THE IMPORTANCE OF DETECTING BOUNDARIES IN MUSIC STRUCTURE ANNOTATION

Joan Serrà[(a)], Meinard Müller[(b,c)], Peter Grosche[(c,d)], and Josep Ll. Arcos[(a)]

[(a)] Artificial Intelligence Research Institute, Spanish National Research Council (IIIA-CSIC), Bellaterra, Barcelona, Spain.

[(b)] Bonn University, Bonn, Germany.

[(c)] Max Planck Institute for Computer Science, Saarbrücken, Germany.

[(d)] Saarland University, Saarbrücken, Germany.

{jserra,arcos}@iiia.csic.es, {meinard,pgrosche}@mpi-inf.mpg.de

EXTENDED ABSTRACT

This document summarizes our submission to the MIREX 2012 [1] Structural Segmentation task [2] and provides pointers to the relevant papers explaining the details of such submission. The method we follow is rooted on first detecting segment boundaries [3] and matching the found segments by state-of-the-art music alignment/similarity algorithms afterwards [4]. In fact, the rationale is that if structure or segment boundaries are sufficiently accurately placed, then the task of structure annotation can be easily solved by employing standard tools for music matching (e.g. [1, 5]). Hence the title of this extended abstract.

Our method starts with the extraction of pitch class profile (PCP) features (also called chromas, cf. [1]). Contiguous PCPs are then joined to form delay coordinate vectors, and a recurrence plot (RP) is constructed (an RP can be thought of as a thresholded self-similarity matrix, cf. [2]). Next, structure features (SF) are obtained by transforming the RP into a time-lag matrix and convolving this with a two-dimensional Gaussian kernel [3]. The differences between consecutive SFs yields a novelty curve whose peaks generally correspond to structural boundaries. Once structural segment boundaries are delimited, we cross compare all possible segments using the previously computed RP, similarly to what is done in music version identification [5] or in audio matching [1]. This process yields all pairwise similarities between segments. Finally, we apply a recursive matrix multiplication process to decide if these similarities are large enough to classify two segments as being the same or not [4].

For a more complete explanation of our method we refer to [3] and [4]. In particular, we strongly recommend ref. [4], as it explains the entire method in detail.

[1] http://www.music-ir.org/mirex/wiki/2012:MIREX_Home

[2] http://www.music-ir.org/mirex/wiki/2012:Structural_Segmentation

ACKNOWLEDGMENTS

JS and JLA acknowledge 2009-SGR-1434 from Generalitat de Catalunya, TIN2009-13692-C03-01 from the Spanish Government, and EU Feder funds. JS also acknowledges JAEDOC069/2010 from Consejo Superior de Investigaciones Científicas. PG and MM have been supported by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University and the German Research Foundation (DFG MU 2682/5-1).

REFERENCES

[1] M. Müller. *Information retrieval for music and motion*. Springer, Berlin, Germany, 2007.

[2] J. Paulus, M. Müller, and A. Klapuri. Audio-based music structure analysis. In *Proc. of the Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 625–636, 2010.

[3] J. Serrà, M. Müller, P. Grosche, and J. Ll. Arcos. Unsupervised detection of music boundaries by time series structure features. In *Proc. of the AAAI Int. Conf. on Artificial Intelligence*, pages 1613–1619, 2012.

[4] J. Serrà, M. Müller, P. Grosche, and J. Ll. Arcos. Unsupervised music structure annotation by time series structure features and segment similarity. *Journal publication*, Submitted, 2012.

[5] J. Serrà, X. Serra, and R. G. Andrzejak. Cross recurrence quantification for cover song identification. *New Journal of Physics*, 11(9):093017, 2009.