ILSP AUDIO MUSIC SIMILARITY 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 an audio music similarity algorithm submitted to the MIREX 2011. Similarity measure is based on scaled versions of the periodicity vectors extracted from the audio signal and can be considered as rhythmic similarity measure.

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. Chroma vectors and the energies of 8 triangular filters in the mel scale are calculated from the harmonic/percussive part of the signal respectively.

2. PERIODICITY ANALYSIS

Feature vectors are differentiated and convolved with a bank of resonators as in [2] in the range of $[40, 250]$ 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. SIMILARITY MEASURE

The similarity measure of an audio excerpt with periodicity vector $\mathbf{T}_1$ with an audio excerpt with periodicity vector $\mathbf{T}_2$ is calculated as:

$$sim(T_1(s), T_2(s)) = \max_{0.8 \le r \le 1.2} (\cos(T_1(s), T_2(rs)) \quad (2)$$

i.e. $\mathbf{T}_2$ is scaled in the range $[0.8,1.2]$. Thus, music pieces with similar rhythm but different tempo will exhibit a high similarity measure.

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.