A GENERALIZED BAYESIAN MODEL FOR TRACKING LONG METRICAL CYCLES IN ACOUSTIC MUSIC SIGNALS

Ajay Srinivasamurthy[⋆], *Andre Holzapfel*[†], *Ali Taylan Cemgil*[‡], *Xavier Serra*[⋆]

[⋆]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[†]Austrian Research Institute for Artificial Intelligence, Vienna, Austria
[‡]Dept. of Computer Engineering, Boğaziçi University, Istanbul, Turkey

ajays.murthy@upf.edu, andre@rhythmos.org, taylan.cemgil@boun.edu.tr, xavier.serra@upf.edu

ABSTRACT

Most musical phenomena involve repetitive structures that enable listeners to track meter, i.e. the tactus or beat, the longer over-arching measure or bar, and possibly other related layers. Meters with long measure duration, sometimes lasting more than a minute, occur in many music cultures, e.g. from India, Turkey, and Korea. However, current meter tracking algorithms, which were devised for cycles of a few seconds length, cannot process such structures accurately. We present a novel generalization to an existing Bayesian model for meter tracking that overcomes this limitation. The proposed model is evaluated on a set of Indian Hindustani music recordings, and we document significant performance increase over the previous models. The presented model opens the way for computational analysis of performances with long metrical cycles, and has important applications in music studies as well as in commercial applications that involve such musics.

Index Terms— Rhythm analysis, Bayesian models, Meter tracking, Particle filters, Hindustani music

1. INTRODUCTION

An important rhythm analysis task in Music Information Research (MIR) is to identify and align the underlying meter within an audio music recording. For instance, beat tracking aims to align audio with the metrical level called the *tactus*, referred to as the metrical level at which a listener taps her foot [1, p.21] (see [2] for a list of beat tracking algorithms). Tracking the meter at the higher level of bar/measure is often referred to as downbeat tracking in MIR. We explore the combined task of beat and downbeat tracking, which we refer to as meter tracking since it aims to align several levels of a known meter to an audio recording of a music performance.

Several approaches have discussed meter tracking in the past [3, 4] and recent approaches in meter tracking have successfully applied Bayesian models [5,6]. Based on the model presented in [6], the task of meter tracking was combined with the determination of the type of meter in [7]. Other strategies involve the usage of deep learning, e.g. in [8], where a set of deep belief networks are trained on various features for the task of downbeat detection.

All methods presented so far in the context of meter tracking, to the best of our knowledge, have been evaluated on metrical cycles of short durations. In specific, the typical duration of a 4/4 measure in popular Eurogenetic music would last from a bit less than 2s to little more than 4s. Longer metrical cycles were reported to cause problems in existing approaches [7]. Interestingly, this upper duration coincides with the limit of a perceptual phenomenon referred to as *perceptual present* [9], and it has been argued that longer metrical cycles might not be perceived as a single rhythmic entity [10]. In tracking such long metrical cycles, listeners often track shorter, but musically meaningful sections of the cycle.

A similar idea was applied by Böck et al. [11], where rhythmic patterns of beat length are learned in order to perform beat tracking. However, the authors assume beats to form an isochronous sequence - an assumption that does not hold for many musics of the world, such as Indian, Turkish, Balkan, or Korean musics. Furthermore, they do not attempt to infer higher metrical levels, i.e. downbeat positions. In this paper, we address for the first time two basic limitations of the existing meter tracking approaches, which are the restrictions to short cycles and isochronous (equally spaced in time) beat sequences. We propose a generalization of our previous models that uses musically meaningful and possibly unequal section length rhythmic patterns in the task of meter tracking, and apply it to Hindustani music. With the new model, we evaluate if using shorter section length rhythmic patterns can improve meter tracking compared to bar (cycle) length rhythmic patterns, in the presence of long metrical cycles.

Hindustani music (HM) is an art music tradition from the Indian subcontinent, which continues to play an important role in the local sociocultural context with significant musicological literature and a large audience. The rhythmic framework in HM is based on cyclic rhythmic modes of certain length called the *tāl*. A cycle of a tāl is divided into isochronous basic time units called *mātrā*. The mātrās of a tāl are grouped into sections, sometimes with unequal time-spans, called *vibhāgs*. The beginning of a cycle (the downbeat) is referred to as *sam* [10]. Figure 1 shows a tāl that is 7 mātrā long, *rūpak tāl*, with three vibhāgs of unequal lengths. The vibhāgs are numbered, with the sam shown with numeral 1. Percussion accompaniment is provided by the tabla, which acts as the timekeeper of the tāl and plays pre-defined rhythmic patterns (called the *ṭhēkā*) for each tāl (for more details see [10, 12–14], and acoustic examples are provided at `http://compmusic.upf.edu/examples-taal-hindustani`).

Hindustani music divides tempo into three main tempo classes (*lay*). Since no exact tempo ranges are defined for these classes, we determined suitable values, measured in mātrās per minute (MPM), in correspondence with a professional Hindustani musician as 10-60 MPM, 60-150 MPM, and >150 MPM for the slow (*vilaṁbit*), medium (*madhya*), and fast (*dṛt*) tempi, respectively. In our experiments, we will examine how the tempo class affects the tracking accuracy, and we will compare with an informed case, in which the tempo class is known.

This work is partly supported by the European Research Council as part of the CompMusic project (ERC grant agreement 267583), and Vienna Science and Technology Fund (WWTF, project MA14-018). The authors thank F. Krebs and S. Böck, Johannes Kepler University, Austria for providing the code for the state of the art model, and K. K. Ganguli, IIT Bombay, India for the help in creating the dataset described in the paper.

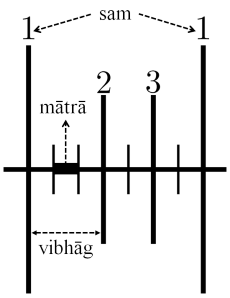

Fig. 1: Rūpak tāl (7 mātrās)

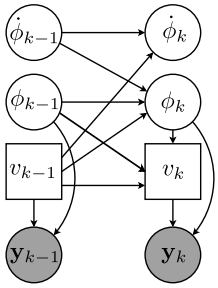

Fig. 2: Section Pointer Model

With a wide range of tempo, cycles as long as a minute, and non-isochronous subdivisions of the cycle, Hindustani music is a suitable case for extending the horizon of the state of the art in meter tracking [15]. There has been some previous work in rhythmic analysis of Hindustani music in meter estimation [16] and tāl recognition [12], but to the best of our knowledge, this is the first work to propose meter tracking for Hindustani music. However, since the proposed model is a generalization of a state of the art model, it can be applied to arbitrary music styles in a straight forward way. Free access for research purposes to all code repositories and datasets is provided on the companion webpage[1]. We begin by describing the model and the inference scheme that we use for meter tracking.

2. MODEL STRUCTURE AND TRAINING

Given an audio recording of a music piece along with the information about its rhythmic mode (tāl), we aim to time align the mātrā and the sam with the recording i.e. the goal is to track a known metrical structure. We propose a Dynamic Bayesian Network (DBN) called the Section Pointer Model (SPM) that is based on the bar pointer model (BPM), which was initially proposed in [17], and then applied in [5–7, 18]. The model is shown in Figure 2, where circles and squares denote continuous and discrete variables, respectively. Gray nodes and white nodes represent observed and latent variables, respectively.

In a DBN, an observed sequence of features derived from an audio signal $\mathbf{y}_{1:K} = [\mathbf{y}_1, \ldots, \mathbf{y}_K]$ is generated by a sequence of latent variables $\mathbf{x}_{1:K} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$, where K is the length of the sequence (number of frames in an audio excerpt). The joint probability distribution of latent and observed variables factorizes as,

$$P(\mathbf{y}_{1:K}, \mathbf{x}_{0:K}) = P(\mathbf{x}_0) \cdot \prod_{k=1}^{K} P(\mathbf{x}_k|\mathbf{x}_{k-1})\, P(\mathbf{y}_k|\mathbf{x}_k) \qquad (1)$$

where, $P(\mathbf{x}_0)$ is the initial state distribution, $P(\mathbf{x}_k|\mathbf{x}_{k-1})$ is the transition model, and $P(\mathbf{y}_k|\mathbf{x}_k)$ is the observation model.

2.1. Latent variables

At each audio frame k, the latent variables describe the state of a hypothetical pointer $\mathbf{x}_k = [\phi_k \;\; \dot{\phi}_k \;\; v_k]$, representing the position in the section, instantaneous tempo, and a section indicator, respectively.

- *Section indicator*: The section indicator variable $v \in \{1, \ldots, V\}$ is an indicator variable that identifies the section (vibhāg) of a tāl, and selects one of the V observation models corresponding to each section length rhythmic pattern learned from data. A tāl might have many sections of different lengths. We denote the number of mātrās in a section v by B_v.

[1] `http://compmusic.upf.edu/icassp-2016-spm`

- *Position in section*: The position within a section is tracked by $\phi \in [0, M_v)$, with ϕ increasing from 0 to M_v and then resetting to 0 to start the next section, where M_v is the length of section v. We set the length of the longest section as M, and then scale the lengths of other sections accordingly.
- *Instantaneous tempo*: Instantaneous tempo $\dot{\phi}$ (measured in positions per time frame) is the rate at which the position variable ϕ progresses through a section at each time frame. The allowed range of the variable $\dot{\phi}_k \in [\dot{\phi}_{\text{min}}, \dot{\phi}_{\text{max}}]$ depends on the frame hop size ($\Delta = 0.02s$ used in this paper), and can be preset or learned from data. In a given section v, a value of $\dot{\phi}_k$ corresponds to a section duration of $(\Delta \cdot M_v/\dot{\phi}_k)$ seconds and $(60 \cdot {}^{B_v \cdot \dot{\phi}_k}/_{(M_v \cdot \Delta)})$ mātrās per minute (MPM).

2.2. Transition and Observation model

In this paper, we assume uniform priors, $P(\mathbf{x}_0)$, on all variables, within the allowed ranges of tempo. Due to the conditional dependence relations shown in Figure 2, the transition model factorizes as,

$$P(\mathbf{x}_k|\mathbf{x}_{k-1}) = P(\dot{\phi}_k|\dot{\phi}_{k-1}, v_{k-1})\, P(v_k|v_{k-1}, \phi_k, \phi_{k-1}) \\ P(\phi_k|\phi_{k-1}, \dot{\phi}_{k-1}, v_{k-1}) \qquad (2)$$

Each term in Eq. (2) is further defined in Eq. (3)–(5).

$$P(\phi_k|\phi_{k-1}, \dot{\phi}_{k-1}, v_{k-1}) = \mathbb{1}_{\phi} \qquad (3)$$

where $\mathbb{1}_{\phi}$ is an indicator function that takes a value of one if $\phi_k = (\phi_{k-1} + \dot{\phi}_{k-1}) \bmod (M_{v_{k-1}})$ and zero otherwise, meaning that the position advances at the rate of the instantaneous tempo variable, and is reset when it crosses the maximum value $M_{v_{k-1}}$ of the section being tracked.

$$P(\dot{\phi}_k|\dot{\phi}_{k-1}, v_{k-1}) \propto \mathcal{N}(\dot{\phi}_{k-1}, \sigma^2_{\dot{\phi}}) \times \mathbb{1}_{\dot{\phi}} \qquad (4)$$

where $\mathbb{1}_{\dot{\phi}}$ is an indicator function that equals one if $\dot{\phi}_k \in [\dot{\phi}_{\text{min}}, \dot{\phi}_{\text{max}}]$ and zero otherwise, restricting the tempo to be between a predefined range. $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. The value of $\sigma_{\dot{\phi}}$ depends on the value of tempo and the length of the section. We set $\sigma_{\dot{\phi}} = \sigma_n \cdot \dot{\phi}_{k-1} \cdot ({}^{M_{v_{k-1}}}/_{M})$, where σ_n is a user parameter that controls the amount of local tempo variations we allow in the music piece.

$$P(v_k|v_{k-1}, \phi_k, \phi_{k-1}) = \begin{cases} \mathbf{A}(v_{k-1}, v_k) & \text{if } \phi_k < \phi_{k-1} \\ \mathbb{1}_v & \text{else} \end{cases} \qquad (5)$$

where $\mathbf{A}(v_i, v_j)$ is the time-homogeneous transition probability from v_i to v_j, and $\mathbb{1}_v$ is an indicator function that equals one when $v_k = v_{k-1}$ and zero otherwise. Section changes are permitted only at the end of the section, and the matrix $\mathbf{A}$ is used to determine the order of the sections as defined in the tāl by allowing only those defined transitions.

The observation model incorporates the same two dimensional spectral flux feature used in [6], and depends on the position (ϕ) and the section indicator (v) variables. Using mātrā and sam annotated training data, section length feature sequences are obtained from each tāl. We then discretize a section into cells that are 1/8$^{\text{th}}$ of a mātrā in length, collect all the features within the cell and compute the maximum likelihood estimates of the parameters of a two component Gaussian Mixture Model (GMM). The discretization and tying of several position states to the same observation model is based on the fact that the features do not change abruptly within the fraction of a mātrā. The observation probability hence is computed as,

Tāl	Structure			Dataset				
	#Mātrās	#Vibhāgs	Vibhāg structure	HMD_l	HMD_s	HMD_f	#Annotated mātrās	#Annotated sam
tīntāl	16	4	4,4,4,4	13	41	54	17142	1081
ēktāl	12	6	2,2,2,2,2,2	32	26	58	12999	1087
jhaptāl	10	4	2,3,2,3	6	13	19	3029	302
rūpak tāl	7	3	3,2,2	8	12	20	2841	406
Total	-	-	-	**59**	**92**	**151**	**36011**	**2876**

Table 1: The Hindustani music collection showing the number of excerpts in each subset - HMD_l , HMD_s and HMD_f . The number of sam and mātrā annotations in HMD_f are also shown. For each tāl, the table also shows the structure of the tāl with the number of mātrās and vibhāgs (sections) in each cycle. The fourth column shows the grouping of the mātrās in a cycle into vibhāgs, and the length of each vibhāg, e.g. each cycle of rūpak tāl shown in Figure 1 has three sections consisting of three, two, and two mātrās, respectively.

$$P(\mathbf{y}|\mathbf{x}) = P(\mathbf{y}|\phi, v) = \sum_{i=1}^{2} c_{\phi,v,i}\,\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_{\phi,v,i}, \boldsymbol{\Sigma}_{\phi,v,i}) \quad (6)$$

where, $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a normal distribution, and for each mixture component i, $c_{\phi,v,i}$, $\boldsymbol{\mu}_{\phi,v,i}$ and $\boldsymbol{\Sigma}_{\phi,v,i}$ are the component weight, mean (2-dimensional) and the covariance matrix (2×2), respectively. Hence, there is an observation GMM for every section and tied position states.

The described section pointer model is a generalization of the previously presented bar pointer model [17] - when a tāl is assumed to have only one section spanning the whole cycle ($V = 1$), we obtain the tracking model presented in [5–7]. Further, by changing the structure of the transition matrix $\mathbf{A}$ to include many tāl section transitions, the model can also be used for determining the type of meter as in [7], but this is beyond the scope of this paper.

3. INFERENCE

The goal of inference is to find a maximum *a posteriori* (MAP) sequence of latent variables ($\mathbf{x}^*_{1:K}$) that maximizes the posterior $P(\mathbf{x}_{1:K}|\mathbf{y}_{1:K})$, which can then be straightforwardly translated into a sequence of downbeat (sam) instants ($\phi^*_k = 0, v^*_k = 1$), section boundaries ($\phi^*_k = 0, v^*_k \neq v^*_{k-1}$), mātrā instants ($\phi^*_k = (i-1) \cdot {}^{M_v}/_{B_v}, i = 1, \ldots, B_v$), and the time varying instantaneous tempo ($\dot{\phi}^*_k$).

For inference on the proposed model, we use an approximate inference scheme called the Auxiliary Mixture Particle Filter (AMPF), which has been shown to be effective for meter tracking [6]. We briefly outline the AMPF, emphasizing on relevant aspects. A detailed description of the algorithm has been presented in [6] and an introduction to particle filtering can be obtained from [19]. It has been shown that this approximate inference scheme performs comparable to the exact inference using a hidden Markov model, while being computationally less demanding [6, 18].

In particle filters, the posterior density $P(\mathbf{x}_{1:K}|\mathbf{y}_{1:K})$ is approximated using a weighted set of points (called particles) in the state space as,

$$P(\mathbf{x}_{1:K}|\mathbf{y}_{1:K}) \approx \sum_{i=1}^{N_p} w_K^{(i)} \delta(\mathbf{x}_{1:K} - \mathbf{x}^{(i)}_{1:K}) \quad (7)$$

Here, $\{\mathbf{x}^{(i)}_{1:K}\}$ is a set of N_p number of points (particles) with associated weights $\{w_K^{(i)}\}$, $i = 1, \ldots, N_p$, and $\mathbf{x}_{1:K}$ is the set of all state trajectories until frame K, while $\delta(x)$ is the Dirac delta function.

Particle filters approximate the posterior pointwise, for which we need a suitable method to draw samples $\mathbf{x}_k^{(i)}$ and compute appropriate weights $w_k^{(i)}$ recursively at each time step. Using Sequential Importance Sampling (SIS) [19] and the transition probability as the proposal distribution as in [6], the particle weights can be recursively computed as,

$$w_k^{(i)} \propto w_{k-1}^{(i)} P(\mathbf{y}_k|\mathbf{x}_k^{(i)}) \quad (8)$$

The SIS algorithm derives samples by first sampling from a proposal, in this case the transition probability and then computes weights according to Eq. (8) using the observation model in Eq. (6). Once we determine the particle trajectories $\{\mathbf{x}^{(i)}_{1:K}\}$, we then select the trajectory $\mathbf{x}^{(i^*)}_{1:K}$ with the highest weight $w_K^{(i^*)}$ as the MAP state sequence.

The SIS algorithm has several limitations that need improvements to be practically useful. The degeneracy problem of the SIS causes only few particles to have non-zero weights after a few steps, a problem that can be overcome by resampling [19] - we use systematic resampling in this paper. In our problem, the posterior is highly multimodal due to multiple possible tempo hypotheses at each time, and we hence applied two additional extensions as in [6]. The first, the Auxiliary Particle Filter (APF) [20] manipulates weights during the resampling step, and the second, called the Mixture Particle Filter (MPF) [21], groups the particles into clusters. Each of these clusters can then, in principle, track a distinct mode in the posterior. The combination of APF and MPF called the AMPF, as proposed in [6], is what we apply here for inference as well.

4. DATASET

For the purpose of this study, we compiled a dataset (HMD_f) that consists of 151 two minute long excerpts of Hindustani music sampled from the CompMusic Hindustani music research corpus [22], a curated collection of commercial audio releases and metadata. The excerpts have a tabla accompaniment and span four popular tāls of Hindustani music. The dataset is described in Table 1 and consists of both vocal and instrumental recordings spanning different lay, artists, and stylistic schools. The audio is stereo and sampled at 44100 Hz. To the best of our knowledge, this is the first sizeable mātrā and sam annotated collection of Hindustani music, and all annotations are publicly available for download from the companion webpage.

For each audio excerpt, the annotations consist of editorial metadata about the tāl, as well as time-aligned metrical annotations of all mātrā and sam instances. The mātrā annotations are accompanied with the mātrā number in the cycle so that the vibhāg (section) boundaries can be easily obtained. The annotations were manually done using Sonic Visualizer [23] by tapping to the excerpt and then correcting them. All annotations were then verified by a professional Hindustani musician.

The dataset contains excerpts with a large tempo range of over five octaves from 10 MPM to 370 MPM, with cycle length varying between 2.3 s and 69.6 s. To study the effect of tempo class on the performance of meter tracking, we form two subsets of the full HMD_f dataset: all the vilaṁbit (slow) lay excerpts into a long cycle subset (HMD_l), and all the madhya (medium) and dṛt (fast) lay excerpts into a short cycle subset (HMD_s).

Tempo class	Subset	Method	Sam tracking					Mātrā tracking				
			tīntāl	ēktāl	jhaptāl	rūpak tāl	**Mean**	tīntāl	ēktāl	jhaptāl	rūpak tāl	**Mean**
Informed	HMD_l	bar	0.464	0.078	0.178	0.630	**0.234**	0.705	0.186	0.656	0.703	**0.406**
		section	0.696	0.161	0.256	0.681	**0.359**	0.793	0.268	0.736	0.799	**0.503**
	HMD_s	bar	0.768	0.935	0.874	0.671	**0.817**	0.916	0.966	0.921	0.806	**0.916**
		section	0.806	0.930	0.949	0.716	**0.850**	0.917	0.968	0.948	0.829	**0.924**
Not informed	HMD_l	bar	0.128	0.032	0.663	0.536	**0.186**	0.370	0.107	0.743	0.690	**0.309**
		section	0.098	0.034	0.691	0.463	**0.173**	0.340	0.118	0.821	0.695	**0.317**
	HMD_s	bar	0.725	0.868	0.932	0.665	**0.787**	0.905	0.936	0.938	0.798	**0.905**
		section	0.776	0.884	0.941	0.706	**0.820**	0.916	0.945	0.956	0.835	**0.919**

Table 2: F-measure values of meter tracking for all combinations of experimental setups, column-1: either using tempo class information in training or not, column-2: the subset (long, or short cycle pieces) used for testing, and column-3: either tracking the whole cycle as a unit (*bar*) - BPM, or in sections as proposed (*section*) - SPM. The column titled "**Mean**" shows the average performance over all the pieces of the subset. For each subset, the value underlined denotes a statistically significant improvement over the value of the other method (in a paired-sample t-test at 5% significance levels).

5. EXPERIMENTS

The experiments aim to compare the performance of meter tracking using bar length (BPM) and the proposed section length (SPM) patterns. The BPM applies the position variable ϕ to the whole tāl cycle, while the proposed SPM applies ϕ to the sections (vibhāg) and imposes a sequential structure as described in Section 2. Performance is monitored for short cycles (HMD_s) and long cycles (HMD_l) separately. Tracking is done for each type of meter (tāl) separately in a two fold cross validation experiment. We will further examine two cases, the tempo-informed case, in which only samples from the same lay group (subset) are used for training, and the uninformed case, in which samples from all lay groups are used for training.

5.1. Parameter learning and evaluation measures

The tempo ranges $[\dot{\phi}_{\text{min}}, \dot{\phi}_{\text{max}}]$ are learned from the training data of each fold, with an additional 20% margin for unseen data. For the SPM, the length of longest section, $M = 1600$, is set for the four mātrā long sections in tīntāl and other section lengths are scaled accordingly. The number of particles is set equal to $N_p = 1500 \cdot V$. For the BPM hence, since $V = 1$, $M = 1600$ corresponds to the whole of the longest cycle(tīntāl) and $N_p = 1500$. For the AMPF, we set $\sigma_n = 0.02$ and the maximum number clusters to 200. The other AMPF parameters are identical to the values used in [6].

A variety of measures are available for evaluating beat and downbeat tracking performance (see [24] for an overview). We chose the F-measure metric that is widely used in beat tracking evaluation. The F-measure (Fmeas) is a number between 0 and 1 computed from correctly detected beats, within a window of ± 70 ms, as the harmonic mean of the precision (the ratio between the number of correctly detected beats and all detected beats) and recall (the ratio between the number of correctly detected beats and the total annotated beats). This beat tracking definition extends to tracking the mātrās (m-Fmeas) and the downbeats/sams (s-Fmeas) as well, with the same tolerances. We tested with other beat evaluation measures applied in MIR, but they did not provide qualitatively different results. For evaluation in this paper, we used the code available at `http://code.soundsoftware.ac.uk/projects/beat-evaluation/` with default settings.

5.2. Results and Discussion

Table 2 depicts the obtained meter tracking results. Each number is the mean performance over three inference runs with the AMPF algorithm. Both the BPM and SPM have good tracking performance for shorter cycles (HMD_s) in the tempo class informed case, with high Fmeas for both sam and mātrā tracking. The proposed SPM performs significantly better than the BPM for most tāls in the HMD_s case.

On the longer cycles (HMD_l), the performance is generally lower, confirming that tracking longer cycles is more challenging than shorter cycles. The SPM provides a significant improvement over BPM in the tempo-informed case, an improvement that vanishes in the non-informed case. Furthermore, the results on HMD_l vary considerably across the tāls. For instance, the mean s-Fmeas in the informed case with SPM for ēktāl is 0.161 while for both tīntāl and rūpak tāl a much smaller decrease compared to HMD_s can be observed (*e.g.* 0.806 to 0.696 for tīntāl). This large disparity is caused, on the one hand, by the sections in ēktāl being all of identical length (two mātrā), which causes higher similarity between sections, and, on the other hand, the extremely low tempo of the vilaṁbit ēktāl pieces; the median tempo of vilaṁbit ēktāl pieces in 13.51 MPM, which means that the *fastest* defined pulsation at the mātrā level of the tāl occurs about every four seconds.

For the tempo uninformed case, the vilaṁbit (slow) pieces have an average s-Fmeas of 0.173 with the SPM, while on the faster non-vilaṁbit pieces 0.820 is reached. This can be compared to the s-Fmeas with SPM in the informed case of 0.359 and 0.850, for long and short cycles, respectively. Therefore, for the short cycles almost no advantage is obtained when providing tempo class information, while for long cycles tempo class information leads to a significant improvement (both for BPM and SPM). This also clearly shows the bias of the models towards higher tempi when the complete range of tempi is allowed. Since tempo class is provided as editorial metadata, it can be used to improve meter tracking performance.

6. CONCLUSIONS

We presented a generalized Bayesian model for meter tracking that enables an improved tracking in cycles of long durations. It was further shown that even in the case of shorter cycles, the results improved over a previously presented tracking model. We demonstrated that providing the information of the tempo class can improve the tracking accuracy especially for cycles of long duration. While the depicted accuracies for long cycles are still quite low, we presume that large improvement can be obtained by replacing the purely onset based observation model by a richer signal representation. Furthermore, in recordings that have two pieces in different tāls and lay, an automatic segmentation might be necessary before tracking. Such segmentation could be performed using, e.g. Bayesian change point detection [25], a problem that needs further exploration.

7. REFERENCES

[1] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, MIT Press Cambridge, 1983.

[2] A. 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, Nov. 2012.

[3] A. P. Klapuri, A. J. Eronen, and J. 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] G. Peeters and H. Papadopoulos, "Simultaneous beat and downbeat-tracking using a probabilistic framework: Theory and large-scale evaluation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1754–1769, 2011.

[5] F. Krebs, S. Böck, and G. Widmer, "Rhythmic pattern modeling for beat- and downbeat tracking in musical audio," in *Proc. of the 14th International Society for Music Information Retrieval (ISMIR) Conference*, Curitiba, Brazil, Nov. 2013, pp. 227–232.

[6] F. Krebs, A. Holzapfel, A.T. Cemgil, and G. Widmer, "Inferring metrical structure in music using particle filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 5, pp. 817–827, May 2015.

[7] A. Holzapfel, F. Krebs, and A. Srinivasamurthy, "Tracking the "odd": Meter inference in a culturally diverse music corpus," in *Proc. of the 15th International Society for Music Information Retrieval (ISMIR) Conference*, Taipei, Taiwan, 2014, pp. 425–430.

[8] S. Durand, J. P. Bello, B. David, and G. Richard, "Downbeat tracking with multiple features and deep neural networks," in *Proc. of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, 2015.

[9] E. Clarke, "Rhythm and timing in music," in *The Psychology of Music*, D. Deutsch, Ed., pp. 473–500. Academic Press, San Diego, 2 edition, 1999.

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

[11] S. Böck, F. Krebs, and G. Widmer, "A multi-model approach to beat tracking considering heterogeneous music styles," in *Proc. of the 15th International Society for Music Information Retrieval (ISMIR) Conference*, Taipei, Taiwan, 2014, pp. 602–607.

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

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

[14] S. Naimpalli, *Theory and practice of Tabla*, Popular Prakashan, 2005.

[15] A. Srinivasamurthy, A. Holzapfel, and X. 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.

[16] S. Gulati, V. Rao, and P. Rao, "Meter detection from audio for indian music," in *Speech, Sound and Music Processing: Embracing Research in India. Lecture Notes in Computer Science, vol. 7172*, S. Ystad, M. Aramaki, R. Kronland-Martinet, K. Jensen, and S. Mohanty, Eds., pp. 34–43. Springer: Berlin Heidelberg, 2012.

[17] N. Whiteley, A. T. Cemgil, and S. Godsill, "Sequential inference of rhythmic structure in musical audio," in *Proc. of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, USA, Apr. 2007, vol. 4, pp. 1321–1325.

[18] A. Srinivasamurthy, A. Holzapfel, A. T. Cemgil, and X. Serra, "Particle Filters for Efficient Meter Tracking with Dynamic Bayesian Networks," in *Proc. of the 16th International Society for Music Information Retrieval (ISMIR) Conference*, Malaga, Spain, Oct. 2015, pp. 197–203.

[19] A. Doucet and A. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, 2009.

[20] A. Johansen and A. Doucet, "A note on auxiliary particle filters," *Statistics and Probability Letters*, vol. 78, no. 12, pp. 1498–1504, 2008.

[21] J. Vermaak, A. Doucet, and P. Pérez, "Maintaining multimodality through mixture tracking," in *Proc. of the 9th IEEE International Conference on Computer Vision*, Nice, France, Oct. 2003, pp. 1110–1116.

[22] A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, and X. Serra, "Corpora for Music Information Research in Indian Art Music," in *Proc. of Joint International Computer Music Conference/Sound and Music Computing Conference*, Greece, Sept. 2014.

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

[24] M. E. P. Davies, N. Degara, and M. D. Plumbley, "Evaluation methods for musical audio beat tracking algorithms," *Queen Mary University of London, Technical Report C4DM-09-06*, 2009.

[25] D. Barber, A. T. Cemgil, and S. Chiappa, *Bayesian time series models*, Cambridge University Press, 2011.