A SUPERVISED APPROACH TO HIERARCHICAL METRICAL CYCLE TRACKING FROM AUDIO MUSIC RECORDINGS

Ajay Srinivasamurthy, Xavier Serra

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

ajays.murthy@upf.edu, xavier.serra@upf.edu

ABSTRACT

A supervised approach to metrical cycle tracking from audio is presented, with a main focus on tracking the tāḷa, the hierarchical cyclic metrical structure in Carnatic music. Given the tāḷa of a piece, we aim to estimate the akṣara (lowest metrical pulse), the akṣara period, and the sama (first pulse of the tāḷa cycle). Starting with percussion enhanced audio, we estimate the akṣara pulse period from a tempogram computed using an onset detection function. A novelty function is computed using a self similarity matrix constructed using frame level audio features. These are then used to estimate possible akṣara and sama candidates, followed by a candidate selection based on periodicity constraints, which leads to the final estimates. The algorithm is tested on an annotated collection of 176 pieces spanning four different tāḷas. Though applied to Carnatic music, the framework presented is general and can be extended to other music cultures with cyclical metrical structures.

Index Terms— Rhythm, Musical meter, Carnatic Music, Metrical Cycles

1. INTRODUCTION

Rhythm in many music cultures is organized based on hierarchical metrical cycles at various time scales. The organization and relationship between these cyclical structures forms an important aspect of meter. Tracking these cycles from audio music pieces is an important and useful task for automatic rhythm description and can provide listeners with an understanding of the temporal structure of the music piece at various levels. These rhythmic structures form the basis for building more complex musical patterns and play an important role in defining rhythmic similarity. They are also useful for a rhythm based segmentation of a piece and can be further used for higher level structural segmentation. Most of these metrical cycles can be described using musically well defined periodic rhythmic events at different metrical levels. In this paper, we propose a framework and describe a specific algorithm to track such events, with a primary focus on Carnatic music.

Carnatic music is an art music tradition from South India. It has a long history, a large repertoire, and significant musicological literature. It exists and continues to evolve in the present day social context with a large number of active composers, musicians and listeners. Carnatic music has a well defined rhythmic framework with minimal ambiguity in terminology and practice. It is thus an excellent music culture to explore automatic metrical analysis and tracking, and hence is the primary focus of this paper. We first provide a brief introduction to rhythm in Carnatic music.

This work is partly supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

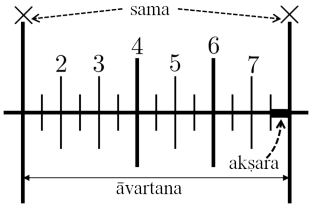

Fig. 1: An āvartana of miśra chāpu tāḷa and its structure shown with the terminology used in this paper. The akṣaras are shown with time ticks, the beats are numbered, and the sama are marked using $\times$.

1.1. Rhythm in Carnatic Music

Rhythm in Carnatic music is based on the framework of tāḷa, which provides a broad structure for repetition of music phrases, motifs and improvisations. A tāḷa consists of fixed length time cycles called *āvartana* which can be referred to as the tāḷa cycle. An āvartana is divided into equidistant basic time units called *akṣaras*, and the first akṣara of each āvartana is called the *sama*. Fig. 1 shows the structure of a popular Carnatic tāḷa called miśra chāpu in which each āvartana consists of 14 akṣaras. The akṣaras are also grouped to give the "beats" of the tāḷa that are displayed by the musician through hand gestures. Sambamoorthy [1] provides a comprehensive description of rhythm in Carnatic music[1].

The sama of an āvartana is often accented, with notable melodic and percussive events. Most phrase changes and improvisations are aligned with the sama. Hence, local changes in melodic, rhythmic and timbral characteristics are possible indicators of sama. However, many phrases span over many āvartanas and hence it is not necessary that every sama has notable events. Since there is no specific indicator for samas, estimating sama locations is a very challenging task. The primary percussion accompaniment in Carnatic music is the Mridangam, a two-sided pitched drum. Mridangam follows the tāḷa closely and hence Mridangam stroke onsets can be used as the primary indicators of the akṣaras. Often, the left(bass)-strokes of Mridangam are used to accentuate important events in the progression of the tāḷa.

1.2. State of the art

Automatic rhythm annotation aims to annotate a piece of music with different events and aspects related to its pulsating rhythm. Most of the prior work has mainly focused on onset detection, tempo and beat tracking, downbeat detection, meter estimation, and drum transcription. Müller et al. [2] report the estimation of beat pulsation (beat tracking) and higher-level rhythmic structures such as the measure

[1] A few tāḷa annotated Carnatic music audio examples can be found at http://compmusic.upf.edu/examples-taala-carnatic

Tāḷa	# Akṣara	$\overline{\text{ISI}} \pm \sigma_{\overline{\text{ISI}}}$	$\overline{\text{APP}}$	# Pieces	$\overline{\text{LEN}}$	Total Minutes (Hours)	# Samas
Ādi	32	5.38 ± 0.6997	0.1680	50	4.85	252.78 (4.21)	2882
Rūpaka	12	2.15 ± 0.2508	0.1790	50	4.62	267.45 (4.46)	7582
Miśra chāpu	14	2.64 ± 0.3394	0.1889	48	6.59	342.13 (5.70)	7795
Khaṇḍa chāpu	10	1.84 ± 0.2788	0.1840	28	4.41	134.62 (2.24)	4387
Total				**176**	**5.06**	**996.98 (16.62)**	**22646**

Table 1: The CompMusic Carnatic Music Rhythm Dataset. The table also shows the number of akṣaras in each āvartana. $\overline{\text{ISI}}$ and $\overline{\text{APP}}$ refer to the median of ISI and APP in the dataset respectively. $\sigma_{\overline{\text{ISI}}}$ is the standard deviation of $\overline{\text{ISI}}$ in the dataset. The number of pieces, median length of each piece $\overline{LEN}$, total size of the dataset and the number of annotated samas is also shown.

length. Approaches such as the one presented by Klapuri et al. [3] aim at estimating structures at several metrical levels at the same time in order to track the metrical grid of a piece. Other approaches focus on specific aspects, such as beat tracking [4], estimating the type of meter [5] in terms of a time signature, or tracking the downbeat locations in a music piece [6, 7], commonly referred to as downbeat detection. Most presented approaches assume the presence of clear cues for a regular metrical grid in the surface rhythm. Hence, they are reported to work reasonably well, especially for European and American popular music. When applied to music with a less clear surface structure, the performance is observed to deteriorate [3, 8]. A recent study that applied the state of the art to Indian and Turkish music showed that analysis methods that exploit the rhythmic structural information are necessary for rhythm description in these music cultures [9].

The work on estimating cyclical metrical structures has been addressed to a very limited extent in Indian Art music. In the context of Hindustani music (Art music culture from North India), approaches for tempo and meter estimation were presented by Gulati et al. [10], transcription of percussive instruments was approached by Chordia [11], and recognition of tāḷas was attempted by Miron [12]. For Carnatic music, Srinivasamurthy et al. [13] proposed to describe meter in terms of the time-span relations at measure, beat and akṣara levels. None of these approaches have considered the task of tracking the different components of a tāḷa in a piece. To the best of our knowledge, this is the first work in tracking components of a tāḷa.

2. PROBLEM DEFINITION

Starting with audio music pieces, we wish to estimate the rhythmic events that will finally lead to a complete description of the hierarchical metrical cycles in the piece. We present a general framework for this task that can incorporate domain specific knowledge about the music and the kind of rhythmic structures present in that music.

We primarily intend to work with large music collections comprising of multiple sources of information. We wish to design an approach that combines all the data sources effectively in the task of rhythm description. Most music collections in Carnatic music contain the editorial metadata of the tāḷa for each piece and hence tāḷa recognition for such a collection is a redundant task. Hence, we propose a supervised approach that uses the tāḷa label (and hence its structure) and then estimates the different events related to the tāḷa in the audio. In this paper, we limit ourselves to estimating the akṣaras and the samas, which can be used towards complete tāḷa tracking.

A major component of describing rhythm is the tempo of the piece, measured with the number of akṣaras per minute (APM). From an audio piece, we estimate an equivalent tempo measure - the akṣara pulse period (APP) defined as $60/\text{APM}$. Carnatic music pieces do not have a pre-notated tempo and the musician is free to choose any tempo. Given the APP, since an āvartana has a fixed number of akṣaras depending on the tāḷa, we can estimate the inter-sama-interval (ISI), the time interval between two adjacent sama pulses or the length of an āvartana. Further, since the pieces are not played to a click track, the APP can vary through a piece and hence we estimate a possibly time varying APP curve from which we can derive a time varying ISI curve.

In summary, we estimate the APP curve, akṣara and sama time locations from an audio music recording. Though the main task of tāḷa tracking encompasses these sub-tasks, each of these sub-tasks is important in itself and provide useful rhythm descriptors.

3. DATASET

The dataset used in this paper is a part of the CompMusic [14] Carnatic music audio collection, which is a research corpus of curated commercial audio releases and metadata with over 397 hours of music and is representative of the present day Carnatic music. Since no rhythm annotated Carnatic audio dataset existed, we compiled a dataset that consists of 176 excerpts of Carnatic music sampled from the CompMusic collection. The pieces span the four most popular tāḷas in Carnatic music in which a majority of pieces are composed. They consist of both vocal and instrumental recordings. The audio is stereo and sampled at 44100 Hz. All the pieces contain a percussion accompaniment, predominantly the Mridangam.

The annotations consist of the editorial metadata about the tāḷa and the time aligned tāḷa annotations including the sama and the beats for each piece. The tāḷa annotations were manually done by a Carnatic percussionist by tapping to the piece and then correcting the taps using Sonic Visualizer [15]. The dataset is representative of the present day performance practice in Carnatic music and spans a wide variety of artists, forms and instruments. The dataset is described in Table 1. From Table 1, we see that though there is no notated tempo in Carnatic music, musicians tend to perform in a narrow range of tempo, as shown by the small variance of $\overline{\text{ISI}}$. To the best of our knowledge, this is the first tāḷa annotated collection of Carnatic music (The annotations are publicly available for download at `http://compmusic.upf.edu/carnatic-rhythm-dataset`).

4. APPROACH

The primary philosophy of the approach presented in the paper is to incorporate specific knowledge of the rhythmic structures we aim to estimate. For estimating the APP, akṣara period curve, and the samas, we first estimate a descriptor computed from the audio that indicates the possible candidates for each musical concept. We then make use of the periodicity and the relationships between these structures to estimate the components. This framework can be generalized to estimating other rhythmic structures by suitably modifying the audio descriptor for the specific music culture and the rhythmic structure under consideration. The algorithm for Carnatic music is

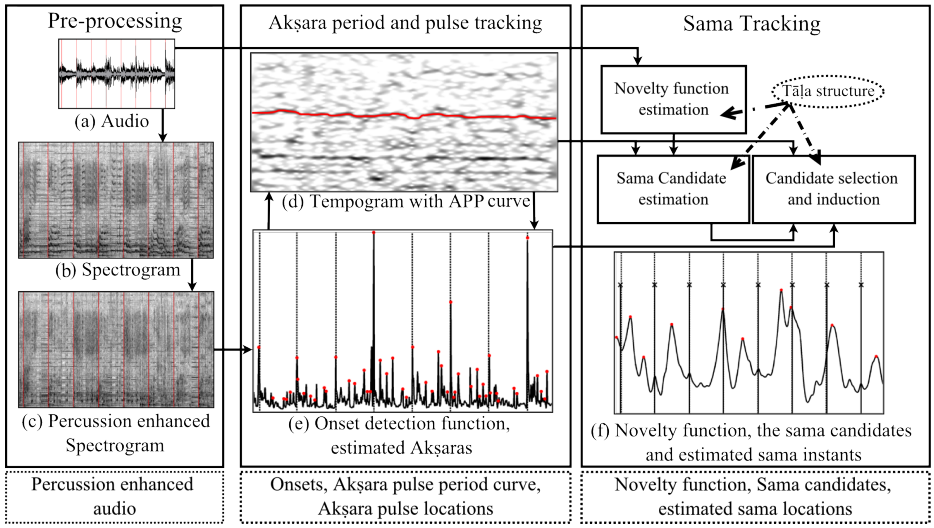

Fig. 2: Block Diagram of the algorithm showing the signal flow and representative illustrations of different stages of the algorithm. The important outputs at each stage are also shown at the bottom. In each panel, the vertical lines that run through the panel indicate the sama ground truth instants. The estimated sama/akṣara candidates are shown with red dots and the estimated sama are shown with $\times$.

explained in detail in this section.

4.1. Pre-processing

The akṣara pulse most often coincides with the onsets of Mridangam strokes. To enhance the Mridangam onsets, percussion enhancement is performed on the downmixed mono audio, as it has been shown to improve beat tracking performance in pieces with predominant vocals [16]. The predominant melody is estimated using the algorithm proposed by Salamon et al. [17] using which, we extract the harmonic component of the signal using a sinusoidal+residual model [18]. The percussion enhanced signal $x_p[n]$, with the harmonic component suppressed, is used for further processing (Fig. 2(c)).

4.2. Akṣara period and pulse tracking

The spectrogram of $x_p[n]$ is used to compute two frame level spectral-flux based onset detection functions [19] computed every 11.6 ms. The first function ($d_a[m]$) uses the whole frequency range of the spectrogram and the other function computes the spectral flux only in the range of 0-120 Hz ($d_l[m]$) and captures the low frequency onsets of the left(bass)-side of the Mridangam.

The function $d_a[m]$ is used to compute a Fourier-based Tempogram [20] computed every 0.25 second using a 8 second long window (Fig. 2(d)). If we denote the time indices at which the tempogram is computed with i, $(1 \leq i \leq T)$, we can track the most predominant APP curve by estimating the best path $\mathbf{P} = \{p_i : i = 1, 2, \cdots, T\}$ through the tempogram matrix $\mathbf{M}$ that provides a balance between tempogram amplitude at time index i, $M_{p_i,i}$, and the local continuity of APP. We define a cost function that is an extended version of the one used by Wu et al. [21] as shown in Eqn. 1.

$$J_1(\mathbf{P}, \theta_1, \theta_2) = \sum_{i=1}^{T} M_{p_i,i} - \sum_{i=1}^{T-1} \left(\theta_1 \left| p_i - p_{i+1} \right| + \theta_2 \, O\left(\frac{p_i}{p_{i+1}} \right) \right) \tag{1}$$

The function $O(p_i/p_{i+1})$ is an extra penalty term to penalize tempo doubling and halving between adjacent frames, and the weights θ_1(=0.01) and θ_2(=1e6) provide different weights to the three terms. Based on our observation from the dataset (Sec. 3), the search for the best path through the tempogram is restricted between the range of 120 to 600 APM. The above cost function is solved using a dynamic programming (DP) based approach to obtain an APP curve that is then corrected for tempo doubling/halving errors, if any, to obtain the final curve $\mathbf{P}^\star$ (Fig. 2(d), $\mathbf{P}^\star$ is shown as a thick red line).

The akṣara pulse locations predominantly lie on strong Mridangam onsets. We estimate the akṣara pulse candidates as the peaks of the function $d_a[m]$. Using these K candidate peaks $\{a_k\}$, $k = 1, 2, 3, \cdots, K$, with locations t_k and peak amplitude w_k, we setup a cost function shown in Eqn. 2 to select the best candidates that provide a balance between the amplitude of these candidates and a periodicity provided the estimated APP. The best set of candidates $\{a_k^\star\} \subset \{a_k\}$ are estimated using a DP approach (Fig. 2(e)).

$$J_2(\{a_k\}, \delta) = \sum_{k \subset \{1,2,\cdots,K\}} \left(w_k + \delta C(t_k, t_{k+1}, \mathbf{P}) \right) \tag{2}$$

The function $C(t_k, t_{k+1}, \mathbf{P})$ is a function that returns a exponentially decaying weight based on the time difference between t_k and t_{k+1} in relation to the local APP, P_{t_k}. The parameter δ(=3) provides a trade-off between the two terms. Using the APP and the tāḷa information, we can obtain the time varying ISI curve for the piece by multiplying the APP by the number of akṣaras in a cycle of the tāḷa.

4.3. Sama Tracking

As described earlier, the samas are often associated with significant melodic, rhythmic and timbral changes. In this article, we explore the use of MFCC (mel frequency cepstral coefficients) as features for timbral characteristics. As a detection function for sama ($d_s[m]$), we use a novelty function computed through the diagonal processing of a self similarity matrix [22] constructed using frame level z-score MFCC features from audio (using audio processing library *Essentia* [23]) as shown in Fig. 2(f). Based on the $\overline{\text{ISI}}$ shown in Table 1, a checkerboard kernel with size of 7, 3, 4, and 3 seconds is used for ādi, rūpaka, miśra chāpu and khaṇḍa chāpu respectively so that the novelty function is computed over about an āvartana.

The peaks of the novelty function $d_s[m]$ indicate a significant change of timbre at that time. Starting with the premise that timbral change is an important indicator of sama location, we use the peaks of the novelty function to estimate sama candidates. We explore two methods to estimate the candidates. In Method-A, to uniformly choose sama candidates throughout a piece, we cut the piece

into segments of length 120, 40, 40 and 30 seconds for ādi, rūpaka, miśra chāpu and khaṇḍa chāpu respectively ($\sim$10 āvartanas), and estimate the top five most prominent peaks in each segment of the piece as sama candidates ($\{s_k^A\}$). We also propose another approach, Method-B, for candidate estimation that enforces a periodicity constraint while estimating sama candidates. Starting from the peaks of $d_s[m]$ and estimated ISI curve, for a specific peak, we induce the tāḷa cycle starting from it. Based on how many other peaks would support such an induced tāḷa is assigned as the weight of the specific peak. We can rank order the peaks using this weight and choose the top ten ranked peaks as the sama candidates ($\{s_k^B\}$). We also create two random baseline methods RB-1 and RB-2 to compare the performance. In RB-1, we use a randomly chosen constant ISI between 1-8 seconds, and a random starting time between 0-2 seconds to induce periodic samas. In RB-2, we use the estimated ISI with 10 randomly chosen akṣara locations from $\{a_k\}$ as sama candidates. RB-1 neither uses the ISI, nor the candidate estimation using $d_s[m]$, while RB-2 uses the estimated ISI but not the candidate estimation using $d_s[m]$.

Starting with the sama candidates from either Method-A or Method-B, for each candidate, we induce the tāḷa cycles based on local ISI period obtained from the ISI curve. For each seed, we search within the next and previous three estimated cycle periods for onset peaks in $d_a[m]$ that support a sama. If we find a supporting onset, we mark it as a sama and then proceed further with the new estimated onset as the new anchor. We stop the induction from a candidate when it does not lead to such a supporting onset. Hence for each candidate, we obtain an estimated sama sequence. Since all candidates are not necessarily sama locations, though the estimated ISI is right, the sequences can have different offsets. The final step of the algorithm is to shift, align and merge these sequences obtained from each candidate. Starting with the longest sama sequence that has been estimated, we merge the other sequences into this based on maximum correlation between the sequences. The merging of these sequences often leads to many sama estimates concentrated around the true location of sama due to small offsets. Since the left bass onsets on the Mridangam are often strong at the samas, all groups of sama estimates that are closer than 1/3rd of ISI are merged into a single sama estimate aligned with the closest left stroke onset obtained from $d_l[m]$. This forms the final set of sama locations $\{s_{t_k}\}$ estimated from the candidates and the onset detection function, as shown in Fig. 2(f) with $\times$.

5. RESULTS AND DISCUSSION

The annotated dataset has annotations only for beats and samas of the piece. From the sama locations, we can obtain the ground truth for ISI curve, and hence the ground truth for APP curve. Since we do not have the ground truth for akṣara locations, we present the results only for APP and sama tracking.

Measure	CML (%)	AML (%)
$\overline{APP}$ estimation	81.25	98.86
APP tracking	80.45	96.26

Table 2: Accuracy of APP and ISI tracking

The performance of APP tracking is measured by comparing the ground truth APP curve with the estimated curve with an error tolerance of 5%. We also report the results for estimation of $\overline{APP}$ computed from the whole APP curve. Further, since there can be tempo doubling and halving errors, we also report the accuracy at the annotated correct metrical level (CML) and then using a weaker measure that allows tempo halving and doubling (AML - allowed metrical levels). The results are presented in Table 2. We see that an acceptable level accuracy is achieved at CML for both $\overline{APP}$ estimation and APP tracking and further, there is not a significant difference between their performances, indicating that the algorithm can track changes in tempo effectively. Even when the APP tracking fails at CML, the algorithm tracks a metrically related APP, as indicated by a high AML accuracy.

Variant	**P**	**R**	**F**	**I** (bits)	**Cand. Est.**
Method-A	29.02	19.01	21.61	1.17	20.46
Method-B	24.60	20.15	21.51	1.25	27.85
RB-1	15.51	17.47	13.73	0.40	-
RB-2	22.76	19.98	20.62	1.11	15.3

Table 3: Accuracy of Sama Tracking. P: Precision, R: Recall, F: f-measure, I: Information Gain. The values are mean performance over the whole dataset expressed in % except for Information Gain, which is in bits. The last column shows the fraction of the estimated sama candidates that are true samas.

For sama tracking, we report the accuracy of estimation with a margin of 7% the annotated $\overline{ISI}$ of the piece. Given the ground truth and the estimated sama time sequence, we use the common evaluation measures used in beat tracking - precision, recall, f-measure and Information Gain [24] to measure the performance. The results are shown in Table 3, which also shows the accuracy of sama candidate estimation. The results for RB-1 and RB-2 show mean performance over 100 and 10 experiments for each piece, respectively.

We see that the performance of sama candidate estimation and sama tracking is poor in general, with samas correctly tracked only in about a fifth of cases. The precision is higher than recall in all cases, and Information Gain is lower than a perceptually acceptable threshold [25]. Both methods perform better than RB-1, but have comparable results with RB-2, with a slightly better f-measure performance (statistically significant in a Mann–Whitney U test at $p < 0.05$). This shows that the estimated ISI is useful for sama estimation, whereas candidate estimation using novelty function is only marginally useful. The poor performance can be mainly attributed to poor sama candidate estimation with either of Method-A or Method-B. This is further substantiated by the fact that Method-B achieves an f-measure of 43.58% and an information gain of 1.70 bits when at least half the estimated candidates are true samas. This clearly shows that the performance of sama tracking crucially depends on sama candidate estimation. There are only four pieces (among all pieces with accurate ISI estimation) in which all the estimated sama candidates are true samas, for which an f-measure of 89.38% and a information gain of 3.51 bits is achieved. This clearly indicates that the novelty function from which the sama candidates were estimated is not a very good indicator of samas, and better descriptors need to be explored.

6. CONCLUSIONS

We presented a general framework for tracking hierarchical metrical cycle from audio and a specific algorithm to estimate the akṣara pulse period, akṣara and sama locations from Carnatic music audio pieces. We also presented a tāḷa annotated Carnatic music dataset that can be used in automatic rhythm annotation tasks. APP tracking performs to an acceptable accuracy for practical applications, while sama tracking is challenging and performs poorly primarily due to poor sama candidate estimation. Though the framework for estimating sama is promising, the novelty function used presently is not a good indicator for sama and hence better audio descriptors need to be explored for the task.

7. REFERENCES

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

[2] Meinard Müller, Daniel P. W. Ellis, Anssi Klapuri, Gaël Richard, and Shigeki Sagayama, "Introduction to the Special Issue on Music Signal Processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1085–1087, 2011.

[3] Anssi P. Klapuri, Antti J. Eronen, and Jaakko T. Astola, "Analysis of the Meter of Acoustic Musical Signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 342–355, 2006.

[4] Dan P. W. Ellis, "Beat Tracking by Dynamic Programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[5] Aggelos Pikrakis, Iasonas Antonopoulos, and Sergios Theodoridis, "Music meter and tempo tracking from raw polyphonic audio," in *Proc. of 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, October 2004.

[6] Matthew E. P. Davies and Mark D. Plumbley, "A spectral difference approach to downbeat extraction in musical audio," in *Proc. of the 14th European Signal Processing Conference (EUSIPCO 2006)*, Florence, Italy, September 2006.

[7] Jason A. Hockman, Matthew E. P. Davies, and Ichiro Fujinaga, "One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass," in *Proc. of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, Porto, Portugal, October 2012, pp. 169–174.

[8] André Holzapfel, M.E.P. Davies, J.R. Zapata, J.L. Oliveira, and F. Gouyon, "Selective Sampling for Beat Tracking Evaluation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2539–2548, November 2012.

[9] 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*, vol. 43, no. 1, pp. 97–117, 2014.

[10] Sankalp Gulati, Vishweshwara Rao, and Preeti Rao, "Meter detection from audio for indian music," in *Speech, Sound and Music Processing: Embracing Research in India: 8th International Symposium, CMMR 2011, 20th International Symposium, FRSM 2011, Bhubaneswar, India, March 9-12, 2011, Revised Selected Papers, Lecture Notes in Computer Science, vol. 7172*, Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, Kristoffer Jensen, and Sanghamitra Mohanty, Eds., pp. 34–43. Springer: Berlin Heidelberg, 2012.

[11] Parag Chordia, *Automatic Transcription of Solo Tabla Music*, Ph.D. thesis, Stanford University, 2005.

[12] Marius Miron, "Automatic Detection of Hindustani Talas," Master's Thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2011.

[13] Ajay Srinivasamurthy, Sidharth Subramanian, Gregoire Tronel, and Parag Chordia, "A Beat Tracking Approach to Complete Description of Rhythm in Indian Classical Music," in *Proc. of the 2nd CompMusic Workshop*, Istanbul, Turkey, July 2012, pp. 72–78.

[14] Xavier Serra, "A multicultural approach in Music Information Research," in *Proc. of 12th International Society for Music Information Retrieval Conference (ISMIR-2011)*, Miami, USA, October 2011, pp. 151–156.

[15] C. Cannam, C. Landone, and M. Sandler, "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files," in *Proceedings of the ACM Multimedia 2010 International Conference*, Florence, Italy, October 2010, pp. 1467–1468.

[16] J. Zapata and Emilia Gómez, "Using Voice Suppression Algorithms to improve Beat Tracking in the Presence Of Highly Predominant Vocals," in *Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, Vancouver, Canada, May 2013, pp. 51–55.

[17] Justin Salamon and Emilia Gómez, "Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759–1770, August 2012.

[18] Xavier Serra, "Musical Sound Modeling with Sinusoids plus Noise," in *Musical Signal Processing*, C. Roads, S. T. Pope, A. Picialli, and G. De Poli, Eds., Studies on New Music Research, pp. 91–122. Swets & Zeitlinger, 1997.

[19] Juan Pablo Bello, L Daudet, S Abdullah, C Duxbury, M Davies, and M Sandler, "A Tutorial on Onset Detection in Music Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, September 2005.

[20] Peter Grosche and Meinard Müller, "Tempogram Toolbox: MATLAB tempo and pulse analysis of music recordings," in *12th International Conference on Music Information Retrieval (ISMIR, late-breaking contribution)*, Miami, USA, 2011.

[21] Fu-Hai Frank Wu, Tsung-Chi Lee, Jyh-Shing Roger Jang, Kaichun K. Chang, Chun-Hung Lu, and Wen-Nan Wang, "A Two-Fold Dynamic Programming Approach to Beat Tracking for Audio Music with Time-Varying Tempo," in *Proc. of 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, USA, 2011, pp. 191–196.

[22] Jonathan Foote, "Automatic audio segmentation using a measure of audio novelty," in *IEEE International Conference on Multimedia & Expo 2000*, Tokyo, Japan, 2000, vol. 1, pp. 452 – 455.

[23] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "ESSENTIA: an Audio Analysis Library for Music Information Retrieval," in *Proc. of 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, Curitiba, Brazil, November 2013, pp. 493–498.

[24] M. F. McKinney, D. Moelants, M. E. P. Davies, and A. Klapuri, "Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms," *Journal of New Music Research*, vol. 36, no. 1, pp. 1–16, 2007.

[25] José R. Zapata, André Holzapfel, Matthew E. P. Davies, Joao Lobato Oliveira, and Fabien Gouyon, "Assigning a Confidence Threshold on Automatic Beat Annotation in Large Datasets," in *Proc. of 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, October 2012, pp. 157–162.