A BEAT TRACKING APPROACH TO COMPLETE DESCRIPTION OF RHYTHM IN INDIAN CLASSICAL MUSIC

Ajay Srinivasamurthy
ajays.murthy@gmail.com

Gregoire Tronel
greg.tronel@gmail.com

Sidharth Subramanian
sidharth.subramanian@gmail.com

Parag Chordia
parag@smule.com

Georgia Tech Center for Music Technology, Atlanta, USA

ABSTRACT

In this paper, we propose a beat tracking and beat similarity based approach to rhythm description in Indian Classical Music. We present an algorithm that uses a beat similarity matrix and inter onset interval histogram to automatically extract the sub-beat structure and the long-term periodicity of a musical piece. From this information, we can then obtain a rank ordered set of candidates for the *tāla* cycle period and the *naḍe* (sub-beat structure). The tempo, beat locations along with the tāla and naḍe candidates provide a better overall rhythm description of the musical piece. The algorithm is tested on a manually annotated Carnatic music dataset (CMDB) and Indian light classical music dataset (ILCMDB). The allowed metrical levels recognition accuracy of the algorithm on ILCMDB is 79.3% and 72.4% for the sub-beat structure and the tāla, respectively. The accuracy on the difficult CMDB was poorer with 68.6% and 51.1% for naḍe and tāla, respectively. The analysis of the algorithm's performance motivates us to explore knowledge based approaches to tāla recognition.

1. INTRODUCTION

Indian classical music has an advanced rhythmic framework which revolves around the concept of *tāla*, where the rhythmic structure is hierarchically described at multiple time-scales. A complete description of rhythm in Indian Classical music traditions - both Hindustani and Carnatic, would need a rhythm model which can analyze music at these different time-scales and provide a musically relevant description. In this paper, we propose a beat tracking approach to rhythm description in Indian music. In specific, we discuss an algorithm to extract short-term and long-term rhythmic structure of a music piece. This information can further be used to extract global rhythm descriptors of Indian classical music.

In western music, a complete rhythm description involves the estimation of tempo, beats, time signature, meter and other rhythmic characteristics. The basic units of rhythm,

called "beats" - correspond to the "foot tapping" time locations in the musical piece. The period between beats describes the tempo period. Describing the structure within the beats (most often called the *tatum* level periodicity) and the longer rhythmic cycles (which most often correspond to phrase boundaries) provide information about the higher level rhythm information such as the time signature and meter. In Indian classical music, rhythm description invariably involves describing the tāla and associated parameters. In this paper, we extend the state of the art beat tracking algorithms to Indian classical music and explore its applicability to Indian music.

We motivate the problem of rhythm description and provide an introduction to rhythm in Indian classical music. We then describe the algorithm and discuss the results.

1.1 Motivation

The notion of rhythmic periodicity refers to a sequence of progressive cycles with distinct rhythmic patterns occurring repeatedly through time. Distinction of these cyclical themes is easily perceived by humans as our ear is able to very efficiently process subtle variations in rhythm, melody and timbre. However while we rely on our intuition to detect and react to musical periodicity, automatic tracking of these cyclical events is a relatively intricate task for an artificially intelligent system.

A rhythm description system has a wide range of applications. The system can be used for music segmentation and automatic rhythm metadata tagging of music. A causal estimation of rhythm would be an advantage for automatic accompaniment systems and for interactive music applications. Multi-scale rhythmic structure estimation would be useful in music transcription. The system described in the paper could be used as a "reverse metronome", which gives out the metronome click times, given a song.

Indian classical music, with its intricate and sophisticated rhythmic framework presents a challenge to the state of the art beat tracking algorithms. Identifying these challenges is important to further develop culture specific or more robust rhythm models. The performance of the current rhythm description systems can also be improved using the ideas from rhythm modeling of non-western traditions.

1.2 Rhythm in Indian Classical Music

The concept of tāla forms the central theme in rhythm modeling of Indian music. The main rhythmic accompaniment in Hindustani music is the Tablā, while its Carnatic counterpart is the *Mṛdaṅgaṁ*. Several other instruments such as the *Khañjira* (the Indian Tambourine), *ghaṭaṁ*, *mōrsiṅg* (the jaw harp) are often found accompanying the Mṛdaṅgaṁ in Carnatic music. We first provide an introduction to rhythm in these two music traditions.

1.2.1 Tāla in Carnatic Music

A tāla is an expression of inherent rhythm in a musical performance through fixed time cycles. Tāla could be loosely defined as the rhythmic framework for a music composition. A tāla defines a broad structure for repetition of music phrases, motifs and improvisations. It consists of fixed time cycles called *āvartanaṁs*, which can be referred to as the tāla cycle period. An āvartanaṁ of a tāla is a rhythmic cycle, with phrase refrains, melodic and rhythmic changes occurring at the end of the cycle. The first beat of each āvartanaṁ (called the *Sama*) is accented, with notable melodic and percussive events.

Each tāla has a distinct division of the cycle period into parts called the *aṅgas*. The aṅgas serve to indicate the current position in the āvartanaṁ and aid the musician to keep track of the movement through the tāla cycle. A movement through a tāla cycle is explicitly shown by the musician using hand gestures, which include accented beats and unaccented finger counts or a wave of the hand, based on the aṅgas of the tāla.

An āvartanaṁ of a tāla is divided into beats, which are sub-divided into micro-beat time periods, generally called *akṣaras* (similar to notes/strokes). The sub-beat structure of a composition is called the *naḍe*, which can be of different kinds (Table 1(b)). The third dimension of rhythm in Carnatic music is the *kāla*, which loosely defines the tempo of the song. Kāla could be *viḷaṁbita* (slow), *madhyama* (medium) and *dhṛta* (fast). The kāla is equivalent to a tempo multiplying factor and decides the number of akṣaras played in each beat of the tāla. Another rhythm descriptor is the *eḍupu*, the "phase" or offset of the composition. With a non-zero eḍupu, the composition does not start on the *sama*, but before (*atīta*) or after (*anāgata*) the beginning of the tāla cycle. This offset is predominantly for the convenience of the musician for a better exposition of the tāla in certain compositions. However, eḍupu is also used for ornamentation in many cases. We focus on the tāla cycle period and the naḍe in this paper.

The rhythmic structure of a musical piece can thus be completely described using the tāla's āvartanaṁ period (P), which indicates the number of beats per cycle, the naḍe (n), which defines the micro-beat structure, and the kāla (k). The total number of akṣaras in a tāla cycle (N) is computed using $N = nkP$. As an example, in an āvartanaṁ period of P = 8 beats (*Ādi* tāla) with tiśra naḍe (n = 3), in dhṛta kāla (k = 4), has $3 \times 4 = 12$ akṣaras played in a beat, with a total of $N = 12 \times 8 = 96$ akṣaras in the one āvartanaṁ.

Carnatic music has a sophisticated tāla system which incorporates the concepts described above. There are 7 ba-

(a)			(b)	
tāla	P	N	naḍe	n
Ādi	8	32	*Tiśra* (Triple)	3
Rūpaka-1	3	12	*Caturaśra* (Quadruple)	4
Rūpaka-2	6	24	*Khaṇḍa* (Pentuple)	5
Miśra Chāpu	7	14	*Miśra* (Septuple)	7
Khaṇḍa Chāpu	5	10	*Saṅkīrṇa* (Nonuple)	9

Table 1. (a) Popular tālas in Carnatic music and their structure (explained in detail in text); (b) Different naḍe in Carnatic music

sic tālas defined with different aṅgas, each with 5 variants (*jāti*) leading to the popular 35 tāla system [1]. Each of these 35 talas can be set in five different naḍe, leading to 175 different combinations. Most of these tālas are extremely rare and Table 1(a) shows the most common tālas with their total akṣaras for a caturaśra naḍe (n=4) and madhyama kāla.

The Mṛdaṅgaṁ follows the tāla closely. It strives to follow the lead melody, improvising within the framework of the tāla. The other rhythmic accompaniments follow the Mṛdaṅgaṁ. The Mṛdaṅgaṁ has characteristic phrases for each tāla, called as *Ṭhēkās* and *Jatis*. Though these characteristic phrases are loosely defined unlike Tablā bols (described next), they serve as valuable indicators for the identification of the tāla and the naḍe.

Percussion solo performance, called a *tani āvartanaṁ* includes the Mṛdaṅgaṁ and other optional accompanying percussion instruments. It is an elaborate rhythmic improvisation within the framework of the tāla. Different naḍe in multiple kālas are played in a duel between the percussionists, taking turns. In this solo, the focus is primarily on the exposition of the tāla and the lead musician helps the ensemble with the visual tāla hand gestures. The patterns played can last longer than one āvartanaṁ, but stay within the framework of the tāla.

1.2.2 Tāl in Hindustani Music

Hindustani music also has a very similar definition of tāl (the ending vowel of a word is truncated in most Hindi words). A tāl has a fixed time cycle of beats, which is split into different *vibhāgs*, which are indicated through the hand gestures of a *thāli* (clap) and a *khāli* (wave). The complete cycle is called an *āvart* and the beginning of a new cycle is called the *sam* [2]. Each tāl has an associated pattern called the *ṭhēkā*. Ṭhēkās for commonly used tāls and a detailed discussion of tāl in Hindustani music can be found in [2], [3], [4], and [5].

Unlike Carnatic music, the tāl is not displayed with visual cues or hand gestures by the lead musician. The Tablā acts as the time-keeper, with the characteristic ṭhēkās defining the āvart cycles. The lead musician improvises based on the tāl cue provided by the Tablā, returning to sam at every phrase. This time-keeping responsibility of Tablā limits the improvisation of Tablā during a composition. However, a Tablā solo performance focuses on the tāl and its exposition, while the lead musician keeps the tāl cycle through repetitive patterns. Since there are no visual tāl cues, the lead musician and the Tablā player take turns to indicate
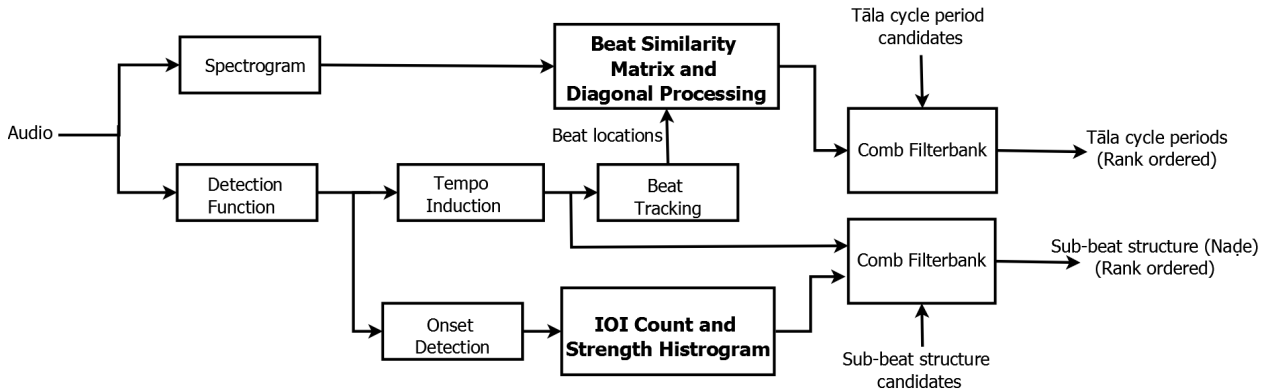
Figure 1. Block Diagram of the system

tāl for the other performer. A Tablā solo aims to expose the variety of rhythms which can be played in the specific tāl, and can be pre-composed or improvised during the performance.

As we see, the complete description of the tāla depends both on the sub-beat structure and the long-term periodicity in the song. The problem of tāla recognition is not well defined since multiple tāla cycle periods, naḍe, and kāla values can lead to the same rhythm for the musical piece. However, even if the tāla label is ambiguous, we can estimate the structure and then find the most probable tāla which corresponds to the rhythmic structure of the song. This needs a knowledge based approach. In the present case, we focus only on estimating the rhythmic structure of the song, without an emphasis on the actual musically familiar label.

1.3 Prior Art

A survey of rhythm description algorithms is provided in [6]. There are several current state of the art tempo estimation and beat tracking algorithms [7], [8], [9]. The problem of estimating the meter of a musical piece has been addressed in [10] and [11]. A beat spectrum based rhythm analysis is described in [12]. The algorithm in this paper is based on [10]. However, these algorithms are not robust to metrical level ambiguity. There has been a few recent attempts of tāla and meter detection for Indian music [2], [13]. There is no current research work that performs an automatic recognition of *tāla* in Carnatic Music [14].

1.4 Proposed Model

The proposed model aims to estimate musically relevant similarity measures at multiple time scales. It is based on the premise that the beats of a song are similar at the rhythmic cycle period and that given the tempo period of the song, the sub-beat structure is indicated by the onsets detected at the sub-beat level in the song. It uses a beat tracker to obtain the tempo and the beat locations. A beat similarity matrix is computed using the beat synchronous frames of the song to obtain the long-term periodicity. A comb filter is then used to rank order the long-term periodicity candidates to estimate the rhythmic cycle period. An inter-
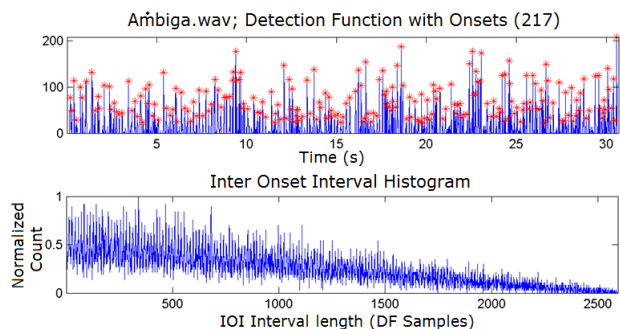

Figure 2. The onsets detected and the IOI Histogram

onset interval histogram is computed from the onsets obtained from the audio signal. Using the tempo estimated from the beat tracker, this IOI histogram is filtered through a comb filterbank to estimate the sub-beat structure. In Carnatic music, coupled with the tempo information, this can be used to obtain the tāla and the naḍe of the musical piece. In Hindustani music, this can used to obtain the tāl. Most often, the tempo, naḍe, and kāla can vary through a composition. But, we focus only on the extraction of global rhythm descriptors of the song in this paper. The algorithm is presented in detail in Section 2.

2. APPROACH

This section describes an algorithm for estimating the sub-beat structure and long-term periodicity of a musical piece. In all our analyses, we use mono audio pieces sampled at 44.1kHz. The block diagram of the entire system is shown in Figure 1.

2.1 Pre-processing

A Detection Function (DF) [8] is first computed from the audio signal $s[n]$, and is a more compact and efficient representation for onset detection and beat tracking. We use a detection function based on spectral flux [7]. The detection function is derived at a fixed time resolution at $t_{DF} = 11.6$ ms and computed on audio signal frames which are 22.64 ms long with 50% overlap between the frames. For each
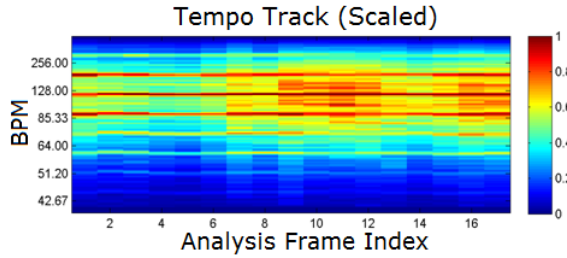
Figure 3. The tempo map over the analysis frames

Figure 4. Beat Similarity matrix of the example song - *Aṁbiga*

frame m, the detection function $\Gamma(m)$ is first smoothed to obtain $\bar{\Gamma}(m)$ and then half wave rectified, as described in [8] to obtain the processed detection function $\tilde{\Gamma}(m)$.

2.2 Onset detection, IOI Histogram and Beat Tracking

The onset detector finds the peaks of the processed detection function $\tilde{\Gamma}(m)$, based on the criteria in [7]. As an additional criterion to ensure only salient onsets are retained, the detected onsets which are less than 5% of the maximum value of $\tilde{\Gamma}(m)$ are ignored. Once the onsets are detected, we compute the inter-onset-interval (IOI) histogram shown in Figure 2. The IOI Histogram $H(m)$ for each $m \in \mathcal{N}$ is a histogram of the number of pairs of onsets detected over the entire song for the given IOI of m DF samples. A peak in the IOI histogram at $m = m'$ indicates a periodicity of detected onsets at that IOI value of m'. This histogram will be used for estimating the sub-beat structure. The detection function $\tilde{\Gamma}(m)$ is used to estimate the tempo of the song using the General State beat period induction algorithm described in [8]. The tempo is estimated over 6 second frames (corresponding to 512 DF samples with a hop size of 128 DF samples). However, instead of a single tempo period for the entire song, we obtain a tempo map over the entire song as shown in Figure 3. The most likely tempo is then obtained by a vote over all the frames. The tempo period thus obtained is τ_p DF samples. The tempo of the song can be estimated from the Tempo period τ_p as in Equation 1.

$$\text{Tempo(bpm)} = \frac{60}{\tau_p \, t_{DF}} \quad (1)$$

It is to be noted that the Rayleigh weighting used in [8] peaks at 120 bpm. This has an influence on the choice of the metrical level at which the sub-beat and the long-term structure is estimated.

A dynamic programming approach proposed by Ellis [9] is used for beat tracking. The inducted tempo period τ_p and the smoothed and normalized (to have unit variance) detection function $\bar{\Gamma}(m)$ are used to track beats at t_i with $1 \leq i \leq N_B$, where N_B is the total number of beats detected in the song.

2.3 Beat Similarity Matrix Diagonal Processing

The spectrogram of the audio $s[n]$ for frame m at frequency bin k is computed as $S(k, m)$ with the same frame size of 22.6 ms and 50% overlap, using a 2048 point DFT. From
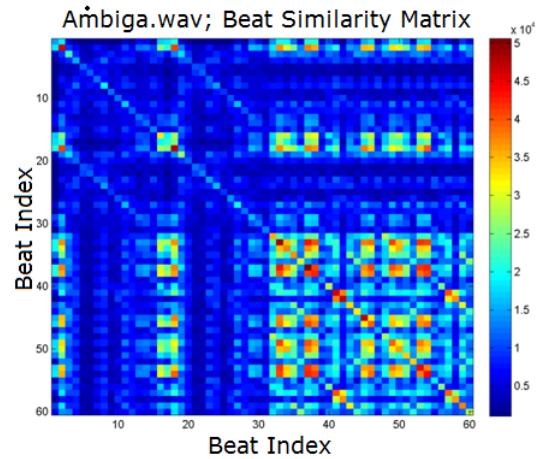
the beat locations t_i, spectrogram is chopped into beat synchronous frames $B_i = \{S_i(k, m)\}$, where for the ith beat B_i, $t_i \leq m < t_{i+1}$ and $t_0 = 1$. The beat similarity matrix (BSM) [10] aims to compute the similarity between each pair of beats B_i and B_j and represent them in a matrix at the index (i, j). The similarity between two beats can be measured in a multiple variety of ways. For simplicity we choose the cross correlation based similarity measure. Since beats can be of unequal length, we first truncate the longer beat to the length of the shorter beat. Also, since the beats could be misaligned, we compute the cross correlation over 10 spectrogram frames and select the maximum. If the length of a beat B_i is τ_i DF samples, with $\tau_{min} = min(\tau_i, \tau_j)$ and for $0 \leq l \leq 10$, the BSM is computed as,

$$BSM(i, j) = \max_l \left[R_l(B_i, B_j) \right] \quad (2)$$

$$R_l(B_i, B_j) = \sum_{p=1}^{\tau_{min}-l} \frac{1}{|\tau_{min} - l|} \left(\sum_{k=1}^{K} S(k, t_{i-1} + p + l) \, S(k, t_{j-1} + p) \right)$$
$$(3)$$

Since spectrogram of a signal is non-negative, the cross-correlation function is estimated as an unbiased estimate of cross-correlation by dividing with $|\tau_{min} - l|$. BSM is symmetric and hence only half of the matrix is computed. To improve computational efficiency, the BSM is computed over only the first 100 beats of the song. The BSM of an example song *Aṁbiga*, a carnatic composition by Sri Purandaradasa is shown in Figure 4.

The diagonals of the BSM indicate the similarity between the beats of the song. A large value on the kth sub-(or supra-) diagonal indicates the similarity of every k beats in the song. Thus we compute the mean over diagonal as,

$$d(l) = \text{mean} \left[\text{diag}(BSM_l) \right] \quad (4)$$

for $1 \leq l \leq L_{\max} = min(N_B, 100)$, where BSM_l refers to the lth sub-diagonal of the BSM. For this computation, $l = 0$ which corresponds to the main diagonal, is ignored. Figure 5 shows a distinct peak at the 16th diagonal for *Aṁbiga*, which is an indicator of rhythmic cycle period.
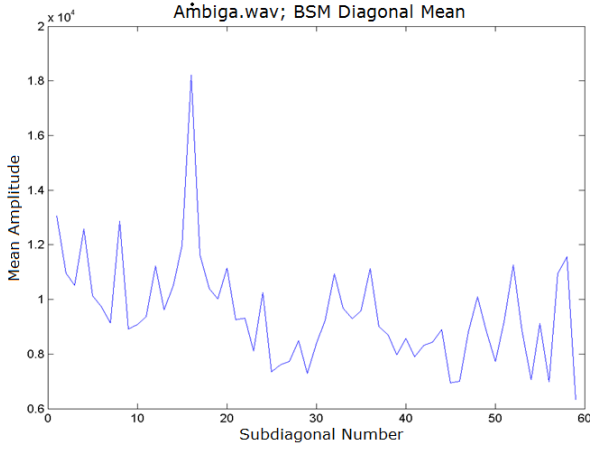
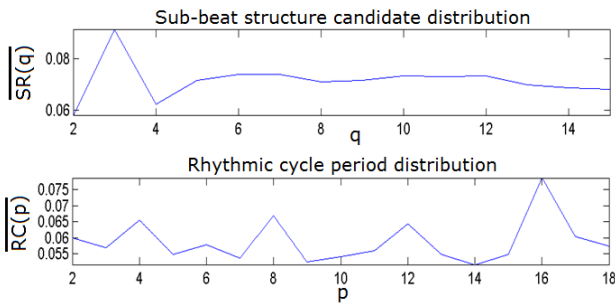**Figure 5**. $d(l)$ - The diagonal mean function

Figure 6. The score of each sub-beat and long-term periodicity candidates

2.4 Estimating long-term rhythmic cycle period

The rhythmic cycle period candidates are tested on the function $d(l)$ using a set of comb filters $C_p(l)$ to obtain a score for each candidate. We test the long-term rhythmic cycle period for the candidates $p = 2, 3, \cdots, 18$. The score $R(p)$ for each p is obtained as in Equations 5 and 6.

$$C_p(l) = \frac{1}{\left\lfloor \frac{L_{max}}{p} \right\rfloor} \sum_{k=1}^{\left\lfloor \frac{L_{max}}{p} \right\rfloor} \delta(l - kp) \qquad (5)$$

$$R(p) = \sum_{l=1}^{L_{max}} C_p(l)d(l) \qquad (6)$$

Here, we define $\delta(n-m) = 1$ if $n = m$ and $\delta(n-m) = 0$ if $n \neq m$. The score is the normalized to obtain a mass distribution over the candidates as in Equation 7 and shown in Figure 6.

$$\overline{R}(p) = \frac{R(p)}{\sum_k R(k)} \qquad (7)$$

The periodicity candidates are rank ordered based on the values of $\overline{R}(p)$.

2.5 Estimating Sub-beat structure

To estimate the sub-beat structure, we use the IOI count histogram $H(m)$. A comb template is used to estimate the periodicity of the IOI histogram at the sub-integral multiples of the tempo period. We use the tempo period τ_p and compute the score for each of the sub-beat candidates $q = 2, 3, \cdots, 15$ using the comb template $D_q(m)$ as,

$$D_q(m) = \frac{1}{qK_1} \sum_{l=1}^{qK_1} \delta\left(m - \left\lfloor \frac{\tau_p}{q} \right\rfloor l\right) \qquad (8)$$

$$S(q) = \sum_m H(m)D_q(m) \qquad (9)$$

where K_1 is the beat periods over which the comb template is computed. In the present work, we set $K_1 = 3$. The score is the normalized to obtain a distribution over the candidates as,

$$\overline{S}(q) = \frac{S(q)}{\sum_k S(k)} \qquad (10)$$

The periodicity candidates are rank ordered based on the values of $\overline{S}(q)$.

3. EVALUATION

The algorithm is tested on approximately 30 second long audio clips reflective of the perceived periodicity of the song. The evaluation of the algorithm is done over two manually annotated collections of songs -

1. Carnatic music dataset (CMDB): Collection of 86 carnatic compositions with a wide set of examples in different tālas with cycle period of 4, 5, 6, 7, 8 and different naḍe - doubles, triples, and pentules.

2. Indian light classical music dataset (ILCMDB): A collection consisting of 58 semi-classical songs based on popular Hindustani rāgs, mainly accompanied by Tablā.

3.1 Evaluation methodology

The scores assigned by the algorithm to each periodicity candidate for a particular song are indicative of the strength of that perceived periodicity. To gauge how well these assigned scores reflect the rhythmic structure of the song, we assign a confidence measure to each possible candidate to indicate the confidence level with which the algorithm predicts that candidate. This allows for a measurable accuracy metric over a range of candidates, all of which are allowable with a certain probability. The confidence measure is defined as,

$$A_{p_c} = \frac{\overline{R}(p_c) - R_{min}}{R_{max} - R_{min}} \qquad (11)$$

where R_{max} is the mode of the $\overline{R}(p)$ distribution and the R_{min} is the minimum probability mass of the distribution. The period p_c corresponds to the annotated period. We define a similar measure A_{q_c} for sub-beat structure candidates using $\overline{S}(q)$.

Another consideration is the metrical level at which accuracy is computed. If accuracy is calculated solely by comparing to the annotated periodicity, it is then a reflection of

Dataset	Examples	Accuracy	CML Accuracy(%)			AML Accuracy(%)		
		Confidence Measure	=1	>0.9	>0.7	=1	>0.9	>0.7
CMDB	86	naḍe	30.23	33.72	36.04	**68.60**	69.77	74.42
		tāla cycle period	9.30	12.79	22.09	**51.16**	55.81	66.28
ILCMDB	60	Sub-beat structure	29.31	31.03	46.55	**79.31**	81.03	86.21
		Long-term cycle period	25.86	31.03	53.45	**72.41**	77.59	84.48

Table 2. Performance of the algorithm on the datasets

how accurately the algorithm detects a single periodicity in the musical structure. Considering that different listeners often perceive rhythm at different metrical levels, and that in many cases, periodicity could be defined as a combination of multiple periods - we define two metrical levels at which we calculate accuracy:

- Correct Metrical Level (CML)
- Allowed Metrical Levels (AML)

CML refers to the periodicity/time signature annotated for each clip by the authors and is hence interpreted as the annotated metrical level. It is also the musically familiar metrical level. In AML, the periodicity could be a factor/multiple of the annotated periodicity to account for metrical level ambiguity. E.g. a periodicity measure of 4 could also be perceived as 8 depending on the rate of counting (i.e chosen metrical level)

At both AML and CML, we compute three accuracy measures at 100% confidence, 90% confidence and 70% confidence over which the algorithm predicts the annotated periodicity. They indicate the number of clips (divided by the total number of clips) with a confidence score equal to 1, >0.9 and >0.7 respectively at the given metrical level.

3.2 Results and Discussion

The performance of the algorithm on the two collections is shown in Table 2. The AML recognition accuracy of the algorithm on ILCMDB is 79.3% and 72.4% for the sub-beat structure and the tāla, respectively. The accuracy on the difficult CMDB was poorer with 68.6% and 51.1% for naḍe and tāla, respectively. As expected, the performance of the algorithm at AML is better than at CML. Further, we see that the sub-beat structure is better estimated than the tāla cycle period.

The poorer performance on CMDB can be attributed to changes in *kāla* (metrical level) through the song and the lack of distinct beat-level similarity in the songs of the dataset. This is quite typical in Carnatic music where the percussion accompaniment is completely free to improvise within the framework of the tāla. The performance is also poor on the songs in odd beat tālas such as *Miśra Chāpu* and *Khaṇḍa Chāpu*. ILCMDB has more stable rhythms with more reliable beat tracking and tempo estimation, leading to better performance.

The sub-beat structure and the long-term periodicity, along with the tempo and the beat locations provide a more complete rhythm description of the song. The presented approach overcomes two main limitations of the beat tracking algorithms. Firstly, the beat tracking algorithms suffer from metrical level ambiguity. The beats tracked might correspond to a different metrical level when compared to the expected metrical level by the listeners. This ambiguity causes the beats to be tracked at a multiple or an integral factor of the required tempo period. Since we estimate at both sub-beat and long-term level, the error in metrical level would correspond to a integer multiple increase (or decrease) of the sub-beat (or long-term) candidate and an integer multiple decrease (or increase) of the long-term (or sub-beat) candidate. This makes the algorithm robust to beat tracking errors. Secondly, the beats locations tracked by the beat tracking algorithm might be offset by a constant value. More specific is the case when the beat tracking algorithm tracks the off-beats. However, since we cross correlate the beats over 10 frames, the effect of tracking off-beats is mitigated.

Further, though metrical level ambiguity due to beat tracking is largely mitigated, the estimated naḍe and tāla cycle periods may not correspond to musically relevant metrical levels as expected by listeners. The perception of metrical levels is largely a subjective phenomenon. There is no absolute metrical level for a music piece in Indian music due to the lack of an absolute tempo. Hence, the expected metrical level varies largely over the listeners. Further the metrical levels can change in a song, making it further difficult to track at the annotated metrical level. Hence knowledge based approaches would be essential to obtain an estimate at the correct metrical level.

The algorithm is based on the implicit assumption that there is a similarity between the beats at the rhythmic cycle period. For certain music pieces where there are no inherent rhythmic patterns or the patterns vary unpredictably, the algorithm gives a poorer performance. The algorithm is non-causal and cannot track any rhythmic period cycle changes in the pieces, though it provides indicators of both rhythmic cycle periods in the tracked tempo.

In this paper, we aimed at estimating the rhythmic structure without assigning any tāla or naḍe label to the songs. This level of transcription and labeling is sufficient for further computer aided processing which need rhythm metadata. However, to use this information as metadata along with a song in applications such as a music browser or a recommendation engine, labels which are more musically familiar and listener friendly need to be generated.

The listeners perceive the tāla and the naḍe through the ṭhēkās and other characteristic phrases played on the Tablā and the Mṛdaṅgaṁ. The audio regions with these constant ṭhēkās can be estimated using locally computed onset interval histograms. These regions might be more suitable for

tempo and beat tracking and might provide better estimates of sub-beat structure. We focused only on global rhythm descriptors. Since these parameters can change through a song, local analysis to track the changes in these descriptors is necessary. Further work in this direction is warranted.

The choice of the weighting function used for tempo tracking plays an important role in the tempo estimated by the algorithm. Presently, the Rayleigh weighting function is set to peak at 120 bpm. However, a further analysis for both Carnatic and Hindustani music for a suitable tempo weighting function would help in tracking the tempo at the expected metrical level. A semi-supervised approach by providing the expected metrical level to the beat tracking algorithm might also lead to better beat tracking performance. Given an estimate of global tempo, we can then obtain a map of local tempo changes which might be useful for rhythm based segmentation. Further, local analysis of tempo changes using onset information would be a logical extension of the algorithm to choose suitable regions for tāla recognition.

4. CONCLUSIONS

In this paper, we proposed a beat tracking based approach for rhythm description of Indian classical music. In particular, we described an algorithm which can be used for tāla and naḍe recognition using the sub-beat and long-term similarity information. The algorithm is quite robust to ambiguity of beat-tracking at the correct metrical level and tracking of the off-beat. The performance of the algorithm is poorer at the correct metrical level as compared to the allowed metrical level. Choice of a suitable tempo weighting function and suitable regions for analysis are to be explored as a part of future work.

Acknowledgments

5. REFERENCES

[1] P. Sambamoorthy, *South Indian Music Vol. I-VI.* The Indian Music Publishing House, 1998.

[2] M. Miron, ``Automatic Detection of Hindustani Talas,'' Master's thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2011.

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

[4] A. E. Dutta, *Tabla: Lessons and Practice.* Ali Akbar College, 1995.

[5] S. Naimpalli, *Theory and practice of tabla*. Popular Prakashan, 2005.

[6] F. Guoyon, ``A Computational Approach to Rhythm Description,'' Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2005.

[7] S. Dixon, ``Evaluation of The Audio Beat Tracking System Beatroot,'' *Journal of New Music Research*, vol. 36, no. 1, pp. 39--50, 2007.

[8] M. E. P. Davies and M. D. Plumbley, ``Context-Dependent Beat Tracking of Musical Audio,'' *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1009--1020, 2007.

[9] D. Ellis, ``Beat Tracking by Dynamic Programming,'' *Journal of New Music Research*, vol. 36, no. 1, pp. 51--60, 2007.

[10] M. Gainza, ``Automatic musical meter detection,'' in *Proceedings of ICASSP 2009*, Taipei, Taiwan, 2009, pp. 329--332.

[11] C. Uhle and J. Herre, ``Estimation of Tempo, Micro Time and Time Signature from Percussive Music,'' in *Proceedings of 6th International Conference on Digital Audio Effects (DAFX-03)*, London, UK, September 2003.

[12] J. Foote and S. Uchihashi, ``The Beat Spectrum: a new approach to rhythm analysis,'' in *Proceedings of the IEEE International Conference on Multimedia and Expo 2001*, Tokyo, Japan, 2001, pp. 881 -- 884.

[13] S. Gulati, V. Rao, and P. Rao, ``Meter detection from audio for Indian music,'' in *Proceedings of International Symposium on Computer Music Modeling and Retrieval (CMMR)*, Bhubaneswar, India, March 2011.

[14] G. K. Koduri, M. Miron, J. Serra, and X. Serra, ``Computational approaches for the understanding of melody in Carnatic Music,'' in *Proceedings of 12th International Society for Music Information Retrieval (ISMIR) Conference*, Miami, USA, October 2011, pp. 263 -- 268.