AUDIO SEGMENTATION OF HINDUSTANI MUSIC CONCERT RECORDINGS

Vinutha T.P. and Preeti Rao
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076, India
{vinutha, prao}@ee.iitb.ac.in

Abstract

Segmentation of an audio piece into its structural components is crucial to many applications of music information retrieval like repetitive structure finding, audio summarization or fast music navigation. This work investigates the approach to automatically locate the points of significant change in the rhythmic structure of the Hindustani music concerts, by analyzing local self-similarity. Since Hindustani concerts use tabla accompaniment, an explicit representation of the rhythm is available in terms of the pattern and timing of tabla strokes. Changes in rhythm can come about from a change in the manifestation of tala, tempo or even a change in the surface stroke pattern. The autocorrelation of a novelty function obtained by the detection of percussion onsets forms the rhythm representation. The methods are applied to synthesized and actual tabla solos to explore the suitability of available methods and to arrive at optimal parameter settings that are signal dependent.

1. Introduction

Audio segmentation can be done based on different attributes of music like, melody, rhythm, timbre and dynamics of the music. In this work, we consider music segmentation based on locating boundaries by detecting changes in the local rhythmic structure of the sections. Segmentation of the *khyal* vocal performance into different sections like alap, bada-khyal, chota-khyal involves finding the boundaries that mark the transition between two structural segments. Explicit cue of gross boundaries of these sections are available in terms of change in the tabla signal at the boundaries. The role of tabla here is of timekeeping within the framework of tala and to be supportive in establishing the mood of the raga by the embellishments. However, the challenge here is the complexity of the polyphonic audio and high degree of variability in the improvisation. Hence, the algorithms are applied to synthesized audio having controlled variability in different aspects of rhythm and actual tabla solos, having tabla as the lead instrument demonstrating maximum rhythmic variability and complexity, to gain an insight into the available approaches, and also to achieve appropriate parameter selection.

This work investigates methods to automatically locate the points of significant change in the rhythmic structure of the Hindustani music concert by analyzing the local self-similarity of a rhythm representation computed with a sliding window over the audio recording. The autocorrelation of a novelty function obtained by the detection of percussion onsets forms the rhythm representation. Most of the earlier audio segmentation work is focused on western music and Chinese music [1] [2] while very few have worked on Hindustani classical music.

2. Rhythm in Hindustani music

In a broader sense rhythm refers to all aspects of musical time patterns, such as, the way syllables of the lyrics are sung, or the way the strokes of the instrumental music are played or the inherent tempo of the melodic piece.The temporal structure of tal in Hindustani music can be described by the sub-beat structure 'vistar', primary pulse level 'matra', the section by 'vibhag' and complete rhythmic cycle, an 'avart'. Indian music has traditionally three main tempos or laya, *vilambit* (slow), *madhya* (medium), and *drut* (fast). Matra rate defines tempo in madhya laya. Theka is the defined structure of a tala. In a performance, the tabla player will deviate from the usual pattern, will increase rhythmic density, permute measures, to constitute *surface rhythm* but within the cycle. Hence, the rhythmic change in a performance can be due to either the tala change or the tempo change or even the change in the manifestation of the tal [3].

A *khyal* vocal performance usually comprises unmetered alap, followed by bada khyal, rendered either in Madhya or vilambit laya and chota khyal section, rendered in drut or ati- drut laya. These bada khyal and chota khyal sections in turn can consist of different sections like, sthayi, antara, sargam or tan segments. And, the tabla solo performance, exhibits the most complex and detailed percussion representation through sections like, peshkar, which is an introductory section, thematic development in kaidas, fast and flowing version in relas and pre-composed sections called gats. Among these, kaida is the most structurally sophisticated compositional form comprising an opening theme and a series of variations based on the opening theme. In these sections the tempo will be maintained same even though rhythmic density will increase after the opening of kaida.

3. Audio Segmentation System

Proposed Audio segmentation system that consists of three main modules as, detection of percussion onsets of an audio, deriving rhythmic structure and locating segment boundaries by similarity measure between rhythmic frames is shown in Fig 3.1. Autocorrelation of the onsets derived from the audio can be used as the rhythmic feature comprising tempo, rhythmic density or the pattern of strokes, either the basic or the improvised one.

Percussion onset, the event that marks the beginning of the transient period can be characterized by the sudden burst of energy or the change in the short time spectrum of the signal. Sub-band spectral flux represents the change in magnitude in the frequency bins, restricted to the desired band. Detection function, SF(n) is ,

$$SF(n) = \sum_{k=1}^{K} H\left[(|X(n,k)| - |X(n-1,k)|)W(k)\right] \tag{3.1}$$

where, X(n,k) is the short time DFT of the audio, H(x) is the half-wave rectifier function, to count only the onsets rather than offsets, and W[k] is the band limiting filter response with unity gain in the desired frequency region. Instead of involving only a pair of adjacent frames in the detection functions in Eq. (3.1), higher order smoothing and differencing involving multiple frames is used in this analysis [4]. To circumvent the masking effects due to simultaneously occurring events in a polyphonic music, the signal is analyzed in band-wise fashion, with the band boundaries as of 0, 150, 1000, 4000 and 8000Hz for B1, B2, B3 and B4 respectively are used taking into consideration the tabla acoustics [5] [6].

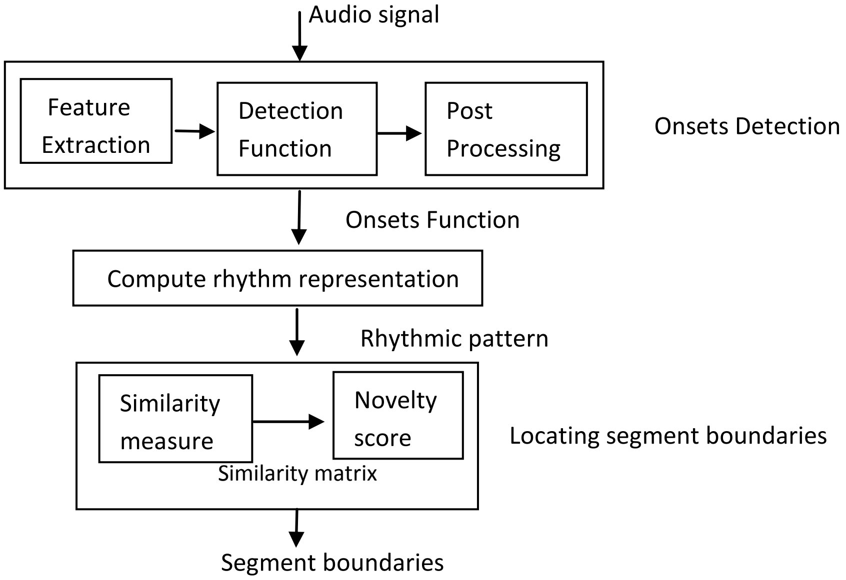

Figure 3.1 Audio segmentation system

A robust approach to capture the salient periodicities and strengths of onsets derived from the audio representing the rhythmic pattern from cycle to cycle is the autocorrelation function (ACF). ACF, $r_n(k)$ of the spectral flux can be expressed as,

$$r_n(k) = \sum_{m=0}^{N-1-k} [sf(n+m)w^{'}(m)][sf(n+m+k)w^{'}(k+m)] \tag{3.2}$$

where, n is the block index, N is the block size and k is the lag. Here, autocorrelation of the spectral flux feature is computed in overlapping blocks of window length comparable to the cycle length, and 0.5s overlap. Only the information between zero and 3s lag is retained as rhythmic feature. Progression of rhythm with time can be represented by the rhythm model rhythmogram, a two dimensional time-pulse representation with lag-time on y- axis, time position on the x-axis and the autocorrelation values visualized as intensity [1][5].

Parameterized rhythmic feature vectors are compared using the correlation distance to arrive at *self-distance matrix* (SDM), a two dimensional representation, that is an indication of the novelty measure that will have high values at the segment boundaries. Correlating the checker board kernel of desired width with this matrix will yield the *novelty score*, whose extrema points to the boundaries.

4. Experimentation and discussion

Database consists of selected Hindustani classical vocal concerts by renowned musicians in Tintal and Ektal are chosen, ensuring that the performances are rendered in different laya. Owing to the complexity of polyphonic nature of vocal concerts, tabla solo performances, having tabla as the lead instrument and having highest rhythmic improvisation are analyzed. To establish the ground truth for the task, segment boundaries in these performances have been marked with careful hearing and ensuring the musical concepts by discussion with musicians.

Different segments appearing in the tabla solo performance by ZH are Peshkar, Kaidas, Gats, Rela and Roun and the acoustic description of this performance is shown in table 1. In Kaida1 slower section of tabla solo performance by U. Zakir Hussain, stroke density is 128 strokes per cycle, with inter Stroke interval (ISI) of 0.22s. It is captured by the novelty function or the onsets function of the frame. Corresponding to this, autocorrelation peaks are at 0.22s and small peaks at 0.08 represent the improvisation strokes as shown in fig 4.1.

Table1: Acoustic descriptions of segments in the tabla solo by ZH

<table>
<tr><th colspan="3">Segment (duration)</th><th>Duration of one cycle</th><th>Strokes per cycle</th><th>Remarks</th></tr>
<tr><td colspan="2">Peshkar (8min) *Slow*</td><td>6.8min</td><td rowspan="2">26s to 24.6s</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td colspan="2">*Fast*</td><td>1.2min</td></tr>
<tr><td rowspan="10">Kaidas</td><td>Kaida1 *S*</td><td>2.7min</td><td rowspan="2">28.3s to 24.9s</td><td rowspan="6">128 strokes in slower section and doubled in faster section</td><td rowspan="8">No change in tempo</td></tr>
<tr><td>(4.1min) *F*</td><td>1.4min</td></tr>
<tr><td>Kaida2 *S*</td><td>0.83min</td><td rowspan="2">27.3s to 25.2s</td></tr>
<tr><td>(2.15min) *F*</td><td>1.72min</td></tr>
<tr><td>Kaida3 S</td><td>1.37min</td><td rowspan="2">28.6s to 25.3s</td></tr>
<tr><td>(2.23min) F</td><td>0.86min</td></tr>
<tr><td>Kaida4 S</td><td>2.2min</td><td rowspan="2">21.7s to 20.9s</td><td rowspan="2">96 strokes in slower section and increased in multiples of 1.5, 3 6 and 8 times in faster sections</td></tr>
<tr><td>(3.6min) F</td><td>1.4min</td></tr>
<tr><td>Kaida5</td><td>50.6s</td><td rowspan="2">16s to 15s</td><td rowspan="2"></td><td rowspan="2">Tempo increased slightly</td></tr>
<tr><td>(7.2min)</td><td>6.3min</td></tr>
<tr><td colspan="3">Gat 1(2.2min)</td><td>4s</td><td></td><td rowspan="3">Tempo has increased much. *Ati-drut* section</td></tr>
<tr><td colspan="3">Rela (1.08min)</td><td>4s</td><td></td></tr>
<tr><td colspan="3">Gat2</td><td>3.7s</td><td></td></tr>
</table>

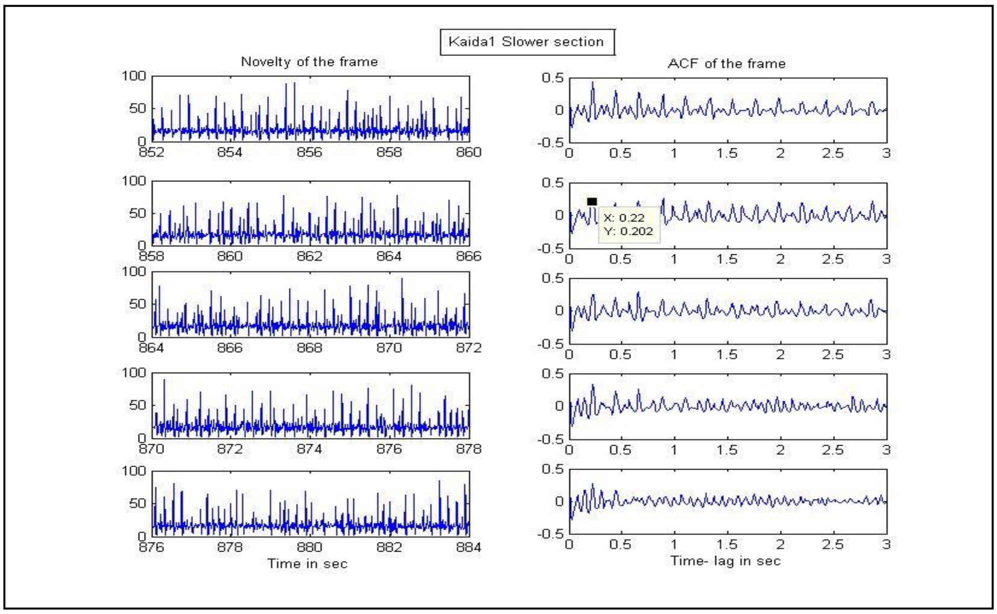

Figure 4.1: Novelty or Onsets function and the ACF of 5 frames in Kaida1

Rhythmic progression of kaidas section of tabla solo by U. Zakir Hussain is shown in fig4.2. In slower sections of kaida1, stroke density is 128 strokes per cycle with ISI of 0.22s and in the faster section the stroke density is doubled, leading to the ISI of 0.1s. This is verified in the rhythmogram as the interval between bright peaks. In Kaidas 1, 2 and 3, stroke density has doubled in the faster section indicating the barabar laya. Different stroke distribution in Kaida4 and its finer segments is also visible. The stroke density has increased in multiples of 1.5, 3, 6 and 8 times in the faster sections.There is a similarity in the fastest section (in multiples of 8 times) of kaida4 with the faster sections of kaida1, kaida2 and kaida3.

Each of the kaidas, kaida1, kaida2 and kaida3 has its own theme and has a slow and fast section and this is reflected in the boundary within each kaida. Slower section of each kaida is similar to the slower section of other two kaidas and hence indicated by the dark regions in the similarity matrix of fig 4.3. Kaida4 having distinct rhythmic distribution is showing dissimilarity with other kaidas. Boundaries of kaida1, kaida2 , kaida3 and kaida4 are identified and also the boundaries of intra segments, faster segments within each kaidas are also indicated by the peaks of the novelty score in fig4.3.

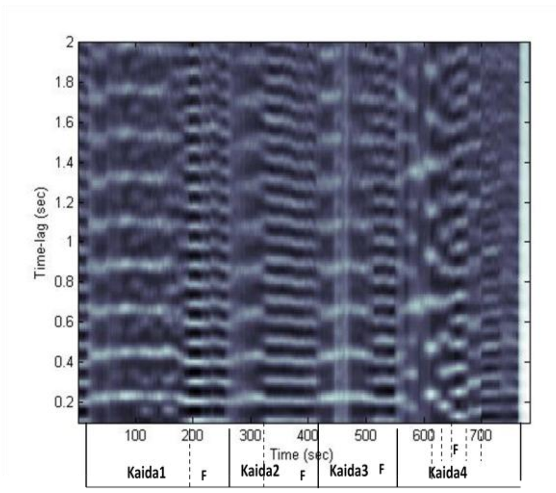

Fig 4.2 Rhythmogram of audio of kaidas of tabla solo by ZH

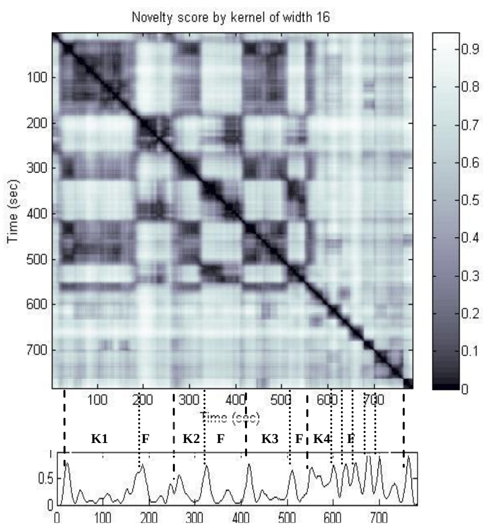

Fig 4.3 Similarity matrix and novelty score of kaidas of tabla solo by ZH

Audio segmentation approach is extended to the polyphonic khyal vocal concerts to get the gross boundaries between major segments, *alap*, *bada khyal* and *chota khayal*. The khyal vocal performance in *raag* Bhoopali by U. Rashid Khan (RK), in Tintal with *bada khyal* rendered in *madhya laya* and *chota khyal* in *drut laya*. Here, tabla signal is not changing in the dynamic section like sargam. Gross boundaries of the sections alap, bada khyal and chota khyal can be observed in the rhythmogram of fig 4.4. Rhythmic structure of the performance in *raag* Deshkar by Kishori Amonkar, in Tintal with b*ada khyal* rendered in *vilambit laya* and *chota khyal* in *drut laya* is shown in fig 4.5. As we can observe the tempo of the performance is increasing progressively in the bada khyal section.

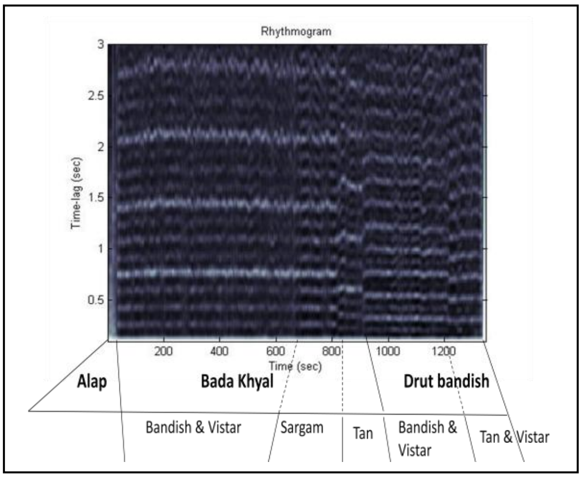

Fig 4.4 Rhythmogram of Khyal vocal concert in Bhoopali by RK

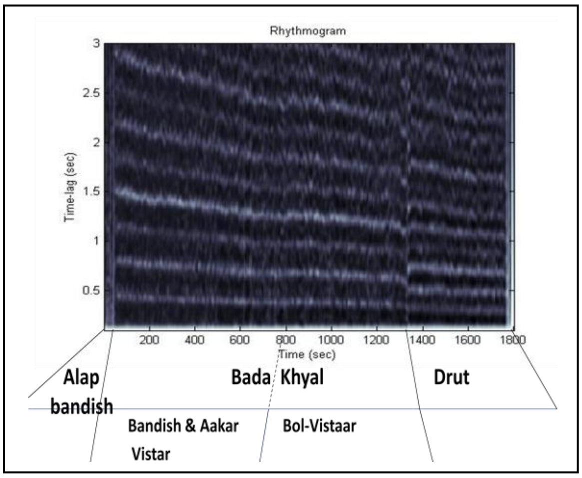

Fig 4.5 Rhythmogram of Khyal vocal concert in Deshkar by KA

5. Conclusions

Audio segmentation algorithm has returned the boundaries of the segments like different kaidas in the complex rhythmic repertoire like tabla solo. Also, the boundaries of sub-segments within each kaida having higher rhythmic density have been identified. Similarity of first three kaidas and the dis-similarity of the fourth kaida having different stroke distribution is revealed fairly well in the similarity matrix. Boundaries of major sections of complex polyphonic vocal concerts have been recognized. Spectral features combined with the auditory processing motivated bi-phasic function achieved good time localization of extracted onsets, particularly for polyphonic audio. ACF has proved to be robust enough to reveal the inherent periodicities and strengths of accents of polyphonic music. To get the boundaries of sub-segments within bada-khyal, like tan, sargam and bol-baat rhythmic structure due to syllable onsets within vocal onsets have to be inspected.

6. References

[1] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in Proceedings of IEEE International Conference on Multimedia and Expo (ICME '00), vol. 1, pp. 452–455, New York, NY, USA, July-August 2000.

[2] K. Jensen, J. Xu, and M. Zachariasen, "Rhythm-based segmentation of popular chinese music," in Proc. Int. Conf. Music Inf. Retrieval (ISMIR), London, U.K., 2005.

[3] M. Clayton, Time in Indian Music : "Rhythm , Metre and Form in North Indian Rag Performance", Oxford University Press, 2000.

[4] D. J. Hermes, "Vowel onset detection," J. Acoust. Soc. of America, 87(2), pp.866-873, 1990.

[5] S. Dixon : " Onset Detection Revisited", In Proceedings of the International Conference on Digital Audio Effects (DAFx'06), Montreal, Canada. 2006

[5] P. Grosche and M. M¨uller, "Extracting predominant local pulse information from music recordings", IEEE Transactions on Audio, Speech, and Language Processing, 19(6):1688–1701, 2011.

[6] S. Gulati and P. Rao, "Rhythm pattern representation for tempo detection in music", Proc. of the First International