Automatic meter extraction from MIDI files
(Extraction automatique de mètres à partir de fichiers MIDI)

Benoit Meudic

Ircam - Centre Pompidou
1, place Igor Stravinsky, 75004 Paris, France
meudic@ircam.fr

Abstract
This paper presents an automatic meter extraction system which relies on auto-correlation coefficients. The input format for music is MIDI, and we assume that a beat and its ocurrences in the musical sequence are known. As output, our algorithm provides a set of possible metric groupings sorted by a confidence criteria. The algorithm has been tested on several pieces of Rameau and results are encouraging.

Keywords
Meter, Rhythm, Music analysis

1 Introduction
According to Cooper et al (1960), “meter is the number of pulses between the more or less regularly recurring accents”. One should remark that in this definition, which will be considered along the article, meter is defined on the assumption that a pulse is known. The main agreed characteristic of meter is its regularity : it is a grouping that can be repeated at least one time in the sequence. Groupings of groupings, if regularly repeated in the sequence, can also be considered as being part of the meter. Thus meter can contain several hierarchical levels. Accents are defined differently by different authors. However, a distinction is often made between metrical accents and others : metrical accents are induced by other accents and then influence our perception of them. The beat is often defined either as the smallest regular grouping of events or as the most perceived one. In the following chapters, we won't focus on possible subdivisions of beat, but only on groupings of beats, being aware that our results will depend on the beat level (note, quarter note etc…) given as input.

If automatic beat extraction from performed music is a topic of active research, little attention has been paid to the study of meter. However, meter is an essential component of rhythm which must be distinguished from the notion of beat (discussions on the relations between beat and meter can be found in Iyer. 1998). The analysis of meter if essential for whom wants to understand musical structure.

Brown (1993) proposes to extract the metric level corresponding to the usual score's signatures directly from an inter-onset sequence (an inter-onset is the duration between two consecutives onsets). Relying on the assumption that "a greater frequency of events occurs on the downbeat of a measure", she proposes to measure it with an auto-correlation method. However, considering the only onsets, she does not take into account all the parameters which participate to our perception of meter. Moreover she assumes that the position of the beginning of the first measure is known, and the method she employs looks for only one repetition of meter in the sequence.

Cambouropoulos (1999) separate the meter extaction task into two phases : the determination of an accentuation structure of the musical sequence, and then the extraction of meter by matching a metrical hierarchic grid onto the accentuation structure. The accentuation structure is determined by gestalt principles of proximity and similarity. One advantage of the method is that contrary to Brown's approach, other parameters than onsets are taken into account in the extraction. However, the matching of a metrical grid with the overall accentuation structure may have drawbacks (it will be dicussed in part 3).

We propose an approach which answers to the drawbacks of the two above methods. It could be seen as a combination of the advantages of thoses methods, but if similar concepts are employed, they are applied to the musical material in a different way.
It is divided in two steps. First we determine a hierarchic structure of beat accents, and then we extract the meter from the hierarchic structure by proposing a new implementation of the method of auto-correlation.

2 choosing a hierarchic structure of accents

In this part, we assume that a sequence of beat segments is given (each segment being a grouping of events). Our goal is to define a hierarchy between the beats according to their propensity to influence our perception of metrical accents. For this, we use the notion of markings.
The marking of a sequence of events is a notion which has been formalised in a theory of rhythm (Lusson 1986) and also in (Mazzola et al 1999). Its is employed without formalism in several music studies (Cooper and Meyer 1960, Lerdhal et al 1983, Cambouropoulos 1999). It consists first in choosing some properties we consider relevant for the sequence and then in weighting the events according to the fact they fulfill or not the property. For instance, considering the property "contains more than three notes", the events containing more than three notes can be weighted 1 and the others 0. Considering several different properties, several weights will be given to each event. Then, if we sum the different weights for each event, we have a measure of the "importance" of the event according to the whole set of properties we have considered. An event which fills all the properties will be high weighted and an event which fill none of the properties will be low weighted. Thus, we measure, with a number, the agreement of a set of properties for each events of a sequence. The events are thus hierarchised.

Of course, the hierarchy depends on the chosen properties. Different properties will provide different hierarchies. That is why we have now to choose relevant properties for our purpose, that is to say detecting beats which make us perceive metrical accents.

Several criteria can be chosen : each beat segment can be marked according to its harmony, its pitch profile, its overall dynamic etc…
For instance, Cambouropoulos (1999) marks the notes according to gestalt principles of proximity and similarity applied to pitches, durations, rests and dynamics.

In our study, we have chosen not to consider the structural relations between the beat segments. Thus, each beat segment was marked considering its own properties, independantly from the properties of the other beat segments. Moreover, the only first event (note or chord) of each beat segment was marked. This drastically reduces the quantity of information which was initially contained in the sequence. Indeed, we wanted in a first approach to validate our method with a minimum set of criteria.

We have considered 5 different markings. The principle we have adopted is to give strong weights to events which combine perceptually important properties (this could be called sonic accents) :

- M1 weights an event proportionnally to its dynamic value
- M2 weights an event proportionnally to its ambitus (interval between pitch extrema)
- M3 weights an event which is followed by a rest
- M4 weights an event proportionnally to its duration
- M5 weights an event proportionnally to the number of notes it contains

For each of the five markings except M3, a weight between 0 and 8 was given to each event of the sequence by scaling the corresponding properties values. M3 which is boolean was given values 0 or 8.

Then, the weights of the markings were added event by event by linear combination. The resulting sequence of weights provided a hierarchic accentuation structure.

3 the detection of groupings in the hierarchised beat sequence

In this part, we assume that a hierarchised sequence of accented beats is given. The problem is to extract meters from this sequence.

3.1 chosen approach

Cambouropoulos (1999) proposes to match a metrical hierarchic grid onto the accentuation structure. The score of the matching for a given grid is the total weights of the accents which coincides with the grid positions. The grid which best fit the accent structure is the one whose different placements onto the accent structure provides "big score changes".

This approach might contain several drawbacks : the criteria of "big score changes" is not clearly defined and thus depends on the user's appreciation. Moreover, the method performs a global measure of the accent strength for each grid positions but do not take in account the variations in the accent structure. One could wonder if an accent profile such as (0 0 1 1 2 2 3 3) would be interpreted as containing a binary meter. Indeed, using the above method, the two scores for the two positions of a binary grid would be the same (0+1+2+3 = 6) so none of the binary grids would be chosen. However, the structure is indeed binary. Finally, the method do not compare different grids (binary, ternary etc…), but different positions for the same grid. Different grid could not even be directly compared because the scores for each grid matching are not meaned which means that for a given accentuation structure, the score for a binary grid will be "a priori" higher than the score for a ternary one (a sequence divided in groups of two contains more elements than a sequence divided in groups of three).

One could wonder if meter can be caracterized by its only positions in a sequence. We think that meter is also characterized by its grouping length which is perceptively salient when compared to other possible groupings lengths.

To answer to theses issues, we propose to extract meter not using a global statistical measure of weights, but using a measure of periodicities. We look for periodic components contained in the accentuation structure.

In order to analyse thoses periodicities, we have chosen the auto-correlation function.
Auto-correlation has already been used in the field of rhythm analysis (Brown (1993), Desain et al (1990)), but it presented some limitations when directly applied to onset sequences : parameters other than onsets were not taken into acount, and the great time deviations resulting from the interpretation of the score could not be always detected. Moreover, when periodicities were detected, the phase (their temporal position in the squence) was not extracted.

Concerning our task, those drawbacks are not important anymore. Indeed, the markings already contain if necessary various informations (events can even be marked according to their structural relations with other events). Moreover, time deviations are not to be considered as the sequence to analyse is composed of regular beats. Lastly, the phase of the meter (ie its position in the sequence) if not provided by auto-correlation, can be determined according to the positions of the highly accented beats.

3.2 definition

The auto-correlation can be defined as follows :
Considering a sequence x[n] of M values (we consider that M is as high as needed), and an integer 0<=m<=M, the auto-correlation A[m] between the sequence x[0..N] and the sequence x[m..m+N] is given by :

$$A[m] = \sum_{n=0}^{N-1} x[n]x[n+m]$$

where N = M-m

The higher the auto-correlation, the higher the similarity between the sequence x[0..N] and the sub-sequence x[m..m+N].

Considering the N+1 values A[0..N] of auto-correlation calculated on the sequence x[0..N], we select the ones which are "local maximum" in a given window centered on their position. Doing this, we select the sub-sequences which are the most correlated with the reference sequence in a given window. The lenght of the window is

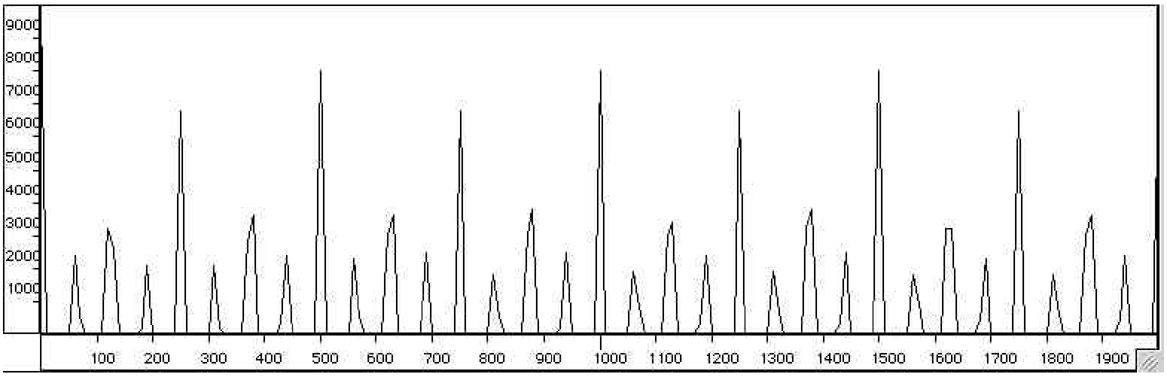

Figure 1. An auto-correlation graph. Horizontally, the sequence of beat accent values. Vertically, the value of auto-correlation. A high value at position p means that sequences x[0..N] and x[p..p+N] are highly correlated.

proportionnal to the position p of the considered sub-sequence. The window is divided in two areas : a small area centered on position p of length 1/3 of p, and a bigest area also centered on p of length p. The position p is considered as a "local maximum" position if the corresponding auto-correlation value is maximum in the small area and at least superior to 1/3 of the values contained in the bigest area.

As result of the auto-correlation, the sequence x[0..N] is associated with its more correlated sub-sequences x[m..m+N], x[n..n+N]. The positions m, n can be seen as the length of the periodic components x[0..m] and x[0..n].

3.3 Application of auto-correlation to our issue

Assuming that we directly compute the auto-correlation onto the sequence of weighted beats Mark[0..N], we would obtain the positions (m, n …) of the more correlated sub-sequences. However, this measure does not look for possible periodicities of other sequences such as Mark[1..1+ N]…Mark[k..k+N] which should be taken into acount in a global analysis of the sequence. Thus, we compute the auto-correlation not only on the sequence Mark[0..0+ N], but also on each sub-sequence Mark[n..N+n]. (a similar method was proposed for measuring musical expression in Desain, 1990). At each step k, the current sub-sequence Mark[k..k+N] is associated with the positions (m, n) of its more correlated sub-sequences Mark[k+m..k+m+N], Mark[k+n..k+n+N]... The values (m, n…) are then interpreted as possible length of meters (measured in number of beats).

As output of the analysis, a list of possible length of meters is proposed for each position k in the beat sequence.

Considering our initial goal, which is the detection of repeated groupings of equal lengths, we sort the different proposed length of meters according to the number of their occurences in the output list.

The first information provided by the sorted list indicates if the sequence is rather binary or ternary. Indeed, if the first proposed length of the sorted list are multiples of three, the sequence can be qualified as ternary, and if the values are multiples of two, the sequence can be qualified as binary.

Assuming for instance that the beats can be grouped by two (binary sequence), the two steps of our algorithm (the marking and the meter extraction) can be applied again, not to the sequence of beats, but to the sequence of the groupings of two beats. The position of the binary grid which determines the position of the groupings in the sequence is chosen so that the addition of the strenght of the events which coincide with the grid is maximum. Then, the accentuation structure is calculated by giving a new accent strenght to each grouping.

Assuming that the first proposed length of the output sorted list is one, then we conclude that there is no higher grouping level for the system of markings we considered.

4 Results

We have analysed the first 35 seconds of 10 of the "Nouvelles Suites de Pieces pour Clavecin" (New Suites of Harpsichord Pieces) from Rameau. Thoses pieces have been selected for their various metric groupings at different levels. The MIDI files which have been analysed are quantized performances. Thus, some indications, which appear in the score such as "tr", will appear as notes in the MIDI file representation. Moreover, some additional notes may also be contained in the MIDI files depending on the performer's interpretation. However, thoses notes do not influence the results. The beat which is considered in the analysis of the MIDI files may not correspond to the beat of the initial score. Indeed, we assume that the MIDI files are performances from which a beat has been automatically extracted. If current beat tracking algorithms do often detect one periodicity which is multiple of the beat, they rarely detect the beat represented in the initial score. One could even wonder if such a detection is possible. Indeed, a composer may have chosen a beat level with its own criteria which do not correspond systematically to the criteria adopted by the beat tracking algorithm. Thus, we did not always considered the same initial beat level as the the score's, in order to show the independancy of our algorithm in regard with this issue.

For each MIDI file, the algorithm proposed several levels of metric groupings. Thoses results are presented in a synthetic way in table 1. The analysing of the results is a difficult task. Indeed, if the measure if often represented in the scores, other metric levels are rarely notated. In our results, we will consider that proposed metric levels which are multiples of the given beat and sub-multiples of the score's measure are relevants. Moreover, levels which are multiples of the measure and which correspond to phrases or motives will also be considered as relevant. Indeed, the segmentation of a musical piece in motives or phrases often corresponds to the metric structure. By the way, we believe that our extraction of different metric levels should be helpful for the detection of phrases and motives.

We will now detail the analysis for one of thoses pieces ("L'indifférente"), which will raise some questions that will be tackled in the discussion part.

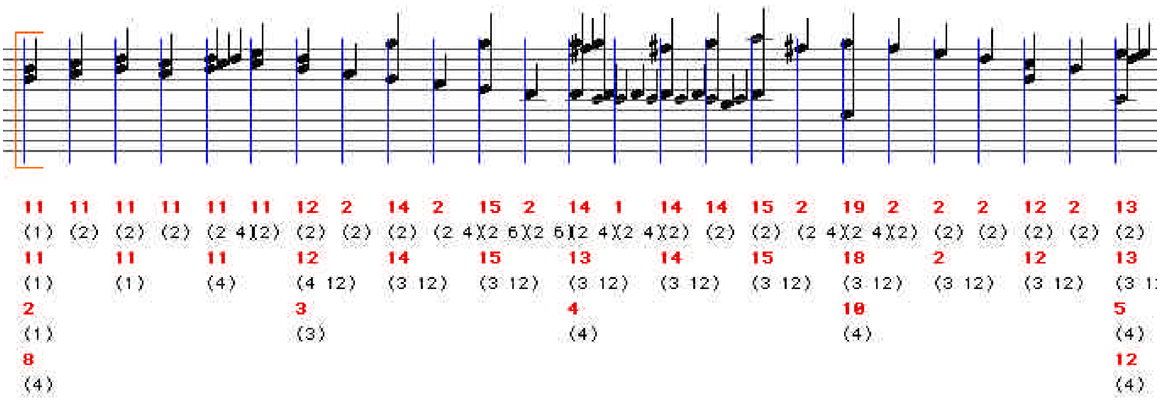

Figure 2. The analysis of the first 7 seconds of "L'indifférente".

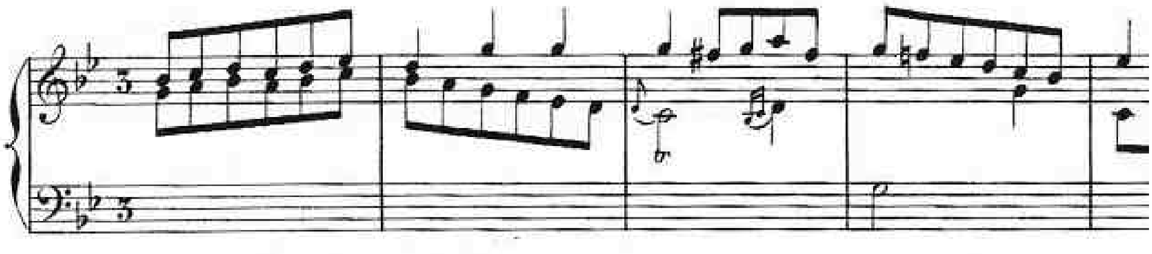

Figure 3 : The score of the first 7 seconds of "L'indifférente"

The first 7 seconds of "L'indifférente" and their analysis are presented in figure 2. The initial beat level which has been chosen corresponds to the eight note of the score.

The score is segmented (vertical blue lines) according to the pulse given as input.

The results of the analysis appear below the score.

Each bold line corresponds to the accentuation structure of one metric level. The first metric level is the beat.

For each metric level, the accentuation structures were calculated as described in section 2 (However, the dynamics were not considered because not provided in the MIDI file).

Under each bold line are the proposed meter length, computed for each event position of the current metric level (as described in section 3).

For instance, starting from the beat level, the first four accent values are 11, 11, 11, 11 and the proposed groupings computed by auto-correlation onto the sequences Mark[0..N], Mark[1..N+1], Mark[2..N+2], Mark[3..N+3] are (1), (2), (2), (2).

The sorted list (not represented in figure 2) of the occurrences of the different proposed groupings for the beat level is :

((2 132) (4 23) (6 5) (10 2) (3 1) (1 1))

The first number of each sub-list is a proposed grouping, and the second number is the total number of sequences Mark[k..k+N] for which the grouping has been proposed.

In this exemple, grouping by 2 is the prefered one with a score of 132. Thus, the upper metric level we choose is the grouping of two beats.

The proposed lists of occurrences of groupings for the four first metric levels are :

Chosen metric level :	Proposed list of groupings :
Beat (eight note in the score)	((2 132) (4 23) (6 5) (10 2) (3 1) (1 1))

2 beats	((3 49) (12 44) (4 17) (6 2) (1 2) (11 1))
6 beats	((4 16) (2 5) (3 1) (1 1))
24 beats	((4 6))
96 beats	…

The metric levels which were represented in the original score of "L'indifférente" correspond quite well with the ones proposed by our algorithm :

Score notation :	Algorithm proposition :
Eight note	beat (given as input)
Quarter note	grouping of 2 beats
Measure (3/4)	grouping of 6 beats
?	grouping of 24 beats
?	grouping of 96 beats

The proposed groupings of 24 and 96 beats, if not represented in the score, are relevant because they correspond to a possible segmentation of the music in phrases and motifs.

We have tested our algorithm on 9 other pieces of Rameau. The results are in table 1. For 7 pieces, the measure level was found, and upper metric levels were proposed. For one of the two other pieces ("Premier Rigaudon"), the double measure level was found (grouping of two measures), and also the phrase level (notated with a double vertical line in the score). For the other piece ("Les tricotets"), the only quarter note level was found starting from the eight note level.

Piece	Proposed groupings	score notation
Allemande	input.beat 4 3	eight note measure ?
Courante	input.beat 3 2	quarter note 1/2 measure measure
Les tricolets	input.beat 2 2 4	eight note quarter note ? ?
Fanfarinette	input.beat 3 2 4	eight note 1/2 measure measure 1/3 phrase
Les trois mains	input.beat 6 2	eight note measure motive
Premier rigaudon	input.beat 8 2	quarter note 2 measures phrase
Sarabande	input.beat 3 3	quarter note measure ?
Gavotte	input.beat 2 3	quarter note measure ?
La triomphante	input.beat 4 6	quarter note measure phrase

Table 1. The results of the analysis of 9 pieces from Rameau.

5 Discussion and Conclusion

We have presented an algorithm which extracts various metric groupings from a MIDI file whose beat is known. The beat do not inevitably corresponds to the beat of the score. It is seen by the algorithm as the lowest periodicity from which different metric levels are calculated recursively. For each metric level, the algorithm outputs a list of possible groupings lengths among which the best groupings are chosen according to a frequency criteria.

In our results, the actual measure level is often proposed (8 pieces on 10). The other proposed metric levels are difficult to evaluate when not notated in score. They sometimes correspond to phrases, motives or double measure. For one piece ("Les tricotets"), the measure level was not reached. This is due to the nature of the piece which is mostly a melody without acompagniement. The only notes corresponding to metric locations do not contain enough information to induce meter, and the pitch contour of the metric segments should be considered in the marking phase.

However, considering the few information which was used to establish the accentuation structure, we consider that our results are quite good and promising.

In a second step, other criteria could be taken into account (for instance the harmonicity of the beat segments). Moreover, the only first event of each beat segment was taken into account, but the others events could also be taken into acount as they also influence our perception of meter. The choice of the markings is a difficult step. One could wonder if the criteria which are relevant for beat extraction are also relevant for meter extraction. For instance, Brower (1993) considers that "the larger timescales associated with meter invoke a different variety of cognition [than our cognition of beat]".

Variations of meters could also be analysed with our algorithm. Indeed, the output list of possible meters contains the evolution of meters along the analysed sequence. Instead of extracting one meter from the global list, we could interpret different areas of stability of the list as different metric sections.

6 Acknowledgments

This research is supported by the european project CUIDADO which aims at developing new content-based technologies for music information retrieval. I want to thank Gérard Assayag for his suggestions about this article.

References

Brown 1993, " Determination of the meter of musicalscores by autocorrelation " *Journal of the Acoustical Society of America 94 (4) 1953-1957 October 1993*

Brower, C. 1993. "Memory and the Perception of Rhythm." *Music Theory Spectrum 15: 19-35.*

Cambouropoulos 1998, " Towards a General Computational Theory of Musical Structure " *The university of Edinburgh, Faculty of Music and Department of Artificial Intelligence, 1998.*

Cooper and Meyer 1960, " The rhytmic structure of music " *Chicago : university of Chicago Press.*

Desain - Siebe de Vos , 1990, " Auto-correlation and the study of Musical Expression " *Proceedings of the 1990 International Computer Music Conference. 357-360.*

Iyer, Vijay S. 1998. "Microstructures of Feel, Macrostructures of Sound: Embodied Cognition in West African and African-American Musics." *Ph.D. dissertation. Berkeley*

Lerdahl Jackendoff 83, " A generative theory of tonal music " *Cambridge : MIT press.*

Lusson 86, " Place d'une théorie générale du rythme parmi les théories analytiques contemporaines " *Analyse Musicale nffi 2, pp. 44-51, 1986.*

Mazzola et al ,1999,"Analyzing Musical Structure and Performance--a Statistical Approach." *Statistical Science. Vol. 14, No. 1, 47-79, 1999*