD CONCERT AUDIO

Vinutha T.P.

Suryanarayana Sankagiri

Kaustuv Kanti Ganguli

Preeti Rao

Department of Electrical Engineering, IIT Bombay, India

prao@ee.iitb.ac.in

ABSTRACT

Hindustani classical instrumental concerts follow an episodic development that, musicologically, is described via changes in the rhythmic structure. Uncovering this structure in a musically relevant form can provide powerful visual representations of the concert audio that is of potential value in music appreciation and pedagogy. We investigate the structural analysis of the metered section (*gat*) of concerts of two plucked string instruments, the sitar and sarod. A prominent aspect of the *gat* is the interplay between the melody soloist and the accompanying drummer (*tabla*). The tempo as provided by the *tabla* together with the rhythmic density of the sitar/sarod plucks serve as the main dimensions that predict the transition between concert sections. We present methods to access the stream of *tabla* onsets separately from the sitar/sarod onsets, addressing challenges that arise in the instrument separation. Further, the robust detection of tempo and the estimation of rhythmic density of sitar/sarod plucks are discussed. A case study of a fully annotated concert is presented, and is followed by results of achieved segmentation accuracy on a database of sitar and sarod *gats* across artists.

1. INTRODUCTION

The repertoire of North Indian (Hindustani) classical music is characterized by a wide variety of solo instruments, playing styles and melodic material in the form of *ragas* and compositions. However, across all these, there is a striking universality in the concert structure, i.e., the way in which the music is organized in time. The temporal evolution of a concert can be described via changes in the rhythm of the music, with homogenous sections having identical rhythmic characteristics. The metric tempo and the surface rhythm, two important aspects of rhythm, characterize the individual sections. Obtaining these rhythm features as they vary with time gives us a rich transcription for music appreciation and pedagogy. It also allows rhythm-base segmentation with potential applications in concert sum-

marization, music navigation. This provides a strong motivation for the rhythmic analysis of Hindustani classical concert audio.

Rhythmic analyses of audio has been widely used for music classification and tempo detection [1–3]. It has also been applied to music segmentation [4,5] although timbre- and harmony-based segmentation are more common. Recently, computational descriptions of rhythm were studied for Indian and Turkish music [6]. Beat detection and cycle length annotation were identified as musically relevant tasks that could benefit from the computational methods.

In this paper, we focus on the Hindustani classical instrumental concert which follows an established structure via a specified sequence of sections, viz. *alap-jod-jhala-gat* [7]. The first three are improvised sections where the melody instrumentalist (sarod/sitar) plays solo, and are often together called the “*alap*”. The *gat* or composed section is marked by the entry of the *tabla*. The *gat* is further subdivided into episodes as discussed later. The structure originated in the ancient style of *dhrupad* singing where a *raga* performance is subdivided unequally into the mentioned temporally ordered sections.

In the present work, we consider concerts of two plucked string instruments, sitar and sarod, which are major components of Indian instrumental music. The two melodic instruments share common origins and represent the fretted and unfretted plucked monochords respectively. Verma et. al. [8] have worked on the segmentation of the unmetred section (*alap*) of such concerts into *alap-jod-jhala* based purely on the tempo and its salience. They use the fact that an increase in regularity and pluck density marked the beginning of *jod*. Higher pluck density was captured via increases in the energy and in the estimated tempo. The transition to *jhala* was marked by a further rise in tempo and additionally distinguished by the presence of the “*chikari*” strings.

In this paper, we focus on the rhythmic analysis and segmentation of the *gat*, or the *tabla*-accompaniment region, into its sections. Owing to differences in the rhythmic structure of the *alap* and the *gat*, the challenges involved in this task are different from those addressed in [8]. In the *gat*, the *tabla* provides a definite meter to the concert by playing a certain *tala*. The tempo, as set by the *tabla*, is also called the metric tempo. The tempo of the concert increases gradually with time, with occasional jumps. While the *tabla* provides the basic beats (*theka*), the melody instrumentalist plays the composition interspersed



© Vinutha T.P., Suryanarayana Sankagiri, Kaustuv Kanti Ganguli, Preeti Rao. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Vinutha T.P., Suryanarayana Sankagiri, Kaustuv Kanti Ganguli, Preeti Rao. “STRUCTURAL SEGMENTATION AND VISUALIZATION OF SITAR AND SAROD CONCERT AUDIO”, 17th International Society for Music Information Retrieval Conference, 2016.

with *raga*-based improvisation (“*vistaar*”). A prominent aspect of instrumental concerts is that the *gat* is characterized by an interplay between the melody instrumentalist and the drummer, in which they alternate between the roles of soloist and timekeeper [7, 9]. The melody instrument can switch to fast rhythmic play (“*layakari*”) over several cycles of the *tabla*. Then there are interludes where the *tabla* player is in the foreground (“*tabla solo*”), improvising at a fast rhythm, while the melody instrumentalist plays the role of the timekeeper by playing the melodic refrain of the composition cyclically. Although both these sections have high surface rhythm, the term “rhythmic density” refers to the stroke density of the *sarod/sitar* [10], and therefore is high only during the *layakari* sections. The values of the concert tempo and the rhythmic density as they evolve in time can thus provide an informative visual representation of the concert, as shown in [10].

In order to compute the rhythmic quantities of interest, we follow the general strategy of obtaining an onset detection function (ODF) and then computing the tempo from it [11]. To obtain the surface rhythm, we need an ODF sensitive to all onsets. However, to calculate the metric tempo, as well as to identify sections of high surface rhythm as originating from the *tabla* or *sarod/sitar*, we must discriminate the *tabla* and *sitar/sarod* stroke onsets. Both the *sitar* and the *sarod* are melodic instruments but share the percussive nature of the *tabla* near the pluck onset. The *tabla* itself is characterized by a wide variety of strokes, some of which are diffused in time and have decaying harmonic partials. This makes the discrimination of onsets particularly challenging.

Our new contributions are the (i) proposal of a *tabla*-specific onset detection method, (ii) computation of the metric tempo and rhythmic density of the *gat* over a concert to obtain a rhythmic description which matches with one provided by a musician, (iii) segmentation of the *gat* into episodes based on the rhythm analysis. These methods are demonstrated on a case study of a *sarod gat* by a famous artist, and are further tested for segmentation accuracy on a manually labeled set of *sitar* and *sarod gats*.

In section 2, we present the proposed *tabla*-sensitive ODF and test its effectiveness in selectively detecting *tabla* onsets from a dataset of labeled onsets drawn from a few *sitar* and *sarod* concerts. In section 3, we discuss the estimation of tempo and rhythmic density from the periodicity of the onset sequences and present the results on a manually annotated *sarod gat*. Finally, we present the results of segmentation on a test set of *sitar* and *sarod gats*.

2. ONSET DETECTION

A computationally simple and effective method of onset detection is the spectral flux which involves the time derivative of the short-time energy [12]. The onsets of both the percussive as well as the string instrument lead to a sudden increase in energy, and are therefore detected well by this method. A slight modification involves using a biphasic filter to compute the derivative [13]. This enhances the detection of *sarod/sitar* onsets, which have a slow decay

in energy, and leads to a better ODF. Taking the logarithm of the energy before differencing enhances the sensitivity to weaker onsets. We hereafter refer to this ODF as the spectral flux-ODF (SF-ODF), and is given by Eq. 1. ($h[n]$ denotes the biphasic filter as in [13])

$$SF-ODF[n] = h[n] * \log\left(\sum_{k=0}^{N/2} |X[n, k]|\right) \quad (1)$$

Figure 1, which contains a *sarod* concert excerpt, illustrates the fact that SF-ODF is sensitive to both *sarod* and *tabla* onsets. In this example, and in all subsequent cases, we compute the spectrum by using a 40ms Hamming window on audio sampled at 16 kHz. The spectrum (and therefore the ODF) is computed at 5 ms intervals. Fig. 1(a) shows the audio waveform where onsets can be identified by peaks in the waveform envelope. Onsets can also be seen as vertical striations in the spectrogram (Fig. 1(b)). SF-ODF is shown in Fig. 1(c). Clearly, SF-ODF is not *tabla*-selective.

In order to obtain a *tabla*-sensitive ODF, we need to exploit some difference between *tabla* and *sarod/sitar* onsets. One salient difference is that in the case of a *tabla* onset, the energy decays very quickly (< 0.1 s). In contrast, the energy of a *sitar/sarod* pluck decays at a much slower rate (> 0.5 s). This difference is captured in the ODF that we propose, hereafter called as P-ODF. This ODF counts the number of bins in a spectral frame where the energy increases from the previous frame, and is given by Eq. 2. This method is similar in computation to the spectral flux method in [12]; we take the 0-norm of the half-wave rectified energy differences, instead of the 2-norm [12] or 1-norm [14]. However, the principle on which this ODF operates is different from the spectral flux ODF. P-ODF detects only those onsets that are characterised by a wide-band event, i.e., onsets that are percussive in nature. Unlike the spectral flux ODF, it does not rely on the magnitude of energy change. In our work, this proves to be an advantage as it detects weak onsets of any instrument better, provided they are wide-band events.

$$P-ODF[n] = \sum_{k=0}^{N/2} \mathbb{1}\{|X[n, k]| > |X[n - 1, k]|\} \quad (2)$$

From Fig. 1(d), we see that P-ODF peaks at the onset of a *tabla* stroke, as would be expected due to the wide-band nature of these onsets. It also peaks for *sarod* onsets, as these onsets have a percussive character. Thus, it is sensitive to all onsets of interest, and can be potentially used as generic ODF in place of SF-ODF, for *sitar/sarod* audio. What is of more interest is the fact that in the region immediately following a *tabla* onset, this count falls rapidly while such a pattern is not observed for *sarod* onsets (see Fig. 1(d)). This feature is seen because of the rapid decrease in energy after a *tabla* onset. In the absence of any activity, the value of the ODF is equal to half the number of bins as the energy changes from frame to frame in a bin due to small random perturbations.

The sharp downward lobe in P-ODF is a striking feature of *tabla* onsets, and can be used to obtain a *tabla*-sensitive

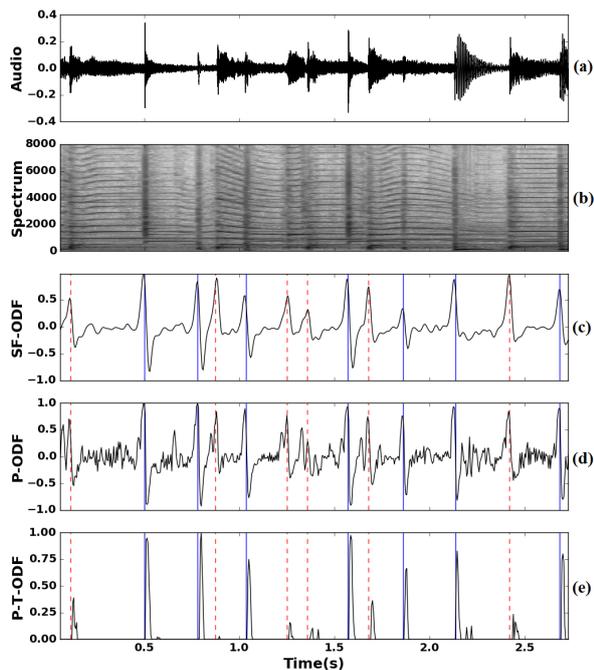


Figure 1: (a) Audio waveform, (b) Spectrogram, (c) SF-ODF, (d) P-ODF and (e) P-T-ODF of an excerpt of a sarod concert. All ODFs normalized. Tabla onsets marked in blue solid lines; sarod onsets marked in red dashed lines

ODF. We normalize the mean-removed function to $[-1,1]$ and consider only the negative peaks of magnitude that exceed the empirically chosen threshold of 0.3. We call our proposed tabla-sensitive ODF as P-T-ODF. An example is shown in Fig. 1(e).

We wish to establish that the P-T-ODF performs better as a tabla-sensitive ODF than other existing methods. The spectral flux method is known to be sensitive to both onsets, and performs poorly as a tabla-sensitive ODF. However, one could hope to obtain better results by computing the ODF on a percussion-enhanced audio. Fitzgerald [15] proposes a median-filter based method for percussion enhancement that exploits the relatively high spectral variability of the melodic component of a music signal to suppress it relative to the more repetitive percussion. We used this method to preprocess our *gat* audio to obtain what we call the enhanced audio signal (tabla is enhanced), and test the SF-ODF on it. With this as the baseline, we compare our P-T-ODF applied to the original audio. In parallel, we wish to justify our claim that the P-ODF is a suitable ODF for detecting sarod/sitar as well as tabla onsets.

We evaluate our ODFs on a dataset of 930 labeled onsets comprising 158 sitar, 239 sarod and 533 tabla strokes drawn from different sections of 6 different concert *gats*. Onsets were marked by two of the authors, by carefully listening to the audio, and precisely locating the onset instant with the aid of the waveform and the spectrogram. We evaluate P-ODF and SF-ODF, derived from the original audio, for detection of all onsets, with SF-ODF serving as a baseline. The obtained ROC is shown in Fig. 2(a). We also evaluate P-T-ODF, derived from the original audio

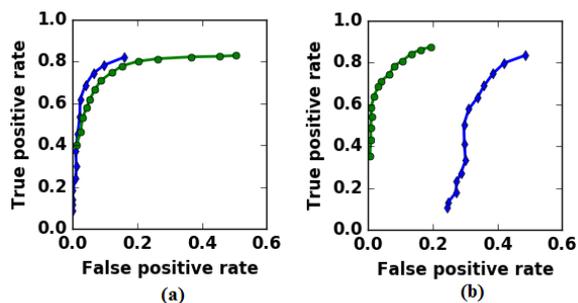


Figure 2: (a) All-onsets ROC for SF-ODF (blue diamonds) and P-ODF (green circles); (b) Tabla-onsets ROC for SF-ODF on enhanced audio (blue diamonds), and P-T-ODF on original audio (green circles)

and compare it with SF-ODF from enhanced audio, for detection of tabla onsets. The corresponding ROC is shown in Fig. 2(b).

We observe that the spectral flux and the P-ODF perform similarly in the all-onsets ROC of Fig. 2(a). A close examination of performance on the sitar and sarod *gats* separately revealed that the P-ODF performed marginally better than SF-ODF on sarod *gats*, while the performance of the spectral flux ODF was better than the P-ODF on the sitar strokes. In the following sections, we use the P-ODF to detect all onsets in sarod *gats* and the spectral flux-ODF on the sitar *gats*. We also note from Fig. 2(b) that the P-T-ODF fares significantly better than the SF-ODF applied on tabla-enhanced signal. The ineffectiveness of Fitzgerald’s percussion enhancement is explained by the percussive nature of both instruments as well as the high variation (intended and unintended) of tabla strokes in performance. We observed that the median filtering did a good job of suppressing the sarod/sitar harmonics in but not their onsets. The P-T-ODF is established as an effective way to detect tabla onsets exclusively in both sarod and sitar *gats*.

3. RHYTHMOGRAMS AND TEMPO ESTIMATION: A CASE STUDY

A rhythm representation of a *gat* can be obtained from the onset detection function by periodicity analysis via the autocorrelation function (ACF) or the DFT. A rhythmogram uses the ACF to represent the rhythmic structure as it varies in time [16]. Abrupt changes in the rhythmic structure can be detected for concert section boundaries. The dominant periodicity at any time can serve as an estimate of the perceived tempo [5, 11]. Our goal is to meaningfully link the outcomes of such a computational analysis to the musicological description of the concert.

In this section, we present the musicological and corresponding computational analyses of a commercially recorded sarod *gat* (*Raga Bahar, Madhyalaya, Jhaptal*) by legendary sarodist Ustad Amjad Ali Khan. The musicological description was prepared by a trained musician on lines similar to the sitar *gat* case study by Clayton [17] and is presented next. The computational analysis involved applying the onset detection methods to obtain a rhythm rep-

resentation that facilitates the detection of the metric tempo and rhythmic density as well as the segmentation of the *gat*.

3.1 Annotation by a Trained Musician

A musician with over 15 years of training in Hindustani classical music made a few passes listening to the audio (duration 14 min) to annotate the *gat* at three levels. The first was to segment and label the sequence of distinct episodes as shown in Table 1. These labels reflect the performers’ (i.e. the sarod and tabla players) intentions as perceived by a trained listener. The next two annotation levels involved marking the time-varying metric tempo and a measure of the sarod rhythmic density. The metric tempo was measured by tapping to the tabla strokes that define the *theka* (i.e. the 10 beats of the *Jhaptal* cycle) and computing the average BPM per cycle with the aid of the Sonic Visualizer interface [18]. The metric tempo is constant or slowly increasing across the concert with three observed instants of abrupt change.

The rhythmic density, on the other hand, was obtained by tapping to the sarod strokes and similarly obtaining a BPM per cycle over the duration of the *gat*. Figure 3 shows the obtained curves with the episode boundaries in the background. We note that the section boundaries coincide with abrupt changes in the rhythmic density. The metric tempo is constant or slowly increasing across the concert with three observed instants of abrupt change. The rhythmic density corresponds to the sarod strokes and switches between being once/twice the tempo in the *vistaar* to four times in the *layakari* (rhythmic improvisation by the melody soloist). Although the rhythmic density is high between cycles 20-40, this was due to fast melodic phrases occupying part of the rhythmic cycle during the *vistaar* improvisation. Since this is not a systematic change in the surface rhythm, it was not labeled *layakari* by our musician. In the tabla solo section, although the surface rhythm increases, it is not due to the sarod. Therefore, the tabla solo section does not appear distinctive in the musician’s markings in Figure 3.

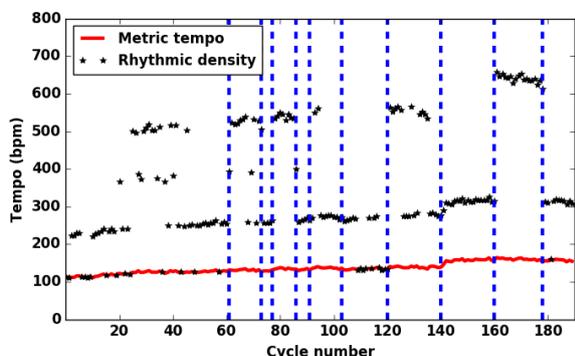


Figure 3: Musicians annotation of tempo and rhythmic density attributes across the *gat*. Dashed lines indicate section boundaries

Sec. No.	Cycles	Time (s)	Label
1	1-61	0-301	Vistaar *
2	62-73	302-356	Layakari
3	74-77	357-374	Vistaar
4	78-86	375-414	Layakari
5	87-91	415-441	Vistaar
6	92-103	442-490	Tabla solo
7	104-120	491-568	Vistaar
8	121-140	567-643	Layakari
9	141-160	644-728	Vistaar #
10	161-178	729-797	Layakari
11	179-190	798-839	Vistaar

Table 1: Labeled sections for the sarod case study. *Tempo increases at 67s & 127s; # also at 657s

3.2 Computational Analysis

3.2.1 Rhythmogram

The onset detection methods of Section 2 are applied over the duration of the concert. We confine our study to two ODFs based on insights obtained from the ROCs of Fig. 2. These are the P-ODF for all onsets and the P-T-ODF for tabla-onsets. Although the P-ODF was marginally worse than spectral flux in Fig. 2(a), it was found to detect weak sarod strokes better while the false alarms were irregularly distributed in time. This property is expected to help us track the sarod rhythmic density better.

The autocorrelation function of the ODFs is computed frame-wise, with a window length of 3 seconds and a hop of 0.5 seconds up to a lag of 1.5 seconds, and is normalized to have a maximum value of 1 in each frame. To improve the representation of peaks across the dynamic range in the rhythmogram, we perform a non-linear scaling of the amplitude of the ACF. For the tabla-centric rhythmogram (from P-T-ODF), we take the logarithm of the ACF between 0.1 and 1; for the generic rhythmogram (from P-ODF), the logarithm is taken between 0.01 and 1 due to its inherently wider dynamic range for peaks. The ACF values below this range are capped to a minimum of -10. This is followed by smoothing in the lag and time axes by moving average filters to length 3 and 10 respectively bringing in short-time continuity.

We thus obtain the two rhythmograms shown in Figures 4 and 5. We note that the P-ODF all-onsets rhythmogram (Figure 4) captures the homogenous rhythmic structure of each episode of *vistaar*, *layakari* and tabla solo, showing abrupt changes at the boundaries. Each section itself appears homogenous except for some spottiness in the sequence of low amplitude ACF peaks at submultiple lags (such as near 0.1s in the region until 300 s).

The tabla-centric rhythmogram (Figure 5), on the other hand, with its more prominent peaks appearing at lags near 0.5s and multiples, is indicative of a metric (base) tempo of around 120 BPM. We clearly distinguish from this rhythmogram, the tabla solo segment (where the tabla surface rhythm shoots up to 8 times the metric tempo). We observe, as expected, that the sarod *layakari* sections are

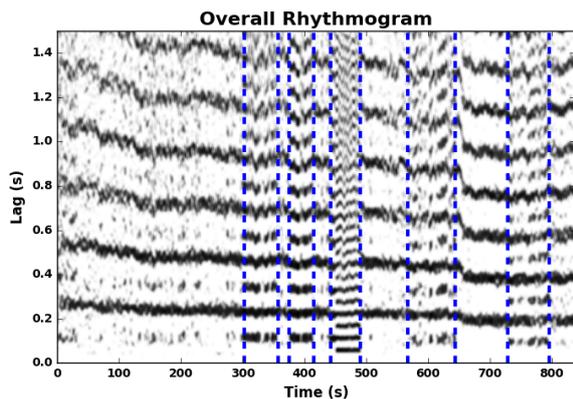


Figure 4: All-onsets rhythmogram from P-ODF

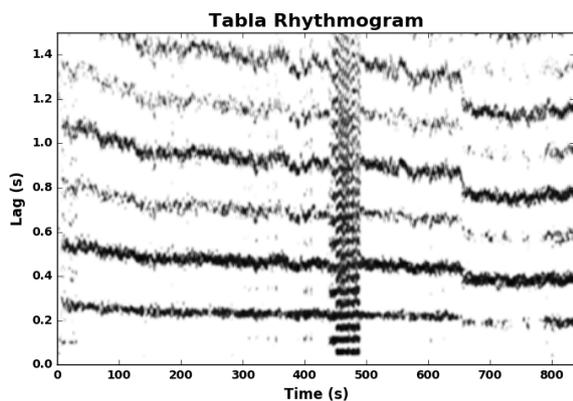


Figure 5: Tabla centric rhythmogram from P-T-ODF

completely absent from the tabla-centric rhythmogram.

3.2.2 Tempo and surface rhythm estimation

The rhythmograms provide interesting visual representations of the rhythmic structure. However a visual representation that is more amenable to immediate interpretation by musicians and listeners would have to parallel the musician's annotation of Fig. 3. We therefore must process the rhythmograms further to extract the relevant attributes of metric tempo and sarod rhythmic density. We present next the frame-wise estimation of these from the ACF vectors of the smoothed rhythmograms of Figs. 4 and 5.

The basic or metric tempo is obtained from the tabla rhythmogram (Fig. 5) by maximizing the mean of the peaks at candidate lags and corresponding lag multiples over the lag range of 50ms to 750ms (1200BPM to 80BPM). The estimated time-varying metric tempo is shown in Fig. 6(a) superposed on the ground-truth annotation (x-axis converted to time from cycles as in Fig. 3). We observe a near perfect match between the two with the exception of the tabla-solo region, where the surface rhythm was tracked. We use our knowledge that the surface rhythm would be a multiple of the metric tempo. Dividing each tempo value by that multiple that maintains continuity of the tempo gave us the detected contour of Fig. 6(a).

The rhythmic density of the sarod is the second musical attribute required to complete the visual representa-

tion. This is estimated from the generic (P-ODF) rhythmogram of Fig. 4 in a manner similar to that used on the tabla-centric version. The single difference is that we apply a bias favouring lower lags in the maximum likelihood tempo estimation. A weighting factor proportional to the inverse of the lag is applied. The biasing is motivated by our stated objective of uncovering the surface rhythmic density (equivalent to the smallest inter-onset interval).

The obtained rhythmic density estimates are shown in Fig. 6(b), again in comparison with the ground truth marked by the musician. The ground-truth markings have been converted to the time axis while smoothening lightly to remove the abrupt cycle-to-cycle variations in Fig. 3. We note that the correct tempo corresponding to the sarod surface rhythm is captured for the most part. The *layakari* sections are distinguished from the *vistaar* by the doubling of the rhythmic density. Obvious differences between the ground-truth and estimated rhythmic density appear in (i) the tabla solo region due to the high surface rhythm contributed by tabla strokes. Since P-ODF captures both the instrument onsets, this is expected. Another step based on the comparison of the two rhythmograms would easily enable us to correct this; (ii) intermittent regions in the 0-300s region of the *gat*. This is due to the low amplitude ACF peaks arising from the fast rhythmic phrases discussed in Sec. 3.1.

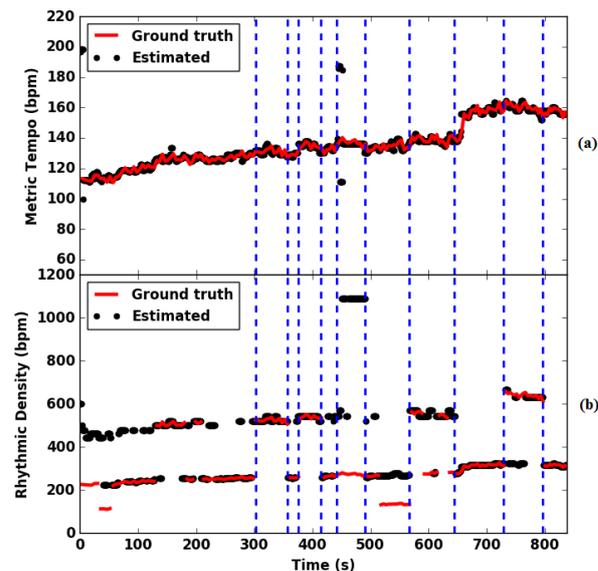


Figure 6: (a) Estimated metric tempo with musician's marked tempo. (b) Estimated rhythmic density with musician's marked rhythmic density

4. SEGMENTATION PERFORMANCE

The all-onsets rhythmogram provides a clear visual representation of abrupt rhythmic structure changes at the section boundaries specified by the ground-truth labels. In order to algorithmically detect the segment boundaries, we resort to the method of the similarity distance matrix (SDM) where peaks in the novelty function derived from diagonal kernel convolution can help identify instants of

change [19]. We treat the ACF at each time frame as a feature vector that contains the information of the local rhythmic structure. We compute the correlation distance between the ACF of every pair of frames across the concert to obtain the SDM. The diagonal of the SDM is then convolved with a checker-board kernel of $25s \times 25s$ to compute the novelty function. Local maxima in the novelty function are suitably thresholded to locate instants of change in the rhythmic structure. Figure 7 shows the SDM and novelty function computed on the rhythmogram of Figure 5 corresponding to the case study sarod *gat*. We observe that all the known boundaries coincide with sharp peaks in the novelty function. The *layakari-vistaar* boundary at 644s is subsumed by the sudden tempo change at 657s due to the minimum time resolution imposed by the SDM kernel dimensions. We next present results for performance of our system on segment boundary detection across a small dataset of sitar and sarod *gats*.

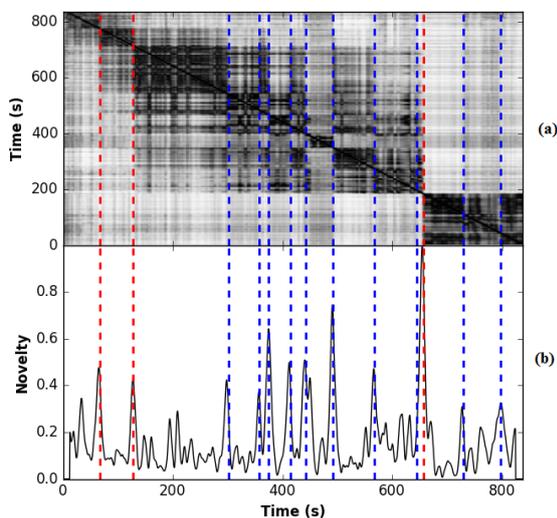


Figure 7: SDM and novelty curve for the case study sarod *gat* (whose rhythmogram appears in Figure 5). The blue dashed lines indicate ground-truth section boundaries as in Table 1. The red dashed lines indicate ground-truth instants of metric tempo jump.

4.1 Dataset

Our dataset for structural segmentation analysis consists of three sitar and three sarod *gats*, by four renowned artists. We have a total of 47 min of sarod audio (including the case study *gat*) and 64 min of sitar audio. Just like the case-study *gat*, each *gat* has multiple sections which have been labelled as *vistaar*, *layakari* and *tabla solo*. Overall we have 37 *vistaar* sections, 21 *layakari* sections and 25 *tabla solo* sections. Boundaries have been manually marked by noting rhythm changes upon listening to the audio. Minimum duration of any section is found to be 10s.

Gat. No.	Dur (min)	Method Used	Hit rate	False Alarms
1	14	P-ODF	13/13	0
2	24	P-ODF	14/14	1
3	9	P-ODF	20/20	2
4	16	SF-ODF	17/17	2
5	21	SF-ODF	11/12	1
6	27	SF-ODF	14/14	4

Table 2: Boundary detection results for 6 *gats*

4.2 Boundary Detection Performance

For each concert, the novelty function was normalised to [0,1] range and peaks above a threshold of 0.3 were taken to indicate boundary instants. We consider the detected boundary as a hit if it lies within 12.5 s of a marked boundary considering our kernel dimension of 25 s. We expect to detect instants where there is either a change in surface rhythm or an abrupt change in the metric tempo. Consistent with our onsets detection ROC study of Section 2, we observed that the P-ODF method gave better segmentation results than the spectral flux for sarod *gats*, while the reverse was true for sitar *gats*. Table 2 shows the corresponding segmentation performance for the sarod (1-3) and sitar (4-6) *gats*. We observe a nearly 100% boundary detection rate with a few false detections in each concert. The false alarms were found to be triggered by instances of *tabla* improvisation (change in stroke pattern) without a change in the metric tempo or basic *theka*.

5. CONCLUSION

Motivated by a compelling visual depiction of the rhythmic structure of a Hindustani classical sitar concert [10], we set about an effort to reproduce automatically, with MIR methods, the manual annotation created by expert musicians. A novel onset detection function that exploited the stroke characteristics of the melodic and percussive instrument, and additionally discriminated the two, proved effective in obtaining rhythm representations that separately captured the structural contributions of the *tabla* and the sitar/sarod. Tempo detection on the separate rhythm vectors provided estimates of the metric tempo and rhythmic density of the sitar/sarod. Segmentation using an SDM on the rhythm vectors provided section boundary estimates with high accuracy. The system now needs to be tested on a large and diverse database of sitar and sarod concerts. Further, given that the rhythmogram contains more information than we have exploited in the current work, we propose to develop methods for section labeling and other relevant musical descriptors.

Acknowledgement: 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). Also, part of the work was supported by Bharti Centre for Communication in IIT Bombay.

6. REFERENCES

- [1] Geoffroy Peeters. Rhythm Classification Using Spectral Rhythm Patterns. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 644–647, 2005.
- [2] Fabien Gouyon, Simon Dixon, Elias Pampalk, and Gerhard Widmer. Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the AES 25th International Conference*, pages 196–204, 2004.
- [3] Klaus Seyerlehner, Gerhard Widmer, and Dominik Schnitzer. From Rhythm Patterns to Perceived Tempo. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 519–524, 2007.
- [4] Kristoffer Jensen, Jieping Xu, and Martin Zachariassen. Rhythm-Based Segmentation of Popular Chinese Music. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 374–380, 2005.
- [5] Peter Grosche, Meinard Müller, and Frank Kurth. Cyclic tempogram—a mid-level tempo representation for music signals. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 5522–5525, 2010.
- [6] Ajay Srinivasamurthy, André Holzapfel, and Xavier Serra. In search of automatic rhythm analysis methods for Turkish and Indian art music. *Journal of New Music Research*, 43(1):94–114, 2014.
- [7] Bonnie C Wade. *Music in India: The classical traditions*, chapter 7: Performance Genres of Hindustani Music. Manohar Publishers, 2001.
- [8] Prateek Verma, T. P. Vinutha, Parthe Pandit, and Preeti Rao. Structural segmentation of Hindustani concert audio with posterior features. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 136–140, 2015.
- [9] Sandeep Bagchee. *Nad: Understanding Raga Music*. Business Publications Inc., India, 1998.
- [10] Martin Clayton. *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*, chapter 11: A case study in rhythmic analysis. Oxford University Press, UK, 2001.
- [11] Geoffroy Peeters. Template-based estimation of time-varying tempo. *EURASIP Journal on Applied Signal Processing*, 2007(1):158–171, 2007.
- [12] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.
- [13] Dik J Hermes. Vowel-onset detection. *Journal of the Acoustical Society of America*, 87(2):866–873, 1990.
- [14] Simon Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digital Audio Effects*, volume 120, pages 133–137, 2006.
- [15] Derry FitzGerald. Vocal separation using nearest neighbours and median filtering. In *IET Irish Signals and Systems Conference (ISSC 2012)*, pages 1–5, 2012.
- [16] Kristoffer Jensen. Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP Journal on Advances in Signal Processing*, 2006(1):1–11, 2006.
- [17] Martin Clayton. Two gat forms for the sitār: a case study in the rhythmic analysis of north indian music. *British Journal of Ethnomusicology*, 2(1):75–98, 1993.
- [18] Chris Cannam, Christian Landone, and Mark Sandler. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1467–1468, 2010.
- [19] Jonathan Foote. Automatic audio segmentation using a measure of audio novelty. In *IEEE International Conference on Multimedia and Expo*, volume 1, pages 452–455, 2000.