

TRACKING THE “ODD”: METER INFERENCE IN A CULTURALLY DIVERSE MUSIC CORPUS

Andre Holzapfel
New York University Abu Dhabi
andre@rhythmos.org

Florian Krebs
Johannes Kepler University
Florian.Krebs@jku.at

Ajay Srinivasamurthy
Universitat Pompeu Fabra
ajays.murthy@upf.edu

ABSTRACT

In this paper, we approach the tasks of beat tracking, down-beat recognition and rhythmic style classification in non-Western music. Our approach is based on a Bayesian model, which infers tempo, downbeats and rhythmic style, from an audio signal. The model can be automatically adapted to rhythmic styles and time signatures. For evaluation, we compiled and annotated a music corpus consisting of eight rhythmic styles from three cultures, containing a variety of meter types. We demonstrate that by adapting the model to specific styles, we can track beats and downbeats in odd meter types like $9/8$ or $7/8$ with an accuracy significantly improved over the state of the art. Even if the rhythmic style is not known in advance, a unified model is able to recognize the meter and track the beat with comparable results, providing a novel method for inferring the metrical structure in culturally diverse datasets.

1. INTRODUCTION

Musical rhythm subordinated to a meter is a common feature in many music cultures around the world. Meter provides a hierarchical time structure for the rendition and repetition of rhythmic patterns. Though these metrical structures vary considerably across cultures, metrical hierarchies can often be stratified into levels of differing time spans. Two of these levels are, in terminology of Eurogenetic music, referred to as beats, and measures. The *beats* are the pulsation at the perceptually most salient metrical level, and are further grouped into *measures*. The first beat of each measure is called the *downbeat*. Determining the type of the underlying meter, and the alignment between the pulsations at the levels of its hierarchy with music performance recordings – a process we refer to as meter inference – is fundamental to computational rhythm analysis and supports many further tasks, such as music transcription, structural analysis, or similarity estimation.

The automatic annotation of music with different aspects of rhythm is at the focus of numerous studies in Music Information Retrieval (MIR). Müller et al [5] discussed

the estimation of the beat (called beat tracking), and the estimation of higher-level metrical structures such as the measure length. Approaches such as the one presented by Klapuri et al [3] aim at estimating structures at several metrical levels, while being able to differentiate between certain time signatures. In [7] beats and downbeats are estimated simultaneously, given information about the tempo and the meter of a piece. Most of these approaches assume the presence of a regular metrical grid, and work reasonably well for Eurogenetic popular music. However, their adaptation to different rhythmic styles and metrical structures is not straight-forward.

Recently, a Bayesian approach referred to as *bar pointer model* has been presented [11]. It aims at the joint estimation of rhythmic pattern, the tempo, and the exact position in a metrical cycle, by expressing them as hidden variables in a Hidden Markov Model (HMM) [8]. Krebs et al. [4] applied the model to music signals and showed that explicitly modelling rhythmic patterns is useful for meter inference for a dataset of Ballroom dance music.

In this paper, we adapt the observation model of the approach presented in [4] to a collection of music from different cultures: Makam music from Turkey, Cretan music from Greece, and Carnatic music from the south of India. The adaption of observation models was shown to be of advantage in [4, 6], however restricted to the context of Ballroom dance music. Here, we extract rhythmic patterns from culturally more diverse data, and investigate if their inclusion into the model improves the performance of meter inference. Furthermore, we investigate if a unified model can be derived that covers all rhythmic styles and time signatures that are present in the training data.

2. MOTIVATION

The music cultures considered in this paper are based on traditions that can be traced back for centuries until the present, and were documented by research in ethnomusicology for decades. Rhythm in two of these cultures, Carnatic and Turkish Makam music, is organized based on potentially long metrical cycles. All three make use of rhythmic styles that deviate audibly from the stylistic paradigms of Eurogenetic popular music. Previous studies on music collections of these styles have shown that the current state of the art performs poorly in beat tracking [2, 9] and the recognition of rhythm class [9]. As suggested in [9], we explore a unified approach for meter inference that can rec-



ognize the rhythmic style of the piece and track the meter at the same time.

The bar pointer model, as described in Section 4, can be adapted to rhythmic styles by extracting possible patterns using small representative downbeat annotated datasets. This way, we can obtain an adapted system for a specific style without recoding and parameter tweaking. We believe that this is an important characteristic for algorithms applied in music discovery and distribution systems for a large and global audience. Through this study, we aim to answer crucial questions: Do we need to differentiate between rhythmic styles in order to track the meter, or is a universal approach sufficient? For instance, can we track a rhythmic style in Indian music using rhythmic patterns derived from Turkish music? Do we need to learn patterns at all? If a particular style description for each style is needed, this has some serious consequences for the scalability of rhythmic similarity and meter inference methods; while we should ideally aim at music discovery systems without an ethnocentric bias, the needed universal analysis methods might come at a high cost given the high diversity in the musics of the world.

3. MUSIC CORPORA

In this paper we use a collection of three music corpora which are described in the following.

The corpus of Cretan music consists of 42 full length pieces of Cretan leaping dances. While there are several dances that differ in terms of their steps, the differences in the sound are most noticeable in the melodic content, and we consider all pieces to belong to one rhythmic style. All these dances are usually notated using a 2/4 time signature, and the accompanying rhythmical patterns are usually played on a Cretan lute. While a variety of rhythmic patterns exist, they do not relate to a specific dance and can be assumed to occur in all of the 42 songs in this corpus.

The Turkish corpus is an extended version of the annotated data used in [9]. It includes 82 excerpts of one minute length each, and each piece belongs to one of three rhythm classes that are referred to as *usul* in Turkish Art music. 32 pieces are in the 9/8-usul *Aksak*, 20 pieces in the 10/8-usul *Curcuna*, and 30 samples in the 8/8-usul *Düyek*.

The Carnatic music corpus is a subset of the annotated dataset used in [10]. It includes 118 two minute long excerpts spanning four *tālas* (the rhythmic framework of Carnatic music, consisting of time cycles). There are 30 examples in each of *ādi tāla* (8 beats/cycle), *rūpaka tāla* (3 beats/cycle) and *mishra chāpu tāla* (7 beats/cycle), and 28 examples in *khanda chāpu tāla* (5 beats/cycle).

All excerpts described above were manually annotated with beats and downbeats. Note that for both Indian and Turkish music the cultural definition of the rhythms contain irregular beats. Since the irregular beat sequence is a subset of the (annotated) equidistant pulses, it can be derived easily from the result of a correct meter inference. For further details on meter in Carnatic and Turkish makam music, please refer to [9].

4. METER INFERENCE METHOD

4.1 Model description

To infer the metrical structure from an audio signal we use the *bar pointer model*, originally proposed in [11] and refined in [4]. In this model we assume that a bar pointer traverses a bar and describe the state of this bar pointer at each audio frame k by three (*hidden*) variables: tempo, rhythmic pattern, and position inside a bar. These hidden variables can be inferred from an (*observed*) audio signal by using an HMM. An HMM is defined by three quantities: A *transition model* which describes the transitions between the hidden variables, an *observation model* which describes the relation between the hidden states and the observations (i.e., the audio signal), and an *initial distribution* which represents our prior knowledge about the hidden states.

4.1.1 Hidden states

The three hidden variables of the bar pointer model are:

- Rhythm pattern index $r_k \in \{r_1, r_2, \dots, r_R\}$, where R is the number of different rhythmic patterns that we consider to be present in our data. Further, we denote the time signature of each rhythmic pattern by $\theta(r_k)$ (e.g., 9/8 for *Aksak* patterns). In this paper, we assume that each rhythmic pattern belongs to a rhythmic class, and a rhythm class (e.g., *Aksak*, *Düyek*) can hold several rhythmic patterns. We investigate the optimal number of rhythmic patterns per rhythm class in Section 5.
- Position within a bar $m_k \in \{1, 2, \dots, M(r_k)\}$: We subdivide a whole note duration into 1600 discrete, equidistant bar positions and compute the number of positions within a bar with rhythm r_k by $M(r_k) = 1600 \cdot \theta(r_k)$ (e.g., a bar with 9/8 meter has $1600 \cdot 9/8 = 1800$ bar positions).
- Tempo $n_k \in \{n_{min}(r_k), \dots, n_{max}(r_k)\}$: The tempo can take on positive integer values, and quantifies the number of bar positions per audio frame. Since we use an audio frame length of $0.02s$, this translates to a tempo resolution of $7.5 (= \frac{60s}{1/4 \cdot 1600 \cdot 0.02s})$ beats per minute (BPM) at the quarter note level. We set the minimum tempo $n_{min}(r_k)$ and the maximum tempo $n_{max}(r_k)$ according to the rhythmic pattern r_k .

4.1.2 Transition model

We use the transition model proposed in [4, 11] with the difference that we allow transitions between rhythmic pattern states within a song as shown in Equation 3. In the following we list the transition probabilities for each of the three variables:

- $P(m_k | m_{k-1}, n_{k-1}, r_{k-1})$: At time frame k the bar pointer moves from position m_{k-1} to m_k as defined by

$$m_k = [(m_{k-1} + n_{k-1} - 1) \bmod (M(r_{k-1}))] + 1. \quad (1)$$
 Whenever the bar pointer crosses a bar border it is reset to 1 (as modeled by the modulo operator).
- $P(n_k | n_{k-1}, r_{k-1})$: If the tempo n_{k-1} is inside the allowed tempo range $\{n_{min}(r_{k-1}), \dots, n_{max}(r_{k-1})\}$,

there are three possible transitions: the bar pointer remains at the same tempo, accelerates, or decelerates:

$$P(n_k|n_{k-1}) = \begin{cases} 1 - p_n, & n_k = n_{k-1} \\ \frac{p_n}{2}, & n_k = n_{k-1} + 1 \\ \frac{p_n}{2}, & n_k = n_{k-1} - 1 \end{cases} \quad (2)$$

Transitions to tempi outside the allowed range are assigned a zero probability. p_n is the probability of a change in tempo per audio frame, and was set to $p_n = 0.02$, the tempo ranges ($n_{min}(r)$, $n_{max}(r)$) for each rhythmic pattern are learned from the data (Section 4.2).

- $P(r_k|r_{k-1})$: Finally, the rhythmic pattern state is assumed to change only at bar boundaries:

$$P(r_k|r_{k-1}, m_k < m_{k-1}) = p_r(r_{k-1}, r_k) \quad (3)$$

$p_r(r_{k-1}, r_k)$ denotes the probability of a transition from pattern r_{k-1} to pattern r_k and will be learned from the training data as described in Section 4.2. In this paper we allow transitions only between patterns of the same rhythm class, which will force the system to assign a piece of music to one of the learned rhythm classes.

4.1.3 Observation model

In this paper, we use the observation model proposed in [4]. As summarized in Figure 1, a Spectral Flux-like onset feature, \mathbf{y} , is extracted from the audio signal (sampled with 44100 Hz) using the same parameters as in [4]. It summarizes the energy changes that are likely to be related to instrument onsets in two dimensions related to two frequency bands, above and below 250 Hz. In contrast to [4] we removed the normalizing step at the end of the feature computations, which we observed not to influence the results.

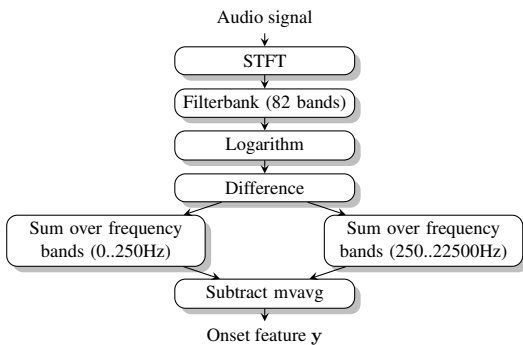


Figure 1: Computing the onset feature \mathbf{y} from the audio signal

As described in [4], the observation probabilities $P(\mathbf{y}_k|m_k, n_k, r_k)$ are modeled by a set of Mixture of Gaussian distributions (GMM). As it is infeasible to specify a GMM for each state (this would result in $N \times M \times R$ GMMs), we make two assumptions: First, we assume that the observation probabilities are independent of the tempo and second, we assume that the observation probabilities only change each 64th note (which corresponds to $1600/64=25$ bar positions). Hence, for each rhythmic pattern, we have to specify $64 \times \theta(r)$ GMMs.

4.1.4 Initial distribution

For each rhythmic pattern, we assume a uniform state distribution within the tempo limits and over all bar positions.

4.2 Learning parameters

The parameters of the observation GMMs, the transition probabilities of the rhythm pattern states, and the tempo ranges for each rhythmic style are learned from the data described in Section 3. In our experiments we perform a two-fold cross-validation, excluding those files from the evaluation that were used for parameter learning.

4.2.1 Observation model

The parameters of the observation model consist of the mean values, covariance matrix and the component weights of the GMM for each 64th note of a rhythmic pattern. We determine these as follows:

1. The two-dimensional onset feature \mathbf{y} (see Section 4.1.3) is computed from the training data.
2. The features are grouped by bar and bar position within the 64th note grid. If there are several feature values for the same bar and 64th note grid point, we compute the average, if there is no feature we interpolate between neighbors. E.g., for a rhythm class which spans a whole note (e.g., *Düyek* (8/8 meter)) this yields a matrix of size $B \times 128$, where B is the number of bars with *Düyek* rhythm class in the dataset.
3. Each dimension of the features is normalized to zero mean and unit variance.
4. For each of the eight rhythm classes in the corpus described in Section 3, a k-means clustering algorithm assigns each bar of the dataset (represented by a point in a 128-dimensional space) to one rhythmic pattern. The influence of the number of clusters k on the accuracy of the metrical inference will be evaluated in the experiments.
5. For each rhythmic pattern, at all 64th grid points, we compute the parameters of the GMM by maximum likelihood estimation.

4.2.2 Tempo ranges and transition probabilities

For each rhythmic pattern, we compute the minimum and maximum tempo of all bars of the training fold that were assigned to this pattern by the procedure described in Section 4.2.1. In the same way, we determine the transition probabilities p_r between rhythmic patterns.

4.3 Inference

In order to obtain beat-, downbeat-, and rhythmic class estimations, we compute the optimal state sequence $\{m_{1:K}^*, n_{1:K}^*, r_{1:K}^*\}$ that maximizes the posterior probability of the hidden states given the observations $y_{1:K}$ and hence fits best to our model and the observations. This is done using the well-known Viterbi algorithm [8].

5. EXPERIMENTS

5.1 Evaluation metrics

A variety of measures for evaluating beat and downbeat tracking performance are available (see [1] for a detailed overview and descriptions of the metrics listed below)¹. We chose five metrics that are characterized by a set of diverse properties and are widely used in beat tracking evaluation.

Fmeas (F-measure): The F-measure is computed from correctly detected beats within a window of ± 70 ms by

$$\text{F-measure} = \frac{2pr}{p+r} \quad (4)$$

where p (*precision*) denotes the ratio between correctly detected beats and all detected beats, and r (*recall*) denotes the ratio between correctly detected beats and the total number of annotated beats. The range of this measure is from 0% to 100%.

AMLt (Allowed Metrical Levels with no continuity required): In this method an estimated beat is counted as correct, if it lies within a small tolerance window around an annotated pulse, and the previous estimated beat lies within the tolerance window around the previous annotated beat. The value of this measure is then the ratio between the number of correctly estimated beats divided by the number of annotated beats (as percentage between 0% and 100%). Beat sequences are also considered as correct if the beats occur on the off-beat, or are double or half of the annotated tempo.

CMLt (Correct Metrical Level with no continuity required): The same as *AMLt*, without the tolerance for off-beat, or doubling/halving errors.

infGain (Information Gain): Timing errors are calculated between an annotation and all beat estimations within a one-beat length window around the annotation. Then, a beat error histogram is formed from the resulting timing error sequence. A numerical score is derived by measuring the K-L divergence between the observed error histogram and the uniform case. This method gives a measure of how much information the beats provide about the annotations. The range of values for the Information Gain is 0 bits to approximately 5.3 bits in the applied default settings.

Db-Fmeas (Downbeat F-measure): For measuring the downbeat tracking performance, we use the same F-measure as defined for beat tracking (using a ± 70 ms tolerance window).

5.2 Results

In Experiment 1, we learned the observation model described in Section 4.2 for various numbers of clusters, separately for each of the eight rhythm classes. Then, we inferred the meter using the HMM described in Section 4.1, again separately for each rhythm class. The results of this experiment indicate how many rhythmic patterns are needed for each class in order to achieve an optimal beat and downbeat tracking with the proposed model.

¹We used the MATLAB code available at <http://code.soundsoftware.ac.uk/projects/beat-evaluation/> with standard settings.

Tables (1a) to (1h) show the performance with all the evaluation measures for each of the eight styles separately. For Experiment 1 (Ex-1), all significant increases compared to the previous row are emphasized using bold numbers (according to paired-sample t-tests with 5% significance level). In our experiments, increasing the number R of considered patterns from one to two leads to a statistically significant increase in most cases. Therefore, we can conclude that for tracking these individual styles, more than one pattern is always needed. Further increase to three patterns leads to significant improvement only in the exceptional case of *Ādi tāla*, where measure cycles with long durations and rich rhythmic improvisation apparently demand higher number of patterns and cause the system to perform worse than for other classes. Higher numbers than $R = 3$ patterns never increased any of the metrics significantly. It is important to point out again that a test song was never used to train the rhythmic patterns in the observation model in Experiment 1.

The interesting question we address in Experiment 2 is if the rhythm class of a test song is a necessary information for an accurate meter inference. To this end, we performed meter inference for a test song combining all the determined rhythmic patterns for all classes in one large HMM. This means that in this experiment the HMM can be used to determine the rhythm class of a song, as well as for the tracking of beats and downbeats. We use two patterns from each rhythm class (except *ādi tāla*), the optimally performing number of patterns in Experiment 1, to construct the HMM. For *ādi tāla*, we use three patterns since using 3 patterns improved performance in Experiment 1, to give a total of $R = 17$ different patterns for the large HMM. The results of Experiment 2 are depicted in the rows labeled Ex-2 in Tables (1a) to (1h), significant change over the optimal setting in Experiment 1 are emphasized using bold numbers. The general conclusion is that the system is capable of a combined task of classification into a rhythm class and the inference of the metrical structure of the signal. The largest and, with the exception of *ādi tāla*, only significant decrease between the Experiment 1 and Experiment 2 can be observed for the downbeat recognition (*Db-Fmeas*). The reason for this is that a confusion of a test song into a wrong class may still lead to a proper tracking of the beat level, but the tracking of the higher metrical level of the downbeat will suffer severely from assigning a piece to a class with a different length of the meter than the test piece.

As described in Section 4.1, we do not allow transitions between different rhythm classes. Therefore, we can classify a piece of music into a rhythm class by evaluating to which rhythmic pattern states r_k the piece was assigned. The confusion matrix is depicted in Table 2, and it shows that the highest confusion can be observed within certain classes of Carnatic music, while the Cretan leaping dances and the Turkish classes are generally recognized with higher recall rate. The accent patterns in *mishra chāpu* and *khanda chāpu* can be indefinite, non-characteristic and non-indicative in some songs, and hence there is a possi-

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	65.9	45.0	57.6	0.89	46.6
	2	91.0	76.6	90.0	1.62	88.6
	3	90.6	77.2	91.1	1.59	86.5
Ex-2	17	85.7	68.7	89.3	1.57	65.1
KL		69.38	41.24	64.60	<u>1.46</u>	-

(a) Turkish Music: Aksak (9/8)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	57.2	33.5	42.2	0.68	37.3
	2	85.2	70.1	82.7	1.51	75.4
	3	83.4	63.3	81.9	1.45	73.7
Ex-2	17	86.6	75.8	87.2	1.64	72.6
KL		70.25	49.52	71.79	<u>1.53</u>	-

(c) Turkish Music: Düyek (8/8)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	49.6	38.9	47.0	0.93	16.5
	2	56.7	44.0	59.5	1.21	32.5
	3	61.6	49.5	65.9	1.40	32.8
Ex-2	17	62.4	40.6	76.7	1.73	21.4
KL		<u>59.42</u>	<u>45.90</u>	64.91	<u>1.53</u>	-

(e) Carnatic music: Ādi (8/8)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	84.1	79.0	79.0	1.54	71.0
	2	93.7	92.2	92.2	2.00	86.4
	3	93.4	91.6	91.6	1.99	89.9
Ex-2	17	90.0	81.6	86.3	1.83	55.0
KL		74.61	44.99	68.71	1.25	-

(g) Carnatic music: Mishra chāpu (7/8)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	71.4	47.8	50.3	0.68	38.9
	2	89.1	75.6	75.6	1.04	48.6
	3	87.7	73.0	73.0	0.99	54.4
Ex-2	17	89.3	74.8	77.5	1.16	41.1
KL		52.77	5.90	59.04	0.77	-

(b) Turkish Music: Curcuna (10/8)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	68.1	60.7	60.8	1.33	59.1
	2	93.0	91.3	91.3	2.25	86.2
	3	92.9	91.0	91.0	2.25	85.8
Ex-2	17	88.8	74.3	92.5	2.24	72.2
KL		35.87	34.42	72.07	1.57	-

(d) Cretan leaping dances (2/4)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	68.2	65.8	71.4	2.04	60.8
	2	82.8	82.5	90.2	2.77	81.9
	3	83.0	82.9	89.5	2.73	80.5
Ex-2	17	77.2	60.6	88.9	2.39	62.0
KL		53.42	29.17	60.37	1.30	-

(f) Carnatic music: Rūpaka (3/4)

	R	Fmeas	CMLt	AMLt	infGain	Db-Fmeas
Ex-1	1	58.9	38.0	41.5	0.70	27.7
	2	94.3	88.9	94.9	2.00	77.3
	3	93.7	88.1	94.3	1.95	78.2
Ex-2	17	90.3	76.0	93.2	2.01	70.6
KL		76.16	57.76	66.34	1.18	-

(h) Carnatic music: Khanda chāpu (5/8)

Table 1: Evaluation results for each rhythm class, for Experiment 1 (separate evaluation per style, shown as Ex-1), and Experiment 2 (combined evaluation using one large HMM, shown as Ex-2). The last row in each Table, with row header as KL, shows the beat tracking performance using Klapuri beat tracker. For Ex-1, bold numbers indicate significant change compared to the row above, for Ex-2, bold numbers indicate significant change over the best parameter setting in Ex-1 (bold R parameter), and for KL the only differences to Ex-2 that are not statistically significant are underlined.

bility of confusion between the two styles. Confusion between the three cultures, especially between Turkish and Carnatic is extremely rare, which makes sense due to differences in meter types, performance styles, instrumental timbres, and other aspects which influence the observation model. The recall rates of the rhythm class averaged for each culture are 69.6% for Turkish music, 69.1% for the Cretan music, and 61.02% for Carnatic music. While the datasets are not exactly the same, these numbers represent a clear improvement over the cycle length recognition results depicted in [9] for Carnatic and Turkish music.

Finally, we would like to put the beat tracking accuracies achieved with our model into relation with results obtained with state of the art approaches that do not include an adaption to the rhythm classes. In Table 1, results of the algorithm proposed in [3], which performed generally best among several other approaches, are depicted in the last rows (KL) of each subtable. We underline those results that do not differ significantly from those obtained in Experiment 2. In all other cases the proposed bar pointer model

performs significantly better. The only rhythm class, for which our approach does not achieve an improvement in most metrics is the ādi tāla. As mentioned earlier, this can be attributed to the large variety of patterns and the long cycle durations in ādi tāla.

6. CONCLUSIONS

In this paper we adapted the observation model of a Bayesian approach for the inference of meter in music of cultures in Greece, India, and Turkey. It combines the task of determining the type of meter with the alignment of the downbeats and beats to the audio signal. The model is capable of performing the meter recognition with an accuracy that improves over the state of the art, and is at the same time able to achieve for the first time high beat and downbeat tracking accuracies in additive meters like the Turkish Aksak and Carnatic mishra chāpu.

Our results show that increasing the diversity of a corpus means increasing the number of the patterns, *i.e.* a larger

	Turkish			Greek	Carnatic				Recall
	Aksak	Düyek	Curcuna	Cretan	Ädi	Rüpaka	M.chäpu	K.chäpu	
Aksak	21	7	2	2					66
Düyek		23	2	5					77
Curcuna	1	3	13	2				1	65
Cretan	3	5		29	3	2			69
Ädi					14	8	1	7	47
Rüpaka					3	19	1	7	63
M.chäpu					2	1	16	11	53
K.chäpu						4	1	23	82
Precision	84	61	76	76	64	56	84	47	

Table 2: Confusion matrix of the style classification of the large HMM (Ex-2). The rows refer to the true style and the columns to the predicted style. The empty blocks are zeros (omitted for clarity of presentation).

amount of model parameters. In the context of the HMM inference scheme applied in this paper this implies an increasingly large hidden-parameter state-space. However, we believe that this large parameter space can be handled by using more efficient inference schemes such as Monte Carlo methods.

Finally, we believe that the adaptability of a music processing system to new, unseen material is an important design aspect. Our results imply that in order to extend meter inference to new styles, at least some amount of human annotation is needed. If there exist music styles where adaptation can be achieved without human input remains an important point for future discussions.

Acknowledgments

This work is supported by the Austrian Science Fund (FWF) project Z159, by a Marie Curie Intra-European Fellowship (grant number 328379), and by the European Research Council (grant number 267583).

7. REFERENCES

- [1] M. Davies, N. Degara, and M. D. Plumbley. Evaluation methods for musical audio beat tracking algorithms. *Queen Mary University of London, Tech. Rep. C4DM-09-06*, 2009.
- [2] A. Holzapfel and Y. Stylianou. Beat tracking using group delay based onset detection. In *Proceedings of ISMIR - International Conference on Music Information Retrieval*, pages 653–658, 2008.
- [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*, 14(1):342–355, 2006.
- [4] 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 Conference (ISMIR-2013)*, Curitiba, Brazil, nov 2013.
- [5] M. Müller, D. P. W. Ellis, A. Klapuri, G. Richard, and S. Sagayama. Introduction to the Special Issue on Music Signal Processing. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1085–1087, 2011.
- [6] G. Peeters. Template-based estimation of tempo: using unsupervised or supervised learning to create better spectral templates. In *Proc. of the 13th International Conference on Digital Audio Effects (DAFX 2010)*, Graz, Austria, 2010.
- [7] 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*, 19(6):1754–1769, 2011.
- [8] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.
- [9] A. Srinivasamurthy, A. Holzapfel, and X. Serra. In search of automatic rhythm analysis methods for Turkish and Indian art music. *Journal for New Music Research*, 43(1):94–114, 2014.
- [10] A. Srinivasamurthy and X. Serra. A supervised approach to hierarchical metrical cycle tracking from audio music recordings. In *Proc. of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2014)*, pages 5237–5241, Florence, Italy, May 2014.
- [11] N. Whiteley, A. Cemgil, and S. Godsill. Bayesian modelling of temporal structure in musical audio. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR-2006)*, Victoria, 2006.