MIREX 2009 "MUSIC STRUCTURE SEGMENTATION" TASK: IRCAMSUMMARY SUBMISSION

Geoffroy Peeters
Ircam Sound Analysis-Synthesis Team - CNRS STMS

ABSTRACT

This extended abstract details a submission to the Music Information Retrieval Evaluation eXchange (MIREX) 2009 for the "Structure Segmentation" task. The system named ircamsummary performs both structure estimation and summary generation. Detailed description of the system can be found in [1] [2] and [3]. We briefly summarized them below.

1. IRCAMSUMMARY IMPLEMENTATION

Ircamsummary is both a Matlab-compiled software, a pcode function or a C++ software and library running under Linux, Windows-XP and Mac-OS-X. It performs music structure estimation using the "state" or the "sequence" approach and music audio summary generation using various strategies. Ircamsummary uses the library ircambeat for beat/downbeat synchronous analysis.

2. IRCAMSUMMARY ALGORITHM DESCRIPTION

The flowchart of ircamsummary is represented in Figure 1.

Ircamsummary is a software allowing structure estimation both as "state" representation and "sequence" representation. It also allows the generation of music audio summary using various strategies.

Four independent sets of features are extracted from the signal: MFCC, Chroma features, Spectral Flatness/ Crest measures, Spectral Valley/ Contrast factors. They are processed separately in the remaining.

Various temporal modelling can then be applied to the instantaneous features. In the MIREX09 submission, a simple beat-synchronous mean-value modelling has been used. Other temporal modelling such as Dynamic features modelling or membership to hidden states are possible.

Four distance matrices are computed from the four feature sets. After processing (stretching) the four distance matrices are combined together using a weighted sum.

In the case of "state" representation (MIREX09 submission), the resulting distance matrix is used to perform a rough segmentation of the signal.
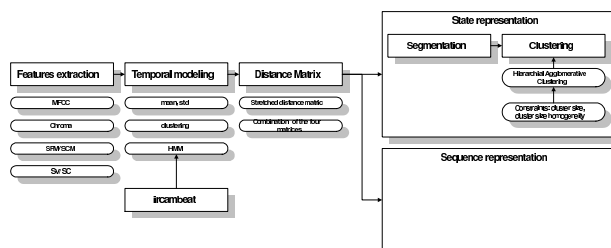
Figure 1. Flowchart of ircamsummary

The resulting segments are then progressively merged using a hierarchical agglomerative clustering. A specific distance is used in order to consider both segments similarity and dissimilarity while enhancing the diagonal ("sequence" approach). Two constraints are added for the linkage: a) cluster size and b) cluster size homogeneity.

3. MIREX09 RESULTS AND DISCUSSIONS

The MIREX task on "Music Structure Segmentation" was created in 2007 during the Euro MIREX in Vienna, Austria (see http:// www.music-ir.org/ mirex2007/ index.php/ 2007_Euro_MIREX_Planning_Meeting for details). Because this task is difficult to evaluate (in terms of definition of what is music structure, in terms of definition of performance measures and in terms of test-set availibility), it couldn't take place in MIREX 2007 and 2008. Therefore MIREX 2009 is the first edition of this task. It should therefore be considered as a first attempt of the task definition.

3.1 Experiment

The test-set was provided by Matthias Mauch from Queen Mary University of London. It consists of 297 popular music songs annotated into structure. Few information about the test-set were provided to the participants in terms of track content or in terms of type of annotations. Also no training or development set has been provided. We just know that part of the test-set consists of tracks from Carole King, Michael Jackson, Queen, The Beatles and Zweieck; that the annotations were made by students supervised by Matthias Mauch; that the annotations were rounded to the nearest bar and that they include labels such as silence, intro, outro, chorus, verse and half-verse.

Eleven performance measures have been used which can be grouped into five main categories:

overSegScore and underSegScore are the normalized conditional entropy proposed by [4]

pwF, pwPrecision, pwRecall are the pair-wise F-measure, precision and recall [5]

R is the rand clustering index

Fmeasure@0.5,3s, precRate@0.5,3s and recRate@0.5,3s are the F-measure, precision and recall of the segmentation obtained through the structure. Precision Window of 0.5s and 3s are used. Note that these measures do not depend on the labelling of the segments.

medianTrue2claim (medianClaim2true) is the median distance from an annotated (estimated) segment boundary to the closest found (annotated) boundary expressed in seconds. Note that these measures do not depend on the labelling of the segments.

Details about the MIREX-09 "Music Structure Segmentation" task can be found at http:// www.music-ir.org/ mirex/ 2009/ index.php/ Structural_Segmentation

3.2 Results

Results are indicated into Table 1 for the various participants to the task and the various performance measures.

Because the results obtained strongly depends on the way annotations is performed (we discuss deeply the subject of music structure annotation in [6] and [7]), and because this was not clearly defined for the first edition of the MIREX 2009 task, it should be noted that no conclusions can be derived from the results presented in the following.

The only conclusion we can draw is that - for this specific test-set and this specific annotations -

- using our internal performance measure, which is the summation of overSegScore (measuring song over-segmentation) and underSegScore (measuring song under- segmentation), ircamsummary (GP) would rank third: PK (Paulus): 1.36, MND (Mauch which provided the test-set): 1.33, GP: 1.27, ANO1 (anonymous): 1.24, ANO2 (anonymous): 1.19.

- using the R measure, ircamsummary (GP) would also rank third: PK: 0.79, ANO1: 0.76, GP: 0.75, MND: 0.74, ANO2: 0.72.

The second conclusion we can draw is that - for this specific test-set and this specific annotations -, ircamsummary over-segmented the track. For this reason

- underSegScore is higher than overSegScore,

- pwPrecision is higher than pwRecall (the denominator of pwPrecision being smaller in case of over-segmentation),

- recRate@[0.5, 3] is higher than precRate@[0.5, 3] (over- segmentation produces insertion),

- medianTrue2claim is smaller than medianClaim2true (because of over- segmentation and therefore insertion, the median distance between a target annotation and an estimated segment boundary is smaller than the opposite).

*EvalMeasure	overSegScore	underSegScore	pwF	pwPrecision	pwRecall	R	Fmeasure@0.5s	precRate@0.5s	recRate@0.5s	Fmeasure@3s	precRate@3s	recRate@3s	medianTrue2claim	medianClaim2true
ANO1	0,6	0,6	0,6	0,6	0,6	0,8	0,2	0,2	0,2	**0,6**	0,5	**0,7**	**1,6**	3,5
ANO2	0,6	0,6	0,6	0,5	0,7	0,7	0,1	0,1	0,1	0,6	**0,6**	0,6	2,5	**3,2**
GP	0,6	0,7	0,5	0,6	0,5	0,8	0,2	0,1	0,3	0,5	0,4	0,7	2,1	3,5
MND	**0,7**	0,6	**0,6**	0,6	**0,7**	0,7	0,2	0,2	**0,4**	0,4	0,3	0,7	2,2	3,4
PK	0,6	**0,8**	0,5	**0,7**	0,5	**0,8**	**0,3**	**0,2**	0,3	0,5	0,5	0,6	2,4	3,5

Table 1. MIREX-09 "Music Structure Segmentation" results

Details about the results of the tasks can be found at http:// www.music-ir.org/ mirex/ 2009/ index.php/ Music _Structure_Segmentation_Results

4. CONCLUSION

This extended abstract reviewed ircamsummary submission for the "Music Structure Segmentation" task of MIREX-09.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] G. Peeters, A. Laburthe, and X. Rodet. Toward automatic music audio summary generation from signal analysis. In *Proc. of ISMIR*, pages 94–100, Paris, France, 2002.

[2] G. Peeters. Deriving musical structures from signal analysis for music audio summary generation: Sequence and state approach. In U.K. Wiil, editor, *CMMR 2003 (LNCS 2771)*, Lecture Notes in Computer Science, pages 142–165. Springer-Verlag Berlin Heidelberg 2004, 2004.

[3] G. Peeters. Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach. In *Proc. of ISMIR*, Vienna, Austria, 2007.

[4] H. Lukashevich. Toward quantitative measures of evaluating song segmentation. In *Proc. of ISMIR*, Philadeplhia, USA, 2008.

[5] M. Levy and M. Sandler. Structural segmentation of musical audio by constrained clustering. *IEEE transactions on audio, speech, and language processing*, 16(2):318–326, 2008.

[6] G. Peeters and E. Deruty. Toward music structure annotation. In *Proc. of ISMIR (Late-Breaking news category)*, Kobe, Japan, 2009.

[7] G. Peeters and E. Deruty. Is music structure annotation multi-dimensional ? a proposal for robust local music annotation. In *Proc. of Learning the Semantics of Audio Signals (LSAS)*, Graz, Austria, 2009.