RHYTHMIC PATTERN MODELING FOR BEAT AND DOWNBEAT TRACKING IN MUSICAL AUDIO

Florian Krebs, Sebastian Böck, and Gerhard Widmer
Department of Computational Perception
Johannes Kepler University, Linz, Austria
florian.krebs@jku.at

ABSTRACT

Rhythmic patterns are an important structural element in music. This paper investigates the use of rhythmic pattern modeling to infer metrical structure in musical audio recordings. We present a Hidden Markov Model (HMM) based system that simultaneously extracts beats, downbeats, tempo, meter, and rhythmic patterns. Our model builds upon the basic structure proposed by Whiteley et. al [20], which we further modified by introducing a new observation model: rhythmic patterns are learned directly from data, which makes the model adaptable to the rhythmical structure of any kind of music. For learning rhythmic patterns and evaluating beat and downbeat tracking, 697 ballroom dance pieces were annotated with beat and measure information. The results showed that explicitly modeling rhythmic patterns of dance styles drastically reduces octave errors (detection of half or double tempo) and substantially improves downbeat tracking.

1. INTRODUCTION

From its very beginnings, music has been built on temporal structure to which humans can synchronize via musical instruments and dance. The most prominent layer of this temporal structure (which most people tap their feet to) contains the approximately equally spaced *beats*. These beats can, in turn, be grouped into *measures*, segments with a constant number of beats; the first beat in each measure, which usually carries the strongest accent within the measure, is called the *downbeat*. The automatic analysis of this temporal structure in a music piece has been an active research field since the 1970s and is of prime importance for many applications such as music transcription, automatic accompaniment, expressive performance analysis, music similarity estimation, and music segmentation. However, many problems within the automatic analysis of metrical structure remain unsolved. In particular, complex rhythmic phenomena such as syncopations, triplets, and swing make it difficult to find the correct phase and period of downbeats and beats, especially for systems that rely on the assumption that beats usually occur at onset times. Considering all these rhythmic peculiarities, a general model no longer suffices.

One way to overcome this problem is to incorporate higher-level musical knowledge into the system. For example, Hockman et al. [12] proposed a genre-specific beat tracking system designed specifically for the genres hardcore, jungle, and drum and bass. Another way to make the model more specific is to model explicitly one or several *rhythmic patterns*. These rhythmic patterns describe the distribution of note onsets within a predefined time interval, e.g., one bar. For example, Goto [9] extracts bar-length drum patterns from audio signals and matches them to eight pre-stored patterns typically used in popular music. Klapuri et al. [14] proposed a HMM representing a three-level metrical grid consisting of tatum, tactus, and measure. Two rhythmic patterns were employed to obtain an observation probability for the phase of the measure pulse. The system of Whiteley et al. [20] jointly models tempo, meter, and rhythmic patterns in a Bayesian framework. Simple observation models were proposed for symbolic and audio data, but were not evaluated on polyphonic audio signals.

Although rhythmic patterns are used in some systems, no systematic study exists that investigates the importance of rhythmic patterns for analyzing the metrical structure. Apart from the approach presented in [17], which learns a single rhythmic template from data, rhythmic patterns to be used for beat tracking have so far only been designed by hand and hence depend heavily on the intuition of the developer.

This paper investigates the role of rhythmic patterns in analyzing the metrical structure in musical audio signals. We propose a new observation model for the HMM-based system described in [20], whose parameters are learned from real audio data and can therefore be adapted easily to represent any rhythmic style.

2. RHYTHMIC PATTERNS

Although rhythmic patterns could be defined at any level of the metrical structure, we restrict the definition of rhythmic patterns to the length of a single measure.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
© 2013 International Society for Music Information Retrieval.

2.1 Data

As stated in Section 1, strong deviations from a straight on-beat rhythm constitute potential problems for automatic rhythmic description systems. While pop and rock music is commonly concentrated on the beat, Afro-Cuban rhythms frequently contain syncopations, for instance in the *clave* pattern – the structural core of many Afro-Cuban rhythms. Therefore, Latin music represents a serious challenge to beat and downbeat tracking systems.

The ballroom dataset [1] contains eight different dance styles (Cha cha, Jive, Quickstep, Rumba, Samba, Tango, Viennese Waltz, and (slow) Waltz) and has been used by several authors, for example, for genre recognition [6, 18]. It consists of 697 [2] 30 seconds-long audio excerpts (sampled at 11.025 kHz) and has tempo and dance style annotations. The dataset contains two different meters (3/4 and 4/4) and all pieces have constant meter. The tempo distributions of the dance styles are displayed in Fig. 4.

We have annotated both beat and downbeat times manually. In cases of disagreement on the metrical level we relied on the existing tempo and meter annotations. The annotations can be downloaded from `https://github.com/CPJKU/BallroomAnnotations`.

2.2 Representation of rhythmic patterns

Patterns such as those shown in Fig. 1 are learned in the process of inducing the likelihood function for the model (cf. Section 3.3.3), where we use the dance style labels of the training songs as indicators of different rhythmic patterns. To model dependencies between instruments in our pattern representations, we split the audio signal into two frequency bands and compute an onset feature for each of the bands individually as described in Section 3.3. To illustrate the rhythmic characteristics of different dance styles, we show the eight learned representations of rhythmic patterns in Fig. 1. Each pattern is represented by a distribution of onset feature values along a bar in two frequency bands.

For example, the *Jive* pattern displays strong accents on the second and fourth beat, a phenomenon usually referred to as *backbeat*. In addition, the typical *swing* style is clearly visible in the high-frequency band. The *Rumba* pattern contains a strong accent of the bass on the 4th and 7th eighth note, which is a common bass pattern in Afro-Cuban music and referred to as *anticipated bass* [15]. One of the characteristics of *Samba* is the shuffled bass line, a pattern originally played with the *Surdo*, a large Brazilian bass drum. The pattern features bass notes on the 1st, 4th, 5th, 9th, 12th, and 13th sixteenth note of the bar. *Waltz*, finally, is a triple meter rhythm. While the bass notes are located mainly on the downbeat, high-frequency note onsets are also located at the quarter and eighth note level of the measure.

[1] The data was extracted from `www.ballroomdancers.com`.

[2] One of the 698 original files was found duplicated and was removed.

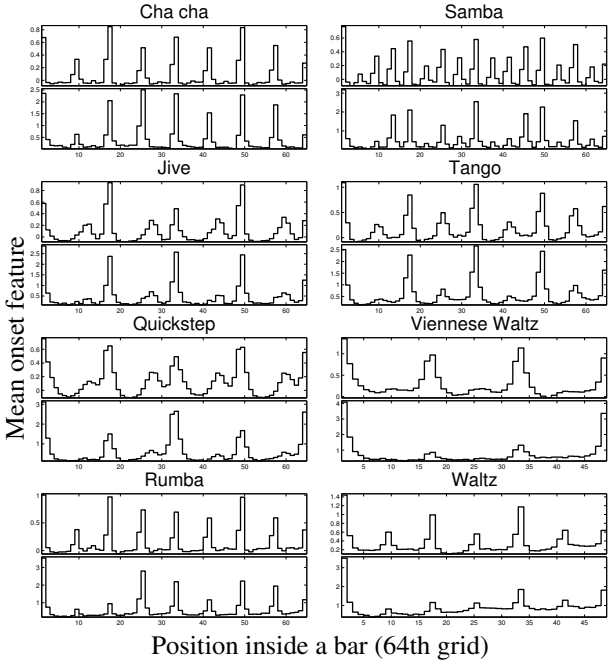

Figure 1. Illustration of learned rhythmic patterns. Two frequency bands are shown (Low/High from bottom to top).

3. METHOD

In this section, we describe the *dynamic Bayesian network* (DBN) [16] we use to analyze the metrical structure. We assume that a time series of *observed* data $\mathbf{y}_{1:K} = \{\mathbf{y}_1, ..., \mathbf{y}_K\}$ is generated by a set of unknown, *hidden* variables $\mathbf{x}_{1:K} = \{\mathbf{x}_1, ..., \mathbf{x}_K\}$, where K is the length of an audio excerpt in frames. In a DBN, the joint distribution $P(\mathbf{y}_{1:K}, \mathbf{x}_{1:K})$ factorizes as

$$P(\mathbf{y}_{1:K}, \mathbf{x}_{1:K}) = P(\mathbf{x}_1) \prod_{k=2}^{K} P(\mathbf{x}_k|\mathbf{x}_{k-1}) P(\mathbf{y}_k|\mathbf{x}_k) \quad (1)$$

where $P(\mathbf{x}_1)$ 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 proposed model is similar to the model proposed by Whiteley et. al [20] with the following modifications:

- We assume conditional dependence between the tempo and the rhythmic pattern (cf., Section 3.2), which is a valid assumption for ballroom music as shown in Fig. 4.
- As the original observation model was mainly intended for percussive sounds, we replace it by a Gaussian Mixture Model (GMM) as described in Section 3.3.

3.1 Hidden variables

The *dynamic bar pointer model* [20] defines the state of a hypothetical bar pointer at time $t_k = k \cdot \Delta$, with $k \in \{1, 2, ..., K\}$ and Δ the audio frame length, by the following discrete hidden variables:

1. Position inside a bar $m_k \in \{1, 2, ..., M\}$, where $m_k = 1$ indicates the beginning and $m_k = M$ the end of a bar;

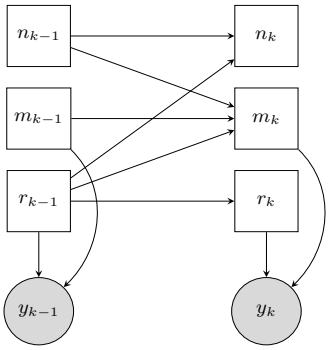

Figure 2. Dynamic Bayesian network; circles denote continuous variables and rectangles discrete variables. The gray nodes are observed, and the white nodes represent the hidden variables.

2. Tempo $n_k \in \{1, 2, ..., N\}$ (unit $\frac{\text{bar positions}}{\text{audio frame}}$), where N denotes the number of tempo states;

3. Rhythmic pattern $r_k \in \{r_1, r_2, ..., r_R\}$, where R denotes the number of rhythmic patterns.

For the experiments reported in this paper, we chose $\Delta = 20$ ms, $M = 1216$, $N = 26$, and R (the number of rhythmic patterns) was 2 or 8 as described in Section 4.2. Furthermore, each rhythmic pattern is assigned to a meter $\theta(r_k) \in \{3/4, 4/4\}$, which is important to determine the measure boundaries in Eq. 4. The conditional independence relations between these variables are shown in Fig. 2.

As noted in [16], any discrete state DBN can be converted into a regular HMM by merging all hidden variables of one time slice into a 'meta-variable' $\mathbf{x}_k$, whose state space is the Cartesian product of the single variables:

$$\mathbf{x}_k = [m_k, n_k, r_k]. \tag{2}$$

3.2 Transition model

Due to the conditional independence relations shown in Fig. 2, the transition model factorizes as

$$\begin{aligned} P(\mathbf{x}_k|\mathbf{x}_{k-1}) = \quad & P(m_k|m_{k-1}, n_{k-1}, r_{k-1}) \times \\ & \times P(n_k|n_{k-1}, r_{k-1}) \times P(r_k|r_{k-1}) \end{aligned} \tag{3}$$

where the three factors are defined as follows:

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

  $$m_k = [(m_{k-1} + n_{k-1} - 1) \bmod (N_m \cdot \theta(r_{k-1}))] + 1. \tag{4}$$

  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 $\{n_{min}(r_{k-1}), ..., n_{max}(r_{k-1})\}$, there are three possible transitions: the bar pointer remains at the same tempo, accelerates, or decelerates:
  if $n_{min}(r_{k-1}) \leq n_{k-1} \geq n_{max}(r_{k-1})$,

  $$P(n_k|n_{k-1}) = \begin{cases} 1 - p_n, & n_k = n_{k-1}; \\ \frac{p_n}{2}, & n_k = n_{k-1} + 1; \\ \frac{p_n}{2}, & n_k = n_{k-1} - 1. \end{cases} \tag{5}$$

  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 the step-size of a tempo change per audio frame was set to one bar position per audio frame.

- $P(r_k|r_{k-1})$
  For this work, we assume a musical piece to have a characteristic rhythmic pattern that remains constant throughout the song; thus we obtain

  $$r_{k+1} = r_k. \tag{6}$$

3.3 Observation model

For simplicity, we omit the frame indices k in this section. The observation model $P(\mathbf{y}|\mathbf{x})$ reduces to $P(\mathbf{y}|m, r)$ due to the independence assumptions shown in Fig. 2.

3.3.1 Observation features

Since the perception of beats depends heavily on the perception of played musical notes, we believe that a good onset feature is also a good beat tracking feature. Therefore, we use a variant of the *LogFiltSpecFlux* onset feature, which performed well in recent comparisons of onset detection functions [1] and is summarized in the top part of Fig. 3. We believe that the bass instruments play an important role in defining rhythmic patterns, hence we compute onsets in low-frequencies (< 250 Hz) and high-frequencies (> 250 Hz) separately. In Section 5.1 we investigate the importance of using the two-dimensional onset feature over a one-dimensional one. Finally, we subtract the moving average computed over a window of one second and normalize the features of each excerpt to zero mean and unity variance.

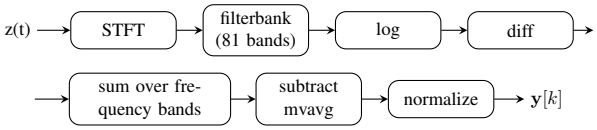

Figure 3. Computing the onset feature $\mathbf{y}$[k] from the audio signal $z(t)$

3.3.2 State tying

We assume the observation probabilities to be constant within a 64th note grid. All states within this grid are tied and thus share the same parameters, which yields 64 (4/4 meter) and 48 (3/4 meter) different observation probabilities per bar and rhythmic pattern.

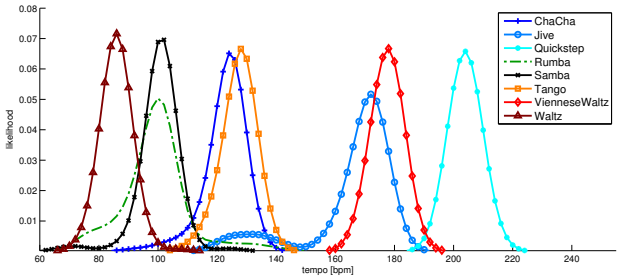

Figure 4. Tempo distributions of the ballroom dataset dance styles. The displayed distributions are obtained by (Gaussian) kernel density estimation for each dance style separately.

3.3.3 Likelihood function

To learn a representation of $P(\mathbf{y}|m,r)$, we split the training dataset into pieces of one bar length, starting at the downbeat. For each bar position within the 64th grid and each rhythmic pattern, we collect all corresponding feature values and fit a GMM. We achieved the best results on our test set with a GMM of $I = 2$ components. Hence, the observation probability is modeled by

$$P(\mathbf{y}|m,r) = \sum_{i=1}^{I} w_{m,r,i} \cdot \mathcal{N}(\mathbf{y}; \mu_{m,r,i}, \Sigma_{m,r,i}), \qquad (7)$$

where $\mu_{m,r,i}$ is the mean vector, $\Sigma_{m,r,i}$ is the covariance matrix, and $w_{m,r,i}$ is the mixture weight of component i of the GMM. Since, in learning the likelihood function $P(\mathbf{y}|m,r)$, a GMM is fitted to the audio features for every rhythmic pattern (i.e., dance style) label r, the resulting GMMs can be interpreted directly as representations of rhythmic patterns. Fig. 1 shows the mean values of the features per frequency band and bar position for the GMMs corresponding to the eight rhythmic patterns $r \in \{$Cha cha, Jive, Quickstep, Rumba, Samba, Tango, Viennese Waltz, Waltz$\}$.

3.4 Initial state distribution

The bar position and the rhythmic patterns are assumed to be distributed uniformly, whereas the tempo state probabilities are modeled by fitting a GMM [3] to the tempo distribution of each ballroom style shown in Fig. 4.

3.5 Inference

We are looking for the state sequence $\mathbf{x}^*_{1:K}$ with the highest posterior probability $p(\mathbf{x}_{1:K}|\mathbf{y}_{1:K})$:

$$\mathbf{x}^*_{1:K} = \arg\max_{x_{1:K}} p(\mathbf{x}_{1:K}|\mathbf{y}_{1:K}). \qquad (8)$$

We solve Eq. 8 using the Viterbi algorithm [19]. Once $\mathbf{x}^*_{1:K}$ is computed, the set of beat and downbeat times are obtained by interpolating $m^*_{1:K}$ at the corresponding bar positions.

[3] The number of components was set to two (PS2), and four (PS8)

4. EXPERIMENTAL SETUP

We use different settings and reference methods to evaluate the relevance of rhythmic pattern modeling for the beat and downbeat tracking performance.

4.1 Evaluation measures

A variety of measures for evaluating beat tracking performance is available (see [3] for an overview). We chose to report continuity-based measures for beat and downbeat tracking as in [4, 5, 14]:

- CMLc (Correct Metrical Level with continuity required) assesses the longest segment of correct beats at the correct metrical level.
- CMLt (Correct Metrical Level with no continuity required) assesses the total number of correct beats at the correct metrical level.
- AMLc (Allowed Metrical Level with continuity required) assesses the longest segment of correct beats, considering several metrical levels and offbeats.
- AMLt (Allowed Metrical Level with no continuity required) assesses the total number of correct beats, considering several metrical levels and offbeats.

Due to lack of space, we present only the mean values per measure across all files of the dataset. Please visit `http://www.cp.jku.at/people/krebs/ISMIR2013.html` for detailed results and other metrics.

4.2 Systems compared

To evaluate the use of modeling multiple rhythmic patterns, we report results for the following variants of the proposed system (PS): PS2 uses two rhythmic patterns (one for each meter), PS8 uses eight rhythmic patterns (one for each genre), PS8.genre has the ground truth genre, and PS2.meter has the ground truth meter as additional input features.

In order to compare the system to the state-of-the-art, we add results of six reference beat tracking algorithms: Ellis [7], Davies [4], Degara [5], Böck [2], Ircambeat [17], and Klapuri [14]. The latter two also compute downbeat times.

4.3 Parameter training

For all variants of the proposed system PS*x*, the results were computed by a leave-one-out approach, where we trained the model on all songs except the one to be tested. The Böck system has been trained on the data specified in [2], the SMC [13], and the Hainsworth dataset [10]. The beat templates used by Ircambeat in [17] have been trained using their own annotated PopRock dataset. The other methods do not require any training.

4.4 Statistical tests

In Section 5.1 we use an analysis of variance test (ANOVA) and in Section 5.2 a multiple comparison test [11] to find

System	CMLc	CMLt	AMLc	AMLt
PS2.1d	**62.2**	**65.8**	**87.6**	**93.1**
PS2.2d	**66.7**	**70.1**	**88.5**	**93.2**
PS8.1d	**76.6**	**79.7**	**87.7**	**92.1**
PS8.2d	**79.5**	**83.0**	**87.6**	**91.6**
PS2	66.7	70.1	**88.5**	**93.2**
PS8	**79.5**	**83.0**	87.6	**91.6**
Ellis [7]	26.7	30.9	65.2	80.2
Davies [4]	57.9	59.2	87.9	89.8
Degara [5]	64.6	66.9	85.3	89.5
Ircambeat [17]	58.1	60.3	86.1	89.6
Böck [2]	65.7	67.7	**92.0**	**94.4**
Klapuri [14]	55.2	57.0	84.9	87.3
PS2.meter	68.0	71.7	88.7	93.7
PS8.genre	89.9	93.7	90.9	94.8

Table 1. Beat tracking performance on the ballroom dataset. Results printed in bold are statistically equivalent to the best result.

statistically significant differences among the mean performances of the different systems. A significance level of 0.05 was used to declare performance differences as statistically relevant.

5. RESULTS AND DISCUSSION

5.1 Dimensionality of the observation feature

As described in Section 3.3.1, the onset feature is computed for one (PS*x*.1d) or two (PS*x*.2d) frequency bands separately. The top parts of Table 1 and Table 2 show the effect of the dimensionality of the feature vector on the beat and downbeat tracking results respectively.

For beat tracking, analyzing the onset function in two separate frequency bands seems to help finding the correct metrical level, as indicated by higher CML measures in Table 1. Even though the improvement is not significant, this effect was observed for both PS2 and PS8.

For downbeat tracking, we have found a significant improvement for all measures if two bands are used instead of a single one, as evident from Table 2. This seems plausible, as the bass plays a major role in defining a rhythmic pattern (see Section 2.2) and helps to resolve the ambiguity between the different beat positions within a bar.

Using three or more onset frequency bands did not improve the performance further in our experiments. In the following sections we will only report the results for the two-dimensional onset feature (PS*x*.2d) and simply denote it as PS*x*.

5.2 Relevance of rhythmic pattern modeling

In this section, we evaluate the relevance of rhythmic pattern modeling by comparing the beat and downbeat tracking performance of the proposed systems to six reference systems.

System	CMLc	CMLt	AMLc	AMLt
PS2.1d	46.9	47.1	70.5	71.1
PS2.2d	**55.5**	**55.7**	**76.2**	**76.5**
PS8.1d	65.4	65.8	80.9	81.8
PS8.2d	**71.1**	**71.5**	**85.3**	**85.9**
PS2	55.5	55.7	76.2	76.5
PS8	**71.1**	**71.5**	**85.3**	**85.9**
Ircambeat [17]	36.5	37.4	57.4	59.4
Klapuri [14]	39.6	40.1	68.1	68.9
PS2.meter	62.1	62.4	84.2	84.6
PS8.genre	82.8	83.1	92.6	92.9

Table 2. Downbeat tracking performance on the ballroom dataset. Results printed in bold are statistically equivalent to the best result.

5.2.1 Beat tracking

The beat tracking results of the reference methods are displayed together with PS2 (=PS2.2d) and PS8 (=PS8.2d) in the middle part of Table 1. Although there is no single system that performs best in all of the measures, we can still determine a best system for the CML measures and one for the AML measures separately.

For the CML measures (which require the correct metrical level), PS8 clearly outperforms all other systems. If the correct dance style is supplied as in PS8.genre, the performance increases even more. Apparently, the dance style provides sufficient rhythmic information to resolve tempo ambiguities.

For the AML measures (which do not require the correct metrical level), we found no advantage of using the proposed methods over most of the reference methods. The system proposed by Böck, which has been trained on Pop/ Rock music, outperforms all other systems, even though the difference to PS2 (for AMLc and AMLt) and PS8 (for AMLt) is not significant.

Hence, if the correct metrical level is unimportant or even ambiguous, a general model like Böck or any other reference system might be preferable to the more complex PS8. On the contrary, in applications where the correct metrical level matters (e.g., a system that detects beats and downbeats for automatic ballroom dance instructions [8]), PS8 is the best system to chose.

Knowing the meter a priori (PS2.meter) was not found to increase the performance significantly compared to PS2. It appeared that meter was identified mostly correct by PS2 (in 89% of the songs) and that for the remaining 11% songs both of the rhythmic patterns fitted equally well.

5.2.2 Downbeat tracking

Table 2 lists the results for downbeat tracking. As shown, PS8 outperforms all other systems significantly in all metrics. In cases where the dance style is known a priori (PS8.genre), the downbeat performance increases even more. The same was observed for PS2 if the meter was known (PS2.meter). This leads to the assumption that downbeat

tracking (as well as beat tracking with PS8) would improve even more by including meter or genre detection methods. For instance, Pohle et al. [18] report a dance style classification rate of 89% on the same dataset, whereas PS8 detected the correct dance style in only 75% of the cases.

The poor performance of Ircambeat and Klapuri's system is probably caused by the fact that both systems were developed for music comprising a completely different metrical structure than present in ballroom data. In addition, Klapuri's system explicitly assumes 4/4 meter (only true for 522 songs) and relies on the high-frequency content of the signal (that is drastically reduced using a sampling rate of 11.025 kHz) to determine the measure boundaries.

6. CONCLUSION AND FUTURE WORK

In this study, we investigated the influence of explicit modeling of rhythmic patterns on the beat and downbeat tracking performance in musical audio signals. For this purpose we have proposed a new observation model for the system proposed in [20] representing rhythmical patterns in two frequency bands.

Our experiments indicated that computing an onset feature for at least two different frequency bands increases the downbeat tracking performance significantly compared to a single feature covering the whole frequency range.

In a comparison with six reference systems, explicitly modeling dance styles as rhythmic patterns was shown to drastically reduce octave errors (detecting half or double tempo) in beat tracking. Besides, downbeat tracking was improved substantially compared to a variant that only models meter and two reference systems.

Obviously, ballroom music is well structured in terms of rhythmic patterns and tempo distribution. If all the findings reported in this paper also apply to music genres other than ballroom music has yet to be investigated.

In this work, the rhythmic patterns were determined by dance style labels. In future work, we want to use unsupervised clustering methods to extract meaningful rhythmic patterns from the audio features directly.

7. ACKNOWLEDGMENTS

We are thankful to Simon Dixon for providing access to the first bar annotations of the ballroom dataset and to Norberto Degara and the reviewers for inspiring inputs. This work was supported by the Austrian Science Fund (FWF) project Z159 and the European Union Seventh Framework Programme FP7 / 2007-2013 through the PHENICX project (grant agreement no. 601166).

8. REFERENCES

[1] S. Böck, F. Krebs, and M. Schedl. Evaluating the online capabilities of onset detection methods. In *Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR)*, Porto, 2012.

[2] S. Böck and M. Schedl. Enhanced beat tracking with context-aware neural networks. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2011.

[3] 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.

[4] M. Davies and M. Plumbley. Context-dependent beat tracking of musical audio. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3):1009–1020, 2007.

[5] N. Degara, E. Argones Rua, A. Pena, S. Torres-Guijarro, M. Davies, and M. Plumbley. Reliability-informed beat tracking of musical signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, (99):1–1, 2011.

[6] S. Dixon, F. Gouyon, and G. Widmer. Towards characterisation of music via rhythmic patterns. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, Barcelona, 2004.

[7] D. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[8] F. Eyben, B. Schuller, S. Reiter, and G. Rigoll. Wearable assistance for the ballroom-dance hobbyist-holistic rhythm analysis and dance-style classification. In *Proceedings of the 8th IEEE International Conference on Multimedia and Expo (ICME)*, Beijing, 2007.

[9] M. Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, 30(2):159–171, 2001.

[10] S. Hainsworth and M. Macleod. Particle filtering applied to musical tempo tracking. *EURASIP Journal on Applied Signal Processing*, 2004:2385–2395, 2004.

[11] Y. Hochberg and A. Tamhane. *Multiple comparison procedures*. John Wiley & Sons, Inc., 1987.

[12] J. Hockman, M. Davies, and I. Fujinaga. One in the jungle: Downbeat detection in hardcore, jungle, and drum and bass. In *Proceedings of the 13th International Society for Music Information Retrieval (ISMIR)*, Porto, 2012.

[13] A. Holzapfel, M. Davies, J. Zapata, J. Oliveira, and F. Gouyon. Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2539–2548, 2012.

[14] A. Klapuri, A. Eronen, and J. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):342–355, 2006.

[15] P. Manuel. The anticipated bass in cuban popular music. *Latin American music review*, 6(2):249–261, 1985.

[16] K. Murphy. *Dynamic bayesian networks: representation, inference and learning*. PhD thesis, University of California, Berkeley, 2002.

[17] 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*, (99):1–1, 2011.

[18] T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer. On rhythm and general music similarity. In *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR)*, Kobe, 2009.

[19] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[20] N. Whiteley, A. Cemgil, and S. Godsill. Bayesian modelling of temporal structure in musical audio. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, Victoria, 2006.