TEMPO EXTRACTION FOR AUDIO RECORDINGS

Miguel Alonso, Bertrand David and Gaël Richard
GET–Télécom Paris
46 rue Barrault, Paris, 75634
cedex 13, France
{miguel.alonso,bertrand.david,gael.richard}@enst.fr

ABSTRACT

Nowadays, the problem of tempo estimation 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 article presents a technique for estimating and tracking the tempo and beat phase in audio recordings. The proposed method relies on a front end that detects phenomenal accents and their respective time location. The second step consists in calculating the periodicity inherent in the audio signal and it is followed by a dynamic programming stage that tracks the course of these periodicities in time.

Keywords: energy flux, phenomenal accents, dynamic programming.

1 Algorithm description

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 propose 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;
- *beat phase estimation:* finds the time location of the first beat from the beginning of the audio signal.

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 (Scheirer, 1998). Next, $x(n)$ is decomposed into eight-uniform non-overlapping subbands using a maximally decimated polyphase filter bank.

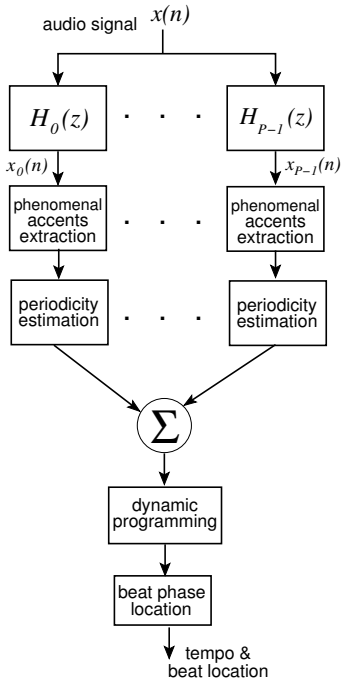

Figure 1: Overview of the system.

Phenomenal accents are detected in every subband signal 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. The output is a *detection function* that exhibits sharp peaks at transients and note onsets.

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). The detection function is decomposed in frames, whose respective periodicity is calculated and stored as rows to form a time–frequency matrix.

Global score	0.689
At least one tempo correct	95.00 %
Both tempi correct	55.71 %
At least one phase correct	25.00 %
Both phases correct	5.00 %
Mean absolute difference of scored saliences	0.239
Runtime (seconds)	2875

Table 1: Mirex contest results for the proposed algorithm

Then, a dynamic programming algorithm is used to determine and track the optimum paths of tempo candidates at each analysis frame. For the Mirex contest, to estimate the tempi ($\mathbb{T}_{1,2}$) of the excerpt, the best paths found by every periodicity method are first time-averaged followed by a single fusion method using a majority rule, i.e., the paths that are common to more than one method have a higher score than those found by a single method. If no consensus among methods is found, the spectral product result is taken as the tempo estimation.

Once the two best tempi $\mathbb{T}_1$ and $\mathbb{T}_2$ have been estimated, the relative salience is computed using the scores obtained for both tempi ($S_{\mathbb{T}_1}, S_{\mathbb{T}_2}$)

$$\text{Salience} = \frac{S_{\mathbb{T}_1}}{S_{\mathbb{T}_1} + S_{\mathbb{T}_2}}. \quad (1)$$

The initial beat phase is found by cross-correlating the summary detection function (obtained after computing the band-wise sum of the detection functions) with a pulse-train signal of tempo $\mathbb{T}_1$. The same operation is repeated for $\mathbb{T}_2$.

The tempo estimation system was submitted as a Matlab (version 6.5, release 13) p-coded file. The processing time under Debian/Linux on a Pentium IV running at 1.6 GHz with 512MB of RAM is approximately 70% of the excerpt's length in seconds.

Contest results. The score obtained by the afore described algorithm during MIREX'05 contest is presented in Table 1. Dataset was composed of 140 files of 30 seconds length each.

ACKNOWLEDGEMENTS

This work was partly supported by the Mexican Council on Science and Technology grant No. 129114 and the French Ministry of Education and Research under the ACI-Music Discover project.

References

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

Miguel Alonso, Gaël Richard, and Bertrand 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.

Eric D Scheirer. Tempo and beat analysis of acoustic music signals. *J. Acoust. Soc. Am.*, 103(1), 1998.