TEMPO ESTIMATION BASED ON MULTIPATH TEMPO TRACKING

Bruno Di Giorgi, Massimiliano Zanoni, Augusto Sarti
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

ABSTRACT

We present a tempo tracking algorithm based on the efficient generation and joint steering of multiple trackers (paths). This solution leads to improved computational efficiency and scalability.

1. ALGORITHM DESCRIPTION

The starting point of the algorithm is the rhythmogram $R(t, \tau)$ computed as described in [1]. $R(t, \tau)$ expresses the salience of the tempo τ at time t. In order to take into account the fact that many individuals can generally perceive different tempi, we make use of a number N of simple trackers. At the end of the tracking algorithm each one of them will give a tempo path hypothesis together with a score value representing the probability of his estimate. The trackers are initialized with different tempi and proceed simultaneously, but they are not fully independent. In fact, at each step their states are supervised by a manager, with the aim of avoiding convergence on a single path. At the end of the tracking stage, each tempo path is summarized by its temporal median. Finally, the most likely pair of trackers is picked by maximizing a "coupling" score and their relative strength is computed.

In Section 1.1 we describe the behavior of a simple tracker. The manager is analyzed in Section 1.2 and the choice of the two tempi is discussed in Section 1.3.

1.1 Simple tracker

The simple tracker exploits the information contained in the rhythmogram $R(t, \tau)$ to track a path of the time-varying tempo. It is initialized at t_0 with a tempo τ_0. At the n-th step it finds the most likely successive tempo τ_{n+1} among a set of candidates for succession. The choice is achieved by maximizing a score function obtained by weighting a transition score between the current tempo and the candidates, and a salience, given by the rhythmogram values of those candidates. Once the tracker reaches the last frame of the rhythmogram it is terminated. By cumulating the score at each step it maintains a self-evaluation value and obtains, once terminated, the final score of its path.

While the appearence is similar to the viterbi decoding or dynamic programming, this technique is inherently different because it does not take into account all different paths but only a small subset of them. In spite of not returning the global optimum path, this method is very efficient and performs quite nicely for the application at hand. Futhermore it lends itself to the extension of increasing the nunber of trackers to rise the accuracy expectation.

1.2 Manager

When running more simple trackers, we desire to better cover the solution space in the search of a more general optimality. The space of solutions, however, is only effectively spanned if the trackers evolve as independently as possible, which does not naturally occur. Whenever two trackers meet, in fact, they stick together unless we introduce specific rules that will promote independence. In particular, every time two paths cross, we successively apply the two following rules:

- "Equalize the past": the current scores of the two trackers are compared, and the path of lower score becomes a copy of the winner's one.
- "Differentiate the future": we perturb one of the two (now equal) paths in order to prevent them from continue following the same tempo path.

1.3 Coupling score

At the end of the tracking stage, the i-th tracker is represented by the temporal median of its tempo path τ_i and a global score of its path s_i. In order to find the optimal pair of estimates, we take into account both these values.

For each pair of trackers (i, j), their tempi τ_i and τ_j are well coupled if they have a small, integer ratio $r = \tau_i/\tau_j$. This is enforced by a score function

$$f(r) = \frac{1}{r(\lfloor r \rceil - r)} \tag{1}$$

where $\lfloor r \rceil$ indicates the nearest integer and division by 0 is avoided by using a threshold floor on the denominator.

The other constraint is given simply by the sum of the two scores $s_i + s_j$. The relative strength of the i-th estimate is given by $s_i/(s_i + s_j)$.

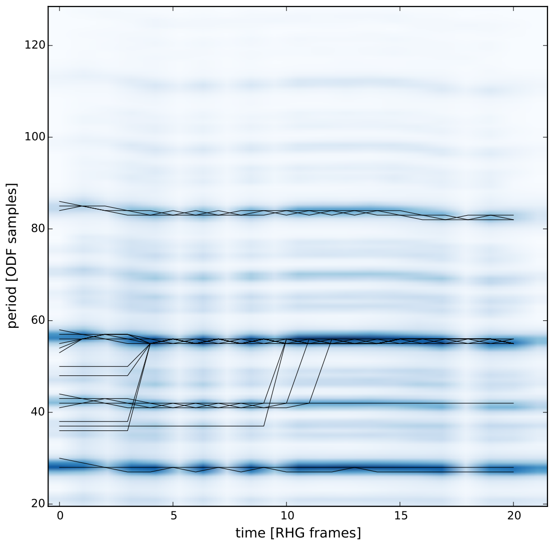

Figure 1: The different paths of each tracker are the black lines, overlayed above the rhythmogram. On the x axis the time is expressed in rhythmogram frames, while on the y axis the tempi are expressed in onset-detection-function samples per beat.

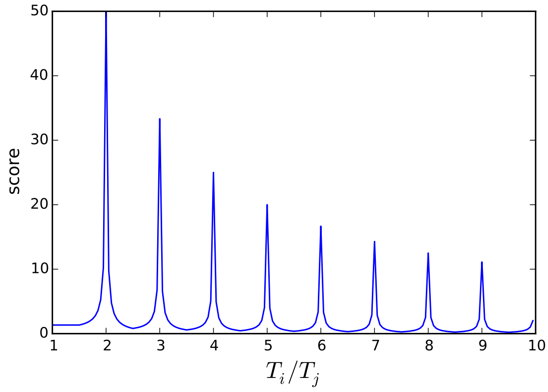

Figure 2: score function $f(r)$

2. REFERENCES

[1] Matthew EP Davies and Mark D Plumbley. Context-dependent beat tracking of musical audio. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3):1009–1020, Mar. 2007.