RHYTHMIC SIMILARITY OF MUSIC BASED ON DYNAMIC PERIODICITY WARPING

Andre Holzapfel and Yannis Stylianou

Institute of Computer Science, FORTH, Greece,
and Multimedia Informatics Lab, Computer Science Department, University of Crete
{hannover, yannis}@csd.uoc.gr

ABSTRACT

This paper introduces a new way to measure rhythmic similarity between two musical pieces using periodicity spectra. In order to detect similarity for pieces of different tempi, the linearity of the warping path between their spectra serves as a measure of their rhythmic similarity. Using a modified kNN classification approach on two datasets, the proposed measure provides comparable classification accuracy (82.1%) to the best of widely used measures (85.5%) for the first dataset; For the second dataset, which is characterized by a large variance of tempi, the proposed measure outperforms all reference measures, reaching an accuracy of 69.0%, while the best of the other measures reaches 53.8%. Moreover, the presented technique works fully automatically, and no information regarding tempo is needed.

Index Terms— Rhythm, similarity, music, information retrieval

1. INTRODUCTION

The use of detecting rhythmic similarity can be easily explained by mentioning two real life examples:
A dance hall DJ has discovered a song in his digitalized collection. He would like to gather a list of rhythmically similar songs in a list, in order to present a homogeneous program. But neither the song he has listened to, neither many songs in his collection have a genre label.
In our second example an ethnomusicologist is trying to get an overview over many hours of field recordings from some country. The recordings are not categorized, while our scientist has only a rough understanding of the musical structures that characterize the recordings. After some hours of listening, the rhythm of a song reminds him of some pieces he listened to at the beginning. But, ... finding them again takes time.
There have been some approaches to solve these kinds of problems. In [1] a cosine measure between beat spectra is used to measure rhythmic distances. These beat spectra are computed using self similarity between the Fourier transform at different time lags. Spectra similar to the beat spectra are used to derive a parametric description of the rhythm in [2], by estimating the tempo and then associating peaks in the spectra with specific note durations. Different descriptions are used in [3], where the tactus is estimated and rhythmic patterns are extracted from the time signal, which are then warped for comparison. The approach presented in [1] has the shortcoming of being applicable only when rhythmically similar pieces are presented without big changes in tempo. The other approaches make some kind of tempo or meter estimation necessary, which is problematic especially for music without strong percussive elements. Different features for dance music classification have been compared in [4], providing a valuable reference for comparison.

When deriving rhythmic descriptors, often some kind of knowledge about the tempo or the placement of bars is assumed. If this information is not available, it has to be estimated from the signal or annotation work has to be performed [4]. However, tempo estimation from musical mixtures is not perfectly possible [5] and hand annotation is time consuming. Thus, an automatic derivation of the metrical structure of complex mixtures for classification purposes is not feasible yet. This makes it worth to describe the rhythm of a piece just by considering the weights of the periodicities found in the sound, without regarding the order of sound events. Such approaches have been presented in [1] and [2]. In [2] the description is based on salient periodicities only, whereas the way to describe the periodicity spectrum in a continuous way was preferred in this paper. This is because the performance character of traditional music causes timing variabilities that are important for a similarity measure and found in the general shape of the spectrum. The weak point of the approach of [1] is that it is not robust when similar pieces are presented with large variations in tempo between them, which is the case in traditional folk music of different regions [6][7]. Thus, in this paper a similarity measure is proposed that is based on the warping of the two periodicity spectra to be compared. For this, a matrix of point wise distances between the spectra is computed, and a minimum cost warping path through this matrix is found. Each point on this path is projected to a reference line that would be the mapping between a piece and the same piece in a different tempo. In this paper, the sum of these projection distances will be referred to as *Dynamic Periodicity Warping* (DPW) dissimilarity measure.
Corresponding to the two real life examples given above, two datasets have been chosen for the tests. The first dataset, D1, has been used in the ISMIR Rhythm Description contest[1] and contains eight classes of ballroom dances. This dataset has also been used in [4]. The second dataset, D2, contains six different traditional dances commonly encountered in the island of Crete (Greece). The within class variation of tempo in D2 is much larger than that of the first database, as will be detailed in Section 4. Percussive instruments are not contained in most of the pieces. The validity of the proposed approach will be judged by regarding classification accuracy, assuming that pieces in one class of dance are rhythmically more similar.
The next Section of this paper will detail the computation of the periodicity spectra. The suggested similarity measure based on DPW of these spectra is described in Section 3.1. Section 3.2 overviews the similarity measures to be compared with DPW. Section 4 describes the datasets, emphasizing the new dataset presented in this paper. Its musical properties are detailed and the differences between the two datasets are pointed out. Section 5 gives the classification results for both data sets. Section 6 concludes the paper.

[1] http://www.iua.upf.es/mtg/ismir2004/contest/rhythmContest/

2. PERIODICITY COMPUTATION

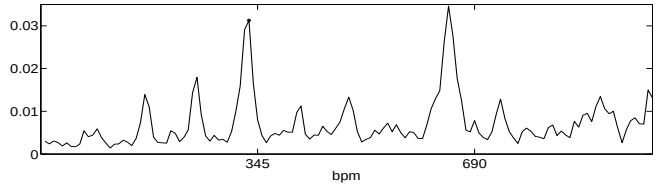
The first step in computing the periodicity spectra is a computation of an onset strength signal, using the same method and parameters as used by Dan Ellis in the MIREX 2006 beat tracking contest[2]. Note that we do not perform a peak picking on this signal thereafter, which would result in a discrete set of impulses, like described in [8]. From a signal processing point of view this onset strength signal, $\mathbf{p}(t)$, is the linear combination of the convolution of N sound events $\mathbf{e}_i$ with impulses at a finite set of lags $\mathcal{K}_i$

$$\mathbf{p}(t) = \sum_{i=1}^{N} \left(\mathbf{e}_i(t) * \sum_{k \in \mathcal{K}_i} \delta(t - kT) \right) \tag{1}$$

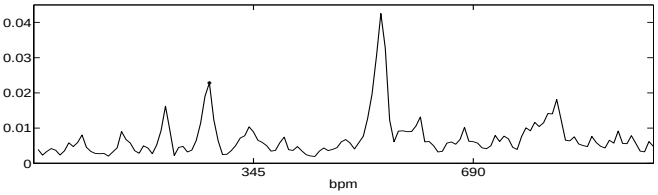
where T is the sampling period of the onset strength signal and '$*$' denotes convolution. Note that in [8], event vectors $\mathbf{e}_i$ have been replaced by scalar values that denote the strength of the onset. While this is desirable when aiming at beat tracking, this loss in detail is considered disadvantageous for classification. By preserving these details, the onset strength signal contains characteristics of the rhythm like *e.g. staccato*, *legato*, or timing accuracy.
The periodicity spectra $\mathbf{P}(f)$ are obtained by computing the magnitude of the Fourier transform of the onset strength signal:

$$\mathbf{P}(f) = \left| \sum_{i=1}^{N} \left(\frac{1}{T} \mathbf{E}_i(f) \sum_{k \in \mathcal{K}_i} \delta(f - \frac{k}{T}) \right) \right| \tag{2}$$

In (2), we only consider $f < 1000bpm$ which approximately equals $16.7Hz$. Higher frequencies do not contribute to the perception of rhythm. The Fourier transform is computed on a hamming window of appropriate length that is supposed to include several realizations of the basic rhythm. In this paper, a window length of eight



Faster Example of Siganos



Slower Example of Siganos

Fig. 1. Two example periodicity spectra from the same class of D2

seconds is chosen, while the hop size is set to half a second. To get a single vector describing the periodicities in a piece, the mean, $\bar{\mathbf{P}}(f)$, of the spectra derived from the song's onset strength signal is computed. In Figure 1 two such mean periodicity spectra are shown for dances of the same class having different tempi. They show a similar shape, but they are not aligned. Thus, an alignment of these description leads to the proposed similarity measurement as described in Section 3.1.
Several parameterizations of $\bar{\mathbf{P}}(f)$ have been tried; its squared magnitude has been modeled by Cepstral Coefficients, which lead to a smoothing of the spectra. This was observed to have the shortcoming of introducing artefacts in the shape of the spectra (*i.e.* smearing). Autoregressive modeling is not appropriate because of the often peaky structure of the spectra, that implies a filter close to instability. Therefore, it has been decided to use the frequency bins of $\bar{\mathbf{P}}(f)$ as inputs to the similarity measurement.

[2] www.music-ir.org/mirex2006/index.php/Audio_Beat_Tracking_Results

3. DISTANCE MEASURES

3.1. DPW Distance

As mentioned in the introduction, a direct comparison of periodicity spectra will be problematic between rhythmically similar pieces with different tempi. To solve this problem, a similarity measure based on Dynamic Periodicity Warping (DPW) is computed as depicted in Figure 2. This warping approach has the advantage that no estimation of metric structure like in [3] is needed. It is based on the idea, that warping a periodicity spectrum of one piece on the periodicity spectrum of exactly the same piece at a different tempo should result in a warping path that has the form of a straight line, a case which will be referred to as *perfect mapping*. Thus, the deviation of a warping path from such a straight line can be used to judge rhythmic similarity.
As depicted in Figure 2, the first step of the proposed warping ap-

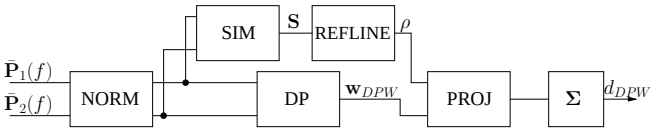


Fig. 2. Schematic of DPW dissimilarity computation

proach normalizes each of the input spectra, $\bar{\mathbf{P}}_1(f)$ and $\bar{\mathbf{P}}_2(f)$, to sum to one. Then both normalized spectra are used to compute a minimum cost warping path $\mathbf{w}_{DPW}$ through the difference matrix $\mathbf{D}$, which contains the pairwise squared difference of the values in $\bar{\mathbf{P}}_1(f)$ and $\bar{\mathbf{P}}_2(f)$: $\mathbf{D}(i,j) = \left(\bar{\mathbf{P}}_1(f_i) - \bar{\mathbf{P}}_2(f_j)\right)^2$. The path is computed using the fast algorithm provided in [9], with step size one.
In order to use this path for a distance measure, a reference path has to be found that represents a perfect mapping. For this, a similarity matrix $\mathbf{S}$ is formed by computing the vector direct product of the two normalized periodicity spectra $\bar{\mathbf{P}}_1(f) \otimes \bar{\mathbf{P}}_2(f)^T$, where $\otimes$ denotes the matrix direct product. This matrix has local maxima wherever maxima in the two vectors coincide. As the detailed description of the datasets in Section 4 shows, tempos of dances from the same class vary only within a certain bandwidth (in bpm). Assuming that two similar dances have similar periodicity spectra, there have to be strong maxima in matrix $\mathbf{S}$: in the main diagonal and in a certain number of sub- and superdiagonals that cover this bandwidth. It was set to 20 bpm in this paper, motivated by the depicted tempo ranges of the dances as shown in Table 1. By localizing the highest local maximum within this tridiagonal matrix, we get an estimation for the relation of the tempi of the two pieces. The straight line ρ, starting from the upper left corner of $\mathbf{S}$ and passing through the local maximum, forms the reference line that would be the mapping path

of two linearly warped periodicity spectra (*perfect mapping*). In the next step, the distance of the orthogonal projection of each point on the warping path $\mathbf{w}_{DPW}$ onto ρ is computed. The sum of these distances gives the DPW dissimilarity d_{DPW} introduced in this paper. Note that for the step matrix in the dynamic programming other possibilities have been tried, such as an adaption of the step cost to the angle of the reference line ρ. As this had no influence on the results, the matrix with step size one has been used throughout the experiments.

3.2. Baseline Distances

In order to examine the properties of the DPW dissimilarity, some widely used metrics will be used for comparison. In Section 5 their behavior will be compared with the DPW dissimilarity in classification tasks. As in [1], the Euclidean distance and the cosine distance will be used, applied to the unwarped periodicity spectra. Regarding the findings in [1], better results for the cosine distance might be expected. In order to justify the proposed measure, it will also be compared with two other measures obtained after warping the spectra. The first is the cost of the warping, d_{Cost}, obtained by summing the local costs on the warping path, as used in [3]. The other is the cosine distance of the two aligned spectra, $d_{CosPost}$.

4. DATA SETS

The first data set, D1, contains 698 songs from eight classes of ballroom dances. Their variation in tempo is relatively small for every class. This makes the dances easily distinguishable, if the true tempo is given, as has been analyzed in [10]. But, as mentioned in the introduction, in this paper it is assumed that no information about tempo is given (besides the estimation of tempo is too unreliable). The minimum number of songs in a class is 60.

The second dataset, D2, contains six dances commonly encountered in the island of Crete: Kalamatianos, Siganos, Maleviziotis, (fast) Pentozalis, Sousta and Chaniotis. Each class contains either fifteen or sixteen instrumental song excerpts of about ten seconds length. As depicted in Table 1 their inner class variances in tempo are much larger than those for D1, where on average the tempo of pieces in the same class does not show a difference of more than 16 bpm. All depicted tempi are the rates of the quarter note. Note that all traditional Cretan dances have a $\frac{2}{4}$ time signature, only Kalamatianos as a dance originating from a different part of Greece has a $\frac{7}{8}$ rhythm. Between the dances there are significant overlaps in tempo. This is even more severe, when considering that Chaniotis is often transcribed in notes of double length, which results its quarter note rate per minute to be from 116 to 158 for this data set. Most of the pieces contain only two kinds of string instruments, while percussive instruments are not contained in most samples. This absence of percussive instruments makes a description of rhythm more difficult, as the onset energies are less peaky.

Table 1. Tempi of D2 (Traditional Dances)

Dance	Tempo Range (♩)
Kalamatianos	116-142
Siganos	93-103
Maleviziotis	132-160
Pentozalis	123-182
Sousta	111-136
Chaniotis	58-79

From a musicological point of view, the length of the excerpts in D2 is sufficient to classify the pieces. As detailed in [11], Cretan dances are characterized by small melodic phrases, *Kontilies*, which extend usually over two or four bars. As such, in the given duration at least one such element will be contained. A *Kontilia* is always connected to a certain dance, but the variety between the melodic phrases of a certain dance style are big, considering tempo and melodic structure. Furthermore it is interesting to note that the dances seem to group into three different tempo classes. According to [12], Chaniotis and Siganos should be related to movement of circular shape, Kalamatianos and Sousta to expressing movements of excitement and Maleviziotis and Pentozalis to smaller movements at a high repetition rate. Indeed, these descriptions fit the movements of the particular dances very well.

Using D2, preliminary listening tests have been conducted to evaluate the ability of a listener to correctly categorize the dances. Six subjects were asked to classify each piece in the dataset after listening to it one time. All subjects are dancers familiar with each of the dances. The average correct classification per class and overall is depicted in Table 2. It can be seen that some of the classes are particularly difficult and the overall accuracy is far from being perfect. Interested readers can listen to two samples on the authors' website[3].

Table 2. Listeners' Mean Classification Accuracies

Kal.	Sig.	Mal.	Pen.	Sous.	Chan.	Mean
93.3	88.9	79.2	45.6	58.3	88.5	75.6

5. EXPERIMENTS

In order to compare the quality of the similarity measures, their accuracies in a k-Nearest Neighbor (kNN) classification was measured. For this, 10 repetitions of 10-fold stratified cross-validations have been performed. A locally weighted k-Nearest Neighbor (wkNN) classifier has been tried as well. The simple kNN classifier just counts the training samples in the k neighborhood for each class, and assigns the test sample to the class which has most training samples in the k neighborhood. Thus, each training sample has equal weight $w_i = 1, i = 1...k$. The locally weighted kNN assigns $w_i = 1 - (d_i/d_{k+1})$ to the i-th training sample, where d_{k+1} is the distance of the $k+1$-nearest neighbor to the test sample. Thus, training samples more far away from the test sample, contribute less to its classification.

Table 3. Classification Accuracies on D1

	wkNN	kNN
Cosine	**85.5** (k=7)	84.5 (k=3)
Euclidean	83.8 (k=6)	82.7 (k=3)
d_{Cost}	72.4 (k=14)	70.7 (k=7)
$d_{CosPost}$	70.7 (k=32)	69.2 (k=17)
d_{DPW}	82.1 (k=11)	80.9 (k=20)

Tables 3 and 4 show the classification accuracies on the two datasets, the size of the neighborhood of these optimal values is denoted in parentheses. It can be seen that using wkNN has a positive effect on both datasets. On D1 the best classification has

[3] www.csd.uoc.gr/∼hannover/Datasets.html

Table 4. Classification Accuracies on D2

	wkNN	kNN
Cosine	53.8 (k=1)	53.8 (k=1)
Euclidean	48.9 (k=1)	48.8 (k=1)
d_{Cost}	51.8 (k=18)	48.5 (k=8)
$d_{CosPost}$	51.1 (k=19)	48.7 (k=12)
d_{DPW}	**69.0** (k=4)	64.4 (k=5)

been achieved using the cosine measure, slightly outperforming the Euclidean distance as well as the DPW dissimilarity. This confirms the results in [1], where the cosine measure was found superior on an exemplary set of pieces. The latter can be assigned to the small tempo variance of the dataset, which limits the gain of warping. Note that, according to the authors' knowledge, the best accuracies presented on this data set without the usage of tempo information is 85.7% [4], where the features contained periodicity descriptions and rhythmic patterns, and adaptive boosting was incorporated in classification. This indicates that due to small intra class variability, this dataset might be too easy.

On D2, the advantage of the DPW based similarity measure is obvious. While the superiority of the cosine measure compared with the Euclidean measure remains the same, the DPW based measure results in a clearly higher accuracy. It is important to notice the small difference of the classification result of 69.0% of the presented system, and the classification accuracy by humans (75.6%) as shown in Table 2.

A closer look at the misclassification using the DPW based measure on D2 is given in form of a confusion matrix in Table 5. The row contains the correct class labels and the columns the prediction of the weighted kNN classifier, that achieved the highest accuracy in Table 4. The strongest confusion is observed between the dances Maleviziotis and Pentozalis, a phenomenon that has been observed also in the listening tests. Compared to the listening tests, the accuracy for the Chaniotis is low. This dance is the most popular nowadays in Crete, and its style varies very much from region to region and from player to player. This causes a big diversity for this dance, which leads to periodicity spectra that vary stronger for this class than for others. The small number of 16 samples in D2 of this dance seems not to be capable of this diversity.

Table 5. Confusion matrix for D2

	Kal.	Sig.	Mal.	Pen.	Sous.	Chan.
Kal.	130	0	0	0	1	19
Sig.	0	140	0	0	10	0
Mal.	0	0	72	80	8	0
Pen.	10	10	26	94	10	0
Sous.	10	7	2	8	129	4
Chan.	33	0	12	20	18	77

The proposed DPW dissimilarity performs better than the cost of warping, d_{Cost}, and the post warp cosine distance, $d_{CosPost}$, on both datasets. This justifies the usage of the proposed measure. The explanation of the worse performance of the cosine measure after warping is the smaller distance of periodicity spectra of different classes after warping. This shows the necessity of considering the path chosen in DPW.

6. CONCLUSION

This paper introduces a similarity measure for musical pieces with similar rhythmic structure, but varying tempo. The discriminant power of the periodicity spectra was demonstrated in dance music classification tasks. The advantage of the proposed method, besides its high discriminant power, is its simple and annotation free computation. A next step in refining this method is a dynamic programming algorithm more adapted to the periodicity spectra. Also the combination of the dissimilarity presented in this paper with other dissimilarities will be subject of experiments. On request, the introduced data set is provided for research proposals.

7. REFERENCES

[1] Jonathan Foote, Matthew D. Cooper, and Unjung Nam, "Audio retrieval by rhythmic similarity," in *Proc. of ISMIR 2002 - 3rd International Conference on Music Information Retrieval*, 2002.

[2] Geoffroy Peeters, "Rhythm classification using spectral rhythm patterns," in *Proc. of ISMIR 2005 - 6th International Conference on Music Information Retrieval*, 2005, pp. 644–647.

[3] Jouni Paulus and A.P. Klapuri, "Measuring the similarity of rhythmic patterns," in *Proc. of ISMIR 2002 - 3rd International Conference on Music Information Retrieval*, 2002.

[4] Simon Dixon, Fabien Gouyon, and Gerhard Widmer, "Towards characterisation of music via rhythmic patterns," in *Proc. of ISMIR 2004 - 5th International Conference on Music Information Retrieval*, 2004.

[5] A. P. Klapuri, A. J. Eronen, and J. T. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Transactions on Acoustics Speech and Signal Processing*, in press.

[6] Irene Loutzaki, "Audio report: Greek folk dance music," *Yearbook for traditional music*, vol. 26, pp. 168–179, 1994.

[7] B.A. Aning, "Tempo change: Dance music interactions in some ghanaian traditions," *Institute of African Studies: Research Review*, vol. 8, no. 2, pp. 41–43, 1972.

[8] A.P. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in *ICASSP '99: Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference*, Washington, DC, USA, 1999, pp. 3089–3092, IEEE Computer Society.

[9] Dan Ellis, *Dynamic Time Warp (DTW) in Matlab*, www.ee.columbia.edu/∼dpwe/resources/matlab/dtw/, 2002.

[10] Fabien Gouyon and Simon Dixon, "Dance music classification: A tempo based approach," in *Proc. of ISMIR 2004 - 5th International Conference on Music Information Retrieval*, 2004.

[11] Irini B. Theodosopoulou, *Methodology of morphological analysis and analytic data of small rhythmic patterns of cretan folk music, (in Greek Language)*, Athens: Kultura, 2004.

[12] Dirk Moelants, "Dance music, movement and tempo preferences," in *Proceedings of the 5th Triennial ESCOM Conference*, Hannover, Germany, 2003.