Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. Geoff Luck & Olivier Brabant (Eds.)

ESTIMATING TEMPO AND METRICAL FEATURES BY TRACKING THE WHOLE METRICAL HIERARCHY

Olivier Lartillot, Donato Cereghetti, Kim Eliard, Wiebke J. Trost, Marc-André Rappaz, Didier Grandjean

Swiss Center for Affective Sciences, University of Geneva, Switzerland
olartillot@gmail.com

Abstract

Meter is known to play a paramount role in the aesthetic appreciation of music, yet computational modelling remains deficient compared to other dimensions of music analysis. Classical audio-based methods detect the temporal repartition of notes, leading to an onset detection curve that is further analysed, in a second step, for periodicity estimation. Current state of the art in onset detection, based on energy and spectral flux, cannot handle complex but common musical configurations such as dense orchestral textures. Our proposed improvement of the flux method can detect new notes while ignoring spectral fluctuation produced by vibrato. Concerning periodicity estimation, we demonstrate the limitation of immediately restricting the range of tempi and of filtering out harmonics of periodicities. We show on the contrary how a complete tracking of a broad set of metrical levels offers a detailed description of the hierarchical metrical structure. One metrical level is selected as referential level defining the tempo and its evolution throughout the piece, by comparing the temporal integration of the autocorrelation score for each level. Tempo change is expressed independently from the choice of a metrical level by computing the difference between successive frames of tempo expressed in logarithmic scale. A new notion of dynamic metrical centroid is introduced in order to show how particular metrical levels dominate at particular moments of the music. Similarly, dynamic metrical strength is defined as a summation of beat strength estimated on dominant metrical levels. The model is illustrated and discussed through the analysis of classical music excerpts.

Keywords: periodicity estimation, onset detection, metrical analysis

1. Introduction

The metrical dimension of music is known to play a paramount role in the aesthetic appreciation of music, including emotion, yet computational modelling remains particularly deficient, compared to other dimensions of music.

Classical audio-based methods detect in a first step the temporal repartition of notes, leading to an onset detection curve that is further analysed, in a second step, for periodicity estimation. This paper follows the same two-step approach, and introduces new methods for each step.

The model has been implemented in *MIRtoolbox* (Lartillot & Toiviainen, 2007) and is available in the new version 1.5 of the toolbox.

2. Onset detection curve

The first step consists in producing an "onset detection curve", which is a temporal curve: musical events are indicated by peaks; the height of each peak is related to the importance of the related event, in terms of energy and/or spectral contrast. There are two main approaches to obtain such onset curve:

One approach consists in extracting an amplitude *envelope* curve, showing the general evolution of energy along time, corresponding to the following command in *MIRtoolbox*:

o = mironsets(a, 'Envelope')

where *a* can be for instance the name of an audio file.

Another approach consists in computing spectral *flux*, i.e. in evaluating the distance

with respect to the global spectral distribution between successive instants:

o = mironsets(a, 'Flux')

Both methods give relevant results for music where notes are sufficiently isolated or accentuated with respect to the background environment, such as in Fig. 1a and 1b. But when dealing with a complex orchestral sound where notes cannot be detected based on global energy changes, such as in Fig 2a and 2b, things get more complex. The *'Envelope'* method does not work since the general envelope indicates the global dynamic change without revealing the note onset position hidden in the polyphonic texture. The *'Flux'* method, on the other hand, can detect notes in the polyphony but may fail in the presence of vibrato.

For that reason, we have developed an improved version of the flux method, called 'Emerge', that is able to detect more notes and in the same time ignore the spectral variation produced by vibrato. More precisely, when comparing two successive frames, for each periodicity, the energy from the new frame that exceeds the energy level *for a range of similar periodicities* from the previous frame is summed. By looking not only at the exact same periodicity in the previous frame, but also similar periodicities, this allows to ignore slight changes of periodicities. For the moment, the frequency tolerance has been simply fixed to an arbitrary value that corresponds to a maximal frequency difference between successive frames of 17 Hz.

The new onset curve is available in MIRtoolbox 1.5 by using the following command:

o = mironsets(a, 'Emerge')

The onset curves related to this new *'Emerge'* method are shown in Fig. 1c and 2c. We can also observe how the choice of onset curve has an impact in the estimation of periodicity, further discussed in the next section, since the *autocorrelogram* (the matrix showing the autocorrelation function frame by frame, as in Figure 3) computed with the new *'Emerge'* method shows more clearly the metrical structure than those computed with the other methods: In Fig. 3a, using the *'Flux'* method, the three metrical levels 1, 2 and 3 are not clearly shown in the autocorrelogram, but still detected by the metrical tracker, whereas they are clearly shown in Fig. 3b, using the *'Emerge'* method. Besides, the *'Emerge'* method clearly shows the subdivision of level 1 into six sub-beats. Notice also how accentuations on the fifth sub-beat (level 5/6) in the second half of the excerpt, shown in Fig. 3b, in a very constant tempo is roughly understood in Fig. 3a as a global tempo increase at level 1.

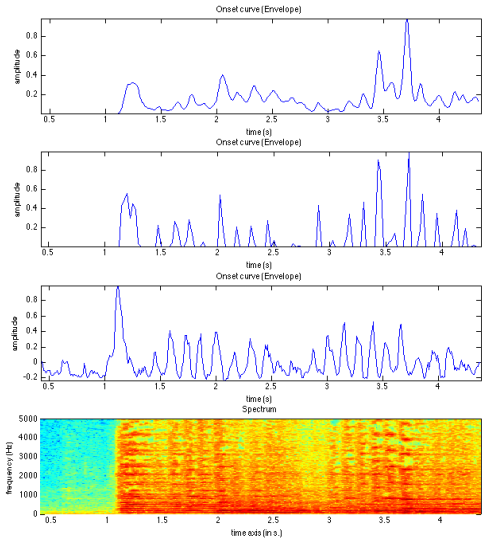

Figure 1. Three different onset curves extracted from the first seconds of a performance of the 3rd movement of C.P.E. Bach's *Concerto for cello in A major*, WQ 172, using the *'Envelope'* (1a), *'Flux'* (1b) and *'Emerge'* (1c) methods, with the detailed spectrogram (1d) used for the *'Emerge'* method.

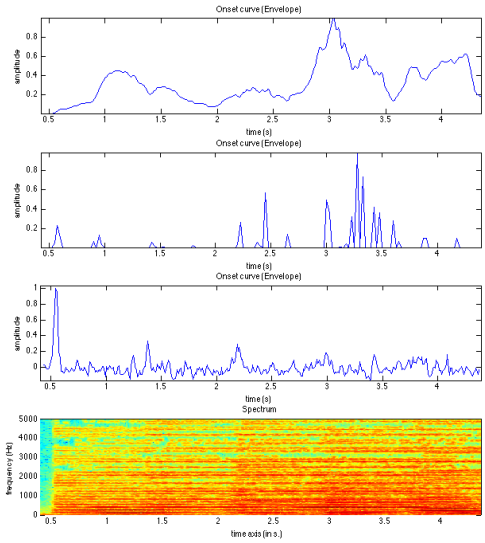

Figure 2. Three different onset curves and spectrogram extracted from the first seconds of a performance of the *Aria* of J.S. Bach's *Orchestral suite No.3 in D minor*, BWV 1068, using the same approaches as in Fig. 1.

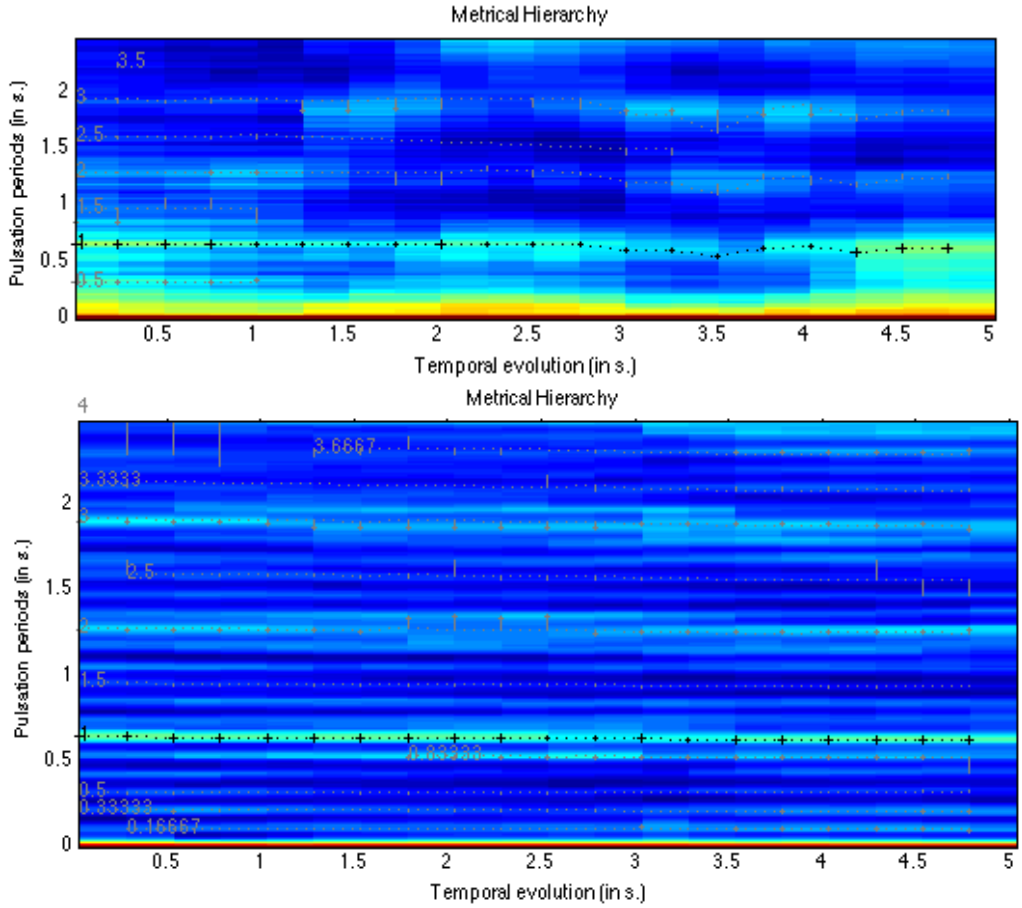

Figure 3. Metrical analysis of the first five seconds of a performance of the *Finale* (*Allegro energico*) of M. Bruch's *Violin Concerto No.1 in G minor*, op.26. Both figures show an *autocorrelogram*: each successive column corresponds to a time frame of 5 s, starting from 0 s and moving every .25 s. For each frame in each column the autocorrelation function is represented, showing periodicities with warm colours (green-yellow) at the given period in s. (on Y-axis). The autocorrelogram (3a) is computed using the *'Flux'* onset curve method, while for the second one (3b) the new *'Emerge'* method is used. On top of the autocorrelogram the metrical structure as tracked by the algorithm is annotated (cf. text for an explanation).

3. Periodicity estimation

Pulsation corresponds to a periodicity in the succession of peaks in the onset curve. This periodicity can be detected through the computation of autocorrelation function on successive large frames (of a few seconds) of the onset curve, such as:

ac = mirautocor(o, 'Frame')
mirpeaks(ac, 'Total', 1)

In the presence of a given pulsation in the musical excerpt that is being analyzed – let's say with a BPM of 120, i.e., with two pulses per second – the autocorrelation function will indicate a high autocorrelation score related to the period .5 s. But generally if there is a pulsation at a given tempo, subdivisions of the pulsation can also be found that are twice slower (1 s), three times slower, etc. For that reason, the autocorrelation function usually shows a series of peaks equally distant for all multiples of a given period. This has close connections with the notion of metrical structure in music, with the hierarchy ordering the levels of rhythmical values such as whole notes, half notes, quarter notes, etc.

One common approach to extract the tempo from the autocorrelation function is to select the highest peak, within a range of beat periodicities considered as most adequate, typically between 40 and 200 BPM, with the possible use of a resonance curve (Toiviainen & Snyder 2003).

This can be performed in *MIRtoolbox* by calling the *mirtempo* operator and toggling off

the *'Metre'* option (further presented in the next sections):

mirtempo(a, 'Metre', 'No')

One main limitation of this approach is that if different metrical levels are emphasized throughout the temporal development (Fig. 4a), the tempo tracking will constantly switch from one BPM value to another one twice slower, twice faster, etc (Fig. 4b). This would happen very often, since in most music, successions of same durations can often be followed by succession of durations twice slower or faster for instance.

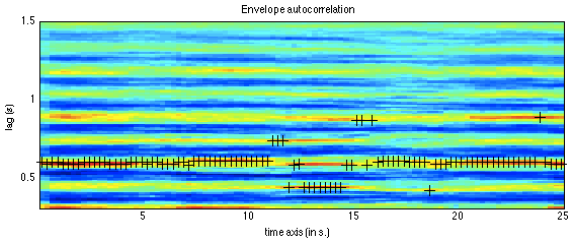

Figure 4a. Metrical analysis of the first seconds of the first movement of a performance of J.S. Bach's *Brandenburg concert No.2 in F Major*, BWV 1047. Traditional tempo extraction approach based on detecting the most dominant pulse frame by frame from the processed autocorrelogram.

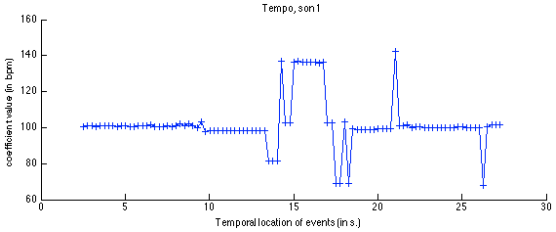

Figure 4b. Tempo curve resulting from the approach presented in Fig. 4a.

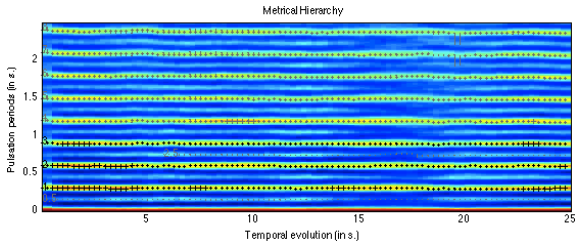

Figure 4c. New method constructing a metrical structure from the unprocessed autocorrelogram.

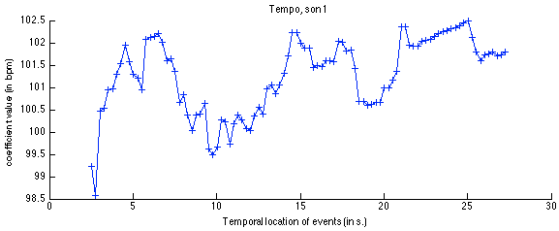

Figure 4d. Tempo curve resulting from the approach presented in Fig. 4b.

4. Metrical structure tracking

As a solution to this problem, we propose to track a large part of the metrical structure, by following in parallel each metrical level separately and combining all the levels in one single hierarchical structure.

The metrical structure of any audio file can be computed and displayed in *MIRtoolbox* 1.5 using the following command:

mirmetre(a)

Examples of metrical structures are shown in Fig. 3. Metrical levels are indicated with lines of crosses that are linked together between successive frames with dotted lines. The level index is indicated on the left of each line. The dominant metrical level, indicated as level 1, is drawn in black, while other levels are shown in light brown. The cross size indicates the related pulse strength, corresponding to the autocorrelation score for that periodicity. If the actual periodicity is deviated from the theoretical harmonic series of periodicities expected from a metrical structure, a vertical line is drawn between the actual and the theoretical periods.

In Fig. 3a, level 1 is subdivided into one sublevel .5, corresponding to a binary rhythm, with multiples 1.5, 2.5, etc. In Fig. 3b, on the contrary, level 1 is subdivided into six subbeats, with its elementary level 1/6, as well as its half slower level 2/6 corresponding to a ternary division of level 1, and finally its three times slower level 3/6 corresponding to a binary division of level 1.

Other examples of metrical structures are given in Fig. 4c, 5a and 6a. We can notice for instance in Fig. 6a that at the middle of the excerpt, the ternary rhythm turns into a binary rhythm for a dozen of seconds.

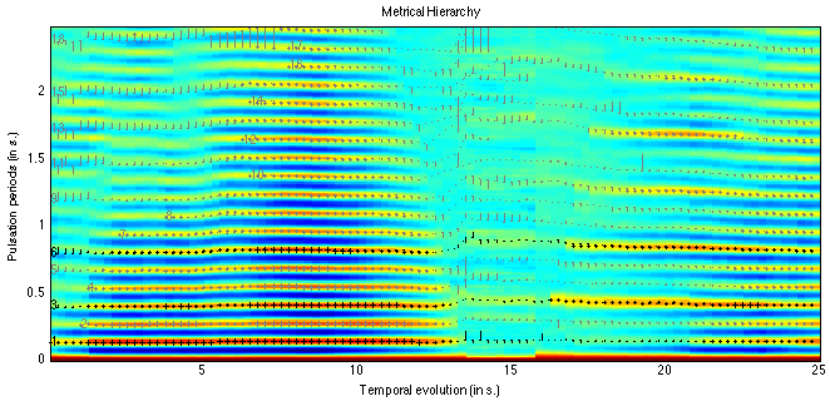

Figure 5a. Autocorrelogram with tracking of the metrical structure for the first seconds of a performance of the 3rd movement of C.P.E. Bach's *Concerto for cello in A major*, WQ 172.

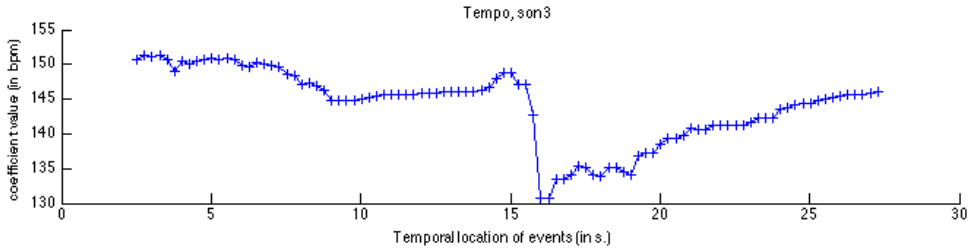

Figure 5b. Corresponding tempo curve.

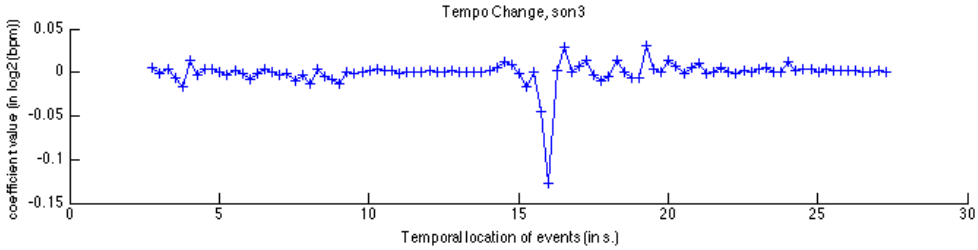

Figure 5c. Corresponding tempo change curve.

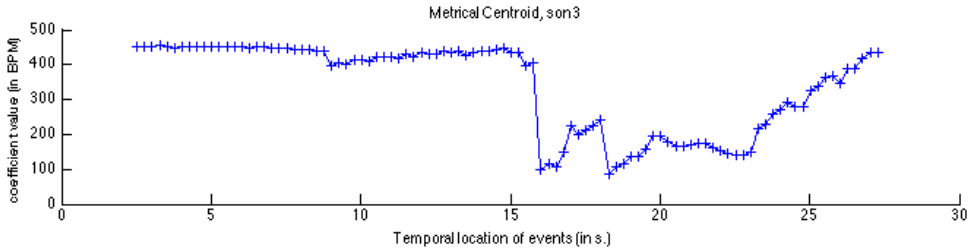

Figure 5d. Corresponding metrical centroid curve.

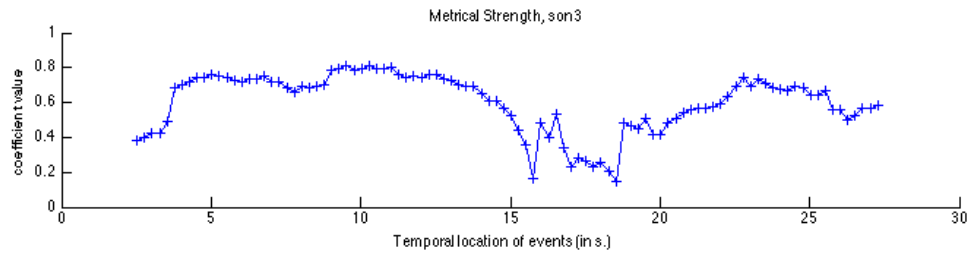

Figure 5e. Corresponding metrical strength curve.

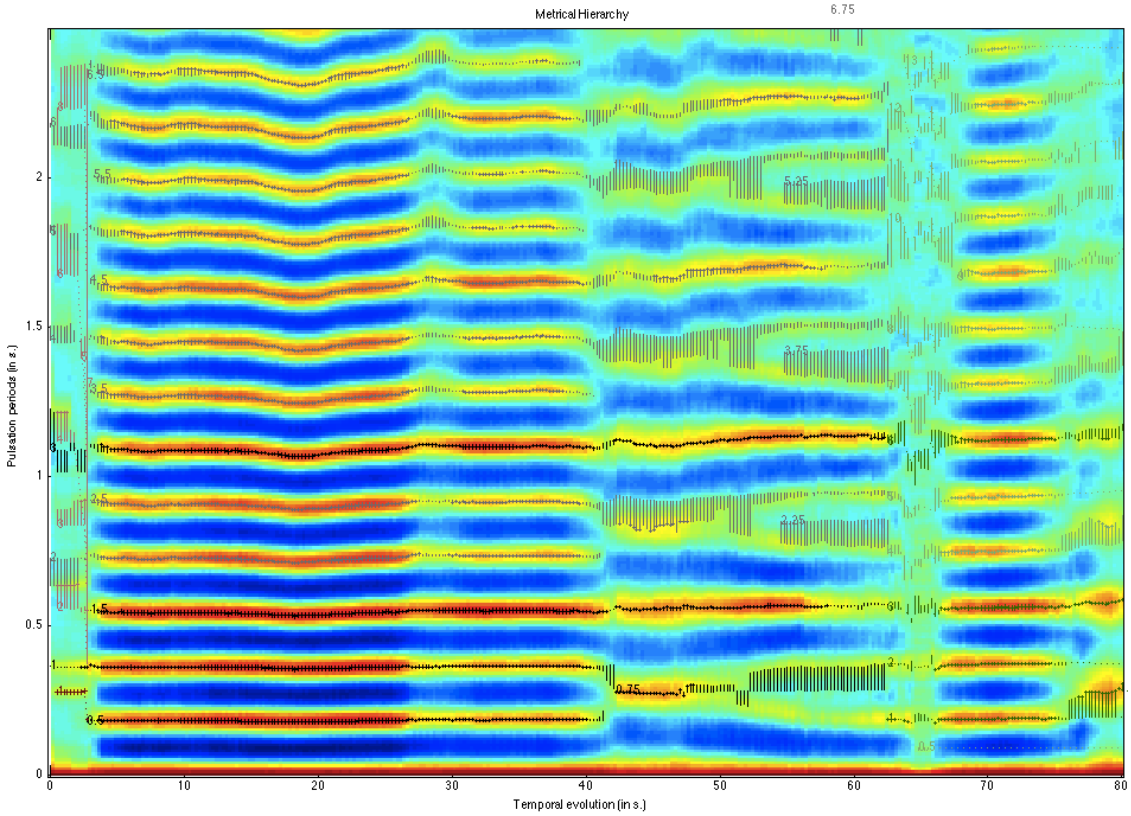

Figure 6a. Autocorrelogram with tracking of the metrical structure for the first 80 seconds of a performance of the *Scherzo* of L. van Beethoven's *Symphony No.9 in D minor, op.125*.

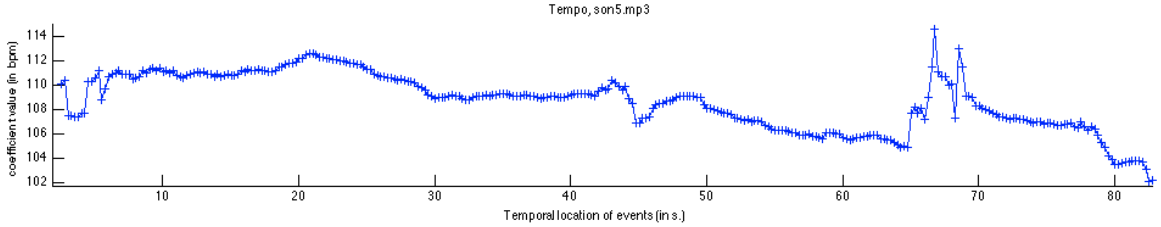

Figure 6b. Corresponding tempo curve.

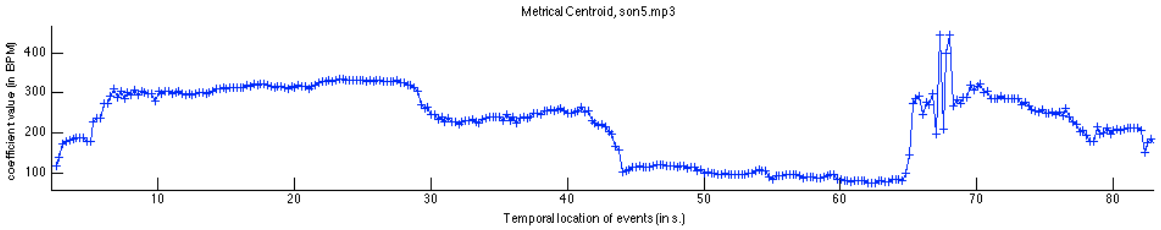

Figure 6c. Corresponding metrical centroid curve.

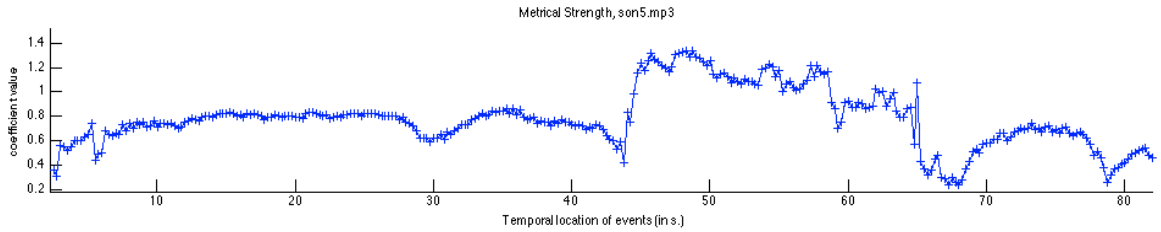

Figure 6d. Corresponding metrical strength curve.

5. Tempo and tempo change

Once the metrical structure has been constructed, one metrical level is chosen as the referential level that defines the tempo and its evolution throughout the piece (Fig. 4d). This metrical level is chosen due to its high degree of saliency and also because it is related to a tempo rate that fits best to the range of best perceived tempos. To each metrical level is associated a global sum by summing up the autocorrelation scores over time for each metrical level separately, and by weighting this total score with a resonance curve (Toiviainen & Snyder 2003) in order to emphasize most easily perceived pulsations. The metrical level with the highest global sum is selected as the main metrical level that defines the tempo.

This approach for tempo estimation based on metrical structure is used by default by the *mirtempo* operator in the new version 1.5 of *MIRtoolbox.* So it can be called simply like this:

mirtempo(a)

Fig. 4 compares the tempo tracking methods. The classical paradigm is based on selecting a preferred range of BPMs in the autocorrelogram, and choosing the maximum autocorrelation score at each frame (Fig. 4a). This leads to a tempo curve with a lot of shifts from one metrical level to nother (Fig. 4b). The new method builds a metrical structure (Fig. 4c), which enables to find coherent metrical levels leading to a continuous tempo curve (Fig. 4d). Other examples of tempo curve are given in Fig. 5b and 6b.

The selection of a main metrical level as referential level for the computation of tempo values remains somewhat subjective. Often neighboring levels (twice faster, twice slower, etc.) could have been selected as well. On the other hand, the dynamic evolution of tempo seems to play a more important role for the listener as it describes how music speeds up or slows down, in parallel along all the metrical levels. Tempo change is expressed independently from the choice of a metrical level by computing the difference between successive frames of tempo expressed in logarithmic scale.

The tempo change curve is computed in *MIRtoolbox* 1.5 using the following command:

mirtempo(a, 'Change')

An example of tempo change is given in Fig. 5c.

6. Dynamic metrical centroid

On the other hand, the fact that particular metrical levels may be more dominant than others at particular moments of the music is an important aspect of the appreciation of rhythm. The common method of selecting the most dominant metrical level at each successive frame is not satisfying, as it would lead to shifts between metrical levels that are somewhat artificial and chaotic. Instead of selecting one single metrical level at each frame, we introduce a new assessment of metrical activity that is based on the computation of the centroid of a range of selected metrical levels. Not all levels found in the autocorrelogram are taken in the computation of the centroid, because there can exist dozens of them, without theoretical limitations.

Only levels corresponding to actual theoretical metrical levels, such as whole notes, half notes, quarter notes, etc., are selected. This selection is performed automatically, so that it can detect whether the metrical structure is binary, ternary, etc. More precisely, for each frame considered in isolation, metrical levels whose strengths (defined by the autocorrelation value at those points) are higher than the strengths of all their underlying metrical sub-levels are selected. This corresponds to metrical levels that are N times faster for all $N > 1$. Indeed, if a given metrical level (let's say level 3) is weaker than one if its underlying metrical sub-level (for instance level 1), this means that the grouping of three pulses at level 1 does not emerge from the succession of such pulses. On the other hand, if that metrical level (3) is stronger than the immediately lower sub-level (2), this means that the grouping of 3 notes has still more importance than the grouping of 2 notes. For that reason we propose to select such metrical levels as well.

For each frame, the dominant metrical levels are selected and the centroid of their periodicity (in seconds) is computed, using as

weights their related autocorrelation scores. A refined version of the algorithm defines the weights as the amount of autocorrelation score at that specific level that exceeds the autocorrelation score of the underlying metrical sub-levels. In this way, any sudden change in the number of selected metrical levels from one time frame to the successive one does not lead to abrupt changes in the metrical centroid curve.

The resulting metrical centroid curve indicates the temporal evolution of the metrical activity. The metrical centroid values are expressed in BPM, so that they can be compared with the tempo values also in BPM. High BPM values for the metrical centroid indicate that more elementary metrical levels (i.e., very fast levels corresponding to very fast rhythmical values) predominate. Low BPM values indicate on the contrary that higher metrical levels (i.e., slow pulsations corresponding to whole notes, bars, etc.) predominate. If one particular level is particularly dominant, the value of the metrical centroid naturally approaches the corresponding tempo value on that particular level.

The metrical centroid is computed in MIRtoolbox 1.5 using the following command:

mirmetroid(a)

Examples of metrical centroid curves are given in Fig. 5d and 6c. We can notice for instance in the metrical structure in Fig. 5a that the emphasis is put first on the fastest level (level 1), followed by a progressive activation of levels 3 and 6 from t = 6 s. Then, after a break around t = 15 s, level 3 becomes dominant. This can be seen in the metrical centroid curve, first focusing on the fastest pulsation (around 450 BPM) on the first half of the excerpt, followed by a focus on the lower pulsations (around 100 and 200 BPM) on the second half.

7. Dynamic metrical strength

Another major description of the metrical activity is assessing its strength, i.e., whether there is a clear and strong pulsation, or even a strong metrical hierarchy, or whether on in the other hand the pulsation is somewhat hidden, unclear, or there is a complex mixture of pulsations. Studies have been carried out in the old paradigm – i.e., one single metrical level detected at a time, as discussed in Section 3. In such case, since there is just one beat or pulsation, the strength is therefore related to that single metrical level (Lartillot et al., 2008). Following one simple traditional approach, beat strength is simply identified with the autocorrelation score of that main metrical level.

We propose a simple generalization of this metrical strength approach by simply summing the autocorrelation scores of the selected dominant levels (using the same selection method as in last section). The metrical strength is increased by any increase of autocorrelation score at any dominant level, or if new dominant levels are added to the selection. Whereas the autocorrelation score is a value lower than 1, the metrical strength can exceed 1.

The metrical strength is implicitly computed when assessing the metrical centroid, and can for that reason be obtained in *MIRtoolbox* as a second output (below: *ms*) of the *mirmetroid* command:

[mc ms] = mirmetroid(a)

Examples of metrical strength curves are given in Fig. 5e and 6d.

8. Collinearity between metrical features

In this section, we evaluate the collinearity of the most important metrical features introduced in this paper (tempo change, metrical centroid, metrical strength) – computed using the current version of the algorithms while writing the paper, i.e., *MIRtoolbox* 1.4.1.4 – to which we add two additional features. The first additional feature is *pulse clarity*, i.e. the strength of the main pulse detected at each frame (Lartillot et al., 2008), obtained in MIRtoolbox using:

mirpulseclarity(a, 'Frame')

The other additional feature is *metrical novelty*, i.e. the novelty curve, computed using the new method (Lartillot et al., 2013), in its beta-version in MIRtoolbox 1.4.1.4, based on the autocorrelogram function *ac* used for the assessment of the metrical structure (cf. §3):

mirnovelty(ac, 'Flux')

All these features were extracted from thirty-six musical excerpts covering a large range

Table 1. Pairwise Pearson product-moment correlation coefficients between the following features: pulse clarity (pc), metrical centroid (mc), metrical strength (ms), novelty of the autocorrelation computed from the onset curve (nv), tempo change (tc). Features are replaced with their square roots.

	pc	mc	ms	nv	tc
pc	r = 1 N=21721 p= ---	r = 0.203 N=21681 p<0.01	r = 0.053 N=21721 p<0.01	r = -0.167 N=21685 p<0.01	r = 0.006 N=21637 p=0.38
mc		r = 1 N=21682 p= ---	r = -0.314 N=21682 p<0.01	r = 0.057 N=21648 p<0.01	r = 0.006 N=21634 p=0.40
ms			r = 1 N=21722 p= ---	r = -0.036 N=21686 p<0.01	r = 0.011 N=21638 p=0.11
nv				r = 1 N=21686 p= ---	r = -0.002 N=21638 p=0.77
tc					r = 1 N=21638 p= ---

of musical styles from baroque to contemporary classical music (Eliard et al., 2013; Eliard & Grandjean, in preparation) with a mean duration of 155.83 ± 10.66 seconds. Frame size of the moving window was fixed to 1 second and the hop factor was fixed to 0.25 seconds. Since these features were Gamma distributed, Pearson product-moment correlation coefficients were computed on their square roots. Correlations were also calculated using pairwise deletion. The results of the correlations are shown in Table 1.

According to Cohen's convention (1988) effects sizes are small, except for the correlation between metrical centroid and metrical strength ($r = -0.314$, $p < 0.01$). Results also suggest that these features are relatively independents.

9. Discussion

This new computational model, freely available in the new version 1.5 of *MIRtoolbox*, enables a relatively robust assessment of tempo and its dynamic evolution in any piece of music. Even more, it offers a very detailed description of the metrical structure, revealing import aspect of the metrical structure that are independent from the tempo dimension: the new concept of metrical centroid (or *metroid*) that we are introducing, and metrical strength.

We may expect high impact of these metrical dimensions to the emotional experience of music, in particular with respect to activity (Russell, 1980) and more particularly to energetic arousal, and maybe tense arousal as well (Thayer, 1989).

Small deviations from the metrical structure in musical performances are a typical means of musical expressivity, and are therefore crucial for the affective impact of music. The methods presented here provide a procedure to quantify these metrical features, which will be very useful in studies that try to better understand the link of musical structure and performance features with emotions in response to music.

The actual tempo level is not so much of importance in this respect, because a same musical excerpt can often be associated with several parallel tempi in harmonic relation (let's say 60 and 120 BPM). Indeed, tempo is rather a musical convention that does not necessarily reflect the listener's subjective experience. We may expect however a positive correlation between a change in tempo and a change in arousal. An interesting complementary descriptor here is metroid, which could be

also correlated with arousal, not only in terms of metroid change, but also in terms of actual metroid value. Metroid is somewhat independent from tempo: a change of metroid can be independent from a change of tempo, as can be seen for instance in Fig. 6b and 6c. Finally metrical strength can have also an important impact in arousal. The relative contribution from these different tempo and metrical descriptors to the two aspects of arousal, energetic and tense, remains an open question.

Tempo and metrical descriptions might also aspect the other main dimension of the emotional appreciation of music, i.e., valence. But a complete study of this aspect might require additional rhythmical and metrical descriptions, related to accentuation in particular, which would depend also on aspects related to dynamics in general, register and timbre.

Tempi and metrical changes as well as the clarity of them are probably crucial to explain the subjective entrainment during music listening and it has been proposed that this phenomenon is important in emotion emergence in music (Juslin et al., 2010). Entrainment, both visceral and motor components, seems to play an important role in the emergence of feelings during musical listening (Labbé & Grandjean, submitted). Moreover, these tempi and metric features might be combined with dynamic subjective musical feelings and/or dynamic perceived emotions in music. Such combination will allow the assessment of causality in the emergence of complex emotions using for example Granger causality measures to investigate how different characteristics of tempi and metrics might be crucial in the emergence of subtle musical feelings such as them described in Geneva Emotion Musical Scales (Zentner, Grandjean, & Scherer, 2008).

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Eliard, K., Cereghetti, D., Lartillot, O. & Grandjean, D. (2013). Acoustical and musical structure as predictors of the emotions expressed by music. *Proceedings of the 3rd International Conference on Music & Emotion*, Jyväskylä, Finland.

Eliard, K. & Grandjean, D. (in prep). Dynamic approach to the study of the emotions expressed by music.

Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In *Handbook of music and emotion: Theory, research, applications, Series in Affective Science*. New York : Oxford University Press.

Labbé, C., & D. Grandjean (submitted). Self - Reported Motor and Visceral Entrainment Predicts Different Feelings during Music Listening.

Lartillot, O., & Toiviainen, P. (2007). MIR in Matlab (II): A toolbox for musical feature extraction from audio. *Proceedings of the International Conference on Music Information Retrieval*, Wien, Austria.

Lartillot, O., Eerola, T., Toiviainen, P., & Fornari, J. (2008). Multi-feature modeling of pulse clarity: Design, validation and optimization. *Proceedings of the International Conference on Music Information Retrieval*, Philadelphia, PA, USA.

Lartillot, O., Cereghetti, D., Eliard, K., & Grandjean, D. (2013). A simple, high-yield method for assessing structural novelty. *Proceedings of the 3rd International Conference on Music & Emotion*, Jyväskylä, Finland.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.

Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*. Oxford University Press, New York, USA.

Toiviainen, P., & Snyder, J.S. (2003). Tapping to Bach: Resonance-based modeling of pulse. *Music Perception*, 21(1), 43-80.

Zentner, M., Grandjean D., & Scherer K. R. (2008). Emotions evoked by the sound of music: Differentiation, classification, and measurement. Emotion. 8(4), 494-521.