METER CLASS PROFILES FOR MUSIC SIMILARITY AND RETRIEVAL

Matthias Robine and Pierre Hanna
LaBRI - University of Bordeaux
351, cours de la Libération
33405 TALENCE Cedex - FRANCE
firstname.name@labri.fr

Mathieu Lagrange
Telecom ParisTech
46, rue Barrault
75634 PARIS Cedex 13 - FRANCE
lagrange@telecom-paristech.fr

ABSTRACT

Rhythm is one of the main properties of Western tonal music. Existing content-based retrieval systems generally deal with melody or style. A few existing ones based on meter or rhythm characteristics have been recently proposed but they require a precise analysis, or they rely on a low-level descriptor. In this paper, we propose a mid-level descriptor: the Meter Class Profile (MCP). The MCP is centered on the tempo and represents the strength of beat multiples, including the measure rate, and the beat subdivisions. The MCP coefficients are estimated by means of the autocorrelation and the Fourier transform of the onset detection curve. Experiments on synthetic and real databases are presented, and the results demonstrate the efficacy of the MCP descriptor in clustering and retrieval of songs according to their metric properties.

1. INTRODUCTION

The amount of digital music is rapidly increasing, and is mostly comprised of Western pop music. New interfaces for browsing, classifying or searching have to be proposed. Content-based retrieval systems generally consider musical properties related to melody or style. But taking into account other characteristics such as rhythm may lead to the development of useful tools for helping users to browse into large databases.

Research regarding rhythmic or metric properties typically involves tempo or meter analysis. Several systems have been developed for improving tempo induction [1,2], meter analysis [3,4], or estimation of the time signature of audio songs [5]. Other works focus on the automatic classification of songs according to their rhythmic properties [6]. In contrast, only a few methods have been presented for retrieving songs by rhythmic similarity. A few existing ones consider a low-level descriptor such as beat spectrum [7], acoustic features [8], or spectral descriptors [9]. Others precisely analyze tempo, meter and/or time signature [5] and are consequently limited by the error analysis.

In this paper, we present a new mid-level descriptor for retrieving music according to the metric properties. The estimation of this descriptor requires neither a complete analysis of the time signature nor the meter of a song, but only the prior knowledge of its tempo. In comparison to a low-level feature, the essential metric information is reduced to a few values which characterize the metric properties of any song. In Section 2, we discuss the notion of metrical structure. The new mid-level descriptor and the associated analysis method are detailed in Section 3. Then, clustering and retrieval experiments are presented in Section 4. Finally, discussion of these results and perspectives are proposed in Section 5.

2. MUSICAL METER

2.1 Metrical Structure

In Western tonal music, the periodic alternation of strong and weak beats leads to a metrical hierarchy known as metrical structure. The symbolic representation of this structure is present in score notation using markings such as a time signature, bar lines, and dynamic accents.

The Generative Theory of Tonal Music (GTTM) [10] proposes a model of the metrical structure, where the meter of a musical piece may be represented by multiple levels of beats. The periodicity of beats is reinforced from level to level, and it is the interaction of the different levels that produces the sensation of meter. This representation is also used by Temperley [11] in his proposal of a preference rule system for meter.

According to [10], while there are five or six metrical levels in a piece, one is particularly central: the *tactus* level. The tactus identifies a perceptually prominent level, with the levels immediately smaller and immediately larger. It refers to the perceived tempo, the internal clock [12]. It corresponds highly with the notated unit time of a musical piece, but it can differs. Lee [13] indicates thus that listeners may revise meter to always get a tactus between 300ms and 600ms.

In the following, we assume that the tactus level corresponds to the tempo and the basic unit time of the music, when it is known or notated. We choose also to use without distinction the terms *tactus* and *beat*. As the metrical structure is hierarchical, levels lower than that of the tactus can be called subtactus levels and represent divisions of the beat. The smallest division is generally called *tatum* or *tick*. Alternatively, higher levels are termed supertactus levels, and contain multiples of the beat duration, including the measure level.

2.2 Time Signature

We can restrict the notion of meter to two levels, the faster of which provides the element, and the slower of which group them. According to Gouyon [14], this is close to the usual description of meter that can be found in a score, given by the time signature and the bar lines.

The time signature consists of 2 integers arranged vertically, e.g. $\frac{4}{4}$ or $\frac{6}{8}$. The upper number of the signature indicates the number of units in a bar, with the value of unit given by the lower number of the signature (*e.g.* 4 for a quarter note). If the upper number is divisible by 3 and the lower by 2, the time is compound and the number of beats per measure is given by the upper number divided by 3. In this case, the unit of time, *i.e.* the beat, is divided triply at the smaller level. Otherwise, in simple time, the upper number indicates the number of beats per measure and the beat is divided duply. Table 1 presents the time signatures mainly used in Western tonal music.

3. METER CLASS PROFILE

We introduce in this section a new descriptor to represent the metrical structure of the music called Meter Class Profile (MCP). We describe the method used for estimating the MCP, and provide examples for illustration of its use.

3.1 Properties

MCP is a real-valued vector providing information of the strength of the different metrical levels within the music. It is centered on the tactus level: the beat multiples, including the measure level, are represented on the left, and the beat subdivisions, including the tatum level, on the right. The choice was made to represent the relative strength of accents at multiple rates of the tempo: 2, 3, 4, 5, 7, 9, and 11, and also at subdivisions of the tempo: $1/2$, $1/3$, $1/4$, $1/6$, $1/8$, $1/12$. MCP is thus a vector of thirteen dimensions.

A MCP corresponding to a $4/4$ time signature would contain high amplitudes for the bin corresponding to 4 times the tempo, *i.e.* beat multiple 4 representing the measure periodicity, for beat multiple 2 (in $4/4$, there are accents every two beats), for beat subdivision $1/2$ (because $4/4$ is in simple time), and perhaps beat subdivisions $1/4$ and $1/8$ (if 16^{th} or 32^{th} notes occur). A few examples are presented in Figure 2.

As the MCP is independent from tempo, its representation does not change with tempo variations. It may be considered as a mid-level descriptor, since the metric information is summarized to a 13-dimension vector without identifying a particular time signature. Additionally, the amplitude ratios from the different metrical levels provide a meaningful way to handle the meter of a musical piece. The MCP could, for example, also indicate a degree of swing, by considering the balance between duple and triple beat subdivisions.

3.2 Estimation

The method proposed here for computing MCP relies on existing analysis methods recently described for estimating tempo [15]. It has been implemented using the MIR toolbox [16]. The main steps are illustrated on Figure 1, with the example of the country song *Wanted* (A. Jackson), annotated with a $3/4$ time signature.

In this paper, we consider only one global MCP per audio musical signal. We thus choose to analyze a large frame of music (length 60 seconds). The meter is assumed to be stationary during this frame. An onset-energy function is first extracted from the audio signal by taking into account spectral energy flux [17]. Then, dominant periodicities (or frequencies) are estimated.

Two types of observations, respectively termed Onset Discrete Fourier Transform (ODFT) or Onset Autocorrelation Function (OAF), are considered. Their complementary properties are discussed for tempo estimation in [2]. In the experiment presented in this paper, the OAF and ODFT are both computed and normalized and the tempo frequency is known.

The analysis method of MCP locates both periodicities corresponding to beat multiples (related to measure) and beat subdivisions. Estimation of multiples and subdivisions is carried out using the complementary properties of the two observations. On one hand, the ODFT of a periodic signal is a set of harmonically related frequencies, and it is difficult to determine predominant frequencies above the tempo frequency. Therefore, we only estimate frequencies lower than the tempo frequency. These frequencies correspond to beat multiples (first part of MCP), in particular that of the measure.

On the other hand, the OAF of a periodic signal is a set of periodically related lags. It is thus difficult to measure predominant periodicities higher than the tempo period. The OAF is only considered for estimating periods lower than the tempo periods. These periods are related to beat subdivisions (second part of the MCP).

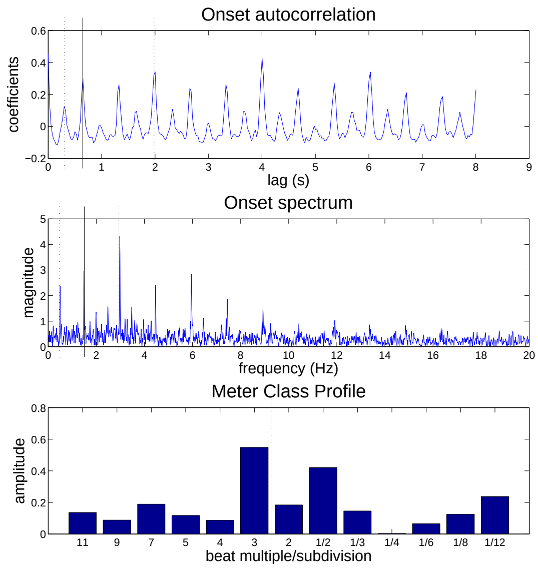


Figure 1. Different stages of the analysis method of Meter Class Profile for the song *Wanted* of A. Jackson (time signature $3/4$): from top to bottom, the autocorrelation function, the spectrum of the onset function and the MCP estimated, showing peaks at beat multiple 3 and beat subdivision $1/2$.

With prior knowledge of the tempo, the first part of the MCP is estimated from the ODFT and the second part is estimated from the OAF. Consequently, period bands corresponding to harmonics of the frequency tempo are analyzed when considering the OAF. Only the six harmonics (2, 3, 4, 6, 8, and 12) are taken into account. The amount of energy of OAF within a thin frequency band around the related periodicities directly determines the amplitude re-

	Duple	Triple	Quadruple	5-uple	7-uple	9-uple	11-uple
Simple	$\frac{2}{1}$ $\frac{2}{2}$ $\frac{2}{4}$ $\frac{2}{8}$	$\frac{3}{1}$ $\frac{3}{2}$ $\frac{3}{4}$ $\frac{3}{8}$	$\frac{4}{1}$ $\frac{4}{2}$ $\frac{4}{4}$ $\frac{4}{8}$	$\frac{5}{1}$ $\frac{5}{2}$ $\frac{5}{4}$ $\frac{5}{8}$	$\frac{7}{1}$ $\frac{7}{2}$ $\frac{7}{4}$ $\frac{7}{8}$	$\frac{9}{1}$ $\frac{9}{2}$ $\frac{9}{4}$ $\frac{9}{8}$	$\frac{11}{1}$ $\frac{11}{2}$ $\frac{11}{4}$ $\frac{11}{8}$
Compound	$\frac{6}{2}$ $\frac{6}{4}$ $\frac{6}{8}$ $\frac{6}{16}$	$\frac{9}{2}$ $\frac{9}{4}$ $\frac{9}{8}$ $\frac{9}{16}$	$\frac{12}{2}$ $\frac{12}{4}$ $\frac{12}{8}$ $\frac{12}{16}$	$\frac{15}{2}$ $\frac{15}{4}$ $\frac{15}{8}$ $\frac{15}{16}$	$\frac{21}{2}$ $\frac{21}{4}$ $\frac{21}{8}$ $\frac{21}{16}$	$\frac{27}{2}$ $\frac{27}{4}$ $\frac{27}{8}$ $\frac{27}{16}$	$\frac{33}{2}$ $\frac{33}{4}$ $\frac{33}{8}$ $\frac{33}{16}$

Table 1. Time signatures. Most common signatures are duple, triple and quadruple time in Western tonal music. Notations 9/2, 9/4 and 9/8 may be used either for a simple measure of 9 pulses, or for a compound measure of 3 pulses.

lated to the beat subdivisions of the MCP. In our implementation, the width of the frequency bands for cumulating the energy in the ODFT and OAF has been set to 5%. In Figure 1, the tempo has been annotated to 1.5Hz and is showed on the OAF with a solid line. When considering periods lower than $\frac{1}{1.5} = 0.66$s, energy is located around period of 0.33s, corresponding to the beat subdivision $1/2$. The contribution to the beat subdivision of the corresponding MCP is thus significant and clearly indicates a simple meter.

The ODFT is considered in a similar way. Only the seven sub-harmonics (2, 3, 4, 5, 7, 9, and 11) of the tempo frequency are considered. Energy within a thin band around these sub-harmonics determines the amplitude related to the beat multiples. In Figure 1, only the energy around frequencies $\frac{1.5}{2}$, $\frac{1.5}{3}$, $\ldots$, $\frac{1.5}{11}$, contributes to the first part of the MCP. In this example, the sub-harmonic $\frac{1.5}{3}$ is significantly predominant and results in a substantial amplitude in the MCP. This high amplitude thus indicates that the song is characterized by 3 beats per measure.

Other examples of MCP computation for real audio songs are shown in Figure 2. In each example, the MCP looks very different, according to their metric properties. In particular, the highest value in the first part of the MCP generally indicates the number of beats per measure, whereas the highest value of the second part is related to the beat subdivision.

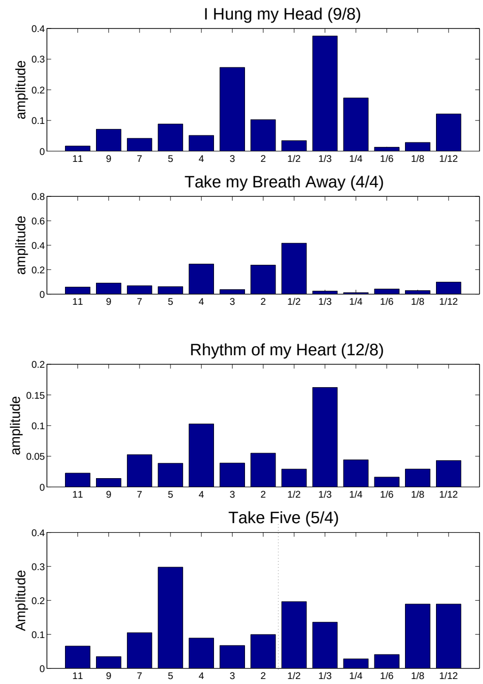


Figure 2. Examples of MCP computed from real audio songs with different time signatures, respectively $9/8$, $4/4$, $12/8$ and $5/4$.

3.3 Distance between MCP

The MCP is proposed for music retrieval purposes. Therefore, a method for computing a matching score between two MCP has to be defined. Several distances are possible, however this is a difficult selection due to the difference in the analysis processes of the two parts of the MCP. We thus propose to consider a global score s as the combination of the two scores s_1 and s_2 obtained with the two parts of the MCP:

$$s = \alpha s_1 + (1 - \alpha) s_2 \quad (1)$$

where α is a fixed weighting value in the interval $[0; 1]$, s_1 the comparison score related to the *meter multiple* part of the MCP, and s_2 the comparison score related to the *meter subdivision*. These two scores s_1 and s_2 are calculated according to correlation:

$$s_i(\text{MCP}_1, \text{MCP}_2) = \frac{c(\text{MCP}_1, \text{MCP}_2)}{\sqrt{c(\text{MCP}_1, \text{MCP}_1)}\sqrt{c(\text{MCP}_2, \text{MCP}_2)}}$$
$$c(\text{MCP}_1, \text{MCP}_2) = \sum_{i=1}^{N} \text{MCP}_1(i)\,\text{MCP}_2(i) \quad (2)$$

where MCP_1 and MCP_2 are MCP vectors of size N.

4. EXPERIMENTS

In this section, experiments are presented that demonstrate the ability of MCP to discriminate songs that have different metric characteristics. The first experiments deal with clustering abilities, and other second concerns song retrieval.

4.1 Databases

Two databases are considered in this paper. The first one is composed of short artificial audio musical pieces (60 seconds long at 16 kHz) that have been synthesized according to different metric properties. The tempo was set to 1 Hz and classes have been constituted considering the time signature. Classes with 2, 3, 4, 5, 7, 9 and 11 beats per measure have been built, in simple time and in compound time. For each class, 2 different distributions of the strong beats in the measure have been chosen, to synthesize 200 pieces. A process has been achieved to randomly add 16^{th} notes in the pieces. The database contains 2800 different files.

The second database is a collection of real pop audio songs indexed using the time signature. The *noise* collection contains 476 simple-meter songs with 2 or 4 beats per measure (sampling rate 44.1 kHz).We constitute another collection of 54 songs with different metric properties. Some of them are in compound time, while others are characterized by a different number of beats per measure. Therefore, 7 different classes are assumed according to 7 different time signatures. The composition of each class is presented in Table 2. Different classes are also deduced from time signatures: one class is composed of 13 songs with 3 beats per measure, another class comprises 24 compound time songs. For all these 54 songs, tempo has been manually annotated.

It is important to notice that all ground truth annotation of the songs from the *noise* database have not been precisely verified; ambiguous meter, large tempo variations, and short-duration time signature changes may result in evaluation errors that may underestimate the quality of the clustering and retrieval systems presented here.

Simple				Compound		
3/4	**5/4**	**7/4**	**11/4**	**6/8**	**9/8**	**12/8**
11	10	7	2	6	2	16

Table 2. Number of songs within each meter class considered for the experiments.

4.2 Clustering

We present here the clustering abilities of the proposed meter feature on the two different databases.

4.2.1 *Evaluation Metrics*

The task can be here reworded as follows: "Do the elements classified together actually belong to the same time signature?". The similarity of the elements of the given database is first computed. As the databases considered in those experiments are of relatively small sizes, we consider an unsupervised clustering scheme (the k-means) to perform the clustering task, *i.e.* each element e_k is given a clustering tag t_k. The correct number of classes is given to the clustering algorithm.

The clustering matrix M is next computed, where each entry is defined as:

$$M(C_x, C_y) = \frac{\#\{(e_k, e_l) | t_k = t_l \wedge e_k \in C_x \wedge e_l \in C_y\}}{\#\{(e_k, e_l) | e_k \in C_x \wedge e_l \in C_y\}} \tag{3}$$

where C_x and C_y are the classes given as ground truth, and (e_k, e_l) is a couple of elements. In order to attenuate the impact of the random initialization of the k-means algorithm, the classification is done 10 times and the mean result over the 10 iterations is considered. The better the classification, the higher the ratio between the diagonal values of M and the remaining of the matrix.

Evaluation over the synthetic database allows us to validate the proposed approach in a controlled environment. Figure 3 depicts the clustering matrix M for the synthetic database. The results are very satisfying in general as most of the values are concentrated on the diagonal, meaning that most of the cluster generated by the k-means algorithm from the features correspond to the actual metrical structureclasses. Typical errors are due to a confusion between simple and compound time, e.g. between 3/4 and 9/8 time signatures, or sometimes between classes for which the numbers of beats per measure have common factors (e.g., 4/4 and 2/4, or 9/4 and 3/4).

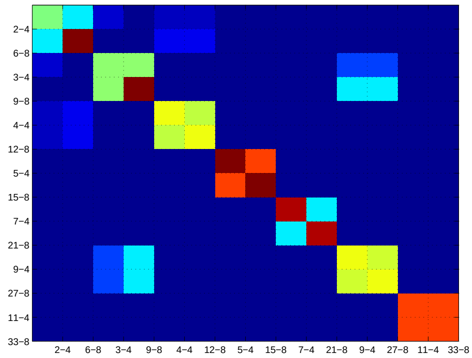


Figure 3. *Clustering matrix over the synthetic database.*

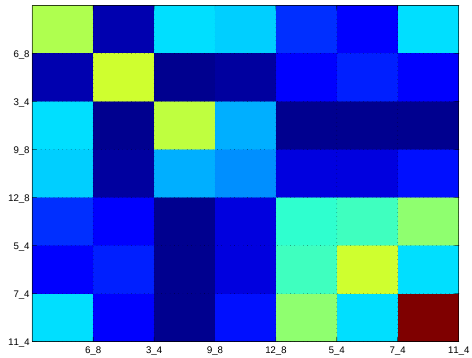


Figure 4. *Clustering matrix over the real database.*

Real data is now considered to evaluate the robustness of the MCP. Evaluations based on clustering allows us to finely analyze the properties of the MCP with respect to the different time signature classes, as shown in Figure 4. We can note that a clear distinction is made between simple and compound time.

4.3 Retrieval

In this section, we propose the evaluation of a music retrieval system based on the MCP. The 54 songs from the real song collection are successively considered as query. The retrieval system computes a similarity score between the query and all the songs of the database. The database is comprised of 530 different songs, including all the queries and all the *noise* collection. These songs are then ranked from most to least similar. For each query, we expect to retrieve all the songs belonging to the class of the query at the top rank. In the following evaluation, results are presented with Precision at Top 1, N and $2N$, in which N denotes the size of the class of the query.

4.3.1 *Synthetic Query*

The first experiments concern retrieval based on a synthetic query. This query is a flat input, *i.e.* a MCP defined by a binary string. For example, if the songs searched have a $3/4$ time signature, the synthetic query is the MCP

$$[0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0]$$

where the only non-null values correspond to the beat multiple 3 and the beat subdivisions $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{8}$ are related to simple time subdivisions. Other experiments only concern the left or right half of the MCP, since retrieval may focus on beat multiples or beat subdivisions. For example, retrieving *compound time* songs is tested by considering only the second part of the MCP with a synthetic query such as:

$$[0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1]$$

which exhibits beat subdivisions $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{12}$.

Table 3 shows the results of the retrieval experiments from those synthetic queries. At the exception of time signatures $11/4$ and $9/8$, one song of the correct class is always correctly retrieved at the first rank. These two exceptions may be explained by the small size the two classes concerned (each comprised of only two songs). Concerning the $11/4$ class, related songs are retrieved at ranks 2 and 11. Concerning the $9/8$ class, related songs are retrieved at ranks 3 and 5. Even if the precision at top 1 is null, the retrieval results are thus quite good.

Average precision at top $2N$ indicates that more than half the songs are generally retrieved within the first $2N$ ranks. Considering one part of the MCP for retrieving songs seems to be effective. For example, a synthetic query allows the retrieval of 75% of the songs of the compound time class at the first N ranks. Almost all the compound time songs are ranked within the first $2N$ best matches. The results of these experiments confirm the quality of the MCP as a metric descriptor for retrieval, since the MCP computed from a real audio file is similar to its time signature properties.

Class	Size	Top 1	Top N	Top 2N
6/8	6	1	0.333	0.500
3/4	11	1	0.364	0.636
9/8	2	0	0.000	0.500
12/8	16	1	0.312	0.375
5/4	10	1	0.800	0.900
7/4	7	1	0.286	0.429
11/4	2	0	0.500	0.500
Total by Class	7	0.714	0.371	0.549
3 beats/mes	13	1	0.846	1.000
5 beats/mes	10	1	0.800	1.000
7 beats/mes	7	1	0.571	0.571
11 beats/mes	2	1	0.500	0.500
Compound	24	1	0.750	0.958

Table 3. Results of the retrieval system based on MCP, considering a synthetic query.

4.3.2 *Audio Query*

We present a second experiment to test the ability of retrieving real songs using a real song as a query. The applications related to these experiments are Query-by-Example systems, which allows users to perform a database search for songs that are similar to a given query song. The similarity here relies on the metric properties, but does not have to be explicitly determined by the user.

Since the query is always retrieved at the first rank, we propose to remove the query from the database. Precision at top N is the number of correct songs retrieved in the first $N-1$ ranks divided by $N-1$, N being the size of the class considered. Precision at top $2N$ is the number of correct songs retrieved in the first $2N-1$ ranks divided by $2N-1$. By using all the songs of each class as a query, N precisions are computed and averaged for each class. Then, the total average is computed by query, or by class. It is respectively denoted *Total by Query*, and *Total by Class*. Such evaluations are respectively named *First Tier* and *Second Tier* in [18].

Table 4 shows the results of the retrieval experiments. The value α determines the weighting between the two parts of the MCP, and has been set to 0.6. The average results by query indicate that 44% of the queries allow the retrieval of one song of the same class at the first rank, 53% of the class is retrieved at the first $2N$ ranks. The accuracy is lower than the results obtained with synthetic queries. This can be explained by the presence of songs in the *noise* database that may be similar to the query. For example, a query with time signature $3/4$ often leads to the retrieval of songs with time signature $9/8$, since the number of beats per measure are the same for each of these two classes and since the beat subdivisions may be varying during the analyzed song (for example in the case of swing). At the opposite, the class $5/4$ leads to the best results: 80% of the correct songs are retrieved.

Moreover, difficulties with annotations of time signatures may lead to errors in evaluation. For example, it is sometimes difficult to discriminate $6/8$ songs from $12/8$ songs. This difficulty is illustrated by the poor results for class $6/8$, whereas the compound time class leads to good results.

Classes	Size	Top 1	Top N	Top 2N
6/8	6	0.000	0.033	0.242
3/4	11	0.727	0.500	0.649
9/8	2	0.000	0.000	0.333
12/8	16	0.562	0.579	0.714
5/4	10	0.700	0.567	0.695
7/4	7	0.000	0.095	0.132
11/4	2	0.000	0.000	0.000
Total by Class	7	0.284	0.253	0.395
Total by Query	54	0.444	0.394	0.529
3 beats/mes	13	1.000	0.686	0.825
Compound	24	0.875	0.784	0.863

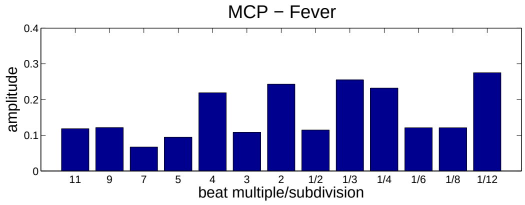
Table 4. Results of the retrieval system based on MCP, considering an audio song as query.

5. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a new mid-level descriptor, related to the beat multiples and subdivisions. Experiments with synthetic and real songs show that considering the MCP allows the retrieval of songs belonging to the same metric class.

When focusing on the time signature, we reduce the search information from the descriptor. Other considerations are also of interest, such as the amplitudes of all the beat subdivisions, which may denote a certain rhythmic complexity of the music. The MCP of one minute of *Fever* by *Ray Charles* is shown in Figure 5. If the time signature of this piece is generally notated 4/4, beat subdivisions at

1/3 are more prevalent than at 1/2. As studied in [19], this is a consequence of the swing. We see here that MCP contains information more complex than the time signature only, and it could thus be used for very specific retrieval purpose.



The estimation of the MCP assumes the prior knowledge of the correct tempo. Its robustness against tempo estimation has to be improved in the future. If the tempo may be automatically estimated, errors are unavoidable and will significantly limit the accuracy of the MCP, and thus the accuracy of the retrieval system.

Furthermore, since metric properties may change during a song, a song may be represented by a sequence of MCP (computed during short frames), in the same way that tonal properties of a song can be represented by a sequence of chromas [20]. Such representation may allow the discrimination of songs with the same metric properties, but with different evolutions with respect to time.

6. ACKNOWLEDGEMENT

The authors would like to thank Jason Hockman for his useful comments during the writing of the paper. This work has been partly funded by the Quaero project within the task 6.4: “Music Search by Similarity” and the French GIP ANR under contracts ANR-06-JCJC-0027-01, DESAM and ANR-JC07-188930, SIMBALS.

7. REFERENCES

[1] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Transactions on Speech and Audio Processing*, 14(5):1832–1844, 2006.

[2] G. Peeters. Template-based estimation of time-varying tempo. *EURASIP Journal on Advances in Signal Processing*, 2007(1), 2007. 14 pages.

[3] J. Brown. Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94(4):1953–1957, 1993.

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

[5] C. Uhle and J. Herre. Estimation of tempo, micro time and time signature from percussive music. In *Proc. of the International Conference on Digital Audio Effects (DAFx*, London, UK, 2003.

[6] S. Dixon, E. Pampalk, and G. Widmer. Classification of dance music by periodicity patterns. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, pages 159–165, Baltimore, 2003.

[7] J. Foote, M.D. Cooper, and U. Nam. Audio retrieval by rhythmic similarity. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, pages 265–266, Paris, France, 2002.

[8] J. Paulus and A. Klapuri. Measuring the similarity of rhythmic patterns. In *Proc. of the International Conference on Musical Information Retrieval (ISMIR)*, pages 150–156, Paris, France, 2002.

[9] A. Holzapfel and Y. Stylianou. A scale transform based method for rhythmic similarity of music. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009.

[10] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Massachussetts, 1985.

[11] D. Temperley. *The Cognition of Basic Musical Structures*. The MIT Press, 2001.

[12] D.J. Povel and P. Essens. Perception of temporal patterns. *Music Perception*, 2:411–440, 1985.

[13] C. S. Lee. *Representing Musical Structure*, chapter The perception of metrical structure: Experimental evidence and a model, pages 59–127. P. Howell, R. West, and I. Cross, London, UK, 1991.

[14] F. Gouyon. *Towards Automatic Description of Musical Audio Signals - Representations, Computional Models and Applications*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2003.

[15] G. Peeters. Rhythm Classification Using Spectral Rhythm Patterns. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 644–647, London, UK, 2005.

[16] O. Lartillot and P. Toiviainen. A Matlab Toolbox for Musical Feature Extraction From Audio. In *Proc. of the International Conference on Digital Audio Effects (DAFx)*, Bordeaux, France, 2007.

[17] M. Alonso, G. Richard, and B. David. Tempo And Beat Estimation Of Musical Signals. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004.

[18] B. De Haas, R. Veltkamp, and F. Wiering. Tonal Pitch Step Distance: A Similarity Measure for Chord Progressions. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, pages 51–56, September 14-18 2008.

[19] K.A. Lindsay and P.R. Nordquist. A technical look at swing rhythm in music. *The Journal of the Acoustical Society of America*, 120(5):3005–3005, 2006.

[20] E. Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, University Pompeu Fabra, Barcelona, Spain, July 2006.