Innovations Syst Softw Eng
DOI 10.1007/s11334-015-0247-5

S.I. : ICACNI 2014

Rhythm analysis of *tablā* signal by detecting the cyclic pattern

Susmita Bhaduri[1] · Orchisama Das[2] · Sanjoy Kumar Saha[3] · Chandan Mazumdar[3]

Received: 26 November 2014 / Accepted: 15 April 2015

Abstract The Indian classical music system follows a cyclic perception as against the linear approach of reductionist concept in Western music. In every *tāla*, there exists a pattern of tangible and intangible events that keeps on repeating in smaller cycles. If such repeating pattern is detected, it will be an important step in the context of rhythm analysis of *hindustañi* music and also for rhythm-based retrieval. In this work, a simple but novel methodology is presented to detect two important rhythmic aspects of *tāla* namely, tempo and *mātrā*. It is focussed on the detection of the repeating structure by analysing the *tablā* signal. The work extends our earlier effort that deals with only the electronic *tablā* signal which is well behaved. In this work, pitfalls of the earlier methodology are analysed and corrective measures are adopted to formulate the improved methodology. The present work computes and tunes the parameters based on the signal content and can work with the signals of wide variety, including the not so well-behaved real *tablā* signal, i.e. the signal captured when *tablā* is played by human artist. Experiment is carried out with a large number of electronic and real *tablā* clips reflecting variety of tempo and *tāla*. Performance of the proposed methodology is also compared with that of the earlier one. Result indicates the superiority and effectiveness of the proposed methodology.

Keywords *mātrā* detection · Tempo detection · Rhythm analysis · *hindustañi* Music · Cyclic pattern

✉ Sanjoy Kumar Saha
sks_ju@yahoo.co.in

Susmita Bhaduri
susmita.sbhaduri@gmail.com

Orchisama Das
orchisamadas@gmail.com

Chandan Mazumdar
chandan.mazumdar@gmail.com

1 Centre for Distributed Computing, CSE Department, Jadavpur University, Kolkata, India

2 Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata, India

3 CSE Department, Jadavpur University, Kolkata, India

1 Introduction

The classical music system of Indian sub-continent is based on two major concepts—*rāga* and *tāla*. *Rāga* describes the melodic or modal aspect of music and *tāla* describes the rhythmic aspect. The rhythmic pattern of any composition in Indian music is described by the term *tāla*, which is composed of cycles of *mātrā*-s. *tāla* roughly correlates with the metres in Western music and also with metres of Sanskrit language. This rhythmic framework based on *tāla* is quite different and complex compared to the Western notions of rhythm. The rhythm in Western music with all its gamuts can be sorted out in the form of a nomenclature, but the rhythm in Indian music, more precisely in *tāla*-s, demands a cognitive human perception that permeates through the whole texture of one's musical experiences. Indian *tāla* is uniquely cyclical as opposed to Western music which is linear. *Tāla* is a cycle of beats centred around the most emphasised beat, called the *sam*, (which is also the first beat of the cycle), that repeats itself in ongoing phases. Western music does not and cannot use such complex beat cycles.

In the context of *hindustānī* rhythm, *tablā* is the most popular percussive instrument. One of the essential requirements of playing and understanding the *tablā* is learning its alphabet. The *dayan* (right hand drum) is made of wood. The *bayan* (left hand drum) is made of iron, aluminium, copper, steel, or clay. When played together, they create regularly spaced amplitude peaks corresponding to each stroke or *bol*. The series of peaks in a *thekā* (characteristic *bol*-pattern and most basic cyclic form of a *tāla*) keeps on repeating centred around the *sam*, due to the cyclic property of *hindustānī tāla*. A *thekā* of a *tāla* can be rendered in innumerable ways with various combination of strokes. However, while playing, there is always a tendency of emphasizing the *sam* in the beginning of each cycle. If this repeating cyclic pattern can be recognised, the analysis of various *hindustānī tāla*-s at different time scales would provide musically relevant information. This would be a positive step towards rhythm information retrieval which eventually verges towards MIR.

In this work, we consider the *mātrā* detection of various *hindustañi* rhythms or *tāla*-s from *tablā*-solo compositions, based on their cyclic recurrence. The rest of the paper is organized as follows. Section 2 provides concept of *tāla* and its cyclic nature. Section 3 denotes a survey of past work. In Section 4, first the method in [1] is analysed for its challenges and, then, the proposed methodology is elaborated. In Section 5, experimental results are presented. The paper is concluded in Sect. 6.

2 Concept of *tāla* and its cyclicity in *hindustañi* music

Hindustañi music is metrically organised and it is called *nibaddh*(bound by rhythm) music. This kind of music is set to a metric framework called *tāla*. Each *tāla* is uniquely represented as cyclically recurring patterns of fixed lengths. This recurring cycle of *tāla* is called *āvart*. The overall time-span of each cycle or *āvart* is made up of a certain number of smaller time units called *mātra*-s. The *mātra*-s of a *tāla* are grouped into sections, sometimes with unequal time-spans, called *vibhāga*-s.*Vibhāga*-s are indicated through the hand gestures of a *tali*(clap) and a *khali*(wave). The beginning of an *āvart* is referred to as *sam*. Ref. Clayton [2].

In the *tāla* system of *hindustañi* music, the actual illustration of *tāla* is done by certain syllables which are the mnemonic names of different strokes corresponding to each *mātra*. These syllables are called *bol*-s. *bol*-s are classified as single or composite *bol*-s.

- *Single bol*: While playing *tablā*, sometimes two *bol*-s are played with a break/distinct discontinuity in between. The signal duration of these two *bol*-s played consecutively, is same as the sum of the duration of individual *bol*-s. Example of single *bol*-s is:*dha, dhi, ta, tin, tun, te, tak, dhe, re*, etc.
- *Composite bol*: When two single *bol*-s get overlapped, it creates a composite *bol*. Composite *bol* has same duration as that of one of the constituent single *bol*-s. Example of composite *bol*-s is:*te-Te,tir-kit, tin-tun, kat-ghe, tra-kra, ta-dha*, etc.

Figure 1 shows the waveform of the single *bol* ***te***. Figure 2 shows the waveform of the composite *bol*-s ***te-Te*** of *tintal*. It is evident from the figures that two single *bol*-s are getting overlapped while generating a composite one.

The first *mātrā* of each cycle or *āvart* is called *sam*. The basic characteristic pattern of *bol*-s and the most basic cyclic form of the *tāla* for the *tablā* are called *thekā* as per David [3].

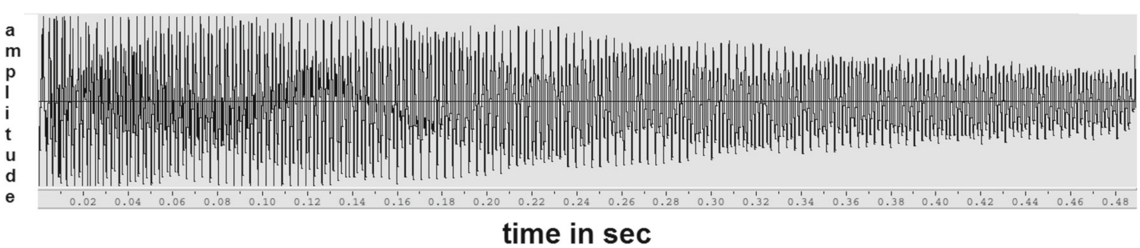

Fig. 1 Single stroke *te*

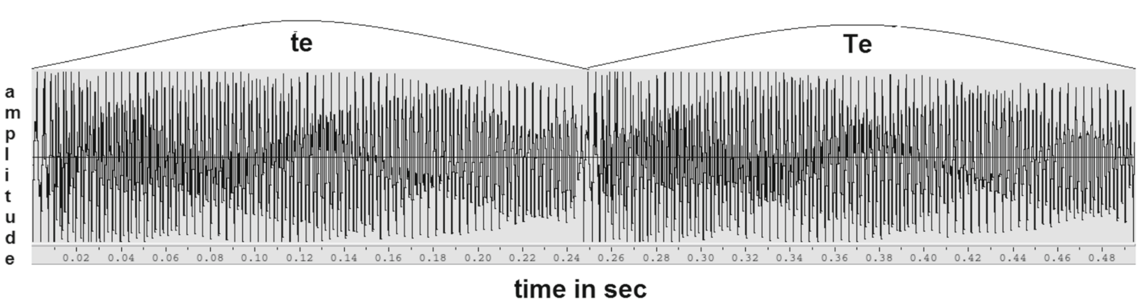

Fig. 2 Composite stroke *te-Te* in *tintal*

Table 1 Description of *jhaptal*, showing the structure and the *thekā*

tālī	+	–	2	–	–	0	–	3	–	–
bol	*dhi*	*na*	*dhi*	*dhi*	*na*	*ti*	*na*	*dhi*	*dhi*	*na*
mātrā	1	2	3	4	5	6	7	8	9	10
vibhāga	1	–	2	–	–	3	–	4	–	–
āvart	1	–	–	–	–	–	–	–	–	–

The *thekā* of *tāla* is cyclically repeated over the entire length of the musical piece in *hindustānī* music. The strong concluding beat or *sam* in a *thekā* carries the main accent and is responsible for creating the sensation of cadence and cyclicity.

For example, in *jhaptal* (refer Table 1), there are four *kriya*-s in its *thekā*, namely two *sasabda kriya*-s or *tali*-s followed by one *nisabda kriya* or *khali*-s and again followed by another *tali*. There are four syllabic groupings or *vibhāga*-s in this *tāla* and it comprises ten *mātrā*-s in total. We can also see the *bol* pattern in its *thekā* as well as in *āvart*.

The next most important concept in *hindustañi* rhythm is *lay*, which governs the tempo or the rate of succession of *tāla*. The *lay* or tempo in *hindustañi* music can vary among *ati-vilambit*(very slow), *vilambit* (slow), *madhya*(medium), *druta*(fast) to *ati-druta*(very fast). Depending on the *lay*, the *bol* may be further subdivided into more pulses that appear in the surface rhythm. Tempo is expressed in beats per minute or BPM.

3 Past work

Rhythm analysis and modeling for Indian music can be traced back to the study of acoustics of Indian drums by Sir C. V. Raman [4]. In this work, the importance of the first three to five harmonics which are derived from the drum-head's vibration mode was highlighted. In the last decade, most of the MIR research on Indian music rhythm has been focused on drum stroke transcription, creative modeling for automatic improvisation of *tablā* and predictive modeling of *tablā* sequences. Bhat [5] extended Raman's work and to explain the presence of harmonic overtones, he applied a mathematical model of the vibration modes of the membrane of a type of Indian musical drum called *mridanga*. Malu and Siddharthan [6] confirmed C.V. Raman's conclusions on the harmonic properties of Indian drums, and the *tablā* in particular. They accredited the presence of harmonic overtones to the central black patch of the *dayan* (the *gaab*).

Goto and Muraoka [7] were first to achieve a reasonable accuracy for tempo analysis on audio signals operated in real time. Their system was based on agent-based architecture and tracking competing meter hypotheses. A computer program based on linear predictive coding (LPC) analysis to recognize spoken *bol*-s, has been developed by Chatwani [8]. Patel et al. [9] performed an acoustic and perceptual comparison of *tablā bol*-s (both, spoken and played). They found that spoken *bol*-s have significant correlations with played *bol*-s, with respect to acoustic features like spectral flux, centroid, etc. It also enables untrained listeners to match the drum sound with corresponding syllables. This gave strong support to the symbolic value of *tablā bol*-s in North Indian drumming tradition. Gillet et al. [10] worked on *tablā* stroke identification. Samudra et al. [11] used cepstrum-based features and the HMM model for their *tablā bol* recognizer. Theory of banded wave guides has been applied to highly inharmonious vibrating structures by Essl et al. [12]. Chordia [13] extended the work of Gillet et al. [10] and implemented different classifiers like neural network, decision trees, multivariate Gaussian model, to create a system that segments and recognizes *tablā* strokes.

Dixon [14] has created a system called BeatRoot for automatic tracking and annotation of beats for a wide range of musical styles. Ellis [5] (2007) describes a beat tracking system which first estimates a global tempo. Davies and Plumbley [15] proposed a context-dependent beat tracking algorithm which handles varying tempos, by providing a two-state model in which the first state tracks the tempo changes. Xiao and Tian [16] correlate tempo and timbre, two fundamental properties of a musical piece, using a parametrized statistical model, based on which a tempo detection method is derived. Holzapfel and Stylianou [17] proposed a rhythm analysis technique for non-Western music, using the scale transform for tempo normalization. Gulati et al. [18] proposed a method for meter detection for Indian music. Miron [19] recently explored automatic recognition of *tāla*-s in *hindustañi* music. A simple method for tempo detection for *hindustañi tāla*-s has been described in [20].

A system for detecting the number of *mātra*-s and tempo of *hindustañi tāla*-s has been presented in [1]. The work is based on the distribution of amplitude peaks and consequent matching of a repetitive pattern present in the sequence of beats with the standard patterns for a sequence of beats in the *thekā* of *tāla*-s of *hindustañi* music. But it suffers from number of drawbacks as elaborated in Sect. 4. The proposed work in this paper is motivated to address those limitations.

4 Proposed methodology

The proposed methodology is the extension of our early effort presented in [1]. It has a number of limitations in the form of its inability to deal with signals of low tempo, *tāla* consisting of composite *bol*-s, etc. Moreover, it was aimed to work with electronic *tablā* signal which is consistent in terms of the periodicity and strength of the beats. In this work, an improved methodology is presented that overcomes the limitations of our early work and can handle the real *tablā* signal captured when an artist plays. For the sake of continuity, methodology discussed in [1] is presented briefly in Sect. 4.1. Weaknesses of the same are analysed in Sect. 4.1.1. Finally, the proposed remedial measures and the improved methodology are detailed in Sect. 4.2.

4.1 Methodology presented in [1]

A *tāla* has a specific number of *mātrā*-s which represents a basic beat pattern or *thekā*. In an audio clip of specific *tāla*, its *thekā* gets repeated over the time. The number of *peaks* in the amplitude envelope of a *tāla* represents the number of *mātrā*-s of the corresponding *thekā*.

In the said work, the first step is to extract the *peaks* from the amplitude envelope of the *tablā* signal. It is done using MIRtoolbox [21]. First of all, the signal is decomposed into number of frequency bands using equivalent rectangular filterband (ERB). From the output signal of each filterband, differential envelope is obtained by applying half wave rectifier. All such envelopes are summed up. Considering the local maxima-s, *peaks* are extracted from the amplitude envelope of the summed up signal. In determining the *peaks*, only those local maxima-s are retained whose amplitudes are higher than that of its neighbouring local minima-s by certain threshold (th). From this *peak signal*, *mātrā* and tempo are detected.

For subsequent analysis, *beats* are extracted from the *peak signal*. *Peak signal* is divided into time-window of t seconds duration. As the *peak signal* is discrete in nature, depending on the tempo a time-window may have zero or multiple *peaks*. The local maxima within each window (if atleast one *peak* exists) are taken as a *beat*. Two *beats* must be well separated in time scale to make it distinguishable to human auditory system. Based on this, t is empirically chosen around 0.1. Ideally, a *beat* corresponds to a *bol*. Once the *beat signal* is obtained, *mātrā* is determined by identifying the basic repetitive pattern of the *bol*-s (*beats*). For a particular *beat*, occurrence of its similar ones is traced over the entire *beat signal*. Two *beats* are considered to be similar if the difference between their amplitudes lies within a threshold th_{bs}. Assuming the amplitudes are normalized within [0, 1], th_{bs} is empirically chosen as 0.025. A *beat* may have multiple periodicity as a *bol* may appear number of times in a rendering of a *tāla*. Suppose, for a *beat* b_i, its similar ones occur after an interval of $(p_1, p_2, \ldots p_k)$. If $m > n$ then $p_m > p_n$ but p_m may or may not be a multiple of p_n. Let p_j be the smallest value in the interval set such that all the intermediate *beats* that follow b_i within the interval p_j also repeat themselves after the same interval. Then, p_j is taken as the periodicity of the basic *beat* pattern or the *mātrā* of the *tāla*. Figure 3 illustrates one example for *dadra tāla* where the first *beat* repeats after an interval of 6 *beats* and the intermediate *beats* are compared with the corresponding ones at the same interval.

From the *beat signal*, tempo is extracted simply as $\frac{N}{T}$, where N stands for number of *beats* in the signal and T is the signal duration in minutes and it is expressed in *beats* per minute (BPM).

4.1.1 Limitations

The performance of the methodology presented in Sect. 4.1 heavily depends on the parameters like t, th, th_{bs} and these are chosen empirically. In this section, we analyse the limitations and diagnose the implications of the parameters in the context of failures.

1. **Missed *peak*-s in the *peak signal***: As discussed in Sect. 4.1, extraction of *peak signal* is based on the selection of *peaks* among the local maxima-s in the summed up signal. It involves a threshold th. *Peaks* are those local maxima-s with amplitude higher than their adjoining local minima-s by the quantity th. By default, th is taken as $0.1 \times l_{\max}$, where $l_{\max}$ is the maximum amplitude

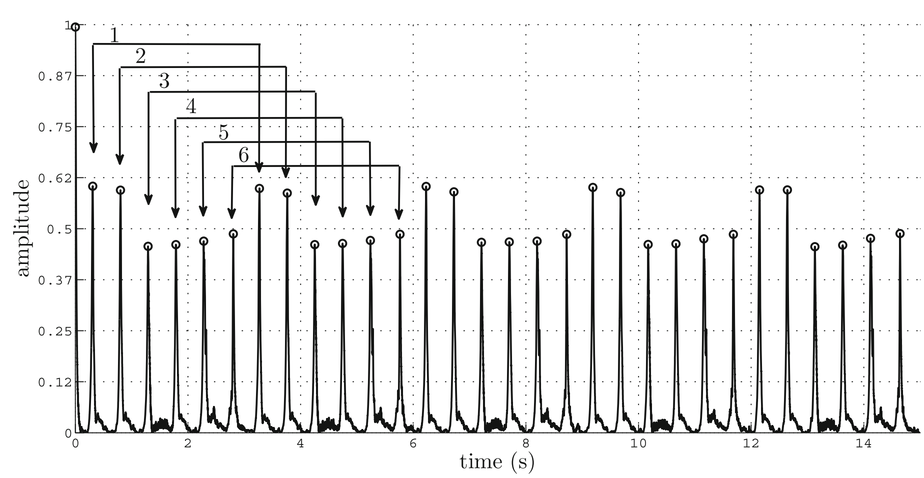

Fig. 3 The process of *mātrā* detection for *dadra tāla*

among the local maxima-s. Thus, the presence of a very high amplitude arising out of noise sets th to a significant value. As a result, some of the local maxima-s may fail to qualify to be a *peak*. Figure 4 illustrates such a scenario. The first *peak* is an outlier and it is considerably higher than the local maxima-s and biases th. As a result the indicated local maxima-s are missed out in the extracted *peak signal*. Looking into the periodicity (time scale), it is desirable to have those *peaks*. Thus, errors crept into the *peak signal* propagates in misjudging the tempo and *mātrā*.

2. **Generation of spurious *beat*-s for low tempo signal:** As discussed earlier, *beats* or *bol*s are extracted from the *peak signal* by dividing in time-window of duration t seconds. Local maxima in each window are taken as a *beat*. A low value of t avoids the miss of any *beat*. On the other hand, for a low tempo signal, it may give rise to spurious *beats*. This is more likely to occur if the signal is noise affected and it is quite common for the recorded real *tablā* signal. Such improper extraction of *beat signal* affects both the detection of tempo and *mātrā*. Figure 5 shows *peak signal* of *dadra tāla* with low tempo where few spurious *peaks* are marked. Visually, it is clear that the desired *beats* appear at a periodicity higher than 0.1 s (it is around 0.75 seconds). As a result, chosen value of $t = 0.1$ results in unwanted spurious *beats*.
3. **Difficulty in handling the composite/absent *bol* in the *tāla***: The *thekā* of a *tāla* can be rendered in innumerable pattern of *beats*. It can have combination of single/composite/absent *bol*-s. A peak in the *beat signal* corresponds to a *bol*. A composite *bol* or absent *bol* culminates into multiple or absence of *beats*, respectively. As a result, computation of tempo becomes erroneous. The process of identification of *beat* pattern for detecting *mātrā* also gets jeopardised.
4. **Failure in detecting the *beat* pattern of a *tāla***: *Mātrā* is detected by identifying the *beat* pattern that repeats in the *tāla*. In the process of doing so, the similarity of the *beats* is adjudged based on the threshold th_{bs}. Two *beats* are taken as similar if their amplitude difference is within th_{bs}. As reported in [1], it is set as 0.025. Such a low value may be justified for ideal signal where strength of the same *bol* is repeated maintaining the high precision. Because of the noise inherent in the process of recording and approximation in the form of filtering to generate the *peak signal*, it is impossible to maintain the consistency of *beat* strength. Thus, the detection of *beat* pattern may fail. In Fig. 5, the *beats* marked as **A** and **B** are the correspond-

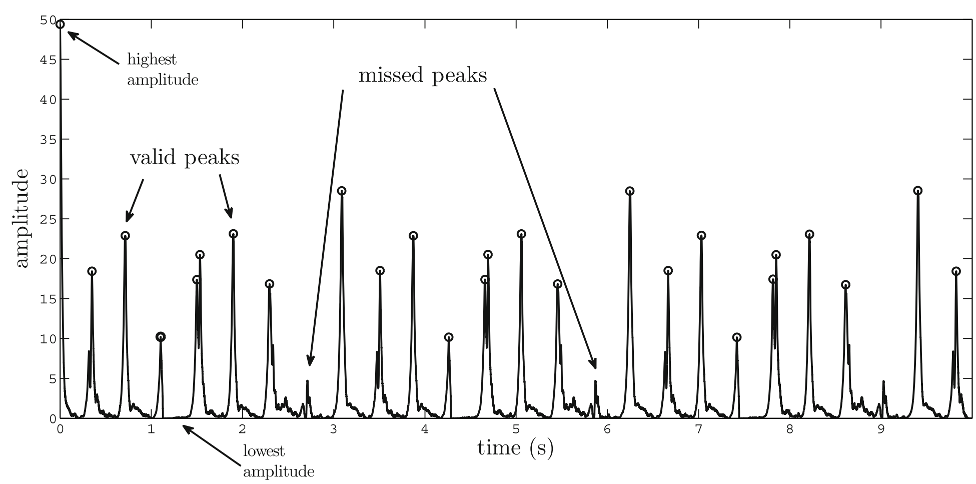

Fig. 4 *Peak signal* generated from envelope of *kaharba tāla*

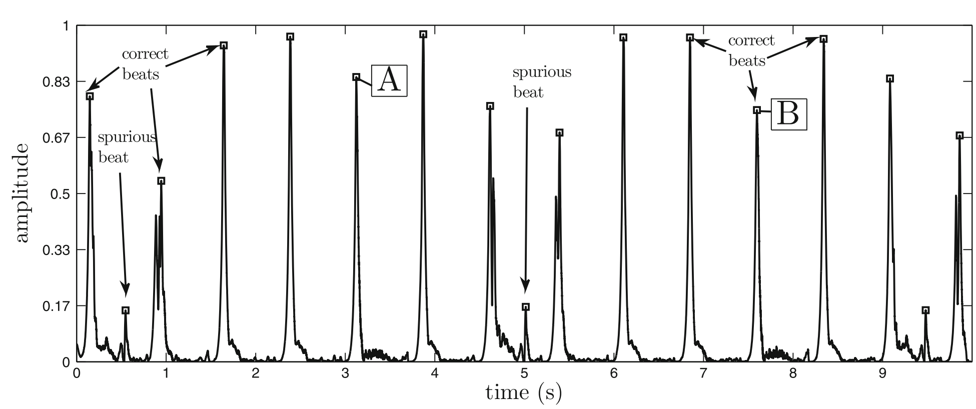

Fig. 5 *Peak signal* of *dadra tāla* with low tempo (80 BPM)

ing *beats* (ignoring the spurious *beat*). But their amplitude difference does not satisfy the threshold criteria.

5. **Failure in handling real *tablā* signals:** In case of the real *tablā* signal, i. e. the signal obtained by recording the performance of an artist playing the instrument, more variety is present. It is impossible to maintain the consistency in terms of the periodicity of the *bol*-s or *beats* played by a human. The fact is also true in the context of strength of the same *bol* played. Moreover, the improvisation in the style of rendering further complicates the scenario. As a result, the issues 1–4 as discussed become more crucial. With the predefined values of the parameters th, t and th_{bs}, it is quite difficult to cope up with the real *tablā* signal. As discussed in [1], while detecting *mātrā*, for a *beat* similar ones are traced. Intermediate *beats* among the similar *beat* pair with smallest periodicity are taken as the pattern template. If the template matches with the next *beat* sequence with same periodicity, then *mātrā* is detected based on the template. Otherwise, the search goes on with the similar *beat* pair with next higher periodicity. But, for each periodicity, template matching is carried out only once. Because of the consistency and well-behaved nature, it may work for electronic *tablā*. For real *tablā* signal, the scenario is not so simple.

4.2 Improved methodology

In this work, our motivation is to propose a methodology for detecting tempo and *mātrā* that can work well on both—electronic and real *tablā* signal. As analysed in Sect. 4.1.1, proper selection of the parameters is the major challenge. Optimal values for the parameters are very much signal dependent. Dependency of th (used in deciding the *peaks*) on the maximum amplitude ($l_{\max}$) makes it vulnerable as it ignores the overall signal and can easily be affected by the outlier. Size for time-window (t) may work well for certain range of tempo, but it may fail for others. The presence of composite *bol* further aggravates the situation. The threshold used to find the similar *beats* (th_{bs}) ignores the variation present in the signal. It has detrimental effect in case of the signals with considerable variation like noise-affected signal or real *tablā* signal. Hence our initial focus is directed towards the selection of these parameters based on the signal content. Once the parameters are tuned, then issues regarding the handling of composite *bol* and real *tablā* signal are addressed.

1. **Selection of th to avoid missed peaks:** The threshold (th) for extracting the peaks is determined as follows.
   - Let l_i be the set of amplitudes of local maxima-s in summed up signal
   - $\mu_l = \text{avg}(l_i)$ and $\sigma_l = \text{Stddev}(l_i)$
   - $th = 0.1 \times \min(l_j)$ where $l_j \in l_i$ and $l_j > \mu_l + \sigma_l$

   It may be recalled that earlier th was taken as $0.1 \times l_{\max}$ where $l_{\max} = \max(l_i)$. In the proposed method of selection, the impact of high amplitude which is actually an outlier is marginalised.
2. **Selection of t to avoid spurious *beats***: The size of time window (t) should vary according to the tempo of the signal. It is used to obtain the *beat signal* from the *peak signal*. In the proposed methodology, *beat signal* is extracted following two steps as presented in [20]. At the first level, candidate *beat signal* is obtained by taking t as 0.1 s. Along with the possible inclusion of spurious *beats*, it also ensures that no valid *beat* is excluded. A refinement is applied on this candidate *beat signal* by dividing it in a number of time-windows of size t seconds and picking up the local maxima-s. At this stage, the concept of *bol* duration is introduced to determine the final value of t. Time interval between two consecutive *beats* in the candidate *beat signal* is taken as the *bol* duration. It varies according to the tempo. A histogram of *bol* duration is formed where the time scale is divided into number of bins. The *peak* of the histograms corresponds to the actual *bol* duration for the signal. Based on the corresponding bin in the time scale, t is computed as the average of the bin boundaries. Thus, t is tuned based on the tempo-content of the signal and is utilized to get rid of spurious *beats*.
3. **Handling of composite or absent *bol***: The issue is addressed by the two-stage process of extraction of *beat signal* as discussed earlier. In reality, a signal is mostly composed of simple *bol*s and a comparatively smaller fraction contributes towards composite or absent *bol*. At the first stage (with $t = 0.1$ s) composite *bol* may give rise to multiple *beats* having relatively smaller *beat* interval. As if, these are the *bol*-s with smaller duration in comparison to that of simple *bol*. At the second stage t, i.e. the time-window is determined based on the histogram of *bol* duration. As the *peak* in the histograms corresponds to duration of simple *bols*-s, t also conforms to that. Thus, the *beats* of composite *bol* are likely to fall in the same window and in the final *beat signal* additional *beats* are removed. In case of absent *bol*, interval between its previous and following *beat* gets increased. As a result, the time span denoting the absent *bol* is likely to get divided with its previous and following window without generating any additional *beat*. Thus, the modified process of extracting *beat signal* from the *peak signal* can handle the issue of composite or absent *bol* satisfactorily.
4. **Selection of th_{bs} to avoid failure in detection *beat* pattern:** In the process of detecting *beat* pattern and thereby to detect *mātrā*, judgement on the similarity of two *beats* plays a major role. Thus, the selection of threshold on

beat similarity-(th_{bs}) is important. The steps to determine the value of th_{bs} is as follows.

- Divide the *beat signal* into two halves
- Initialize the set diff as empty
- For each *beat* b_i in the first half
- Let b_j is most similar to b_i ($i \neq j$)
- diff = diff $\cup\, d_i$ where d_i stands for amplitude difference between b_i, b_j
- $th_{bs} = \max(\text{diff})$

Thus, th_{bs} is determined based on the variations present in the signal. As a result, similar *beats* can be detected in spite of the presence of noise or non-uniformity in terms of the amplitude for the same *bol* in a *tāla*.

5. **Handling of real *tablā* signal:** Unlike the electronic *tablā* signal, the signal of real *tablā* played by human artist shows significant variation. The *beat* interval cannot be maintained in the strict manner. Strength of same *bol* also varies. Style of rendering further adds to the variation. The selection process of the parameter values combats such difficulties considerably. Still, the *beat* pattern detection process demands attention. In case of electronic *tablā* signal, same *bol*-s maintain the consistency in terms of amplitude. *Beat* intervals are also consistent. Hence *beat* pattern can be ascertained by matching the pattern template with its next occurrence. But, in case of real *tablā* signal, complete match of the pattern template may not be achieved. Thus, to determine the *mātrā*, such matching is to be carried out over the entire signal and, finally, a judgement based on the principle of maximum likelihood is followed. The modified process of *beat* pattern matching and *mātrā* detection is detailed as follows.

```
-M = m_1, m_2, ... m_n, be the finite set
of possible mātrā.
-For each m_l ∈ M {
    Divide the beat signal into the k
    equal sized blocks of m_l beat-s.
    count = 0.
    For i = 1 to k−1 {
      For j = 1 to m_l {
        If b_{i_j} and b_{(i+1)_j} match
            then count = count + 1.
       (b_{i_j} and b_{(i+1)_j} are the jth beat of
       ith and (i+1)th block respectively)
      }
    }
    p_{m_l} = count / ((k−1) × m_l), where p_{m_l} stands
    for the probability that the signal
    is of mātrā m_l
}
-p_max = max(p_{m_l}).
-p_max is taken as the mātrā of the
 signal.
```

Different *tāla*-s are of different *mātrā*-s like six, eight, etc. In the proposed methodology, M is the set of such *mātrā*-s. Keeping the variations in the real *tablā* signal in mind, a signal is tested for all the *mātrā*-s. *Beat signal* is divided into blocks having number of *beats* same as the *mātrā* for which it is tested. Thus, a block corresponds to the *beat* pattern. Every pair of consecutive blocks is considered for pattern matching. Corresponding *beats* of the two blocks are compared based on th_{bs}. In case of real *tablā*, all the *beats* may not match. Based on the number of matched *beats*, probability of the event that the signal is of particular *mātrā* is computed. *Mātrā* for which the probability is maximum is considered.

The proposed methodology has been discussed in this section in parts. Finally, broad steps of the complete process of tempo and *mātrā* detection can be summarized as follows:

- Extract *peaks* from the *tablā* signal using MIRtoolbox [21]
  - Decompose the signal in ten frequency bands as per the Klapuri's principle [22] of onset detection based on human auditory system
  - Obtain the summed up signal (S) of amplitude envelope of the individual band
  - Compute th as in item 1 of Sect. 4.2 and extract the *peaks* from S to obtain *peak signal* (P)

Table 2 Description of data

tala	*mātrā*	Tempo range (in BPM)		No of clips		Clip duration (in s)	
		Real *tablā*	Electronic *tablā*	Real *tablā*	Electronic *tablā*	Real *tablā*	Electronic *tablā*
dadra	6	95–400	80–200	280	372	10–50	25–35
kaharba	8	180–415	80–200	722	203	15–85	25–35
jhaptal	10	67–240	80–200	206	390	15–50	25–35
tintal	16	220–225	90–200	366	247	20–75	25–35
rupak	7	85–420	NA	722	NA	15–95	NA

– Obtain candidate *beat signal* (B_c) from P with t as 0.1 s.
– Compute t from B_c as in item 2 of Sect. 4.2
– As discussed in item 2 of Sect. 4.2, obtain *beat signal* (B) from B_c
– Compute $tempo = \frac{60}{t}$ where t stands for the *bol* duration (in s) computed as in item 2 of Sect. 4.2
– Compute th_{bs} from B as in item 4 of Sect. 4.2 and use it to determine *mātrā* as discussed in item 5 of Sect. 4.2

5 Experimental results

In our experiment, we have worked with five different *tāla*-s, namely *dadra*, *kaharba*, *jhaptal*, *tintal* and *rupak*. All the audio clips are recorded with wav recorder in .wav format from the electronic *tablā* and also real *tablā*. The detailed description of the data used is shown in Table 2. Clips are of duration of 10-95 seconds. For each *tāla*, clips are of different tempo. Thus, data reflects variation in terms of duration, tempo and *mātrā* to establish the applicability of proposed methodology in identifying tempo and *mātrā*. It may be noted that the database does not contain the electronic signal for *rupak tāla*.

The proposed methodology is applied on both electronic and real *tablā* signal. Tables 3 and 4 show the confusion matrices for detecting *mātrā* for real and electronic *tablā*, respectively. Figures along the diagonal show the percentage of clips for which *mātrā* is correctly detected. Overall *mātrā* detection performance is shown in Table 5. The confusion matrices show that there is a mis-detection between the pair of *kaharba* and *tintal*. It may be noted that *mātrā* of *kaharba* is exactly half of the *tintal*. Because of the variations present in the beat signal, for such cases similarity to some extent may arise and that results into mis-classification. Novice *tablā* artists can experience the same issue while playing *tablā*, for these kinds of *tāla* pairs. In future, domain knowledge may be considered to verify such pairs further. Table 6 shows the performance of proposed methodology in detecting tempo for different *tāla*-s. It may be noted that a tolerance of $\pm 5\%$ is considered in judging the correctness of tempo. It is clear that the proposed methodology satisfactorily detects *mātrā* and tempo of real and electronic *tablā* clips with wide variety.

Comparison of performance of the proposed methodology and the earlier one [1] is presented in Table 7. As discussed in Sect. 4.1.1, the earlier methodology cannot handle the signals of low tempo, signals with *composite bol* and real *tablā* signal. To verify it experimentally, we have focussed on the clips of low tempo, clips with *composite bol* and clips of real *tablā*. The results are shown in Table 7. It clearly reflects the failure of the earlier methodology [1] and the success of the proposed methodology. It justifies the effectiveness of the improvements proposed over the existing one. It is obvious that as the methodology presented in [1] fails to detect the tempo (as it either misses beats or generates spurious beats), it is bound to fail in *mātrā* detection.

Table 3 Confusion matrix for *matra* detection with real *tablā* clips (all figures in %)

	dadra	*kaharba*	*jhaptal*	*tintal*	*rupak*
dadra	97.50	0.00	2.14	0.00	0.36
kaharba	5.26	83.93	0.97	8.45	1.39
jhaptal	1.94	6.80	83.01	5.34	2.91
tintal	4.64	6.56	4.92	81.42	2.46
rupak	0.55	0.00	13.02	2.08	84.35

Table 4 Confusion matrix for *matra* detection with electronic *tablā* clips (all figures in %)

	dadra	*kaharba*	*jhaptal*	*tintal*
dadra	97.85	0.27	1.61	0.27
kaharba	0.00	90.64	0.99	8.37
jhaptal	0.00	0.00	100.00	0.00
tintal	0.00	7.69	0.40	91.90

Table 5 Overall performance of *mātrā* detection (all figures in %)

Average performance (in %)	
Real *tablā*	Electronic *tablā*
85.24	96.12

Table 6 Performance of tempo detection (all figures in %)

tala	Correct detection (in %)	
	Real *tablā*	Electronic *tablā*
dadra	98.93	99.46
kaharba	99.44	96.06
jhaptal	91.26	100.00
tintal	99.45	100.00
rupak	98.89	NA

Table 7 Comparison of performance for tempo detection (all figures in %)

Type	No. of clips	Tempo (in BPM)	Correct detection (in %)	
			Earlier method	Proposed method
Electronic *tablā* (single *bol*)	59	80	35.59	100.00
	67	90	50.00	100.00
Electronic *tablā* (composite *bol*)	390	80–200	56.92	100.00
Real *tablā*	1574	67–420	21.98	98.28

6 Conclusion

In this work, a novel and robust methodology to detect the important rhythmic parameters like tempo and *mātrā* of *tablā* signal is proposed. An improved methodology based on the concept of *bol* duration is proposed to obtain optimal *beat signal* even in the presence of composite *bol* and noise. In *hindustañi tāla*, the *beat* pattern reflects a cyclic property. This fundamental aspect is utilized to detect *mātrā*. Such detections depend on number of parameters. Tuning of those parameters is signal dependent and non-trivial. In this work, a methodology is also presented for automatic selection of these parameters based on the signal content. It enables to deal with wide variety of signals. Thus, a robust scheme is proposed. Experiment with both types of signals, namely electronic and real *tablā* signal indicates that the proposed methodology can detect the tempo and *mātrā* quite effectively.

Acknowledgments We acknowledge the contribution of Rupak Bhattacharjee(mailto:rupaktabla@gmail.com) for playing and recording various clips of real *tablā*. We have also downloaded few audio clips of real *tablā*(available freely) from the website http://www.tablaradio.com.

References

1. Bhaduri S, Saha S, Mazumdar C (2014) Matra and tempo detection for indic Tala-s. In: Advanced computing and informatics proceedings of the second international conference on advanced computing, networking and informatics (ICACNI-2014), vol 1, pp 213–220
2. Clayton M (2000) Time in Indian music: rhythm, metre and form in North Indian rag performance. Oxford University Press, Oxford
3. Courtney DR (2000) Fundamentals of Tabla. Sur Sangeet Services, 4th edn (October 1, 2013)
4. Raman CV, Kumar S (1920) Musical drums with harmonic overtones. Nature 104(2620): 453–454
5. Bhat R (1991) Acoustics of a cavity-backed membrane: The indian musical drum. J Acoust Soc Am 90:1469–1474
6. Malu SS, Siddharthan A (2000) Acoustics of the indian drum. Technical report, Cornell University
7. Goto M, Muraoka Y (1995) Music understanding at the beat level real-time beat tracking for audio signals. In: IJCAI-95 workshop on computational auditory scene analysis
8. Chatwani A (2003) Real-time recognition of tabla bols. Technical report, Senior Thesis, Princeton University
9. Patel A, Iversen J (2003) Acoustic and perceptual comparison of speech and drum sounds in the north indian tabla tradition: an empirical study of sound symbolism. In: Proceedings of the 15th international congress of phonetic sciences (ICPhS)
10. Gillet O, Richard G (2003) Automatic labelling of tabla signals. In: Proc. ISMIR
11. Samudravijaya K, Shah S, Pandya P (2004) Computer recognition of tabla bols. Technical report, Tata Institute of Fundamental Research
12. Essl G, Serafin S, Cook P, Smith J (2004) Musical applications of banded waveguides. Comput Music J 28:51–63
13. Chorida P (2005) Segmentation and recognition of tabla strokes. In: Proceedings of the ISMIR
14. Dixon S (2007) Evaluation of the audio beat tracking system beatroot. J New Music Res 36:39–50
15. Davies M, Plumbley M (2007) Context-dependent beat tracking of musical audio. IEEE Trans Audio Speech Lang Process 15:1009–1020
16. Linxing Xiao, Aibo Tian WL, Zhou J (2008) Using a statistic model to capture the association between timbre and perceived tempo. In: ISMIR 2008: Proceedings of the 9th international conference of music
17. Holzapfel A, Stylianou Y (2009) Rhythmic similarity in traditional turkish music. In: Proceedings of international conference on music information retrieval (ISMIR)
18. Gulati S, Rao V, Raol P (2011) Meter detection from audio for indian music. In: Proceedings of the international symposium on computer music modeling and retrieval (CMMR), pp 34–43
19. Miron M (2011) Automatic detection of hindustani talas. Technical report, Master's Thesis, Universitat Pompeu Fabra, Barcelona, Spain
20. Bhaduri S, Saha S, Mazumdar C (2014) A novel method for tempo detection of INDIC Tala-s. In: Emerging applications of information technology (EAIT), 2014 fourth international conference of IEEE, pp 222–227
21. Lartillot O, Toiviaine P (2007) A matlab toolbox for musical feature extraction from audio. In: Proceedings of the international conference on digital audio effects
22. Klapuri A (1999) Sound onset detection by applying psychoacoustic knowledge. In: Proceedings of the international conference on acoustics, speech, and signal processing, pp 3089–3092