AUTOMATIC MUSICAL METER DETECTION

Mikel Gainza

Audio Research Group, Dublin Institute of Technology
mikel.gainza@dit.ie

ABSTRACT

A method that automatically estimates the metrical structure of a piece of music is presented. The approach is based on the generation of a beat similarity matrix, which provides information about the similarity between any two beats of a piece of music. The repetitive structure of most music is exploited by processing the beat similarity matrix in order to identify similar patterns of beats in different parts of a piece. This principle proves to be equally effective for the detection of both duple and triple meters as well as complex meters. The use of beat positions and dynamic programming techniques allows tracking similar musical patterns formed by beats with moderate tempo deviations. The robustness of the presented approach is reflected by the results presented, where 361 songs are used in order to compare the presented approach against the use of the autocorrelation function in existing state of the art meter detection methods.

Index Terms— rhythm description, meter detection, similarity measures, dynamic programming

1. INTRODUCTION

The analysis of rhythm is a crucial step in the description and understanding of a musical piece. Rhythm is characterised by patterns of musical units that occur at different hierarchical rhythmic levels. The basic rhythmic units are called beats and the rate of repetition of these beats provides the tempo. The patterns by which the beats are grouped provide the meter of a piece of music. As an example, a song in which beats are grouped in pairs, "one, two, one two, one two..." is denoted as having a duple meter. The subdivision of the beat into two or three rhythmic units provides the meter type of the song, where a simple and compound meter corresponds to a division of the beat into 2 or 3 rhythmic units respectively. Finally, the time signature combines information of both meter and meter type together.

There are numerous algorithms that perform tempo extraction or beat tracking [1, 2]. However, the automatic detection of the time signature or the metrical structure of a piece of music remains a relatively unexplored area [3]. In [4-6] the meter is calculated by tracking the peak in a periodicity function, such as the autocorrelation function, which corresponds to the periodicity of the bar line; [4] extracts the most prominent peak in the periodicity function, [5] investigates all peak pairs for possible beat/bar combinations and [6] tracks the periodicity of peaks corresponding to a duple or triple meter. In [7], the meter is classified by using a discrimimant analysis of the autocorrelation function of onset detection functions weighted by different accent types. However, the material used was limited to MIDI format. In [8], Gouyon estimates a song's meter (duple or triple) by tracking periodicities

of low level features around a beat segment. In [9], Klapuri estimates the position of three different hierarchical metrical units, (tatum, beat and bar), by using a probabilistic method based on musical knowledge. However, the purpose of the method is not the estimation of the global meter of the song. In [10], an audio similarity matrix (ASM) [11, 12] is used in order to track similarities in the audio signal between instants separated by beat and bar duration. Thus, the method assumes that successive notes and bars have similar characteristics. However, existing methods discard similarities between bars located at different points in the music, which is generally the case in music styles such as traditional Irish music and popular music, where verses and choruses repeat at different parts of the song. This limitation is addressed in [13], where a method based on the use of different ASMs seeks repetitions in any two possible musical bars of the musical piece. However, this method has some limitations inherent to the use of ASMs; similar patterns of beats with slight tempo deviations will degrade the similarity measure calculation. In addition, the computational requirements required to generate the ASMs are very expensive. These limitations are addressed by the method presented in this paper, which uses beat positions in order to compute a beat similarity matrix. This representation provides information of the similarity between any two beats, as opposed to the similarities between any two time instants. In addition, tempo deviation problems are addressed by the use of dynamic programming techniques to calculate the beat similarity matrix.

The organization of the paper is as follows; Section 2 describes the different components that comprise the meter detection approach. In Section 3, evaluation results of the meter detection algorithm are introduced, which includes comparisons with state of the art methods. Finally, a discussion of the results obtained and some future work are presented in Section 4.

2. PROPOSED SYSTEM

The different parts of the meter detection system are depicted in Figure 1. Firstly, a spectrogram of the audio signal is generated. Following this, individual audio similarity matrices are computed by comparing the spectrogram frames of every two beats of the piece of music. Next, a beat similarity matrix is built by using similarity measures derived from the individual audio similarity matrices. Finally, the existence of similar patterns of beats is investigated by processing the diagonals of the beat similarity matrix.
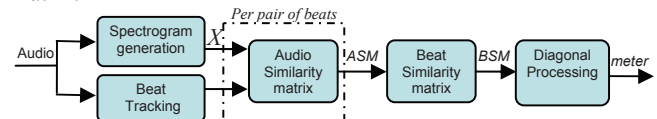

Figure 1: meter detection system block diagram

2.1. Spectrogram generation

A spectrogram is generated from windowed frames of length $L=1024$ samples and a hop size $H=512$ samples, which is equal to half of the frame length.

$$X(m,k) = abs\left(\sum_{n=0}^{L-1} x(n+mH)w(n) * e^{-j(2\pi/N)k.n}\right) \quad (1)$$

where $w(n)$ is a Hanning window that selects an L length block from the input signal $x(n)$, and where m, N and k are the frame index, FFT length and bin number respectively.

In order to improve the computational efficiency of the algorithm, only frequency bins of the spectrogram in the range $k \in \{1...S\}$ are maintained, where S corresponds to the bin located at 5000 Hz.

2.2. Beat tracking

A beat tracking algorithm based on [5] and [2] is used in order to automatically estimate the time instants at which the musical beats occur. As in [5], a complex based onset detector and a periodicity detection method that combines an autocorrelation function with a comb filter are used in order to obtain the beat period τ. Then, dynamic programming is used in order to estimate the sequence of beat locations in the onset detection function separated by τ [2]. The position of the beats is then used in order to compare every two pairs of beats by using corresponding spectrogram frames.

2.3 ASM of every two beats

An individual ASM is built by comparing spectrogram frames of every two beats. In order to improve the efficiency of the algorithm, only frames separated by less than T frames are computed, where T corresponds to 8% of the beat duration. Thus, a frame j of beat x will only be compared with frames $[j\text{-}T,...j\text{+}T]$ of beat y. There are several techniques that measure the similarity between two frames $m=a$ and $m=b$. In [13], the ASM is computed by the using the Euclidian Distance Measure (EDM) as follows:

$$ASM(a,b) = \sum_{k=1}^{S}\left[X(a,k) - X(b,k)\right]^2 \quad (2)$$

In order to remove the dependence on magnitude, the cosine distance measure is used in [10], which is given by:

$$ASM(a,b) = 1 - \frac{\sum_{k=1}^{S}\left[X(a,k) * X(b,k)\right]^2}{\sqrt{\sum_{k=1}^{S}X(a,k)^2} * \sqrt{\sum_{k=1}^{S}X(b,k)^2}} \quad (3)$$

Another widely used technique to compute vector similarities is the Kullback-Leibler (K-L) divergence method [14]:

$$ASM(a,b) = \sum_{k=1}^{N/2} X(a,k) * \log_e \frac{X(a,k)}{X(b,k)} \quad (4)$$

In Figure 2, the spectrogram frames of 2 beats x and y are utilised in order to generate an ASM by using the EDM. As it can be seen in Figure 2, the ASM shows frame similarities within a certain tempo range, where dark blue denotes high similarity.
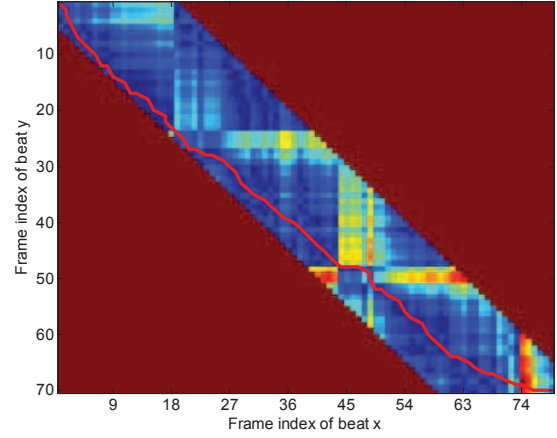

Figure 2: ASM of beat x and beat y

Following this, dynamic programming is used in order to obtain the best path between the top left and the bottom right corners of the ASM which minimises the total similarity cost. Dynamic programming techniques have been used in a large variety of domains, including MIR systems, DNA analysis or automated spelling checking. The principle involves solving a large problem by regarding the problem as the sum of the solution to its recursively solved small problems [15]. In order to find the best path though the ASM of two beats x and y with lengths lx and ly frames respectively, a transition matrix M is generated. Thus, $M_{i,j}$ represents the minimum cost needed to get to the position $[i,j]$ of the ASM from the top left of the matrix:

$$M_{i,j} = ASM_{i,j} + \min(M_{i-1,j-1}, M_{i-1,j}, M_{i,j-1}) \quad (5)$$

The similarity between a pair of beats x and y is given by $S = M_{lx,ly}$. As an example, the best path which minimises the total similarity cost of the ASM shown in Figure 2 is depicted in the same figure. Even though, both compared beats comprise similar note structure, they do not have similar durations, which is due to tempo differences between two renditions of the same beat locations. In addition, the position of one beat was estimated by the beat tracker delayed from its "real" beat time, which complicates the alignment of similar frames of the two beats. As illustrated in Figure 2, these difficulties are overcome by the dynamic programming based method, where the similarities between the notes of both beats were tracked.
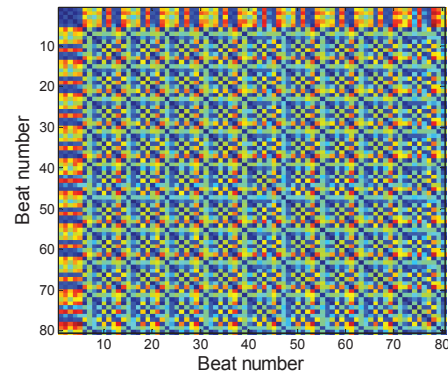

Figure 3: BSM of "Ageispolis" by "Aphex Twin"

2.4. Beat Similarity Matrix (BSM)

The similarity measure S between any pair of beats of a piece of music will be utilised to recursively build a Beat Similarity Matrix. Thus, BSM(x,y) will correspond to the similarity between two beats x and y. As an example, Figure 3 depicts the BSM of a song played in quadruple meter, where the similarities of every two beats are shown. In order to reduce the computational requirements, only one half of the BSM is generated.

2.5 Diagonal processing and meter type calculation

The existence of similar metrical structures in the piece of music is investigated by processing the diagonals of the BSM. Each diagonal represents the similarity between beats separated by a different number of beats. This similarity is measured by calculating the mean of the components of each diagonal of the BSM. Then, the resulting function is inverted in order to build a function d that shows peaks at diagonals in which components exhibit maximal similarity:

$$d_i = mean(diag(ASM_i))$$
$$d = -d + \max(|d|)$$

(6)

where $diag(BSM_i)$ corresponds to the diagonal i of BSM.

In order to resolve ambiguous cases, where similarities in d are not easily discernible, peaks in d are given additional 1.5 weighting. An example of the calculation of d is shown in Figure 4, which is generated by processing the diagonals of the BSM depicted in Figure 3. It can be derived from Figure 4 that the method finds high similarity in musical bars separated by multiples of 8 beats.
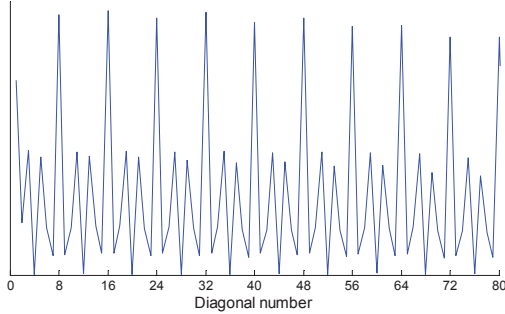

Figure 4: *d* function of Figure 3's example

A large range of meter candidates is considered in the presented meter detection approach. This includes duple and triple meters, which are denoted as $c=2$ and $c=3$, common multiples of duple and triple meters, denoted as $c=4$, 6, and 8 as well as complex meters, denoted as $c=5$, 7, 9 and 11. In order to consider multiples of each meter candidate c, which corresponds to similar beat patterns located at different musical bars, a weighted comb filter is applied to the function d. This resulting function, denoted as T_c, gives more weight to closely separated musical bars as follows:

$$T_c = \sum_{p=1}^{lt} \frac{d(p \times c)}{1 - \frac{p-1}{lt}} \qquad c = 2,...,11 \qquad (7)$$

where lt corresponds to $\lfloor nb/11 \rfloor$ and nb corresponds to the number of beats of the piece of music.

The T_c of Figure 3's example is shown in Figure 5 for all the range of meter candidates. The function shows a distinctive peak at the meter candidate $c=8$, which corresponds to the grouping of 8 beats per musical bar.
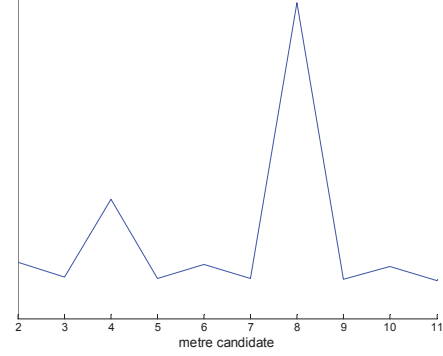

Figure 5: meter detection result example

3. RESULTS

In order to evaluate the performance of the presented approach, a comparative analysis against the standard autocorrelation based meter detection method used in [4-6] is performed. A fair comparison between approaches is ensured by using the same autocorrelation function utilised by the beat tracker introduced in section 2.2, which is denoted as F. Thus, the same beat period τ used by the beat tracker is also used by the autocorrelation function F to calculate the meter for all candidates in the range c. In order to allow deviations in F from perfect multiples of τ, a maximum in F within a region will be used as meter prominence for each meter candidate:

$$ACF_c = \max(F[c*\tau - 2,...,c*\tau + 2]) \qquad (8)$$

where max(E) corresponds to the maximum value of a region E

The presented approach is evaluated by using the 3 different similarity measures of Equations 2 to 4. Since humans perceive rhythm at different rates and the automatic detection of the beat period τ is prone to both half and double errors, meters $c=2$ and $c=8$ in both the song databases and in the results provided by the compared approaches were set to $c=4$. The same principle applies to meter $c=6$, which was set to $c=3$. Firstly, the meter detection methods are evaluated by using a database of duple and triple meters. Next, a database of solely triple meters is used. Finally, a database of complex meters is used.

o **Db1: Database of [9] (mainly quadruple meters)**

In [9], a database of 320 manually annotated musical bar and beats is used. The same database of songs is used in the presented evaluation as follows; the meter of each song in the database is annotated by dividing the median of the difference of the bar annotations by the median of the difference of the beat annotations. Ambiguous results were double checked by the paper's author and the resulting annotations were rounded. The distribution of meters in Db1 is 314 of songs in duple or quadruple meter and the remaining 6 songs in triple meter. The percentage of good detections is shown in Table 1, where EDM, CD and KL denote the use of Euclidian Distance, cosine distance and Kullback-Leibler distance measures respectively in the presented approach and ACF$_c$ denote the use of the autocorrelation function.

ACF$_c$	EDM	CD	KL
95. 07%	94.68 %	95.65 %	88.12 %

Table 1: Percentage of good meter detections for Db1

○ **Db2: Database of triple meters**

In order to investigate the robustness of the algorithms in estimating triple meters, a new database comprised by the 6 songs in [8] played in triple meter and additional 20 songs in triple meter was built. The results are shown in Table 2, where as in Table 1 the results do not greatly differ between the compared methods.

ACF_c	EDM	CD	KL
73.08%	76.92 %	76.92 %	76.92 %

Table 2: Percentage of good meter detections for Db2

○ **Db3: Database of complex meters**

A database of 21 complex meters was also used in order to evaluate the approaches with songs having different metrical patterns. This includes songs played in 5, 7, 9 and 11 beat meters. The results are shown on Table 3, where it can be seen that the presented approach provides significantly better results than the ACF_c for the three similarity measures.

ACF_c	EDM	CD	KL
44.12%	70.59%	67.65%	70.59%

Table 3: Percentage of good meter detections for Db3

4. CONCLUSIONS

A system that detects the meter of a piece of music has been presented. The system has been evaluated by using three different similarity measures in the generation of the ASM, for three different databases and compared against the standard autocorrelation based meter detection method. By considering Db1 and Db2, the results provided by the ACF and the presented approach do not significantly vary in the detection of duple and triple meters. However, the ACF_c had difficulties to track complex meters, which are commonly formed by combining duple and triple meters together. As an example, the autocorrelation function of "Superconductor" by Rush is depicted in Figure 6. This song is played in septuple meter, which can be seen as a "3+4" meter. As it can be seen in the figure, the duple periodicity and its related multiples are more prominent in the autocorrelation function than the 7 beats periodicity. Since the presented approach is purely based on similarity, the similarity of bars in the piece of music will be accurately tracked regardless of the metrical structure of the beats of the bars. This is illustrated in Figure 6, where the meter detection of "Superconductor" by "Rush" is shown for all the range of meter candidates.

The EDM proved to be the most consistent of the three similarity measures across the three different databases. The development of a system that combines the autocorrelation function with the presented approach should warrant future work.
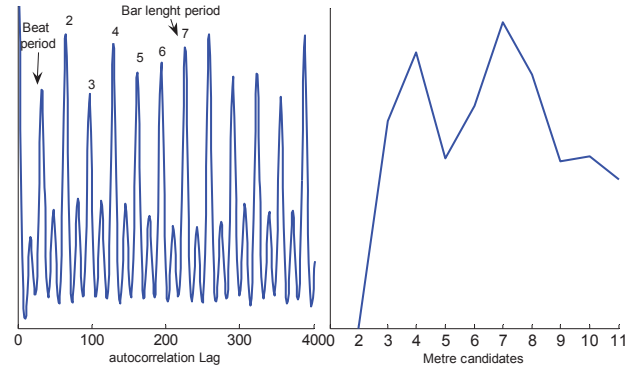
5. ACKNOWLEDGEMENTS

Figure 6: autocorrelation func. (left plot) and *Tc* meter detection (right plot) of "Superconductor" by "Rush"

6. REFERENCES

[1] F. Gouyon, A. Klapuri, et al., "An experimental comparison of audio tempo induction algorithms," *IEEE Trans.on Audio, Speech and Language Processing*, 2006.

[2] D. P. W. Ellis, "Beat Tracking by Dynamic Programming," *Journal of New Music Research,* vol. 36 pp. 51-60., 2007.

[3] F. Gouyon and S. Dixon, "A review of automatic rhythm description systems," *Computer Music Journal*, vol. 29, pp. 34–54, 2005.

[4] J. C. Brown, "Determination of the meter of musical scores by autocorrelation," *Journal of the Acoustical Society of America*, vol. 4, pp. 1953-1957, 1993.

[5] M. Davies and M. D. Plumbley, "Context-dependent beat tracking of musical audio," IEEE Transactions on Audio, Speech and Language Processing, 2007.

[6] S. Dixon, E. Pampalk, et al., "Classification of dance music by periodicity patterns," presented at 4th Int. Conference on Music Information Retrieval, 2003.

[7] P. Toiviainen and T. Eerola, "Autocorrelation in meter induction: The role of accent structure," *The Journal of the Acoustical Society of America*, vol. 119, 2006.

[8] F. Gouyon and P. Herrera, "Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors," AES 114 th Convention, 2003.

[9] A. Klapuri, A. Eronen, et al., "Analysis of the meter of acoustic musical signals," *IEEE Trans. Speech and Audio Processing*, 2004.

[10] A. Pikrakis, I. Antonopoulos, et al., "Music Meter And Tempo Tracking From Raw Polyphonic Audio," 5th Int Conference on Music Information Retrieval, 2004.

[11] J. Foote and S. Uchihashi, "The beat spectrum: a new approach to rhythm analysis," 2001.

[12] J. Foote, "Visualizing Music and Audio using Self-Similarity," ACM Multimedia, Orlando, 1999.

[13] M. Gainza and E. Coyle, "Time Signature Detection by Using a Multi-Resolution Audio Similarity Matrix," Audio Eng. Society 122nd Convention, Viena, 2007.

[14] S. Kullback and R. A. Leibler, "On information sufficiency," *The Annals of Mathematical Statistics* vol. 22, pp. 79–86, 1951.

[15] G. Navarro and M. Raffinot, *Flexible Pattern Matching in Strings: Practical On-Line Search Algorithms Texts and Biological Sequences*: Cambridge Univ. Press, 2002.