TEMPO EXTRACTION FOR AUDIO RECORDINGS

Miguel Alonso, Bertrand David and Gaël Richard
GET–Télécom Paris
46 rue Barrault, Paris 75634
cedex 14, France
{miguel.alonso,bertrand.david,gael.richard}@enst.fr

ABSTRACT

Nowadays, the problem of estimating the tempo of audio recordings receives a large amount of attention from the automatic audio processing community. Applications for this task include automatic playlist generation, synchronization of audio tracks, computer based music transcription, music information retrieval... This paper briefly presents a technique for estimating and tracking the tempo of audio recordings. This approach relies on a front end that estimates phenomenal accents (onsets) and their respective time location. The second step consists in a periodicity detection block that calculates the beat rate of the audio signal and it is followed by a dynamic programming stage that performs tempo tracking. This

Keywords: energy flux, phenomenal accents, dynamic programming.

1 Description of the first algorithm

This year we submit two algorithms. The first one corresponds to the same proposal submitted last year for Mirex'05 with some minor modifications. In addition, we also submit another algorithm based on a similar principle, but using more recent versions of the corresponding system components.

It is assumed that the beat of the audio signal is relatively constant, at least during the duration of the tempo analysis window.

The system that we proposed can be divided into four major steps:

- *phenomenal accent detection:* also called onset detection, refers to locating discrete temporal events in an audio stream where there is a marked change in any of the perceived psychoacoustical properties of sound: loudness, timbre and pitch (Lerdahl and Jackendoff, 1983);
- *periodicity estimation:* consists in detecting the rate at which phenomenal accents appear;
- *periodicity tracking:* this part is carried out by a dynamic programming algorithm that finds the best periodicity path through succesive tempo analysis frames;
- *Path selection and tempo salience estimation:* this module selects the two main tempi from the various periodicity paths found by the tracking algorithm, then it estimates the relative salience of the lower tempo.

The general overview of the system is presented in Figure 1. The algorithm works as follows: the input signal is resampled at a lower frequency ($f_s = 22,050$ Hz) to obtain $x(n)$. This is done in order to reduce the computational burden, knowing that the rhythmic properties of the original audio remain unaltered. Next, $x(n)$ is decomposed into eight-uniform non-overlapping subbands using a maximally decimated polyphase filter bank.

A musical stress profile indicating the potential location of phenomenal accents is computed for each subband signal. This was done using the system presented in (Alonso et al., 2005), where a perceptually plausible power envelope is calculated. Then its derivative is computed using an efficient differentiator filter and a detection function that bears onsets as peaks is obtained.

The periodicity of the detection function is obtained using three methods widely employed in pitch estimation: the summary autocorrelation function, the spectral sum and the spectral product. The procedure followed is explained in (Alonso et al., 2004). After processing each tempo analysis window, the output of every method is stored in its respective time–frequency matrix. Then, a dynamic programming algorithm is used to determine and track the optimum paths of tempo candidates in each analysis frame. For the Mirex contest, to compute the tempo ($\mathbb{T}$) of the excerpt under analysis, the optimum paths found by every periodicity method are first time-averaged followed by an inter-algorithm average.

In the last part, the two main tempi are selected from the set of paths computed by the tracking algorithm. The selection criteria combines path energy (salience) and *a priori* information about the human preferences concerning the beat period as suggested by Moelants (2002) and McKinney and Moelants (2004).

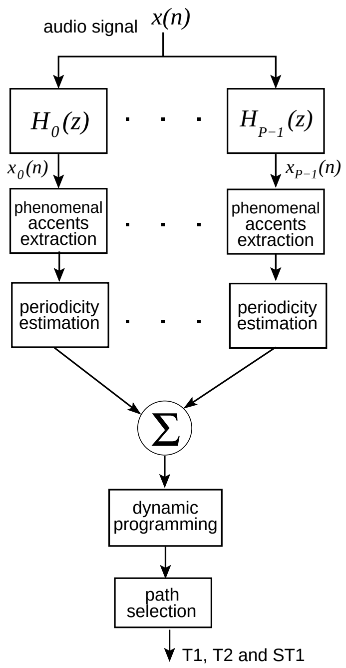

Figure 1: Overview of the system.

2 Description of the second algorithm

This approach bears close resemblance to the algorithm described above. In this variant, the musical stress estimation block has been improved. Instead of computing the traditional spectrogram, this variant uses a reassigned version to enhance simultaneously the resolution in time and frequency.

This algorithm uses only the spectral sum for periodicity estimation, instead of the three technies used in the previous approach. Additionally, the periodicity tracking block has also been upgraded to improve its robustness to tempo variations. Finally, the tempi selection block uses a slightly different scheme for the selection of periodicity paths since only period estimations from one are available.

References

M. Alonso, B. David, and G. Richard. Tempo and beat estimation of music signals. In *Proc. Int. Symposium on Music Inf. Retriev (ISMIR)*, 2004.

M. Alonso, G. Richard, and B. David. Extracting note onsets from musical recordings. In *Proc. IEEE Int. Conf. on Multimedia & Expo (ICME)*, 2005.

F. Lerdahl and R. Jackendoff. *A generative theory of tonal music*. MIT Press, Cambridge, Massachusetts, 1983.

M. F. McKinney and D. Moelants. Extracting the perceptual tempo from music. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2004.

D. Moelants. Preferred tempo reconsidered. In *Proc. of the 7th Int. Conf. on Music Perception and Cognition*, pages 580–583, 2002.