ILSP TEMPLATE-BASED AUDIO TEMPO ESTIMATION ALGORITHM FOR MIREX 2011

Aggelos Gkiokas[1,2], Vassilis Katsouros[1] and George Carayannis[2]

{agkiokas, vsk}@ilsp.gr, gcara@ilsp.athena-innovation.gr

[1] Institute for Language and Speech Processing / "R.C Athena"
[2] National Technical University of Athens

ABSTRACT

This paper describes a tempo estimation system submitted to the MIREX.2011 Two main feature classes are extracted by utilizing percussive/harmonic separation of the audio signal, in order to extract filterbank energies and chroma features from the respective components. Periodicity analysis is carried out by the convolution of feature sequences with a bank of resonators. Target tempo is estimated by comparing the resulting periodicity vector to template vectors.

1. FEATURE EXTRACTION

1.1 Pre-analysis

The constant Q transform (CQT) of the audio signal is calculated on the whole input signal, using 12 bins per octave, with 25Hz and 5kHz minimum/maximum frequencies respectively (Q value equals to 17), and a Hanning window with half overlap. Frequency bins are aligned to the western scale musical pitches. The frequency bins are rescaled by bicubic interpolation/decimation to have equal frames per time unit (200 frames/s), resulting the log-frequency spectrogram $\mathbf{S} = \{S_{i,f}\}$ where i and f denote the time and frequency bin indices respectively.

1.2 Chroma and Filterbank Energies

The percussive/harmonic separation algorithm presented in [1] is applied to the CQT of the signal. From the harmonic/percussive part the chroma vectors and the energies of 8 triangular filters in the mel scale are calculated respectively.

2. PERIODICITY ANALYSIS

Feature vectors are differentiated and convolved with a bank of resonators as in [2] in the range of $[30, 500]$ bmp, resulting $\mathbf{TG}^{fl}$ and $\mathbf{TG}^{ch}$ periodicity vectors for filterbank energies and chroma features respectively. To estimate the global periodicity vector $\mathbf{T}_{gl}$ for the whole excerpt $\mathbf{TG}^{fl}$ and $\mathbf{TG}^{ch}$ are summed across all segments and then multiplied:

$$T_{gl}(t) = (\sum_{s} TG^{fl}(t,s))(\sum_{s} TG^{ch}(t,s)) \qquad (1)$$

3. TEMPO ESTIMATION

Tempo is estimated by comparing periodicity vector $\mathbf{T}_{gl}$ with template periodicity vectors each one annotated with a tempo value. Template vectors where extracted from Ballroom and Songs datasets of the ISMIR 2004 Tempo Induction Contest. For each excerpt of the collection annotated with tempo T the corresponding periodicity vector $\mathbf{T}_n$ is calculated. Afterwards vector $\mathbf{T}_n$ is rescaled in the range $[0.8,1.2]$:

$$\mathbf{T}_n^r(s) = \mathbf{T}_n(rs) \qquad (2)$$

and the corresponding tempo for template $\mathbf{T}_n^r$ is set to T / r.

For each target excerpt the cosine similarity of the corresponding periodicity vector $\mathbf{T}_{gl}$ and all template vectors $\mathbf{T}_k$ is calculated. The tempi of the 50 most similar template vectors are clustered to two classes. The centers of the classes are the resulting tempi.

4. REFERENCES

[1] FitzGerald D. "Harmonic/Percussive Separation Using Median Filtering", *Proceedings of the 13th International Conference on Digital Audio Effects,* Graz, Austria, 2010.

[2] Gkiokas A., Katsouros V. and Carayiannis G., "Tempo Induction Using Filterbank Analysis and Tonal Features", *Proceedings of the 11th International Conference on Music Information Retrieval,* Utrecht, Netherlands, August 2010.