Hindawi Publishing Corporation
EURASIP Journal on Audio, Speech, and Music Processing
Volume 2008, Article ID 846135, 12 pages
doi:10.1155/2008/846135

Research Article

Tango or Waltz?: Putting Ballroom Dance Style into Tempo Detection

Björn Schuller, Florian Eyben, and Gerhard Rigoll

Institute for Human-Machine Communication, Technische Universität München, Arcisstraße 21, 80333 München, Germany

Correspondence should be addressed to Björn Schuller, schuller@tum.de

Received 31 October 2007; Revised 14 February 2008; Accepted 14 March 2008

Recommended by Sen Kuo

Rhythmic information plays an important role in Music Information Retrieval. Example applications include automatically annotating large databases by genre, meter, ballroom dance style or tempo, fully automated D.J.-ing, and audio segmentation for further retrieval tasks such as automatic chord labeling. In this article, we therefore provide an introductory overview over basic and current principles of tempo detection. Subsequently, we show how to improve on these by inclusion of ballroom dance style recognition. We introduce a feature set of 82 rhythmic features for rhythm analysis on real audio. With this set, data-driven identification of the meter and ballroom dance style, employing support vector machines, is carried out in a first step. Next, this information is used to more robustly detect tempo. We evaluate the suggested method on a large public database containing 1.8 k titles of standard and Latin ballroom dance music. Following extensive test runs, a clear boost in performance can be reported.

1. INTRODUCTION

Music Information Retrieval (MIR) has been a growing field of research over the last decade. The increasing popularity of portable music players and music distribution over the internet has made worldwide, instantaneous access to rapidly growing music archives possible. Such archives must be well structured and sorted in order to be user friendly. For example, many users face the problem of having heard a song they would like to buy but not knowing its bibliographic data, that is, title and artist, which is necessary to find the song in conventional (online) music stores. According to Downie in [1], almost three fourths of all MIR queries are of bibliographic nature. The querying person gives information he or she knows about the song, most likely genre, meter, tempo, lyrics, or acoustic properties, for example, tonality and demands information about title and/or artist. In order to have machines assist in building a song database queryable by features such as tempo, meter, or genre, intelligent Information Retrieval algorithms are necessary to automatically extract such high-level features from raw music data. Many works exist that describe or give overviews over basic MIR methods, for example, [2–8]. Besides tonal features, the temporal features play an important role. Tempo, meter, and beat locations form the basis for segmenting music and thus for further feature extraction such as chord change detection or higher level metrical analysis, for example, as performed in [9]. Because of its importance, we will primarily focus on robust tempo detection within this article.

Currently existing state-of-the-art tempo detection algorithms are—generally speaking—based on methods of periodicity detection. That is, they use techniques such as autocorrelation, resonant filter banks, or onset time statistics to detect the tempo. A good comparison and overview is given in [10]. However, very little work exists that combines various low-level detection methods, such as tempo induction, meter recognition, and beat tracking into a system that is able to use features from all these subtasks to perform robust high-level classification tasks, for example, ballroom dance style or genre recognition, and in turn use the classification results to improve the low-level detection results. Only few, such as [11, 12], present data-driven genre and meter recognition. Other methods, such as [13, 14], use rhythmic features only for specific tasks, like audio identification, and do not use rhythmic features in a multistep process to improve results themselves.

A novel approach that aims at robust, data-driven rhythm analysis primarily targeted at database applications is presented in this article. A compact set of low-level rhythmic features is described, which is highly suitable for

discrimination between duple and triple meter as well as ballroom dance style classification. Based on the results of data-driven dance style and meter recognition, the quarter-note tempo can be detected very reliably reducing errors, where half or twice of the true tempo is detected. Beat tracking at the beat level for songs with an approximately constant tempo can be performed more reliably once the tempo is known—however, it will not be discussed in this article. A beat tracking method, that can be used in conjunction with the new data-driven rhythm analysis approach, is presented in [15]. Although the primary aim of the presented approach is to robustly detect the quarter-note tempo, the complete procedure is referred to as rhythm analysis, because meter and ballroom dance style are also detected and used in the final tempo detection pass.

The article is structured as follows. In Section 2, an introduction to tempo detection, meter recognition, and genre classification is given along with an overview over selected related work. Section 3 describes the novel approach to improved data-driven tempo detection through prior meter and ballroom dance style classification. The results are presented in Section 4 and compared to results obtained at the ISMIR 2004 tempo induction contest before the conclusion and outlook in Section 5.

2. RELATED WORK

Tempo induction, beat tracking, and meter detection methods can roughly be divided into two major groups. The first group consists of those that attempt to explicitly find onsets in the first step (or use onsets obtained from a symbolic notation, e.g., MIDI), and then deduct information about tempo, beat positions, and possibly meter by analyzing the interonset intervals (IOIs) [9, 16–21]. The second group contains those that extract information about the tempo and metrical structure prior to onset detection. Correlation or resonator methods are mostly used for this task. If onset positions are required, onset detection can then be assisted by information from the tempo detection stage [2, 4–6, 8, 22].

The more robust methods, especially, for database applications, are those from the second group. However, we will first explain the concept of onset detection used in the methods of the first group, as we believe it is a very intuitive way to approach the problem of beat tracking and tempo detection.

Before we start explaining the tempo induction methods, we take a look at some music terminology regarding meter. The metrical structure of a musical piece is composed of multiple hierarchical levels [23], where the tempo of each higher level is an integer multiple of the tempo on the lowest level. The latter is called *tatum* level. The level at which we tap along when listening to a song is the *pulse* or *beat* level. Sometimes this tempo is referred to as the quarter-note tempo. The *bar* or *measure* level corresponds to the bar in notated music, and the period of its tempo gives the length of a measure. The relation between measure and beat level is often referred to as time signature or more generally the meter.

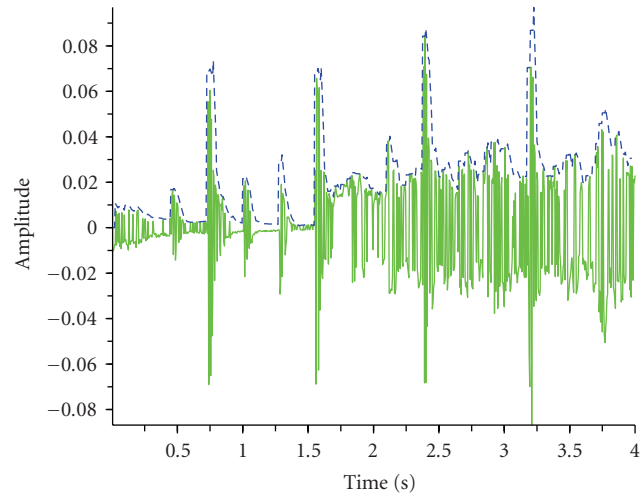

Figure 1: Waveform and envelope (dashed line) of 4 seconds from "OMD—Maid of Orleans."

To get familiar with the concept of onset detection, on which the first group of algorithms is based, let us assume that a beat basically corresponds to a sudden increase in the signal (energy) envelope. This is a very simplified assumption, which is valid only for music containing percussion and strong beats. There are basically two methods for computing an audio signal envelope (depicted in Figure 1) suitable for onset detection of a signal $x[n]$.

(1) Full-wave rectification and lowpass filtering of x followed by down sampling to approximately 100 Hz.

(2) Dividing the signal into small windows having a length around 20 milliseconds with approximately 50% overlap and then calculating the RMS energy of each window by averaging $x[n]^2$ over all n in the window. This can be followed by an additional lowpass filter for smoothing purposes.

The first order differential of the resulting (energy) envelope is then computed (Figure 2). A local maximum in the differential of the envelope corresponds to a strong rise in the envelope itself. By picking peaks in the differential that are above a certain threshold (e.g., the mean value or a given percentage of the maximum of the differential over a certain time window) some onsets can be located. The magnitude, or strength, of the onset is related to the height of the peak.

In [5], Scheirer states that the amplitude envelope does not contain all rhythmic information. Multiple nonlinear frequency bands must be analyzed separately and the results are to be combined at the end. To improve the simple onset detection introduced in the last paragraph, the signal can be split into six nonlinear bands using a bandpass filter bank. Onsets are still assumed to correspond to an increase in the amplitude envelope, not of the full-spectrum signal, but now of each bandpass signal. Therefore, for each bandpass signal the same onset detection procedure as described above can be performed. This results in onset data for each band. The data of the six bands must be combined. This is done by

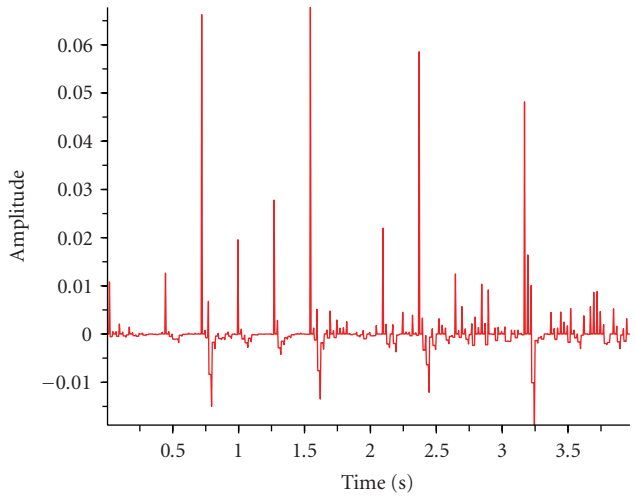

Figure 2: Differential of envelope of 4 seconds from "Maid of Orleans."

adding the onsets of all bands and combining onsets that are sufficiently close together. Such a multiple band approach gives better results for music, where no strong beats, such as base drums in electronic dance music, are present. A more advanced discussion of onset detection in multiple frequency bands is presented in [24].

All methods presented up to this point are based on detecting a sudden increase in signal energy. In recent years, phase based [25] or combined energy/phase approaches [26] introduced by Bello et al. have been shown to give better results than energy-only approaches. Basically, onset detection incorporating phase and energy, that is, operating in the complex domain, bases on the assumption that there is both a notable phase deviation and an energy increase when an onset occurs. Yet, to preserve the general and introductory nature of this overview and focus more on tempo detection, we will not go into details on these techniques.

For tempo detection from onset data mainly a histogram technique is used in the literature [2, 18]. The basic idea is the following: duration and weight of all possible IOIs are computed. Similar IOIs are grouped in clusters and the clusters are arranged in a histogram. From the weights and the centers of the clusters the tempo of several metrical levels can be determined. Dixon in [2] uses a simple rule-based method. Seppänen in [18] uses a more advanced method. He extracts only the tatum pulse level (fastest occurring tempo) directly from the IOI histogram, by picking the cluster with the center corresponding to the smallest IOI. Features in a window around each tatum pulse are extracted. Using Bayesian pattern recognition, the tatum pulses are classified with respect to their perceived accentuation. Thus, the beat level is detected by assuming that beats are more accented than offbeat pulses. Although Seppänen's work stops at the tatum level, the score level could be detected in the same way, assuming that beats at the beginning of a score are more accented than beats within.

We will now take a look at the second group of algorithms that attempt to find the tempo without explicitly detecting onsets. Still it is assumed that rhythmic events such as beats, percussion, or note onsets correspond to a change in signal amplitude in a few nonlinear bands. Again we start with either the envelopes or the differentials of the envelopes of the six frequency bands but omit the step of peak picking. To keep this overview general the term "detection function" [26] will be used in the ongoing, referring to either the envelope, its differential or any other function related to perceivable change in the signal.

The beat level tempo, which is what we are interested in at this point, can be viewed as a periodicity in the envelope function. A commonly used method to detect periodicities in a function is autocorrelation [8, 27]. The periodic autocorrelation is computed over a small window (10 seconds) of the envelope function. The index of the highest peak in the autocorrelation function (ACF) indicates the strongest periodicity. However, as findings in [28] suggest, the strongest periodicity in the signal may not always be the dominant periodicity perceived. The findings suggest an interval of preferred tapping linked to a supposed resonance between our perceptual and motor system. Still, as a first guess, which will work fairly well on music with strong beats in the preferred tapping range, the highest peak can be assumed to indicate the beat level tempo. We also have to combine the results from all bands. The simplest way is to add up the ACF of all bands and pick the highest peak in the summary ACF (SACF). Determining the tempo for each band and choosing the tempo that was detected in the majority of bands as the final tempo is an alternative method. Dixon describes a tempo induction method based on autocorrelation in [2]. Uhle et al. use autocorrelation for meter detection in [8].

An alternative to autocorrelation is a resonant filter bank consisting of resonators tuned to different frequencies (periodicities), first introduced for beat tracking by Scheirer in [5]. The detection function is fed to all resonators and the total output energy of each resonator is computed. In analogy to the highest autocorrelation peak, the resonator with the highest output energy matches the songs periodicity best and thus the beat level tempo is assumed to be its resonance frequency. As explained in the last paragraph, this assumption does not fully match our perception of rhythm. This is one reason why it is so difficult, even for most of state-of-the-art systems, to reliably detect the tempo on the beat level. Octave errors, that is, where double/triple or half/third the beat level tempo is detected, are very common according to [10]. Even human listeners in some cases do not agree on a common tapping level.

All the methods introduced so far require the extraction of a detection function. Publications exist discussing how such a detection function can be computed, considering signal processing theory [26] and applying psychoacoustic knowledge [24]. In order to bypass the issue of selecting a good detection function, a different periodicity detection approach as was introduced for tempo and meter analysis by Foote and Uchihashi [4] can be used. This approach is based on finding self-similarities among audio features. First, the audio data is split into small (20–40 milliseconds) overlapping windows. Feature vectors containing, for example, FFT

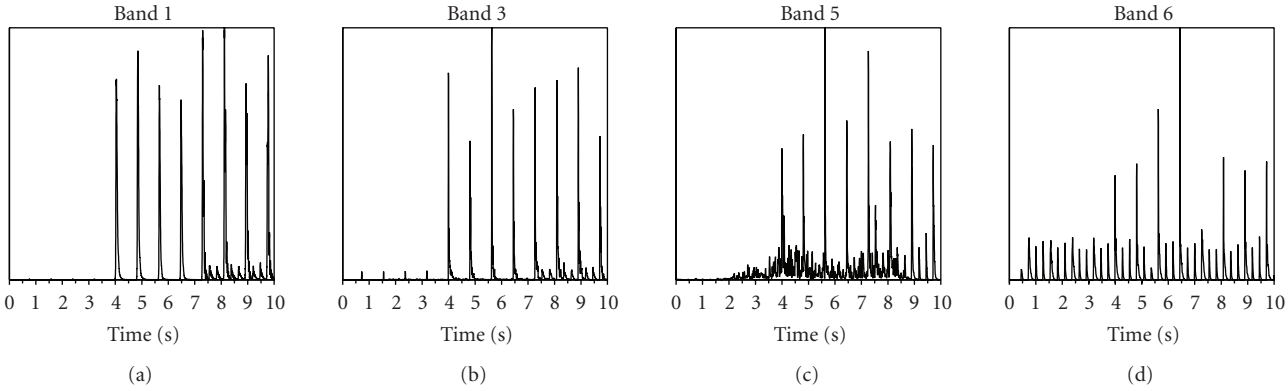

Figure 3: Differentials of frequency band envelopes from 10 seconds of "Maid of Orleans."

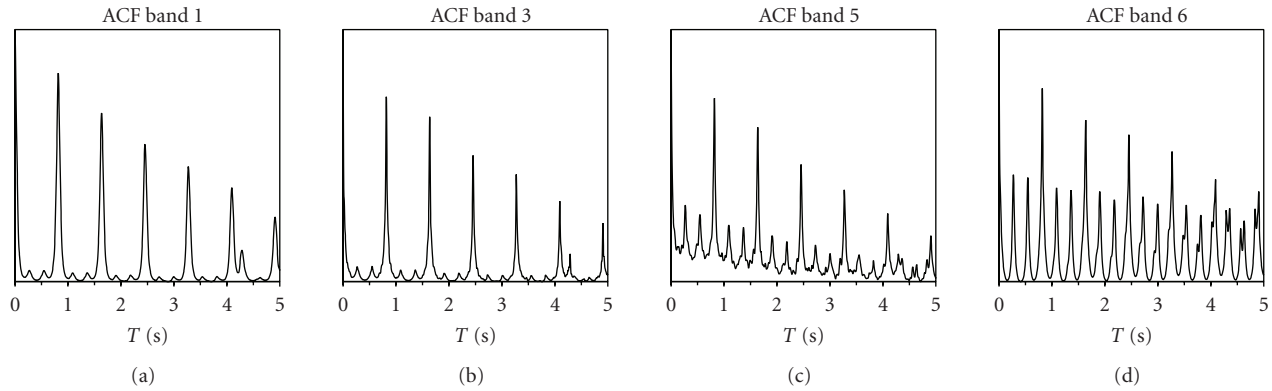

Figure 4: Periodic ACF of band envelope differentials from 10 seconds of "Maid of Orleans."

coefficients or MFCC [29] are extracted from these windows and a distance matrix **D** is computed by comparing every vector with all the remaining vectors via a distance measure or cross-correlation.

Using (1), a so called beat spectrum [4] B can be computed from the distance matrix **D**. This beat spectrum is comparable to the ACF or the output of the resonant filter bank in the previously discussed methods;

$$B(\mathrm{IOI}) = \sum_{k=1}^{K} \mathbf{D}_{k,k+\mathrm{IOI}}. \tag{1}$$

While still the choice of the feature set might have an influence on the performance, this method has an advantage over computing the ACF of a detection function. In computing the correlation or distance of every feature vector to every other feature vector all possible relations between all features in all feature vectors are accounted for. Detection functions for separate frequency bands can only account for (temporal) relations within each band. If the detection function is a sum over all bands, for example, relations between the frequency bands are accounted for, but only in a very limited way. This case would correspond to reducing the feature vector to one dimension by summing its elements before computing the distance matrix.

However, computing distance matrices is a very time consuming task and might thus not be applicable to live applications, for example, that demand real-time algorithms. For most mainstream music, it can be assumed that the sensation of tempo corresponds to a loudness periodicity, as can be represented by a single detection function or a set of detection functions for a few subbands. Therefore, even though in our opinion the distance matrix method seems to be the theoretically most advanced method, it is not used in the rhythm analysis method presented in the following.

In the remaining part of this overview section we will give a very short overview over selected meter detection and ballroom dance style and genre recognition methods.

Various work exists on the subject of genre recognition, for example, [30, 31]. The basic approach is to extract a large number of features representing acoustic properties for each piece of music to be classified. Using a classifier trained on annotated training data, the feature vectors extracted from the songs are assigned a genre. Reference [30] extracts features related to timbral texture, rhythmic content and pitch content. The rhythmic features are extracted from the result of autocorrelation of subband envelopes. As classifiers Gaussian mixture models (GMMs) and K-nearest-neighbour (K-NN) are investigated, a discrimination rate of 61% for 10 musical genres is reported. Reference [31] investigates the use of a large open feature sets and automatic feature selection combined with support vector machines as classifiers. A success rate of 92.2% is reported for discrimination between 6 genres.

The subject of ballroom dance style recognition is relatively new. Gouyon et al. have published a data-driven approach to ballroom dance style recognition in [12]. They test various features extracted from IOI histograms using 1-NN classification. The best result is achieved with 15 MFCC like descriptors computed from the IOI histogram. 90.1% accuracy is achieved with these descriptors plus the ground truth tempo by 1-NN classifiers. Without ground truth tempo, that is, only the 15 descriptors, 79.6% accuracy is reported.

Meter detection requires tempo information from various metrical levels. Klapuri et al. introduce an extensive method to analyze audio on the tatum, pulse, and measure level [6]. For each level, the period is estimated based on periodicity analysis using a comb filter bank. A probabilistic model encompasses the dependencies between the metrical levels. The method is able to deal with changing metrical structures throughout the song. It proves robust for phase and tempo on the beat level, but still has some difficulties on the measure level. The method is well suited for, in depth, metrical analysis of a wide range of musical genres. For a limited set of meters, for example, as in ballroom dance music the complexity can be reduced—at the gain of accuracy—to binary decisions between duple or triple periods on the measure level. Gouyon et al. assume a given segmentation of the song on the beat level and then focus on a robust discrimination between duple and triple meter [11] on the measure level. For each beat segment, a set of low-level descriptors is computed from the audio. Periodic similarities of each descriptor across beats are analyzed by autocorrelation. From the output of the autocorrelation, a decisional criterion M is computed for each descriptor, which is used as a feature in meter classification.

3. RHYTHM ANALYSIS

A data-driven rhythm analysis approach is now introduced, capable of extracting rhythmic features, robustly identifying duple and triple meter, quarter-note tempo and ballroom dance style basing on 82 rhythmic features, which are described in the following sections.

Robustly identifying the quarter-note or beat level tempo is a challenging task, since octave errors, that is, where double or half of the true tempo is detected, are very common. Therefore, a new tempo detection approach, based on integrated ballroom dance style recognition, is investigated.

The tatum tempo [8, 18], that is, the fastest tempo, presents the basis for extracting rhythmic features. A resonator-based approach, inspired by [5], is used for detecting this tatum tempo and extracting features containing information about the distribution of resonances throughout the song.

The features are used to decide whether the song is in duple or triple meter. Confining the metrical decision to a binary one was introduced in [11]. For dance music, the discrimination between duple and triple meter has the most practical significance. Identifying various time signatures, such as 2/4, 4/4, and 6/8 is a more complicated task and of less practical relevance for ballroom dance music. The rhythmic features are further used to classify songs into 9 ballroom dance style classes. These results will be used to assist the tempo detection algorithm by providing information about tempo distributions collected from the training data for the corresponding class. For evaluation 10-fold stratified cross-validation is used. This is described in more detail in Section 3.5.

3.1. Comb filter tempo analysis

The approach for tatum tempo analysis discussed in this article is based on Scheirer's multiple resonator approach [5] using comb filters as resonators. His approach has been adapted and improved successfully in other work for tempo and meter detection [6, 10, 32]. The main concept is to filter the envelopes or detection functions (see Section 2) of six nonlinear frequency bands through a bank of resonators. The resonance frequency of the resonator with the highest output energy is chosen as tempo. The comb filters used here are a slight variation of Scheirer's filters. In the following paragraphs, there will be a brief theoretical discussion of IIR comb filters and a description of the chosen filter parameters.

In the ongoing, the symbol θ will be used to denote a tempo. The tempo is specified as a frequency having the unit BPM (beats per minute). If an index IOI is appended to the symbol θ, it is indicated that the tempo is given as IOI period in frames.

A comb filter adds a signal itself to a delayed version of the signal. Every comb filter is characterized by two parameters: the delay (or period, which is the inverse of the filters resonance frequency) τ and the gain α.

For tempo detection IIR comb filters are used as described in the discrete time domain by (2),

$$y[t] = (1-\alpha)\cdot u[t] + \alpha \cdot y[t-\tau]. \tag{2}$$

The filter has a transfer function in the z-domain given by (3),

$$H(z) = \frac{1-\alpha}{1-\alpha\cdot z^{-\tau}}. \tag{3}$$

The frequency response $H(z)$ for two exemplary values of α is depicted in Figure 6.

To achieve optimal tempo detection performance, an optimal value for α must be determined. Scheirer's [5] method of constant half-energy time by using variable gain α depending on τ has not proven well in our test runs. Instead, we use a fixed value for α. When choosing this value, we have to consider small temporary tempo drifts occurring in most music performances. So the theoretically optimal gain $\alpha \to 1$ cannot be used. We conducted test runs with multiple values for α in the range from 0.2 to 0.99. Best results were obtained with $\alpha = 0.7$.

3.2. Feature extraction

The comb filters introduced in the previous section are used to extract the necessary features for ballroom-dance style recognition, meter recognition, and tempo detection.

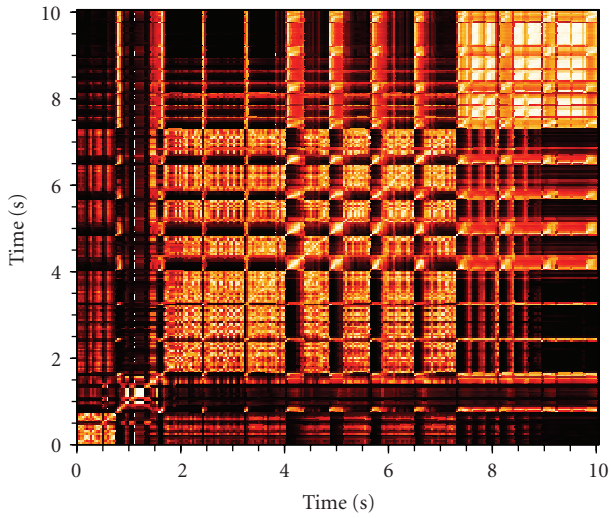

Figure 5: Distance matrix for 10 seconds from the beginning of "Maid of Orleans" (OMD). White spots have a high correlation (or low distance) and black spots a low correlation (or high distance).

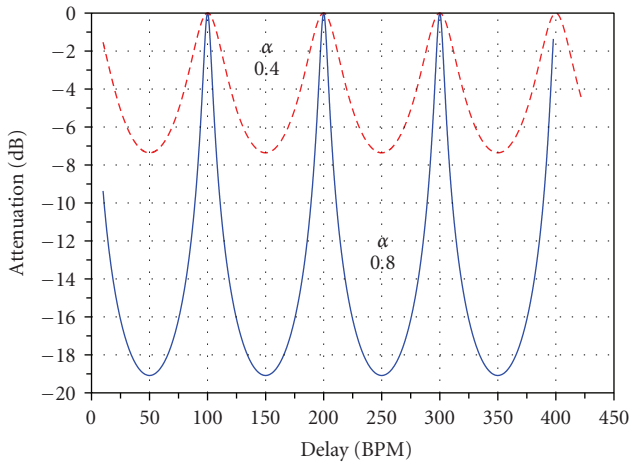

Figure 6: Frequency responses of IIR comb filters with gain of $\alpha = 0.8$ and $\alpha = 0.4$.

The key concept is to set up comb filter banks over a much broader range than used by [5] in order to include higher metrical layers. The resulting features describe the distribution of resonances among several metrical layers, which provides qualitative information about the metrical structure.

To effectively reduce the number of comb filters required, we exploit the fact that in music performances several metrical layers are present (see Section 2). In a first step the tempo on the lowest level, the tatum tempo, is detected. It is now assumed that all possibly existing higher metrical levels can only have tempi that are integer multiples of the tatum tempo. This is true for a wide variety of musical styles.

3.2.1. Preprocessing

The input data is down sampled to $f_s = 11.025\,\text{kHz}$ and converted into monophonic by stereo-channel addition in order to reduce computation time. The input audio of length L_i seconds is split into $N_{\text{frames}} = 100 \cdot L_i$ frames of $N_{s,\text{block}} = 256$ samples with an overlap of 0.57, resulting in a final envelope frame rate of 100 fps (frames per second). A Hamming window w_x is applied to each frame and a fast Fourier transform (FFT) of the frame is computed, resulting in 128 FFT coefficients.

By using N_{mel} overlapping triangular filters, equidistant on the mel-frequency scale, the 128 FFT coefficients are reduced to N_{mel} envelope samples of N_{mel} nonlinear bands. These triangular filters are the same as used in speech recognition for the computation of MFCC [29].

Such a small set of frequency bands, still covering the whole human auditory frequency range, contains the complete rhythmic structure of the musical excerpt, according to experiments conducted in [5].

The envelope samples $x_{\nu,i}$ of each mel-frequency band ν are converted to a logarithmic representation according to the following equation:

$$x_{\nu,i,\log} = 10.0 \cdot \log(x_{\nu,i} + 1.0). \tag{4}$$

The envelopes $\underline{x}_\nu$ of the mel-frequency bands are then lowpass filtered by convolution with a half-wave raised cosine filter with a length of 15 envelope samples, equal to 150 milliseconds. The impulse response of the filter is given in (5). This filter preserves fast attacks, but filters noise and rapid modulation, most as in the human auditory system,

$$h_{\text{rclp}}(i) = \cos\left(\frac{\pi i}{15}\right) + 1, \quad i \in [1; 15]. \tag{5}$$

Of each lowpass filtered mel-frequency band envelope ν a weighted differential d_ν is taken according to (6). For a sample $x_{\nu,i}$ at position i a moving average is calculated over one window of 10 samples to the left of sample $x_{\nu,i}$ (left mean $\overline{x}_{\nu,i,l}$) and a second window of 20 samples to the right of sample $x_{\nu,i}$ (right mean $\overline{x}_{\nu,i,r}$),

$$d_\nu(i) = (x_{\nu,i} - \overline{x}_{\nu,i,l}) \cdot \overline{x}_{\nu,i,r}. \tag{6}$$

This method is based on the fact that a human listener perceives note onsets as more intense if they occur after a longer time of lower sound level and thus are not affected by temporal post-masking caused by previous sounds [33]. The weighting with the right mean $\overline{x}_{\nu,i,r}$ incorporates the fact that note duration and total note energy play an important role in determining the perceived note accentuation [18].

3.2.2. Tatum features

For detecting the tatum tempo θ_T, an IIR comb filter bank is used consisting of 57 filters, with gain $\alpha = 0.7$ and delays ranging from $\tau_{\min} = 18$ to $\tau_{\max} = 74$ envelope samples. This filter bank is able to detect tatum tempos in the range from 81 to 333 pulses per minute. The range might need adjustments when very slow music is processed, that is, music with no tempo faster than 81 pulses per minute.

The weighted differential d_ν of each mel-frequency band envelope ν is fed as input $\underline{u}_\nu$ to each filter $h_{\nu,\tau}$ having a delay

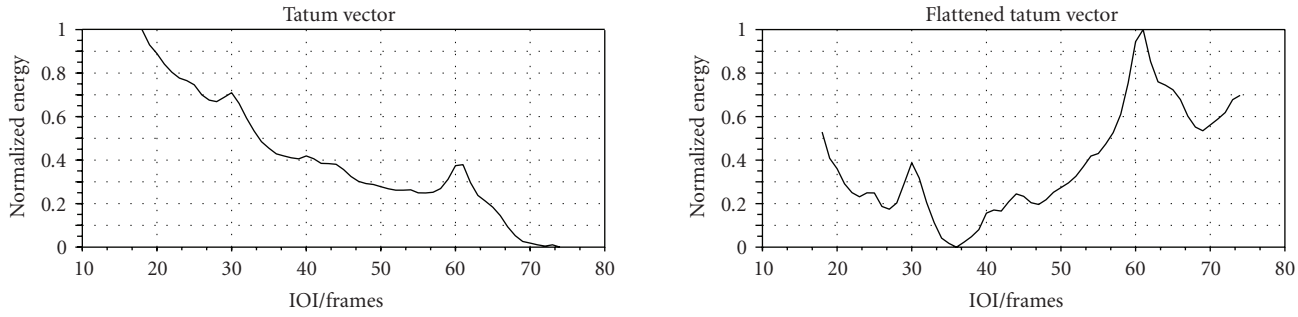

Figure 7: Plots of $\underline{T}'$ (a) and flattened tatum vector $\underline{T}$ (b) for "Celine Dion - My Heart Will Go On".

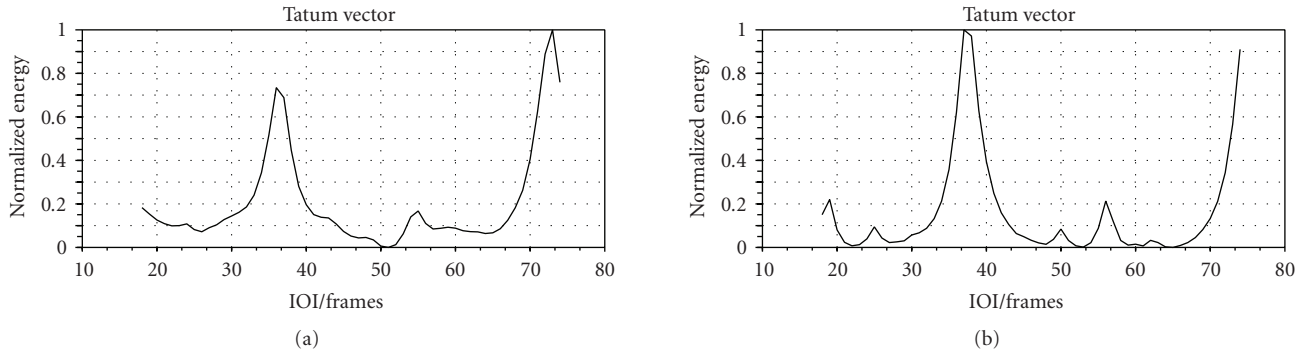

Figure 8: Plots of $\underline{T}'$ for "Moon River (Waltz, triple meter)" (a) and "Hit the Road Jack (Jive, duple meter)" (b).

of τ. The filter output for band ν, frame n and filter $h_{\nu,\tau}$ is referred to as $y_n^{(\tau,\nu)}$. The total energy output $\underline{T}'(\tau - \tau_{\min} + 1)$ over all bands is computed for each filter $h_{\nu,\tau}$ with (7),

$$\underline{T}'(\tau - \tau_{\min} + 1) = \sum_{\nu=0}^{N_{\text{mel}}} \sum_{n=0}^{N_{\text{frames}}} y_n^{(\tau,\nu)}. \tag{7}$$

The result of this step is the not flattened tatum vector $\underline{T}'$ with 57 elements $\underline{T}'(\tau - \tau_{\min} + 1)$, where τ is in the range from 18 to 74. Examples of $\underline{T}'$ for three songs are plotted in Figures 7 and 8.

From $\underline{T}'$ three additional features are extracted that reveal the quality of the peaks.

(i) T_{ratio} is computed by dividing the highest value by the lowest.

(ii) T_{slope} is the fraction of the first value over the last value.

(iii) T_{peakdist} is computed as mean of the maximum and minimum value normalized by the global mean.

These features correspond to how clearly visible the peaks of the vector $\underline{T}'$ are, and how flat $\underline{T}'$ is (see Figures 7 and 8). Since the employed comb filters tend to higher resonances at higher tempos for songs with little rhythmic content (Figure 7), the vector is adjusted, that is, flattened, by considering the difference between the average of the first 6 values and the average of the last 6 values. From the resulting flattened tatum vector $\underline{T}$ the two most dominant peaks are picked as follows. Firstly, all local minima and maxima are detected, then for each maximum its apparent height D is computed by taking the average of the maximum minus its left and right minimum. The indices of the two maxima with the greatest apparent height D are considered possible tatum candidates ($\theta_{T1,\text{IOI}}$ and $\theta_{T2,\text{IOI}}$). For each candidate $\theta_{T1/2,\text{IOI}}$ a confidence $C_{T1/2,\text{IOI}}$ is computed as follows:

$$C_{T1/2} = D_{T1/2} + \underline{T}(\theta_{T1/2,\text{IOI}}). \tag{8}$$

The candidate $\theta_{T1/2,\text{IOI}}$ for which the confidence $C_{T1/2}$ is maximal is called the final tatum tempo θ_T in the ongoing. Conversion from the IOI period $\theta_{T,\text{IOI}}$ of the final tatum tempo to the final tatum tempo in BPM (θ_T) is performed by the following equation:

$$\theta_T = \frac{6000}{\theta_{T,\text{IOI}}}. \tag{9}$$

The 63 tatum features consisting of θ_T, θ_{T1}, θ_{T2}, T_{ratio}, T_{slope}, T_{peakdist}, and the tatum vector $\underline{T}$ with 57 elements constitute the first part of the rhythmic feature set. A major difference to some existing work is the use of the complete tatum vector in the feature set. Reference [30] uses rhythmic features for genre classification. However, from a beat histogram, which is loosely comparable to the tatum vector (both contain information about the periodicities), only a small set of features is extracted, only considering the two highest peaks and the sum of the histogram.

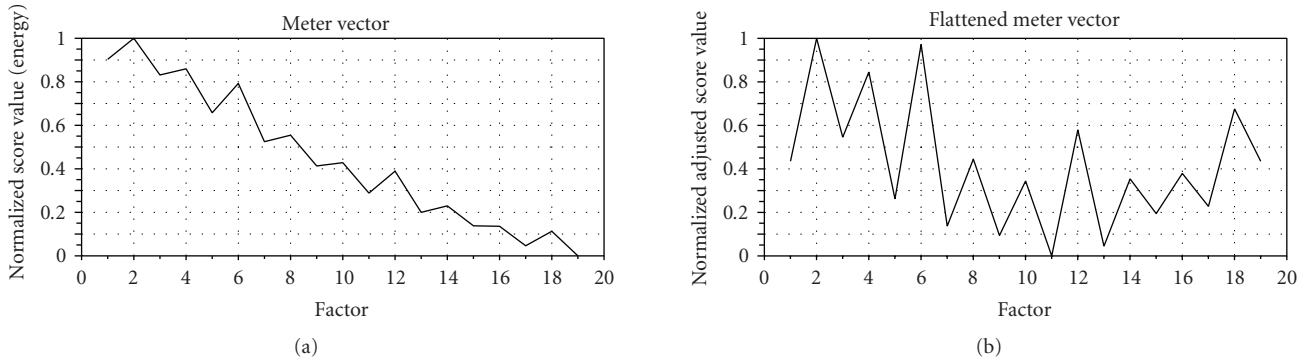

Figure 9: Plots of not flattened meter vector $\underline{m}'$ (a) for "Moon River (Waltz)" and (flattened) meter vector $\underline{m}$ (b).

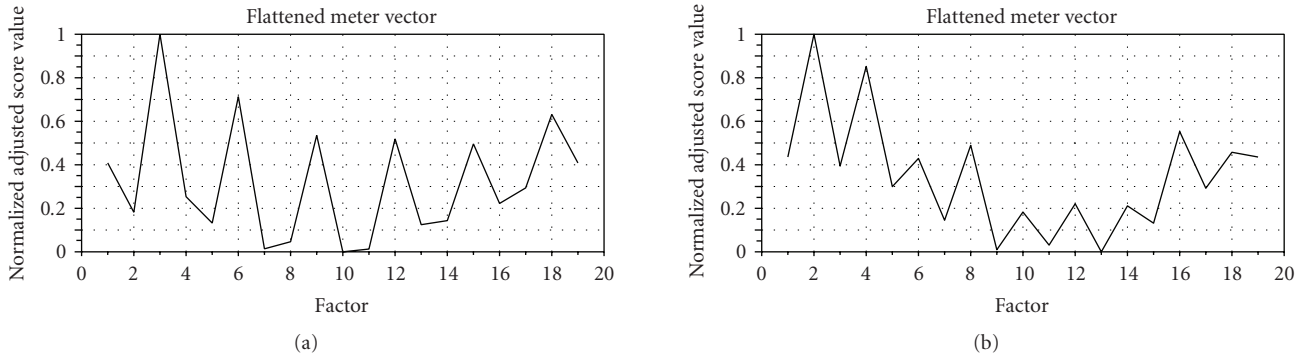

Figure 10: Plots of meter vector $\underline{m}$ for "Maid Of Orleans (3/4)" (a) and "Hit the Road Jack (4/4)" (b).

3.2.3. Meter features

The tatum features only contain information from a very small tempo range, hence, they are not sufficient when one is interested in the complete metrical structure and other tempi than the tatum tempo. Thus, features that contain information about tempo distributions over a broader range are required. These are referred to as meter features, although they do not contain explicit information about the meter.

A so called meter vector $\underline{m}$ is introduced. This vector shows the distribution of resonances among 19 metrical levels, starting at, and including the tatum level.

Each of the 19 elements m_i of vector $\underline{m}$ is a normalized score value of the tempo $\theta_T \cdot i$, indicating how well the tempo $\theta_T \cdot i$ resonates with the song. To compute m_i, first an unnormalized score value m_i' is computed by setting up a comb filter bank for each value of $i \in [1; 19]$. Each filter bank consists of $2i + 1$ filters with delays from $(\theta_{T,\mathrm{IOI}} \cdot i - i)$ to $(\theta_{T,\mathrm{IOI}} \cdot i + i)$. As in Section 3.2.2 the total energy output of each filter in the bank is computed and the maximum value is assigned to m_i'. The delay τ of the filter with the highest total energy output is saved as adjusted tempo $\theta_{i,\mathrm{IOI}}$ belonging to m_i'. The vector consisting of the 19 elements m_i' is the not flattened meter vector $\underline{m}'$. Exemplary plots of $\underline{m}'$ are given in Figures 9 and 10;

$$m_i' = \max_{j \in [-i;+i]} \left(\sum_{\nu=0}^{N_{\mathrm{mel}}} \sum_{n=0}^{N_{\mathrm{frames}}} y_n^{(\theta_T \cdot i + j, \nu)} \right). \tag{10}$$

As the same problem with higher resonances of higher tempi as exists for the tatum vector (see Section 3.2.2) also exists for $\underline{m}'$ (see Figure 9), the vector $\underline{m}'$ is flattened in the same way as the tatum vector by taking into account the difference $m_{19}' - m_1'$. The resulting vector is the flattened meter vector $\underline{m}$, referred to simply as meter vector. For accurate meter vector computation a minimal input length $L_i = \tau_{\max} \cdot 19 \approx 14\,\mathrm{s}$ is required, since the higher metrical levels correspond to very slow tempi and thus large comb filter delays.

The 19 elements of the meter vector $\underline{m}$, without further processing or reduction, constitute the second part of the rhythmic feature set. We would like to note at this point, that no explicit value for the meter (i.e., duple or triple) is part of the meter features. In the ongoing the reader will learn how the meter is detected in a data-driven manner using support vector machines (SVMs).

3.3. Feature selection

A total of 82 features has been described in the previous two sections, including all 19 meter vector elements m_i and the 63 tatum features, namely θ_T, θ_{T1}, θ_{T2}, T_{ratio}, T_{slope}, T_{peakdist} plus all 57 elements of tatum vector $\underline{T}$ (see Table 1). These features will be referred to as feature set FS_R in the ongoing. Basing on our experience in [31, 32], SVMs with a polynomial Kernel function of degree 1 are used for the following classification tasks. The SVMs are trained using a sequential minimum optimization (SMO) method as described in [34].

In order to find relevant features for meter and ballroom dance style classification, the BRD dataset is analyzed for each of these two cases by performing a closed-loop

Table 1: Overview over all 82 rhythmic features. Feature set FS_R.

Tatum features	tatum vector $\underline{T}$ (57 el.)
	tatum candidates θ_{T1}, θ_{T2} [BPM]
	final tatum tempo θ_T [BPM]
	T_{ratio}, T_{slope}, T_{peakdist}
Meter features	Meter vector $\underline{m}$ (19 el.)

Table 2: Mean μ, standard deviation σ, minimum and maximum tempo in BPM for each class, and complete set BRD.

Tempo [BPM]	μ	σ	min	max
All	128.5	38.7	68	208
Cha Cha	122.0	6.5	92	136
Foxtrot	114.8	2.1	104	116
Jive	165.9	11.5	124	176
Quickstep	200.7	6.7	153	208
Rumba	97.7	8.3	76	141
Samba	100.7	8.8	68	202
Tango	127.4	3.2	112	136
Viennese Waltz	177.1	2.3	168	186
Slow Waltz	86.2	1.7	72	94

hill-climbing feature selection employing the target classifier's error rate as optimization criterion, namely, sequential forward floating search (SVM-SFFS) [31].

The feature selection reveals the following feature subset FS_M to yield the best results for meter classification: T_{ratio}, meter vector $\underline{m}$ elements 4, 6, 8, 16, and the tatum vector $\underline{T}$.

For ballroom dance style classification the feature selection reveals the following feature subset FS_D to yield the best results: meter M (see Section 3.5), T_{ratio}, T_{slope}, T_{peakdist}, meter vector $\underline{m}$ elements 4–6, 8, 11, 12, 14, 15, 19, and the tatum vector $\underline{T}$ excluding elements 21 and 29.

3.4. Song database

A set of 1855 pieces of typical ballroom and Latin dance music obtained from [35] is used for evaluation. A more detailed list of the 1855 songs can be found at [36]. The set covers the standard dances Waltz, Viennese Waltz, Tango, Quick Step, and Foxtrot, and the Latin dances Rumba, Cha Cha, Samba, and Jive giving a total of 9 classes. The songs have a wide range of tempi ranging from 68 BPM to 208 BPM. 30 seconds of each song are available, which were converted from a real audio like format to 44.1 kHz PCM, so the preprocessing from Section 3.2.1 can be applied. In total length however, this set corresponds to 5 days of music. The distribution among dance styles is depicted in Table 3. This set is abbreviated BRD in the ongoing. Ground truth statistics about the tempo distribution for the whole set and in each dance style class are given in Table 2.

For the BRD dataset, the ground truth of tempo and dance style is known from [35]. The ground truth regarding duple or triple metrical grouping is also implicitly known from the given source because it can be deduced from the dance style. All Waltzes have triple meter, all other dances have duple meter. Tempo ground truths are not manually double checked as performed in [10], therefore errors among the ground truths might be present. Results with manually checked ground truths might improve slightly. This is further discussed near the end of Section 4.

3.5. Data-driven meter and ballroom dance style recognition

From the abstract features in set FS_R (see Section 3.3) meter and quarter-note tempo have to be extracted. While data-driven meter recognition by SVM yields excellent results, data-driven tempo detection is a complicated task because tempo is a continuous variable. An SVM regression method was investigated, but has not proven successful. The method was not able to correctly identify tempi within a tolerance of only a few percent relative BPM deviation. A hybrid approach is used therefore the data is divided into a small number of classes representing tempo ranges. The ranges are allowed to overlap slightly. As the database described in Section 3.4 already has one of nine ballroom dance styles assigned to each instance, the dance styles are chosen as the tempo classes, since music of the same dance style generally is limited to a specific tempo range. This is confirmed by other work, which uses tempo ranges to assign a ballroom dance style [2, 37].

In three consecutive steps (see Figure 11) meter, ballroom dance style, and quarter-note tempo are determined for the whole dataset in a 10-fold stratified cross validation (SCV) as described in the following.

(1) The feature set FS_R is extracted for all instances in the dataset. The 1855 instances are split into training and test splits for 10 stratified folds. An SVM model for meter classification is built on each training split using the feature subset FS_M. The model is used to assign a meter M (duple or triple) to the instances in each test split. Doing this for all 10 folds, the meter M can be determined for the whole dataset by SVM classification.

(2) The meter M, from the previous step, is used as a feature in feature set FS_D (see Section 3.3) for ballroom dance style classification. The same 10-fold procedure as was used for meter classification in step 1 is performed in order to assign a ballroom dance style to all instances in the BRD dataset.

(3) With the results of both meter and ballroom dance style classification, it is now possible to quite robustly detect the quarter-note tempo. The following section describes the novel tempo detection procedure in detail.

3.6. From ballroom dance style to tempo

For the training data of each of the 10 folds introduced in the previous section, the means $\mu_{q/T}$ and variances $\sigma^2_{q/T}$ of the distributions of quarter-note tempi (ground truths) and tatum tempi θ_T are computed for each of the 9 ballroom dance styles. No ground truth for the tatum tempo is

Table 3: Results obtained on dataset BRD for meter M, quarter-note tempo θ_q, and ballroom dance style (BDS).

Accuracy [%]	ChaCha	Foxtrot	Jive	Quickst.	Rumba	Samba	Tango	V. Waltz	Waltz	MEAN
Instances no.	211	245	138	242	217	188	185	136	293	
Meter	99.1	97.6	97.8	99.6	90.8	98.9	98.4	97.8	94.2	**96.9**
Tempo	97.2	93.9	97.1	96.3	90.3	93.6	94.1	92.6	81.8	**92.4**
Tempo octave	94.8	93.5	90.6	87.6	81.6	86.2	93.5	91.2	81.8	**88.5**
BDS precision	93.0	94.7	90.4	87.9	78.2	89.8	88.0	94.0	88.3	**89.1**
BDS recall	87.7	95.5	88.4	90.1	77.9	84.0	91.4	91.9	93.2	**89.1**
BDS F_1	90.2	95.1	89.4	89.0	78.1	86.8	89.7	92.9	90.7	**89.1**

available, so the automatically extracted tatum tempo (see Section 3.2.2) from step (1) in Section 3.5. is used. Results might improve further if ground truth tatum information were available, since correct tatum detection is crucial for correct results.

For the test data in each fold the tempo is detected with the following procedure. Using the two tatum candidates θ_{T1} and θ_{T2} extracted in step (1) in Section 3.5, the final tatum for the instances in the test split in each fold now is chosen based upon the statistics estimated from the training data. The Gaussian function $G(\theta_{T1/2})$ (11) is used instead of the confidence $C_{T1/2}$ (see Section 3.2.2). Parameters μ and σ^2 are set to the values of μ_T and σ_T^2 for the corresponding ballroom dance style (assigned in step (2) in the previous subsection),

$$G(\theta) = \exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right). \tag{11}$$

Now the candidate $\theta_{T1/2}$ for which the function $G(\theta_{T1/2})$ is maximal is chosen as the final tatum tempo θ_{T*}. Based upon this new tatum, a new flattened meter vector $\underline{m}^*$ is computed for all instances as described in Section 3.2.3.

The new meter vector $\underline{m}^*$ is used for detection of the quarter-note tempo. Each element m_i^* is multiplied by a Gaussian weighting factor $G(\theta_i)$. The parameters μ and σ^2 in (11) are now set to the values μ_q and σ_q^2 of the corresponding ballroom dance style. θ_i indicates the tempo the meter vector element m_i^* belongs to (see Section 3.2.3).

Next, the index $i_{\max}$, for which the expression $m_i^* \cdot G(\theta_i)$ is maximized, is identified. The tempo $\theta_{i_{\max}}$ belonging to index $i_{\max}$ is the detected quarter-note (beat level) tempo θ_q.

4. RESULTS

Results for tempo detection with and without prior ballroom dance style recognition are compared in Table 4. The tempo thereby is detected as described in Section 3.6, except that without dance style only one predefined Gaussian for the tempo distribution is applied, instead of using the distributions determined for each dance style.

By the results in Table 4, it can be clearly seen that the number of instances, where the correct tempo octave is identified, increases by almost 20% absolute, when incorporating the ballroom dance style recognized in step (2). When assuming an optimal ballroom dance style recognition, that is, when ground truth ballroom data is used instead of the recognition results, the tempo octave is identified correctly in almost all cases, where the tempo is identified correctly. With the new data-driven approach to tempo detection, accuracies for the quarter-note tempo are improved by approximately 5% absolute for Waltz and over 10% for Viennese Waltz, compared to previous work on the same dataset [15]. On 88% of all instances the correct tempo octave was identified, which is remarkble, considering the wide range of tempi of the dataset.

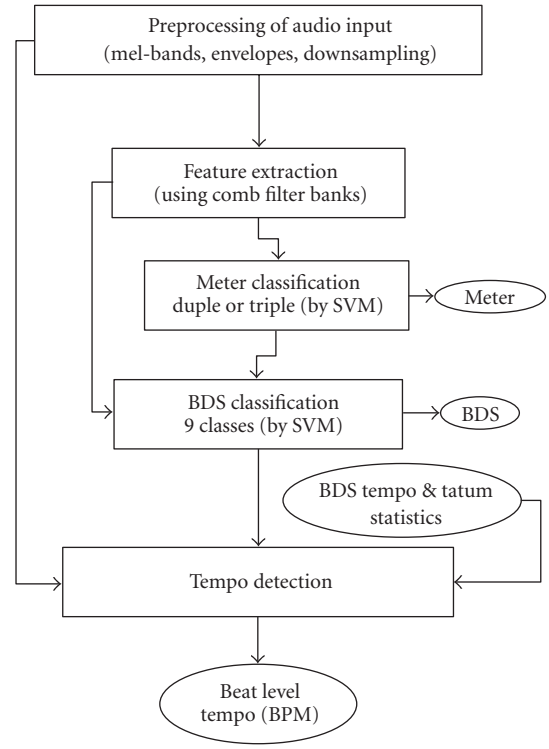

Figure 11: Steps for data-driven tempo detection basing on meter and ballroom dance style recognition.

Table 4: Comparison of tempo detection without (w/o BDS), with incorporated ballroom dance style recognition (w BDS) and using ground truth ballroom classes to simulate optimal BDS recognition (gt BDS).

Accuracy [%]	w/o BDS	w BDS	gt BDS
Tempo	88.8	92.4	93.1
Octave	70.0	88.5	93.0

Table 5: Results on set BRD_{04} for tempo detection without (w/o BDS), with incorporated ballroom dance style recognition (w BDS) and using ground truth ballroom classes (gt BDSs).

Accuracy [%]	w/o BDS	w BDS	gt BDS
Tempo (acc. 1)	88.8	**93.0**	**92.8**
Octave (acc. 2)	69.8	**86.9**	**92.4**

Detailed final results, after applying all the steps from Section 3.5 through Section 3.6, are depicted in Table 3. The tolerance for tempo detection hereby is 3.5% relative BPM deviation to maintain consistency with previous publications [32]. We would like to note that ballroom dance style recognition has been performed completely without using the quarter-note tempo as a feature.

In [2], Dixon et al. use a rule-based approach for dance style classification basing on simple tempo ranges. However, results on a large dataset are not reported. In [12], Gouyon et al. test a data-driven approach on a subset of the BRD dataset. They evaluate multiple feature sets and different classifiers. Using ground truth of tempo and meter from [35] with a K-nearest neighbour classifier, they report an accuracy of 82.3%. Using the same ground truths and SVM instead of k-NN, we achieve 84.6% of correctly classified instances. With a set of 15 MFCC-like features, comparable to our 82 rhythmic features, Gouyon et al. achieve accuracies of 79.6%. Using SVM on the rhythmic features introduced in this article, the ballroom dance style recognition results improve by almost 10% absolute to 89.1%.

Meter detection results improve by approximately 2% over those reported by Gouyon et al. in [11]. However, different datasets and classifiers are used, so results cannot be properly compared. Comparing meter detection results with those reported by Klapuri et al. [6] is not feasible because in our article meter detection is restricted to a simple binary decision due to the main focus being on tempo detection incorporating ballroom dance style recognition. Klapuri et al. describe more in detail, multilevel tempo and meter analysis system.

At ISMIR 2004 a tempo induction contest was held comparing state-of-the-art tempo induction algorithms. The results are reported in [10]. To show the reader how our data-driven tempo induction approach compares to the algorithms of the contest participants, we have conducted a test run on the publicly available ballroom dance set used in the contest (referred to as set BRD_{04} in the ongoing, obtainable at [38]). This set approximately is a subset of the BRD dataset. The tempo ground truth of this set was manually double checked. Two accuracies are evaluated in [10], namely accuracy 1 which corresponds to tempo correct in this article, and accuracy 2, which corresponds to the percentage of correctly identified tempo octaves. Table 5 shows the results obtained on this dataset. The winner of the ISMIR contest is an algorithm by Klapuri et al. which achieves 91.0% accuracy 1 and 63.2% accuracy 2 on the BRD_{04} set. Scheirer's algorithm, on which our comb filter tatum detection stage is loosely based, was also evaluated in the contest. It achieves 75.1% accuracy 1 and 51.9% accuracy 2 on the same dataset. The novel approach presented in this article outperforms Scheirer's algorithm by 17.9% absolute and Klapuri's algorithm by 2.0% absolute regarding accuracy 1 and 35.0% and 23.7% absolute, respectively regarding accuracy 2. These results are the best reported so far. Still, it is to note that tests were only performed on ballroom dance data. In future work, other datasets such as the song set from [10] or the MTV set from [32] must be assigned ground truth tempo range classes, in order to evaluate performance with other data than ballroom songs. Yet already, good results on ballroom dance music are practically useable, for example, for virtual dance assistants [15].

5. CONCLUSION AND OUTLOOK

Within this article, an overview over basic and current approaches for rhythm analysis on real audio was given. Further, a method to improve over today's robustness by combining tempo detection, rhythmic feature extraction, meter recognition, and ballroom dance style recognition in a data-driven manner was presented. As opposed to other work, ballroom dance style classification is carried out first, and significantly boosts performance of tempo detection. 82 rhythmic features were described and their high usefulness for all of these tasks was demonstrated.

Further applications for these features, ranging from general genre recognition to song identification [13], or measuring rhythmic similarity [39], must be investigated. Preliminary test runs for discrimination between 6 genres (Documentary, Chill, Classic, Jazz, Pop-Rock, and Electronic) on the same dataset, and with same test-conditions as used in [31] indicate accuracies of up to 70% using only the 83 rhythmic features.

It will further be investigated if adding other features, such as those described by [8, 12], or [13] can further improve results for all the presented rhythm analysis steps. Moreover, the data-driven tempo detection approach will be extended to nonballroom music, for example, popular and rock music.

Overall, automatic tempo detection on real audio—also outside of electronic dance music—has matured to a degree, where it is ready for multiple intelligent Music Information Retrieval applications in everyday life.

REFERENCES

[1] J. Downie, "Music information retrieval," *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 295–340, 2003.

[2] S. Dixon, E. Pampalk, and G. Widmer, "Classification of dance music by periodicity patterns," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR '03)*, pp. 159–165, Baltimore, Md, USA, October 2003.

[3] N. Hu, R. B. Dannenberg, and G. Tzanetakis, "Polyphonic audio matching and alignment for music retrieval," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '03)*, pp. 185–188, New Paltz, NY, USA, October 2003.

[4] J. Foote and S. Uchihashi, "The beat spectrum: a new approach to rhythm analysis," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '01)*, pp. 881–884, Tokyo, Japan, August 2001.

[5] E. D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *Acoustic Society of America*, vol. 103, no. 1, pp. 588–601, 1998.

[6] A. P. Klapuri, A. J. Eronen, and J. T. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Transactions on Speech and Audio Processing*, vol. 14, no. 1, pp. 342–355, 2006.

[7] N. Orio, "Music retrieval: a tutorial and review," *Foundations and Trends in Information Retrieval*, vol. 1, no. 1, pp. 1–90, 2006.

[8] C. Uhle, J. Rohden, M. Cremer, and J. Herre, "Low complexity musical meter estimation from polyphonic music," in *Proceedings of the 25th International Conference on the Audio Engineering Society (AES '04)*, pp. 63–68, London, UK, June 2004.

[9] M. Goto and Y. Muraoka, "Real-time rhythm tracking for drumless audio signals—chord change detection for musical decisions," in *Proceedings of of IJCAI-97 Workshop on Computational Auditory Scene Analysis (CASA '97)*, pp. 135–144, Nagoya, Japan, August 1997.

[10] F. Gouyon, A. P. Klapuri, S. Dixon, et al., "An experimental comparison of audio tempo induction algorithms," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1832–1844, 2006.

[11] F. Gouyon and P. Herrera, "Determination of the Meter of musical audio signals: seeking recurrences in beat segment descriptors," in *Proceedings of the 114th Convention of the Audio Engineering Society (AES '03)*, Amsterdam, The Netherlands, March 2003.

[12] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, "Evaluating rhythmic descriptors for musical genre classification," in *Proceedings of the 25th International Conference on the Audio Engineering Society (AES '04)*, pp. 196–204, London, UK, June 2004.

[13] F. Kurth, T. Gehrmann, and M. Muller, "The cyclic beat spectrum: tempo-related audio features for time-scale invariant audio identification," in *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR '06)*, pp. 35–40, Victoria, Canada, October 2006.

[14] D. Kirovski and H. Attias, "Beat-ID: identifying music with beat analysis," in *Proceedings of the International Workshop on Multimedia Signal Processing (MMSP '02)*, pp. 190–173, St. Thomas, Virgin Islands, USA, December 2002.

[15] F. Eyben, B. Schuller, and G. Rigoll, "Wearable assistance for the ballroom-dance hobbyist—holistic rhythm analysis and dance-style classification," in *Proceedings of IEEE International Conference on Multimedia & Expo (ICME '07)*, pp. 92–95, Beijing, China, July 2007.

[16] M. Goto and Y. Muraoka, "A real-time beat tracking system for audio signals," in *Proceedings of the International Computer Music Conference (ICMC '95)*, pp. 171–174, Banff, Canada, September 1995.

[17] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, 2001.

[18] J. Seppänen, "Computational models of musical meter recognition," M.S. thesis, Tampere University of Technology, Tampere, Finland, 2001.

[19] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," *Journal of New Music Research*, vol. 30, no. 1, pp. 39–58, 2001.

[20] S. Hainsworth and M. Macleod, "Beat tracking with particle filtering algorithms," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '03)*, pp. 91–94, New Paltz, NY, USA, October 2003.

[21] M. Alonso, G. Richard, and B. David, "Tempo and beat estimation of musical signals," in *Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR '04)*, pp. 158–163, Barcelona, Spain, October 2004.

[22] W. A. Sethares and T. W. Staley, "Meter and periodicity in musical performance," *Journal of New Music Research*, vol. 30, no. 2, pp. 149–158, 2001.

[23] A. P. Klapuri, "Musical meter estimation and music transcription," in *Proceedings of the Cambridge Music Processing Colloquium*, Cambridge University Press, Cambridge, UK, March 2003.

[24] A. P. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '99)*, pp. 3089–3092, Phoenix, Ariz, USA, March 1999.

[25] J. P. Bello and M. Sandler, "Phase-based note onset detection for music signals," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 5, pp. 441–444, Hong Kong, April 2003.

[26] C. Duxbury, J. P. Bello, M. Davies, and M. Sandler, "Complex domain onset detection for musical signals," in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx '03)*, pp. 90–93, London, UK, September 2003.

[27] J. C. Brown, "Determination of meter of musical scores by autocorrelation," *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1953–1957, 1993.

[28] L. van Noorden and D. Moelants, "Resonance in the perception of musical pulse," *Journal of New Music Research*, vol. 28, no. 1, pp. 43–66, 1999.

[29] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[30] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[31] B. Schuller, F. Wallhoff, D. Arsic, and G. Rigoll, "Musical signal type discrimination based on large open feature sets," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 1089–1092, Toronto, Canada, July 2006.

[32] B. Schuller, F. Eyben, and G. Rigoll, "Fast and robust meter and tempo recognition for the automatic discrimination of ballroom dance styles," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. 217–220, Honolulu, Hawaii, USA, April 2007.

[33] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer, New York, NY, USA, 2nd edition, 1999.

[34] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, Calif, USA, 2nd edition, 2005.

[35] Ballroomdancers.com. Preview audio examples of ballroom dance music, November 2006, https://secure.ballroomdancers.com/music/style.asp/.

[36] Songlist brd data-set, 2008, http://www.mmk.ei.tum.de/~sch/brd.txt.

[37] F. Gouyon and S. Dixon, "Dance music classification: a tempo-based approach," in *Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR '04)*, Barcelona, Spain, October 2004.

[38] Ballrom data-set, 2004, http://mtg.upf.edu/ismir2004/contest/tempoContest/node5.html.

[39] J. Paulus and A. P. Klapuri, "Measuring the similarity of rhythmic patterns," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR '02)*, pp. 150–156, Paris, France, October 2002.