INFORMED AUTOMATIC METER ANALYSIS OF MUSIC RECORDINGS

Ajay Srinivasamurthy[*] ajays.murthy@upf.edu

Andre Holzapfel[†] holzap@kth.se

Xavier Serra[*] xavier.serra@upf.edu

[*]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

[†]Media Technology and Interaction Design Department, KTH Royal Institute of Technology, Stockholm, Sweden

ABSTRACT

Automatic meter analysis aims to annotate a recording of a metered piece of music with its metrical structure. This analysis subsumes correct estimation of the type of meter, the tempo, and the alignment of the metrical structure with the music signal. Recently, Bayesian models have been successfully applied to several of meter analysis tasks, but depending on the musical context, meter analysis still poses significant challenges. In this paper, we investigate if there are benefits to automatic meter analysis from additional *a priori* information about the metrical structure of music. We explore informed automatic meter analysis, in which varying levels of prior information about the metrical structure of the music piece is available to analysis algorithms. We formulate different informed meter analysis tasks and discuss their practical applications, with a focus on Indian art music. We then adapt state of the art Bayesian meter analysis methods to these tasks and evaluate them on corpora of Indian art music. The experiments show that the use of additional information aids meter analysis and improves automatic meter analysis performance, with significant gains for analysis of downbeats.

1. INTRODUCTION

Automatic meter analysis of a music recording aims at determining different components of its metrical structure such as the type of meter, the tempo, the beats and downbeats. It is an important Music Information Research (MIR) task that provides useful musically relevant metadata not only for enriched listening, but also for preprocessing of music for several higher level tasks such as section segmentation, structural analysis and defining rhythm similarity measures. Initial approaches to meter analysis explored individual tasks of meter analysis, such as tempo estimation [8,9], beat tracking [5,13], time signature estimation [15] and downbeat tracking [10,14]. Recent approaches consider a joint estimation of several of these components and have successfully applied Bayesian models to jointly estimate beat and downbeats using rhythmic patterns learned from onset detection features [1, 11, 12]. Recent interest has also been to explore neural networks for beat and downbeat tracking with several musically inspired features and network topologies [7]. Despite the recent success, meter analysis still poses significant challenges depending on the musical context [18, 20].

In this paper, we investigate the potential to improve meter analysis methods by providing them with additional prior information about the underlying metrical structure. This is a research problem we define as *informed meter analysis*, referring to a class of analysis tasks that utilize some form of additional information about the underlying metrical structure of the piece. Apart from building meter-aware analysis methods, informed meter analysis is motivated by its potential applications and the need for improved meter analysis performance. It is hypothesized that information available as metadata or obtainable from an expert user can be effectively utilized to significantly improve meter analysis performance. Such informed approaches can help to establish a focus in the space of possible solutions by the incorporation of *a priori* information, supporting meter analysis especially in the context of computationally challenging samples. Some informed meter analysis tasks have been studied before, such as the task of downbeat tracking from a set of known beats [10]. However, there has been no formal treatment of the problem, which is the focus of this paper.

Carnatic and Hindustani music are Indian Art Music (IAM) traditions from Southern and Northern parts of the Indian subcontinent, respectively. Both these musics have a long history of performance and continue to thrive in current sociocultural contexts. While the two musics differ in performance practices, they share similar melodic and rhythmic concepts. The rhythmic framework is based on cyclic metrical structures called the tāḷa in Carnatic music (CM) or tāl in Hindustani music (HM), which provide a broad structure for repetition of music phrases, motifs and improvisations. A cycle of a tāḷa (or tāl) is divided into isochronous beats (called the mātrā in HM), which are grouped into possibly unequal length sections. The beginning of a cycle (the downbeat) is referred to as sama (sam in HM). Given the central importance of tāḷa in defining rhythmic structures, meter analysis in the context of IAM aims to time-align and tag a music recording with tāḷa related events and metadata. Clayton [3] and Sambamoorthy [16] provide an in depth discussion of rhythm in Hindustani and Carnatic music, respectively.

With significant improvisation and expressive timing, a wide range of tempo and cycles as long as a minute, IAM has been shown to pose several challenges to automatic meter analysis [20]. Further, large and continuously growing archives of IAM are available with varying amounts of tāḷa related metadata [17]. In this paper, we use corpora of IAM as a challenging case to explore the potential of informed meter analysis, and include a set of Ballroom dances to enable comparison with other styles.

2. INFORMED METER ANALYSIS

Different kinds of prior information about the underlying metrical structure can be made available to analysis algorithms. In the following subsections, we describe specific informed analysis tasks and emphasize different practical scenarios for each task. At the outset, we assume that some basic information about the music piece is available for all informed analysis tasks. We assume that the music tradition is known, and that the rhythm class (tāḷa) of the piece is from a set of known (from musicological literature) tāḷas. Further, we assume we know the range of tempo generally used in a music culture. A piece of IAM is performed in a single tāḷa (rare exceptions exist, but outside the scope of regular performance practice) and most commercial releases are segmented so that an audio recording is a single piece. However, there are cases when an entire concert or parts of concert with multiple pieces (and hence possibly different tāḷas) are stored in a single audio recording. We assume that such a recording has been segmented into individual pieces of music with a single tāḷa. The case of change of tāḷas within a recording is not addressed.

Finally, for better readability, we use the commonly used terminology of tempo, beats and downbeats in the paper, while we carefully note that the equivalence of these terms across different music cultures cannot be assumed.

2.1 Meter Inference (`Inference`)

Meter inference aims for a complete meter analysis of a recording starting with no prior information. Given an audio music recording, meter inference task aims to estimate the rhythm class (or meter type or tāḷa), time-varying tempo, beats and downbeats. Meter inference in IAM aims to recognize the tāḷa/tāl, estimate the time varying tempo (measured as the inter beat interval), the beat and the sama/sam (downbeat) locations. It is the least informed and most difficult task owing to the large range of tempi and different tāḷas. While meter inference is the only applicable task with unlabeled collections of music, it is often the case that some tāḷa related information is available or can be inferred, e.g. from the editorial metadata of a music piece. Most of commercially released music in both Carnatic and Hindustani music has the name of the tāḷa in editorial metadata. Even within a live concert, the musician often announces the tāḷa of a piece and hence tāḷa recognition is a redundant task. However, meter inference can be used as a baseline task to understand the complexity of uninformed meter analysis.

2.2 Meter Tracking (`Track`)

Given an audio music recording and its rhythm class (or meter type or tāḷa), meter tracking aims to estimate the time varying tempo, beat and downbeat locations. Meter tracking in IAM aims to track the time varying tempo, beats and the sama from an audio music recording, given the tāḷa. Assuming that the tāḷa, and hence the metrical structure is known in advance is a fair and practical assumption making meter tracking the most relevant meter analysis task for IAM.

2.3 Informed Meter Tracking

Informed meter tracking is meter tracking in which some additional information apart from the tāḷa is available. The additional information could be in the form of a tempo range, a few instances of beats and downbeats annotated, or even partially tracked metrical cycles. The additional metadata could come from manual annotation or as an output of other automatic algorithms, e.g. the median tempo of a piece can be obtained from a standalone tempo estimation algorithm, or some melodic analysis algorithms might output (with a high probability) some beats/downbeats as a byproduct.

From a practical standpoint, while it is prohibitively resource intensive to manually annotate all the beats and downbeats of a large music collection, it might be possible to seed the meter tracking algorithms with the first few beats and downbeats. For a musician or even an expert listener, it would be easy to tap some instances of the beat and sama/downbeats, which could then be used to automatically track meter in the whole recording. In specific, we explore three variants of informed meter tracking, with varying levels of available information:

Sama-informed meter tracking (`SI-Track`**)** task in which a few instances of sama/downbeat of the piece are provided as an additional input to the meter tracking algorithm. An example downbeat is expected to help the algorithm to better align the audio to the underlying meter. We only explore the use of first downbeat of the piece, without any knowledge of tempo.

Tempo-informed meter tracking (`TI-Track`**)** task in which the median tempo (or a narrow range of tempi) of the piece is provided as an additional input to the meter tracking algorithm. Providing the median tempo is hypothesized to help reduce metrical level errors - tracking the metrical cycles at the correct metrical level instead of tracking half and double cycles. The median tempo can be obtained manually or through other automatic tempo estimation algorithms [8, 22].

Sama-Tempo-informed meter tracking (`STI-Track`**)** task in which the median tempo and a few downbeat locations in the excerpt are provided as additional inputs to the meter tracking algorithm. We only explore the use of median tempo value and the first downbeat of the music piece provided to the meter tracking algorithm.

The informed meter tracking tasks formulated in this section are relevant and designed to require minimal human effort to provide the necessary additional information. In a best case scenario, the most informed `STI-Track` task can be applied to a music piece by listening to just the first few

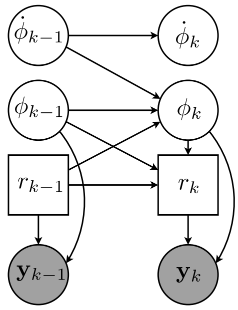

Figure 1: The bar pointer model for meter analysis. The circles and squares denote continuous and discrete variables, respectively. Grey nodes and white nodes represent observed and latent variables, respectively.

seconds of the piece and marking two consecutive downbeats. An estimate of the initial tempo can be obtained the two downbeats and used by the analysis algorithm. Finally, the various tasks were described using terminology of IAM, but they are applicable to any music with hierarchical metrical structures that can be described with beats, downbeats and rhythm patterns.

3. METER ANALYSIS MODEL

To compare different informed analysis tasks, we use and adapt a state of the art Dynamic Bayesian Network (DBN). Referred to as bar pointer model (BP-model) [23], has been successfully applied for meter analysis in different music cultures [11, 19]. We describe the model briefly while a detailed description is presented in [12]. We then explain how it can be adapted to the informed analysis tasks.

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 hidden (latent) variables $\mathbf{x}_{1:K} = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$, where K is the length of the feature sequence (number of audio frames). The joint probability distribution of hidden and observed variables factorizes as,

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

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. The structure of the BP-model in Figure 1 shows the conditional dependence relations between the variables.

3.1 Hidden Variables

At each audio frame k, the hidden variable vector $\mathbf{x}_k$ describes the state of a hypothetical bar pointer $\mathbf{x}_k = [\phi_k \;\; \dot{\phi}_k \;\; r_k]$, representing the bar position, instantaneous tempo and a rhythmic pattern indicator, respectively.

Rhythmic pattern indicator: The rhythmic pattern variable $r \in \{1, \ldots, R\}$ is an indicator variable to select one of the R observation models corresponding to each bar (cycle) length rhythmic pattern of a rhythm class that are learned from training data. Each pattern r corresponds to a rhythm class (or meter type or tāḷa) and has an associated length of cycle M_r and number of beat (or mātrā) pulses B_r.

Bar position: The bar position $\phi \in [0, M_r)$ variable tracks the progression through the bar and indicates a position in the bar at any audio frame. The variable traverses the whole bar and wraps around to zero at the end of the bar to track the next bar.

Instantaneous tempo: Instantaneous tempo $\dot{\phi}$ is the rate at which the bar position variable progresses through the bar at each frame, measured in bar positions per time frame.

3.2 Transition and Observation Model

The initial state distribution $P(\mathbf{x}_0)$ can be used to incorporate prior information about the metrical structure of the music into the model. Given the conditional dependence relations in Figure 1, the transition model factorizes as,

$$P(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = P(\phi_k \mid \phi_{k-1}, \dot{\phi}_{k-1}, r_{k-1}) P(\dot{\phi}_k \mid \dot{\phi}_{k-1}) \\ P(r_k \mid r_{k-1}, \phi_k, \phi_{k-1}) \quad (1)$$

The individual terms of the equation can be expanded as,

$$P(\phi_k \mid \phi_{k-1}, \dot{\phi}_{k-1}, r_{k-1}) = \mathbb{1}_{\phi} \quad (2)$$

where $\mathbb{1}_{\phi}$ is an indicator function that takes a value of one if $\phi_k = (\phi_{k-1} + \dot{\phi}_{k-1}) \bmod(M_{r_{k-1}})$ and zero otherwise. The tempo transition is given by,

$$P(\dot{\phi}_k \mid \dot{\phi}_{k-1}) \propto \mathcal{N}(\dot{\phi}_{k-1}, \sigma^2_{\dot{\phi}_k}) \times \mathbb{1}_{\dot{\phi}} \quad (3)$$

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 $\sigma_{\dot{\phi}_k}$ depends on the value of tempo to allow for larger tempo variations at higher tempi. We set $\sigma_{\dot{\phi}_k} = \sigma_n \cdot \dot{\phi}_{k-1}$, where σ_n $(= 0.02)$ is a user parameter that controls the amount of local tempo variations we allow in the music piece.

The transition probability of pattern indicator variable $P(r_k \mid r_{k-1}, \phi_k, \phi_{k-1})$ is governed by $\mathbb{A}$, a $R \times R$ time-homogeneous transition matrix where $\mathbb{A}(i, j)$ is the transition probability from r_i to r_j. However, since the rhythmic patterns are one bar (cycle) in length, pattern transitions are allowed only at the end of the bar ($\phi_k < \phi_{k-1}$).

The observation model is identical to the one used in [12], and depends only on the bar position and rhythmic pattern variables, without any influence from tempo. To model rhythm patterns, we compute spectral flux feature from audio in two frequency bands (Low: ≤ 250 Hz, High: > 250 Hz). Using beat and downbeat annotated training data, the audio features are grouped into bar length patterns on a bar discretized into 64$^{\text{th}}$ note cells. A k-means clustering algorithm then assigns each bar of the dataset to one of the R rhythmic patterns. All the features within the cell of each pattern are collected and maximum likelihood estimates of the parameters of a two component Gaussian Mixture Model (GMM) are obtained. The observation probability within a 64$^{\text{th}}$ note cell is assumed to be constant and is computed as,

$$P(\mathbf{y} \mid \mathbf{x}) = P(\mathbf{y} \mid \phi, r) = \sum_{i=1}^{2} \pi_{\phi,r,i}\, \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_{\phi,r,i}, \boldsymbol{\Sigma}_{\phi,r,i})$$

where, $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a normal distribution of the two dimensional feature $\mathbf{y}$. For the mixture component i, $\pi_{\phi,r,i}$, $\boldsymbol{\mu}_{\phi,r,i}$ and $\boldsymbol{\Sigma}_{\phi,r,i}$ are the component weight, mean (2-dim.) and the covariance matrix (2×2), respectively.

3.3 Inference in BP-model

The goal of inference with the BP-model is to estimate a hidden variable sequence $\mathbf{x}^*_{1:K}$ that maximizes the posterior probability $P(\mathbf{x}_{1:K} \mid \mathbf{y}_{1:K})$ given an observed sequence of features $\mathbf{y}_{1:K}$. The sequence $\mathbf{x}^*_{1:K}$ can then be translated into a sequence of downbeat (sama) instants ($t^*_k \mid \phi^*_k = 0$), beat instants ($t^*_k \mid \phi^*_k = i \cdot {}^{M_{r^*}}/_{B_{r^*}}$, $i = 1, \ldots, B_r$), local instantaneous tempo ($\dot{\phi}^*_k$) and rhythmic patterns (r^*).

In this paper, we use an approximate particle filter [6] based inference scheme called the Auxiliary Mixture Particle Filter (AMPF), which has been shown to be effective for meter analysis [12]. In a particle filter, the posterior is estimated pointwise by approximating it using a weighted set of points (known as particles) in the state space as,

$$P(\mathbf{x}_{1:K} \mid \mathbf{y}_{1:K}) \approx \sum_{i=1}^{N_p} w_K^{(i)} \delta(\mathbf{x}_{1:K} - \mathbf{x}_{1:K}^{(i)}) \qquad (4)$$

Here, $\{\mathbf{x}_{1:K}^{(i)}\}$ is a set of points (particles) with associated weights $\{w_K^{(i)}\}$, $i = 1, \ldots, N_p$, $\mathbf{x}_{1:K}$ is the set of all state trajectories until frame K, $\delta(x)$ is the Dirac delta function, and N_p is the number of particles. The AMPF algorithm includes several enhancements to make it suitable for inference with the BP-model, a detailed description of which has been presented in [12].

3.4 BP-model and AMPF for Informed Meter Analysis

The AMPF algorithm on the BP-model is generic and can be adapted to be applicable to the informed meter analysis tasks described in Section 2. For meter inference, the rhythm class (tāḷa) can be estimated by allowing rhythmic patterns of different lengths from different rhythm classes to be present in the model, as used by [12]. For meter tracking tasks, we assume that the rhythm class is known and all rhythm patterns belong to that class, i.e. $M_r = M$ and $B_r = B \,\forall\, r$.

The initial state distribution $P(\mathbf{x}_0)$ and the initialization of the particle filter system are modified to suit the informed meter tracking tasks. A uniform initialization over all allowed states is used for `Inference` and `Track` tasks, while a narrower informed initialization is done for informed meter tracking. For `TI-Track` task, we use the median ground truth tempo of the music piece being tracked and initialize the tempo variable $\dot{\phi}$ within a tight bound allowing for 10% variation in tempo around the median value. This enables the tracking algorithm to restrict the tempo variable within the tight tempo range and track the correct tempo at the right metrical level. For `SI-Track` task, the provided sama instance is used to initialize the bar position variable ϕ to zero at the related time position. For `STI-Track` task, both the tempo and bar position variables are initialized appropriately using the given information. The tracking algorithm hence gets the tempo and the beginning of the cycle in the piece, tracking the remaining beats and downbeats.

4. EXPERIMENTS

The experiments aim to compare performance across different informed meter analysis tasks and investigate the advantage of the additional prior information they utilize. While the focus of experiments is on Indian music, we also report the results on a collection of Ballroom dances to evaluate the extensibility of the informed analysis tasks. Furthermore, reproducibility will be ensured by providing free access for research purposes to all code repositories and datasets on the companion webpage, which also provides additional resources and music examples. [1]

Dataset	#Pieces	#Ann.	#Sama
CMR	118	28725	5560
HMR_s	92	32731	2572
HMR_l	59	3280	304
Total (IAM)	269	64736	8436

Table 1: The Carnatic (CMR) and Hindustani ($\text{HMR}_\text{l}$and HMR_s) music datasets showing the number of pieces, sama and beat/mātrā annotations.

4.1 Music Datasets

For the experiments, we use rhythm annotated datasets of Carnatic and Hindustani music (described in Table 1) that have been previously used for evaluating automatic meter analysis algorithms. The Carnatic music rhythm dataset (CMR dataset) [19] includes 118 two minute long excerpts of Carnatic music sampled from commercial releases. The recordings span four commonly used tāḷas with different number of beats in a cycle, with a total duration of 236 minutes. The dataset consists of audio, manually annotated time-aligned markers indicating the progression through the tāḷa cycle, and the associated tāḷa related metadata.

The Hindustani music rhythm dataset consists of 151 two minute long excerpts of Hindustani music sampled from the CompMusic Hindustani music research corpus [21], a curated collection of commercial audio releases and metadata. The excerpts span four popular tāls of Hindustani music that are structurally different and of different lengths. For each audio excerpt, the annotations consist of editorial metadata about the tāl, as well as time-aligned metrical annotations of all beat and sam instances.

The dataset consists of excerpts with a wide tempo range from 10 MPM (mātrās per minute) to 370 MPM. Hindustani music divides tempo into three main tempo classes (lay). Since no exact tempo ranges are defined for these classes, we determined suitable ranges 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 (dr̥t) tempi, respectively. The tempo class of a piece has a significant effect on meter analysis due to the wide range of possible tempi. To study any effects of the tempo class, the full Hindustani dataset is divided into two other subsets - the long cycle duration subset called the HMR_l dataset consisting of vilaṁbit pieces and the short cycle duration subset HMR_s dataset with madhya and the dr̥t lay pieces. The complete collection of Carnatic and Hindustani music datasets together is called IAM dataset.

[1] http://compmusic.upf.edu/informed-meter-tracking

In addition to Indian art music, we evaluate the tasks on a set of Ballroom dances, which includes beat and bar annotations of audio recordings of several dance styles sourced from `BallroomDancers.com` [9, 11]. The ballroom dataset contains eight different dance styles (Cha cha, Jive, Quickstep, Rumba, Samba, Tango, Viennese Waltz, and (slow) Waltz) and has been widely used for several MIR tasks such as genre classification, tempo tracking, beat and downbeat tracking [1, 9, 12]. It consists of 698 thirty second long audio excerpts and has tempo and dance style annotations. The dataset contains two different meters (3/4 and 4/4) and all pieces have constant meter.

4.2 Evaluation Measures

We evaluate the tasks through the relevant meter components they estimate - meter type, tempo, beats and downbeats. We evaluate only the applicable components that are not assumed to be known *a priori* in an informed task (e.g. meter type is known in `Track` task and hence only tempo, beats and downbeats are evaluated).

A variety of measures are available for evaluating beat and downbeat tracking [4]. We chose the f-measure ($\mathfrak{f}$) metric that is widely used in beat tracking evaluation. Other measures were applied in addition during the experiments, but did not add further detail and hence are not reported. It is a number between 0 and 1 computed from estimated and ground truth annotation sequences as the harmonic mean of the precision and recall measures. The definition extends to tracking both the beat/mātrās ($\mathfrak{f}_b$) and the downbeats/samas ($\mathfrak{f}_s$). For `Inference` and `Track` tasks, we additionally report the results of median tempo estimation, comparing the median estimated tempo and the median annotated ground truth tempo with a 5% error margin. For `Inference` task, the algorithms also detect the rhythm class (or tāḷa) and hence the accuracy of this detection is also reported.

4.3 Experimental Setup

Experiments are done separately on each of the three IAM datasets (CMR, HMR_s, HMR_l) and the Ballroom dataset. To compute the f-measure in CMR, HMR_s and Ballroom datasets, an error tolerance window of 70 ms is used between the annotation and the estimated beat/sama. The computation of f-measure with HMR_l dataset is an exception, where a bigger margin window is allowed. Since cycles are of long duration in HMR_l dataset and current evaluation approaches were not designed with such long cycles in mind, an error tolerance window of 70 ms is very tight. To account for the length of the cycle in the error margin, a 6.25% median inter annotation interval is used as the tolerance window, as used in many other beat tracking evaluations (e.g. by [10]). This choice of a larger allowance window also corroborates well with the observation that in vilaṁbit pieces of the HMR_l dataset, there can be significant freedom in pulsation and that larger timing deviations go unnoticed since the pieces are not rhythmically dense. It can be argued that the beat pulsation in vilaṁbit pieces is beyond the duration of what is called the perceptual present [2], and can therefore not be considered to belong to metrical structure. However, it is to be noted that the allowance used in this paper is a compromise and better evaluation measures that can handle these complexities are to be developed.

Dataset	CMR	HMR_s	HMR_l	IAM	Ballroom
Accuracy	68	63	27	57	89

Table 2: Tāḷa recognition accuracy (%) in `Inference` task. Time signature recognition accuracy is reported for Ballroom dataset.

The tempo ranges for initialization of AMPF in `Inference`, `Track` and `SI-Track` tasks are learned from training data of each fold and an additional 20% margin is added to extend to unseen data. However, if the learned ranges are beyond the minimum and maximum tempo limits of each music culture, we set it to the minimum or the maximum. We use one rhythmic pattern per tāḷa (or dance style). Hence, we use $R = 1$ for meter tracking, when a known meter is being tracked, while $R = 4$ (8 in Ballroom dataset) is used for meter inference, with one pattern per tāḷa/rhythm. We use the number of bar positions, $M_r = 1600$ for the longest rhythmic pattern we encounter in the dataset and scale all other pattern lengths accordingly. For the AMPF algorithm, we use 1500 particles per rhythm pattern, with other parameters identical to those used in [12]. A hop size of 20 ms is used to compute the two dimensional spectral flux feature.

4.4 Results and Discussion

The results in Table 2 and Figures 2-3 summarize the performance across different datasets and informed analysis tasks. All results are reported as the mean performance over three runs in a 2-fold (equal size) cross validation experiment on each dataset. The results are presented for each dataset as an average over the pieces in all the tāḷas (or rhythm classes). Table 2 shows the tāḷa recognition accuracy for the Indian music datasets (and time signature estimation accuracy for Ballroom dataset) from the `Inference` task. Figure 3 shows the median tempo estimation accuracy for different datasets in the `Inference`, `Track`, and `SI-Track` tasks, where median tempo is not known *a priori*. The beat and downbeat f-measure values are reported for all the informed analysis tasks in Figure 2. We use a paired-sample t-test to assess statistically significant differences in beat and downbeat tracking performance by pooling the results of Indian music datasets.

Table 2 shows a similar performance with the CMR and HMR_s datasets, but is significantly poor for the long cycle subset of Hindustani music (HMR_l dataset). Whereas in the Carnatic and Hindustani music datasets, each tāḷa has a distinct length, the eight rhythm classes in the Ballroom data are assigned to only two time signatures reducing the task to a classification task between 3/4 and 4/4 time signatures. Ballroom dataset hence shows the best recognition performance.

Tāḷa recognition accuracy affects tempo estimation, as seen in Figure 3 with a poor tempo estimation performance within the HMR_l dataset. Median tempo estimation accuracy is similar for CMR and HMR_s datasets. Tempo es-

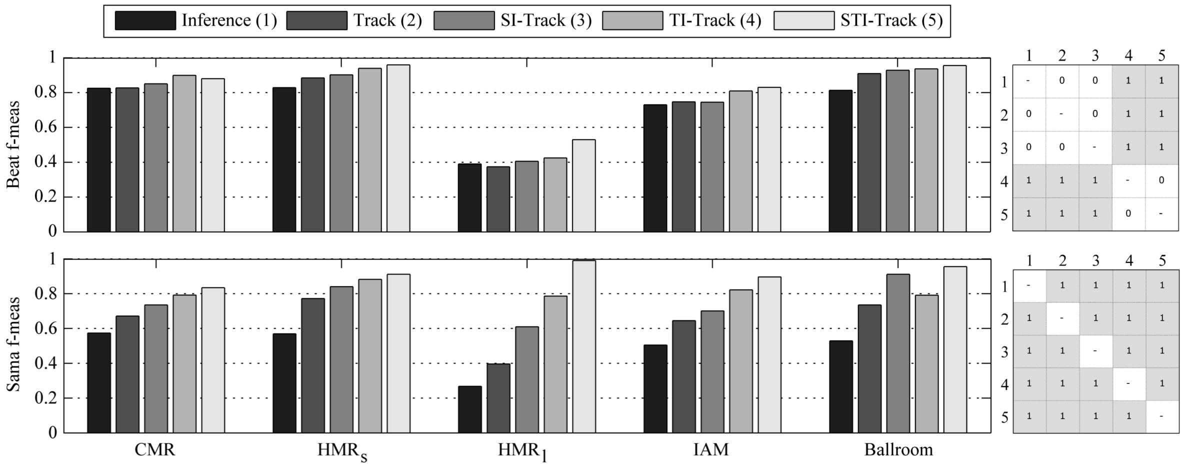

Figure 2: Beat and sama (downbeat) tracking results showing the f-measure as bar plots for different datasets and informed analysis tasks. The matrix on the right shows the results of a significance test between analysis tasks (numbers 1-5 correspond to tasks in the legend) for the IAM dataset. A box with numeral 1 indicates a statistically significant difference in a paired-sample t-test (at $p = 0.05$) while numeral 0 indicates a difference that is not statistically significant.

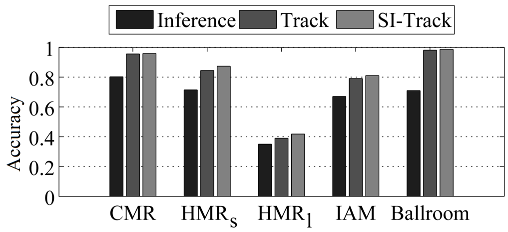

Figure 3: Median tempo estimation accuracy in the `Inference`, `Track` and `SI-Track` tasks.

timation accuracy improves for `Track` task compared to `Inference` task, showing the utility of knowing the meter type in estimating the correct tempo. However, additional downbeat information in `SI-Track` task does not add much to tempo estimation, with marginal or no further improvement. Ballroom dataset shows the best tempo estimation performance except for `Inference` task, where wrong estimations of the rhythm class leads to poorer tempo estimation.

The beat f-measure ($\mathfrak{f}_b$) results in Figure 2 across different informed analysis tasks shows a marginal improvement with informed tracking tasks, but statistically significant improvements are observed only with `TI-Track` and `STI-Track` tasks for IAM datasets, when median tempo is known *a priori*. This shows that the tempo information is more relevant than tāḷa and sama information to improve beat tracking performance for Indian art music. The biggest gains in informed meter analysis are seen in sama f-measure ($\mathfrak{f}_s$), with significant improvements achieved with more informed analysis tasks. For the pooled IAM dataset, starting with a $\mathfrak{f}_s = 0.51$ with `Inference` task, `STI-Track` task achieves $\mathfrak{f}_s = 0.82$, showing the benefit and the utility of both tempo and sama information in informed meter analysis for a more difficult task of downbeat estimation.

For Ballroom dataset, compared to the `Track` task, we observe that downbeat tracking performance for `SI-Track` improves more over `TI-Track` task. This indicates that downbeat information is more important than tempo information. It is perhaps due to the fact that Ballroom dances have a stable tempo and clear repeated rhythmic patterns. Accurate tempo estimation is achieved even without prior tempo information (Figure 3), and hence downbeat information is more useful.

A comparison of performance across datasets shows that CMR, HMR_s and Ballroom datasets have similar trends of improvement in both beat and sama (downbeat) tracking with informed tracking tasks. The largest gains however are obtained with the long cycle HMR_l dataset, which improves from a poor $\mathfrak{f}_s = 0.26$ (`Inference`) to $\mathfrak{f}_s = 0.99$ (`STI-Track`). While we note that a larger error margin and fewer sama examples in the long cycle dataset contribute to this high performance, the overall results considering all datasets and tasks conclude that the use of tempo and sama information enhances the capabilities of automatic meter analysis algorithms to track downbeats.

5. CONCLUSIONS

Starting with a hypothesis that automatic meter analysis performance can be improved by utilizing additional information about meter or tempo of a piece, we formulated relevant informed meter analysis tasks that can incorporate varying levels of prior information about the meter type, tempo and downbeat position. An evaluation on corpora of Indian art music and Ballroom dances showed the utility of prior information for automatic meter analysis, where tempo information is useful for beat tracking and the tempo and downbeat information was shown to be useful for downbeat tracking. We also showed that with minimal effort by a potential user of an annotation system, a high accuracy in tempo, beat and downbeat estimation can be achieved through informed meter analysis algorithms. Evaluation of informed analysis tasks in the paper was done through individual components of meter (tempo, beat, downbeat). In future work, we plan to develop unified meter analysis evaluation measures that take into account the hierarchical structure of musical meter.

6. ACKNOWLEDGMENTS

This work is partly supported by the European Research Council as part of the CompMusic project (ERC grant agreement 267583). Ajay Srinivasamurthy is currently with Idiap Research Institute, Martigny, Switzerland.

7. REFERENCES

[1] S. Böck, F. Krebs, and G. Widmer. A Multi-model Approach to Beat Tracking Considering Heterogenous Music Styles. In *Proc. of the 15th Intl. Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 602–607, Taipei, Taiwan, October 2014.

[2] E. Clarke. Rhythm and timing in music. In Diana Deutsch, editor, *The Psychology of Music*, pages 473–500. Academic Press, San Diego, II edition, 1999.

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

[4] M. E. P. Davies, N. Degara, and M. D. Plumbley. Evaluation Methods for Musical Audio Beat Tracking Algorithms. *Technical Report C4DM-TR-09-06, Queen Mary University of London*, October 2009.

[5] M. E. P. Davies and M. D. Plumbley. Context-Dependent Beat Tracking of Musical Audio. *IEEE Trans. on Audio, Speech, and Language Processing*, 15(3):1009–1020, March 2007.

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

[7] 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 Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)*, Brisbane, Australia, May 2015.

[8] D. Ellis. Beat Tracking by Dynamic Programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[9] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Trans. on Audio, Speech and Language Processing*, 14(5):1832–1844, 2006.

[10] J. Hockman, M. E. P. Davies, and I. Fujinaga. One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass. In *Proc. of the 13th Intl. Society for Music Information Retrieval Conference (ISMIR 2012)*, pages 169–174, Porto, Portugal, October 2012.

[11] F. Krebs, S. Böck, and G. Widmer. Rhythmic Pattern Modeling for Beat- and Downbeat Tracking in Musical Audio. In *Proc. of the 14th Intl. Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 227–232, Curitiba, Brazil, November 2013.

[12] F. Krebs, A. Holzapfel, A. T. Cemgil, and G. Widmer. Inferring Metrical Structure in Music Using Particle Filters. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, 23(5):817–827, May 2015.

[13] 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*, 36(1):1–16, 2007.

[14] G. Peeters and H. Papadopoulos. Simultaneous Beat and Downbeat-Tracking Using a Probabilistic Framework: Theory and Large-Scale Evaluation. *IEEE Trans. on Audio, Speech and Language Processing*, 19(6):1754–1769, 2011.

[15] A. Pikrakis, I. Antonopoulos, and S. Theodoridis. Music meter and tempo tracking from raw polyphonic audio. In *Proc. of the 5th Intl. Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, October 2004.

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

[17] X. Serra. Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project. In *Proc. of the 53rd AES Intl. Conference on Semantic Audio*, London, January 2014.

[18] A. Srinivasamurthy. *A Data-driven Bayesian Approach to Automatic Rhythm Analysis of Indian Art Music*. Doctoral dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2016.

[19] A. Srinivasamurthy, A. Holzapfel, Ali Taylan Cemgil, and X. Serra. Particle Filters for Efficient Meter Tracking with Dynamic Bayesian Networks. In *Proc. of the 16th Intl. Society for Music Information Retrieval Conference (ISMIR 2015)*, pages 197–203, Malaga, Spain, October 2015.

[20] 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*, 43(1):97–117, 2014.

[21] A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, and X. Serra. Corpora for Music Information Research in Indian Art Music. In *Proc. of 13th Sound and Music Computing Conference*, pages 1029–1036, Athens, Greece, September 2014.

[22] A. Srinivasamurthy and X. Serra. A Supervised Approach to Hierarchical Metrical Cycle Tracking from Audio Music Recordings. In *Proc. of the 39th IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, pages 5237–5241, Florence, Italy, May 2014.

[23] N. Whiteley, A. T. Cemgil, and S. Godsill. Sequential Inference of Rhythmic Structure in Musical Audio. In *Proc. of the IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, volume 4, pages 1321–1325, Honolulu, USA, April 2007.