MIREX 2013 - MUSIC STRUCTURAL SEGMENTATION TASK: IRCAMSTRUCTURE SUBMISSION

Florian Kaiser
STMS IRCAM-CNRS-UPMC
1 Place Igor Stravinski
75004 Paris
France
kaiser@ircam.fr

Geoffroy Peeters
STMS IRCAM-CNRS-UPMC
1 Place Igor Stravinski
75004 Paris
France
peeters@ircam.fr

ABSTRACT

This extended abstract introduces the music structure algorithm submitted by Ircam to the MIREX 2013 structural segmentation task. The algorithm estimates structural information that relate to both timbral and harmonic context variations. Homogeneous sections within this description of the audio signal are then estimated by means of the Non-negative Matrix Factorization of the corresponding similarity matrix and a structural segmentation is produced.

1. SIGNAL DESCRIPTION

Low level audio features are first extracted on the audio signal. Timbral information is described by means of the following descriptors of the spectral content: MFCCs, Spectral Centroid, Spread and Skewness, and Spectral Flatness. These features are normalized and embedded in a similarity matrix computing the features pairwise distance by means of the cosine distance. A second similarity matrix that relate to the harmonic content is then computed. Therefore chroma feature frames are extracted over the whole audio signal and split into subsequences of frames of a couple of seconds, usually between 4 and 10s. Each subsequence is then modeled by means of a Multi-Probe Histogram (MPH) that probes dominant pitch classes transitions between adjacent feature frames [5]. Such a temporal modeling of the chroma vectors allows to describe the evolution of the tonal context in the audio piece that is highly relevant for the structural segmentation task. The length of the chroma subsequences for the Multi-probe Histograms computation is automatically adapted over the audio signal [2]. Therefore, time instants of potential strong harmonic changes are estimated in order to reduce the length of the subsequences when needed. The method thus allows to model musical patterns of variable lengths within a same music piece. Histograms are then embedded in a similarity matrix and the timbre-related and harmony-related similarity matrices are merged together.

In parallel, a chroma based recurrence plot as in [6] is computed.

2. STRUCTURAL SEGMENTATION

Temporal segmentation fuses two segmentation approaches. First, boundaries are detected with a kernel-based approach as in [1] but extended to further boundary transitions types [3]. This is applied on the timbre-related similarity matrix. Circular time lag matrix segmentation as in [6] is computed separately. For both segmentations, acoustic distance between segments is computed and embedded in a segments distance matrix. The sum of these two matrices then serves as the final representation and temporal segmentation is obtained applying the novelty kernel approach [1] to it.

Temporal segments are then merged together according to the musical structure with the algorithm proposed in [4]. The method is based on the observation that information in similarity matrices is highly redundant over time in the hypothesis of high homogeneity in the acoustical content of structural sections. An ideal musical structure of homogeneous structural entities is indeed represented in the similarity matrix as a sequence of uniform blocks. The whole structural information is then contained in a few raws or columns, i.e. since similarity matrices are symmetric. Intuitively, the similarity matrix is thus ideally spanned by a much lower dimensional basis, with a dimensionality that relates to the number of states in the musical structure. The task of music structural segmentation then becomes a dimension reduction of the similarity matrix problem that we perform by means of its Non-negative Matrix Factorization (NMF). Structural segmentation is obtained by applying hierarchical clustering on the temporal segments projected on this new basis. This NMF based clustering is applied the merged timbre- and harmony-related similarity matrix. The number of segments to form is estimated in a similar manner as in [7].

3. ACKNOWLEDGMENTS

This work was partly supported by the Quaero Program funded by Oseo French agency.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
© 2013 International Society for Music Information Retrieval.

4. REFERENCES

[1] Jonathan Foote. Automatic audio segmentation using a measure of audio novelty. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2000.

[2] Florian Kaiser and Geoffroy Peeters. Adaptive temporal modeling of audio features in the context of music structure segmentation. In *Proceedings of the 10th international workshop on Adaptive Multimedia Retrieval*, Copenhagen, Denmark, October 2012.

[3] Florian Kaiser and Geoffroy Peeters. Multiple hypotheses at multiple scales for audio novelty computation in music. In *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, 2013.

[4] Florian Kaiser and Thomas Sikora. Music structure discovery in popular music using non-negative matrix factorization. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, aug 2010.

[5] Florian Kaiser and Thomas Sikora. Multi-probe histograms: A mid-level harmonic feature for music structure segmentation. In *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx)*, Paris, France, September 2011.

[6] Joan Serra, Meinard Müller, Peter Grosche, and Josep Ll. Arcos. Unsupervised detection of music boundaries by time series structure features. In *AAAI International Conference on Artificial Intelligence*, 2012.

[7] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistics. *Journal of the Royal Statistical Society, series B*, 63:411–423, 2001.