MIRTEMPO 1.6: TEMPO ESTIMATION BY TRACKING A COMPLETE METRICAL STRUCTURE USING A RICH ONSET DETECTOR

Olivier Lartillot
Finnish Centre of Excellence in Interdisciplinary Music Research
olartillot@gmail.com

ABSTRACT

This paper describes the general properties of the system submitted to the MIREX 2013 tempo estimation tasks.

1. INTRODUCTION

The tempo estimation model has been built within the *MIRtoolbox* platform [1] [1]. The new improvements presented in this paper will be made available in the upcoming version 1.6 of the toolbox.

2. ONSET CURVE

We use our new *'Emerge'* onset detector released in *MIRtoolbox* 1.5 that can handle vibrato and dense textures [2]. The method is based on an improvement and generalization of the flux method that look at particular time / frequency region and can tolerate spectral fluctuations of limited frequency range.

A high-resolution spectrogram is computed from the signal, with 50 ms frame length and 10 ms hop, a frequency range from 0 to 5000 Hz and a frequency resolution of 0.1 Hz. 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.

In *MIRtoolbox* 1.5 and later, this corresponds to the following *mironsets* command:

o = mironsets('mysong', 'Emerge', 'Detect', 0)

3. PERIODICITY ESTIMATION

Tempo is estimated by computing autocorrelation functions, on a moving window of frame length 5 seconds and hop factor 5%, for a range of time lags that corresponds to a tempo range between 24 and 500 BPM. The autocorrelation curve is normalized so that the autocorrelation at zero lag is identically 1.

In *MIRtoolbox*, this corresponds to the following commands:

o = mirframe(o, 5, .05)
a = mirautocor(o, 'Min', 60/500, *'NormalWindow', 0)*

One interesting problem with autocorrelation functions is that a lag can be selected as prominent because it is found often in the signal although the lag is not repeated successively. We propose a simple solution based on the following property: For a given lag to be repeated at least twice, the periodicity score associated with twice the lag should have a high probability score as well. This heuristics can be implemented as a single post-processing operations applied to the autocorrelation function, removing all periodicity candidate that to not have stronger periodicity at twice its lag.

4. PEAK PICKING

A peak picking is applied to the frame-by-frame autocorrelation functions. The beginning and the end of the autocorrelation curves are not taken into consideration for peak picking as they do not correspond to actual local maxima.

A given local maximum will be considered as a peak if its distance with the previous and successive local minima (if any) is higher than this threshold .05. This distance is expressed with respect to the total amplitude of the input signal. This distance of .05 is hence equivalent to 5 % of the distance between the global maximum and the minimum of the input signal [3].

The peak position and amplitude are estimated more precisely using quadratic interpolation.

In *MIRtoolbox*, this would correspond to the following *mirpeaks* command:

p = mirpeaks(a, 'Total', Inf, 'Threshold', 0, 'Contrast', .05, 'NoBegin', 'NoEnd', 'Normalize', 'Local')

5. TRACKING THE WHOLE METRICAL HIERARCHY

In the presence of a given pulsation in the musical excerpt that is being analyzed – lets say with a BPM of 120, i.e., with two pulses per second – the periodicity function will

[1] http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials /mirtoolbox

indicate a high periodicity score related to the period .5 s. But generally if there is a pulsation at a given tempo, multiples of the pulsation can also be found that are twice slower (1 s), three times slower, etc. For that reason, the periodicity 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.

We track large part of the metrical structure, by following in parallel each metrical level separately and combining all the levels in one single hierarchical structure. In this metrical hierarchy, a limited number of metrical levels are detected as dominant levels, for particular periods of time in the piece of music being analyzed. Dominant levels might sometimes correspond to what previous approaches consider as tactus and bar beats.

The internal model of metrical hierarchy considers that pulse lags of individual metrical levels are in exact integer relation one with the others. Apart from the first dominant metrical level discovered i_0, each metrical level i is dependent on another metrical level i_{r_i}: its theoretical pulse lag $\widehat{\tau^i}$ is at any time instant n a multiple or division of its referential metrical level:

$$\widehat{\tau_n^i} = \widehat{\tau_n^{i_{r_i}}} \times m^i \text{ or } \widehat{\tau^i} = \widehat{\tau_n^{i_{r_i}}}/d^i \tag{1}$$

The pulse lags of the entire metrical hierarchy at a time instant n is therefore conditioned solely by the pulse lag $\widehat{\tau_n^{i_0}}$ of one single level i_0, associated with the first dominant level discovered.

$$\widehat{\tau_n^i} = \widehat{\tau_n^{i_0}} \times l^i \tag{2}$$

5.1 Causal algorithm

The analysis is causal: the whole process is carried out for each successive time instant, during which all the levels of the metrical hierarchy are tentatively mapped with the peaks of the periodicity curve at that given time frame n, i.e. real lag values of the form τ_n^i are given to the different levels i. In the same time, the theoretical set of values given by equation 2 are updated so that they map as closely as possible with the real values.

For each successive time frame n, peaks k in the periodicity function are considered in decreasing order of periodicity score p_k.

Each peak k, related to a periodicity lag t_k is tentatively mapped to one metrical level i. For that aim, a succession of tests is carried out, using a particular threshold δ_k, for comparing the lags of the peak and of the metrical levels. This threshold δ_k depends on the min-max normalized periodicity score p_k^*:

$$\delta_k = \delta_0 \times (1 + (1 - p_k^*)^2) \tag{3}$$

because stronger pulses are supposed to continue more smoothly the metrical levels than weaker ones.

- We first try to map this peak to one dominant metrical level $i \in D$:

$$i_k^D = \arg\min_{i \in D} \left| t_k - \tau_*^i \right| \tag{4}$$

where τ_*^i indicates the current periodicity lag value associated with level i, it can be τ_n^i if there has already been a peak at the current time frame n associated with that level, or else its value at the most recent frame where a peak was found $\tau_{n-m}^i, m < M$.

That association is confirmed if the peak is close enough to that dominant metrical level:

$$\left| t_k - \tau_*^{i_k^D} \right| < \delta_k \text{ and } \left| \log_2 \left(\frac{t_k}{\tau_*^{i_k^D}} \right) \right| < .2 \tag{5}$$

$$\implies \tau_n^{i_k^A} = t_k \tag{6}$$

- If the previous test does not succeed, we try to associate the peak to any currently active metrical level $i \in A$:

$$i_k^A = \arg\min_{i \in A} \left(\min(|t_k - \tau_*^i|, |t_k - t_{k-1}|) \right) \tag{7}$$

  - If no peak has been integrated into the metrical hierarchy at the current time frame n, the chosen metrical level is candidate to become dominant, which would be likely only if this integration is particularly smooth:

$$\left| t_k - \tau_*^{i_k^A} \right| < .1 \text{ and } \left| \log_2 \left(\frac{t_k}{\tau_*^{i_k^A}} \right) \right| < .2 \tag{8}$$

$$\implies \tau_n^{i_k^D} = t_k \tag{9}$$

    If this succeeds, the chosen metrical level is considered as dominant if its current peak periodicity score is sufficiently high and if its referential metrical level is also already dominant:

$$p_k > \theta \text{ and } i_{r_{i_k^A}} \in D \implies i_k^A \in D \tag{10}$$

  - In the other cases, this integration can be considered under a loosen condition:

$$\left| t_k - \tau_*^{i_k^A} \right| < \delta_k \implies \tau_n^{i_k^A} = t_k \tag{11}$$

    If the current peak has a pulse lag t_k that is closer to the theoretical peak than the currently registered metrical level lag $\widehat{\tau_n^{i_k^A}}$ is, then the metrical level is updated:

$$\left| t_k - \tau_n^{i_k^A} \right| < \left| \widehat{\tau_n^{i_k^A}} - \tau_n^{i_k^A} \right| \implies \tau_n^{i_k^A} = t_k \tag{12}$$

If that same peak is sufficiently strong ($p_k > .1$), we check whether it initiates a new metrical level:

- For all the slower metrical levels, we find those that have a theoretical pulse lag that is in integer ratio with the peak lag:

$$i \in A, \min \left(\frac{\widehat{\tau_n^i}}{t_k} \bmod 1, 1 - \left(\frac{\widehat{\tau_n^i}}{t_k} \bmod 1 \right) \right) < \epsilon \tag{13}$$

where ϵ is set to to .02 if no other stronger peak in the current time frame n has been identified with the metrical hierarchy, and else to .2 in the other case.

If we find several of those slower levels in integer ratio, we select the fastest one, unless we find a slower one with a ratio defined in equation 13 that would be closer to 0.

- Similarly, for all the faster metrical levels, we find those that have a theoretical pulse lag that is in integer ratio with the peak lag:

$$i \in A, \min\left(\frac{t_k}{\widehat{\tau_n^i}} \bmod 1, 1 - \left(\frac{t_k}{\widehat{\tau_n^i}} \bmod 1\right)\right) < \epsilon \tag{14}$$

where ϵ is set to to .02 if no other stronger peak in the current time frame n has been identified with the metrical hierarchy, and else to .2 in the other case.

- If we have found both a slower and a faster level, we select the one with stronger periodicity score.

- This gives us a referential metrical level i_R, upon which our new discovered metrical level i_N will be based. The level index l_{i_N} of the new metrical level is defined as:

$$l_{i_N} = l_{i_R} * \left[\frac{t_k}{\widehat{\tau_n^{i_R}}}\right] \tag{15}$$

Finally, if the strongest periodicity peak in the given time frame n is not associated with any level of the metrical hierarchy, a new metrical hierarchy is created, with a single metrical level related to that peak. These multiple metrical hierarchies live parallel existences, and the algorithm continues by tentatively mapping the peaks of the periodicity curve on these multiple hierarchies in parallel. Mechanisms have also been conceived to fuse multiple hierarchies whenever it turns out that they belong to a single hierarchy.

Once all the peaks p_k of a given time frame n have been considered, the theoretical pulse lags are updated based on the new empirical data collected.

For lack of space, the details of the models are not given in this paper. Values used for some parameters defined in this section: $\delta_0 = .1, \theta = .15, M = 10$.

Complete examples of metrical structures are shown and discussed in [2].

In *MIRtoolbox* 1.5 and later, this metrical analysis can be performed by simply calling the new *mirmetre* operator.

6. TEMPO-RELATED METRICAL LEVEL SELECTION

If several metrical hierarchies have been constructed on a given musical except, the metrical hierarchy covering the largest temporal span is selected for the definition of the tempo.

For the selected metrical hierarchy, to each metrical level is associated a numerical score, computed as a summation across frames of the related periodicity score for each frame, each frame-related score being decreased by a certain amount depending of the deviation of the BPM of that individual peak with respect to the global BPM of the metrical hierarchy.

Then we construct all possible metrical structures, made of a series of levels that have integer ratio, and that could be related to the idea of tactus/tatum/bar decomposition. We choose the metrical structure that yields the best overall score (obtained by summing the score related to each selected level).

For the selected metrical structure, we finally select two most dominant levels with periodicity between 30 and 300 BPMs by choosing the two levels with best scores. Before selection, these scores are first weighted by a resonance curve that indicates a preference for periodicities closer to 120 BPMs.

For each of the two selected metrical levels, the final tempo value is obtained by taking the median of the BPM values collected across frames.

7. REFERENCES

[1] O. Lartillot, and P. Toiviainen: "A Matlab Toolbox for Musical Feature Extraction From Audio," *Proceedings of the International Conference on Digital Audio Effects (DAFx 2007)*, 2007.

[2] Olivier Lartillot, Donato Cereghetti, Kim Eliard, Wiebke J. Trost, Marc-Andr Rappaz, and Didier Grandjean: "Estimating tempo and metrical features by tracking the whole metrical hierarchy," *Proceedings of the 3rd International Conference on Music & Emotion (ICME 2013)*, 2013.

[3] O. Lartillot, *MIRtoolbox User's Manual*, http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox