# FAST AND ROBUST METER AND TEMPO RECOGNITION FOR THE AUTOMATIC DISCRIMINATION OF BALLROOM DANCE STYLES

*Björn Schuller, Florian Eyben, and Gerhard Rigoll*

Institute for Human-Machine Communication
Technische Universität München, Germany
[sch, eyb, ri]@mmk.ei.tum.de

## ABSTRACT

Fast and robust recognition of a song's meter, and quarter note tempo is crucial in many Music Information Retrieval tasks dealing especially with large databases or real-time musical stream processing. We therefore introduce a novel approach that is capable of extracting musical meter features and tempo in beats per minute. The method is extendable in order to return the locations of beat onsets suitable for example for beat synchronization or musical audio segmentation. We use a simplified psychoacoustic model to split the input into audible frequency bands and two phase comb filtering on those bands to find the quarter note tempo and metrical structure. Based on these features we discriminate the nine classic ballroom dance styles and duple or triple meter by Support-Vector-Machines as exemplary application. Test-runs are carried out on a public Ballroom Dance Music database containing 1.8k titles and the public MTV-Europe Most Wanted 1981-2000 to demonstrate the high effectiveness for popular music with respect to meter, tempo and ballroom dance style recognition.

***Index Terms***— Musical Tempo, Musical Meter, Genre Recognition, Ballroom Dance Style Recognition, BPM Detection

## 1. INTRODUCTION

Almost all processing of musical signals at a higher level requires information about the song's tempo and meter. Finding a song's progression is drastically simplified when one knows at which instants chord changes are to be expected. Several musical tracks of different tempos can be synchronized and blended at one tempo level. Especially for classifying ballroom dance music, which is the main target within this work, tempo and meter features play the key role.

There are numerous works that aim at finding the tempo of the beat which corresponds to the quarter note level [1-6]. However, only few [5] deal with the metrical structure above this level. The approach introduced herein aims at a combined fast and robust extraction of tempo and meter features and subsequent classification of ballroom dance styles. Likewise, a large music database can be automatically annotated by tempo and dance style for Music Information Retrieval purposes.

## 2. GENERAL APPROACH

Meter detection requires tempo independent information about the song's rhythmic structure [7]. For this it is necessary to find the song's quarter note tempo reliably. Other approaches to meter detection [5] manually adjust the automatically extracted tempos. Our approach solemnly relies on finding multiple tempos in a song and comparing how well they resonate with this song. Thus, tempo information is an essential product of our meter recognition approach.

There are mainly three different approaches for tempo detection: using correlation methods [2,3], detecting note onsets and then finding the most common inter-onset interval (IOI) [1,5] and a multiple resonator approach with comb filters [4]. Our tempo extraction is based on [4], with a few improvements and performance enhancements, as we require a larger tempo search range, which implies computing more comb filters. As it is very hard, even for a human listener, to determine if the quarter notes of a song (beats) are better grouped by 2, 4 or 8, we simplify the meter extraction problem to a simple duple or triple decision [5]. We believe this is sufficient even for further processing on higher levels that relies on our metrical data. The algorithms mentioned in [5] and [6] try to determine the metrical grouping by explicitly identifying downbeats. A downbeat thereby is a stronger accented beat, which usually indicates the first beat in a meter. However, the task of reliably finding a downbeat is challenging, even for the non-experienced human listener. Moreover, downbeats may not occur regularly at the beginning of a meter.

Our method therefore focuses on reliably classifying the meter into the two classes, duple and triple, without any knowledge about note onsets, beat positions or downbeat locations. It relies merely on finding a base tempo called Tatum [6], and analyzing how well integer multiples of this Tatum resonate with a large part of the song. The Tatum thereby corresponds to a tempo of at least the quarter note tempo or higher. In a later stage of the algorithm, after the quarter note tempo is known, it is possible to find the correct phase of the quarter notes, by looking at the filter output [4] and tracking the phase over the whole song to sort out errors.

## 3. FEATURE EXTRACTION

In the process of extracting the two main features, tempo and meter, several other features are extracted, that will be used in the later classification step.

### 3.1. Preprocessing

The input data is down sampled to 11.025 kHz and converted into monophonic by stereo-channel addition in order to reduce computation time. Audio is split into frames of 256 samples with an overlap of approximately 0.57, resulting in a final envelope frame rate of 100 fps. A Hamming window is applied to each frame before computing the FFT. The amplitude of each of the 128 FFT frequency bands is weighted according to our hearing, which is most sensitive to frequencies around 3.4 kHz. By using 12 overlapping triangular filters, equidistant on the Mel-Frequency scale, the 128 bands are reduced to 12 nonlinear bands. According to [4] the envelope information of such a small set of bands, covering the whole audible frequency range, still contains the complete rhythmic structure of the musical piece.

The band envelopes are then converted to dB and low pass filtered by convolving with a half wave raised cosine filter of length 15. This preserves fast attacks, but filters noise and high frequencies, most as in the human ear. Of the filtered band envel.opes a weighted differential $d_{rel}$ is taken. This differential is computed in the following way for a sample $o_i$ at position $i$: A moving average is calculated over one window of 10 samples to the left of sample $o_i$ (left mean $\bar{o}_{i,l}$) and a second window of 20 samples to the right of sample $o_i$ (right mean $\bar{o}_{i,r}$). The differential then is:

$$d_{rel}(i) = \left(o_i - \overline{o}_{i,l}\right) \cdot \overline{o}_{i,r} \qquad 1)$$

This method is derived from the fact, that a human listener perceives note onsets louder if they occur after a longer time of lower sound level. The weighting with the right mean $\bar{o}_{i,r}$ incorporates the fact that note duration and total note energy play an important role in determining the phenomenal note accent [6].

### 3.2. Tatum features

The Tatum grid, according to [6], is the lowest metrical level of a song. It represents the highest tempo present in the song and therefore the lowest inter-onset-interval. For finding the Tatum tempo we use a comb filter bank, similar to [4], consisting of 57 filters, with gain 0.8 and delays ranging from $\tau$=18 to $\tau$=74 envelope samples. The differential $d_{rel}$ of each Mel-frequency-band envelope is processed by the filter bank, and the total energy over all bands of the output of each filter is computed. This value for each filter is stored in what we call the Tatum tempo vector $\underline{T}$, as depicted in Fig. 1. From this vector $\underline{T}$ five additional features are extracted considering the quality of the peaks in the vector: *T-max* and *T-mean* are the maximum and mean values of the Tatum vector. *T-ratio* is computed by dividing the highest value by the lowest. *T-slope* is computed by dividing the first value by the last value. *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 Tatum candidate peaks are and how flat the Tatum vector $\underline{T}$ is (see Fig. 1).

Since our comb filters tend to higher resonances at higher tempos on songs with little rhythmic content (Fig 1, b), the vector is adjusted by considering the difference between the average of the first 6 values and the average of the last 6 values. From the resulting vector the two most dominant peaks are picked as follows: Firstly, all local minima and maxima are found, then for each maximum its apparent height 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 are considered possible Tatum (abbreviated *T* in the ongoing) candidates (*T1* and *T2*). To decide which of the two candidates fits best, we comb filter the band envelopes at tempo multiples of 3 and 4 times of each Tatum candidate and then decide for which Tatum candidate the total filter output summed over both multiples is maximal. This candidate is called the final Tatum *Tf* in the following.

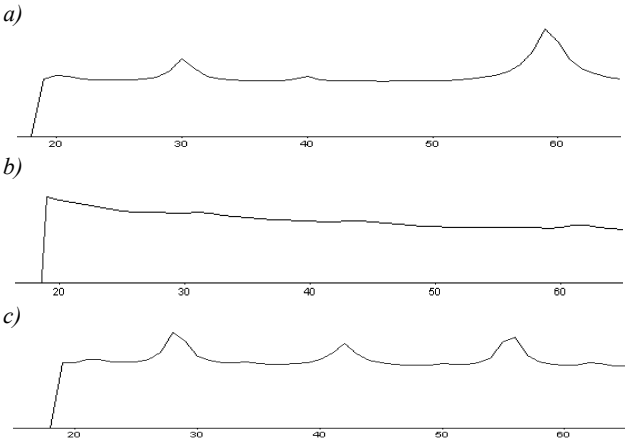



**Fig. 1.** – Plots of 57 dimensional Tatum tempo vectors $\underline{T}$ for the songs *a)* Robbie Williams – Rock DJ, *b)* Celine Dion – My Heart will go on, *c)* OMD – Maid of Orleans. Axes are labeled with the delay in envelope samples of the comb filter corresponding to a Tatum vector element.

### 3.3. Meter and Tempo Features

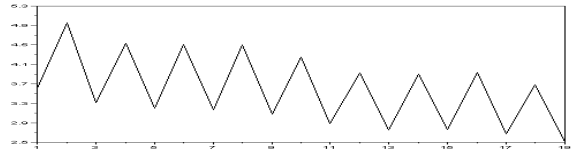
The meter vector $\underline{m}$ is computed by setting up narrow comb filter banks centered on integer multiple tempos of the final Tatum *Tf*. The width of the filter bank, i.e. the number of filters, is identical to two times the factor the Tatum is multiplied by plus one, in order to account for roundoff inaccuracies of the Tatum. For each filter bank the filter with the highest energy output is selected as the tempo and the total energy of this filter over all bands is the tempo score value in the meter vector at the position of the multiplication factor. The tempo score value indicates how well this certain tempo resonates with the song. It is sufficient to use Tatum multiples from 1 to 19, since all important multiples for meter classification are included in this range. The advantage of this two-step method over setting up comb filters for the total tempo search range is a reduction to a third of necessary comb filters to speed-up computation. With our meter vector, as depicted in Fig. 2., it is fairly easy to reliably determine the metrical structure. Two sums $s_2$ and $s_3$ are calculated:

$$s_2 = \tfrac{1}{3} \cdot \left[\underline{m}(4) + \underline{m}(8) + \underline{m}(16)\right] \qquad (2)$$

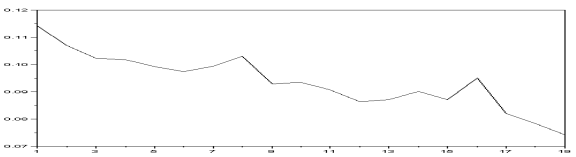
$$s_3 = \tfrac{1}{4} \cdot \left[\underline{m}(3) + \underline{m}(6) + \underline{m}(9) + \underline{m}(18)\right] \qquad (3)$$

The greater of the two sums determines the base meter. If $s_3$ is greater, then the base meter is triple (3/4 or 6/8 time signature), otherwise it is duple (4/4 or 2/4 time signature).

*a)*



*b)*



*c)*

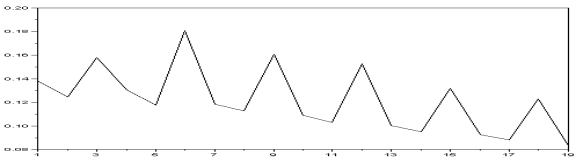


**Fig. 2.** - Meter vectors $\underline{m}$ for the songs referenced in Fig. 1. Clearly visible is the triple meter for the song Maid of Orleans (bottom).

For the quarter note tempo we define bounds within which the tempo is expected. For duple base meter the range reaches from 60 BPM to 143 BPM and from 75 BPM to 240 BPM for triple base meter. As the quarter note tempo we choose the highest scored tempo in the meter vector within these ranges. It is likely that at this point we make so called octave errors. These will not be of any further concern for us since they are not relevant for the meter classification and also a human listener sometimes cannot decide for sure if he taps along at double or half the actual tempo.
Phase information for the quarter note tempo can be derived by analyzing peaks in the output of this tempo's comb filter.
The base meter in most cases refers to the grouping between Tatum and the next higher metrical level. However, if an incorrect Tatum is found, e.g. one third of the true Tatum, the quarter note tempo may still be detected correctly, but the base meter will likely be wrong, e.g. triple instead of duple. To still be able to find the correct meter in such cases, we compare if three times the quarter note tempo has a higher value in the meter vector than two and four times the quarter note tempo, and then decide on triple, or else on duple final meter.

## 4. DANCE AND POPULAR MUSIC DATABASES

As a first database we choose the top 10 songs per year of the MTV Euro Most Wanted from 1981 to 2000. Likewise 200 songs in total are contained in this set abbreviated MTV in the ongoing. This musical selection is a good example of typical popular music from diverse genres as Rock, Hip Hop, Electronic Dance Music, Ballads or Pop. Yet, only 4 pieces with ¾ meter are contained while the rest is in common time.
We therefore chose a secondary set of 1,855 pieces of typical ballroom and Latin dance music to be found at [8], covering the Standard Dances Waltz, Viennese Waltz, Tango, Quick Step, and Foxtrot, and the Latin Dances Rumba, Cha Cha, Samba and Jive. 30 seconds of each song are available, which we converted to 44.1 kHz PCM so that the same preprocessing is used for both sets. The distribution among dance styles is depicted in tab. 1. This set is abbreviated BRD in the ongoing.

**Tab. 1:** Distribution of dance styles within BRD database

| Waltz | Viennese Waltz | Tango | Quick Step |
|---|---|---|---|
| 293 | 136 | 185 | 242 |

| Foxtrot | Rumba | Samba | Cha Cha | Jive |
|---|---|---|---|---|
| 245 | 217 | 188 | 211 | 138 |

The ground truth of tempo, ballroom dance style and meter for the MTV data-set was manually annotated by two professional ballroom dance DJs. For the BRD data-set the ground truth of tempo, dance style and meter is known from [8].

## 5. CLASSIFICATION

With the described features we now aim at classifying meter and ballroom dance style. As classifier we employ Support Vector Machines (SVM) with a polynomial Kernel function basing on our experience in Musical Genre Discrimination [9]. We firstly analyze the BRD dataset by performing a closed-loop Hill-climbing feature selection employing the target classifier's error rate as optimization criterion, namely Sequential Forward Floating Search (SVM-SFFS) [9]. This reveals the following features to yield the best results for dance style classification (*FS-D* in the ongoing): quarter note tempo, *T1* and *T2*, base meter, Tatum tempo vector $\underline{T}$, *T-ratio*, *T-slope*, *T-peakdist*, and meter vector $\underline{m}$ elements 2, 3, 4, 8, 10, 12, 14-18. To evaluate the effectiveness of our automatically extracted features combined with the ground truth of tempo and meter, a second feature set (*FS-Dgt*) is introduced with the following features: ground truth tempo and meter, Tatum tempo vector $\underline{T}$, *T-ratio*, *T-slope*, *T-peakdist*, and meter vector $\underline{m}$ elements 2, 3, 4, 8, 10, 12, 14-18. Features relevant for meter classification (*FS-M* in the ongoing) are the quarter note tempo, *T1* and *T2*, final meter, base meter, Tatum tempo vector $\underline{T}$, *T-ratio*, *T-slope*, *T-peakdist*, and meter vector $\underline{m}$ elements 2-19. The data is standardized before training the classifier.

## 6. RESULTS AND DISCUSSION

Test runs have been carried out on the two introduced sets MTV and BRD. The accuracy of the features tempo, base meter and final meter compared to the ground truth of tempo and meter is evaluated. Hereby the tolerance for tempo detection is 3.5% relative BPM deviation. In a 10-fold stratified cross validation (SCV) we evaluate the effectiveness for dance style and meter classification by SVM. The BRD data-set is further used as training set for dance style classification, to test on the MTV set.

| Accuracy [%] | Jive | Samba | Rumba | Cha | Fox | QuickS | Waltz | VWaltz | Tango | MEAN |
|---|---|---|---|---|---|---|---|---|---|---|
| **Tempo (±3.5%)** | 96.4 | 72.9 | 78.3 | 92.9 | 86.5 | 91.7 | 47.0 | 68.9 | 93.5 | **79.5** |
| **Final meter** | 99.3 | 75.0 | 96.3 | 98.6 | 96.3 | 94.2 | 38.9 | 78.5 | 97.8 | **84.1** |
| **Tempo + final meter** | 96.4 | 72.9 | 78.3 | 92.9 | 86.5 | 90.1 | 29.0 | 63.7 | 93.5 | **76.1** |
| **Base meter** | 92.3 | 65.4 | 86.2 | 96.2 | 78.8 | 83.9 | 79.9 | 86.7 | 82.7 | **83.1** |
| **BDS Recall** | 85.5 | 78.7 | 65.0 | 84.4 | 88.6 | 84.3 | 93.2 | 85.3 | 81.6 | **83.3** |
| **BDS Precision** | 91.5 | 85.1 | 71.6 | 89.0 | 87.5 | 82.6 | 79.4 | 89.2 | 81.2 | **83.4** |
| **BDS $F_1$-Measure** | 88.4 | 81.8 | 68.1 | 86.6 | 88.0 | 83.4 | 85.7 | 87.2 | 81.4 | **83.2** |

**Tab. 2.** - Accuracy for tempo and meter features. Recall, Precision and $F_1$-Measure of ballroom dance style (BDS) classification. Set BRD.

Average computation time for the feature extraction of 120s excerpts from the MTV dataset is 11.6s and 2.8s for the 30s excerpts from the BRD set, roughly corresponding to a rtf of 0.1 on a P4-Mobile with 1.4 GHz.

Using 120s excerpts from the MTV data-set, the quarter note tempo is identified correctly on 178 of the songs resembling 89% recognition accuracy. On 177 of these 178 songs, the meter is identified correctly resulting in 88.5% songs with both – meter and tempo – assigned correctly within the named tolerance. Overall, meter alone was correctly recognized with 95.0% accuracy on the MTV database. No false assignments were observed for the 4 pieces of ¾ meter. Using only 30s excerpts, a loss in accuracy was observed: On only 72% of the songs tempo and meter was identified correctly. This demonstrates that for reliable large database applications larger portions of a song are beneficial.

The excerpts from the BRD database available to us are only 30s long, so no test runs with longer portions could be run, for which we would expect the results to further improve. Detailed results for the BRD database are found in tab. 2. Meter is thereby determined in a rule-based manner as described in sec. 3.3. Results for data-driven meter classification with the feature set *FS-M* and SVM are depicted in tab. 3.

| Accuracy | Triple Meter Recall | Duple Meter Recall |
|---|---|---|
| **95.0%** | 90.7% | 96.3% |

**Tab. 3.** - Meter classification by SVM, 10-fold SCV. Set BRD.

83.7% average recall rate is achieved when using only the ground truth of tempo and meter from [8] for classifying the nine dance styles. This shows that dance style recognition is not solemnly a problem of tempo and meter but nevertheless highly dependent on those features.

Using our feature set *FS-D*, consisting only of our automatically extracted features, we achieve 83.3% correctly classified instances, nearly the same result as with the ground truth of tempo and meter mentioned above. For this feature set, results are shown in tab. 2.

Using the feature set *FS-Dgt*, results improve compared to using only the ground truth of tempo and meter: 93.1% of all instances are classified correctly. This demonstrates that our features provide a more advanced rhythmic structure of the song compared to the sheer ground truth of tempo and meter.

Finally, we test our approach on typically aired pop-music in a cross-validation by training with the BRD-set and testing with the set MTV to test the applicability in a challenging real-world scenario. Thereby 64.9% correctly assigned dance styles are observed. However, this number can be greatly increased, as often recognized dance styles fit as well, e.g. Foxtrot instead of Quickstep.

## 7. CONCLUSION AND OUTLOOK

Within this paper we effectively demonstrated recognition of ballroom dance style, tempo, and meter on two musical databases. Especially on typical dance-music high accuracies in this respect can be reported. Novel data-driven meter recognition thereby outperformed rule-based proceeding. In future work we aim at testing on further styles as Classical Music, Jazz, and world music as Oriental or African folklore to widen the spectrum of meters.

## 8. REFERENCES


[1] M. Goto, Y. Muraoka, "A real-time beat tracking system for audio signals," *Proc. of the Int. Computer Music Conf.*, 1995.

[2] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," *Journal of New Music Research*, Vol. 30, 2001.

[3] J. Foote, S. Uchihashi, "The Beat Spectrum: A new approach to rhythm analysis," *Proc. ICME 2001*, 2001.

[4] E. Scheirer, "Tempo and Beat Analysis of Acoustic Musical Signals," *Acoustic Society of America*, 103(1), p. 588-601, 1998.

[5] F. Gouyon, P. Herrera, "Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors," *Proc. of Audio Engineering Society*, 114th Convention, Amsterdam, The Netherlands, 2003.

[6] J. Seppänen, "Computational models of musical meter recognition," *M. Sc. Thesis*, Tampere Univ. of Technology, 2001.

[7] F. Gouyon, S. Dixon, E. Pampalk, G. Widmer, "Evaluating rhythmic descriptors for musical genre classification," *Proc. AES 25th Int. Conf.*, London, UK, 2004.

[8] http://www.ballroomdancers.com, 2006.

[9] B. Schuller, F. Wallhoff, D. Arsic, G. Rigoll : "Musical Signal Type Discrimination Based on Large Open Feature Sets," *Proc. ICME 2006*, p. 1089-1093, Toronto, Ontario, 2006.