STRUCTURAL ANALYSIS OF HARMONIC FEATURES USING STRING MATCHING TECHNIQUES

Benjamin Martin, Pierre Hanna, Matthias Robine and Pascal Ferraro
LaBRI - Université de Bordeaux
benjamin.martin@labri.fr

ABSTRACT

We focus on the harmonic content of music pieces, considering Harmonic Pitch Class Profiles as audio features to segment audio pieces. The features are compared using Optimal Transposition Index and correlation adapted measures, ensuring robustness to timbre variations or detuning issues. No other audio feature is used for describing audio pieces. The identification of approximate repetitions is handled by string matching techniques applied to musical content. We define a new iterative inference algorithm detecting major repetitions, based on several existing string alignment algorithms. The techniques employed ensure robustness to temporal changes or local transpositions between structural parts.

1. AUDIO FEATURES

We chose Harmonic Pitch Class Profile (HPCP) as audio harmonic features [2], for these provide an accurate and robust pitch and harmonic analysis. The audio piece is analysed frame-by-frame, and a 36-dimensional HPCP feature is computed for each frame. Frames have a constant length, and none synchronization with beat nor any other feature is performed: the method works on a single type of audio features. To compare HPCP vectors one another, we consider the method of the Optimal Transposition Index (OTI) computation [4]. Comparing two feature vectors $\overrightarrow{h_1}$ and $\overrightarrow{h_2}$ consists in transposing (shifting) $\overrightarrow{h_2}$ for every possible transposition value and correlating it with $\overrightarrow{h_1}$. The OTI corresponds to the transposition index applied for the best correlation score.

2. STRING MATCHING TECHNIQUES

We use an extension of the local alignment algorithm [5] by Allali *et al.* [1] that allows detecting approximately similar parts even in presence of local transpositions. The algorithm computes k dynamic programming matrix M such that $M^k[i][j]$ contains the alignment scores between the substrings $u[1 \cdots i]$ and $v[1 \cdots j]$ transposed k bins up. The maximum value in the dynamic programming matrices gives the similarity score between the aligned substrings. The ending indexes of these substrings in their respective strings are defined by the row and column of this score in the matrix, and their beginning indexes are obtained by traceback in the matrix [3].

3. HIERARCHICAL INFERENCE

We consider that music audio is a highly structured information, organized as a hierarchical organization of approximate repetitions. The algorithm described in 1 tries to infer the repetitions tree structure of the piece by descending in the tree levels, from the root to the leaves. Constraints are required over the minimum length of structural segments. At each step, the most salient repetitive sections are assigned a label. A section that was already assigned a label can not be re-segmented unless a high similarity is found between a subsection of it and a non-labelled section. A segment is defined only if at least one approximate repetition of such a segment is found in the piece, except in the final step of the algorithm where all non-segmented sections are assigned a different label. The transitivity phase ensures that at the end of each iteration of the inference algorithm, the resulting segmentation lies on a single level of the hierarchical tree.

4. REFERENCES

[1] Julien Allali, Pascal Ferraro, Pierre Hanna, and Costas Iliopoulos. Local transpositions in alignment of polyphonic musical sequences. In Nivio Ziviani and Ricardo Baeza-Yates, editors, *String Processing and Information Retrieval*, volume 4726 of *Lecture Notes in Computer Science*, pages 26–38. Springer Berlin / Heidelberg, 2007.

[2] E. Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, 2006.

[3] Dan Gusfield. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, New York, NY, USA, 1997.

[4] J. Serrà, E. Gómez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech and Language Processing*, 16:1138–1151, 08/2008 2008.

[5] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1):195–197, March 1981.

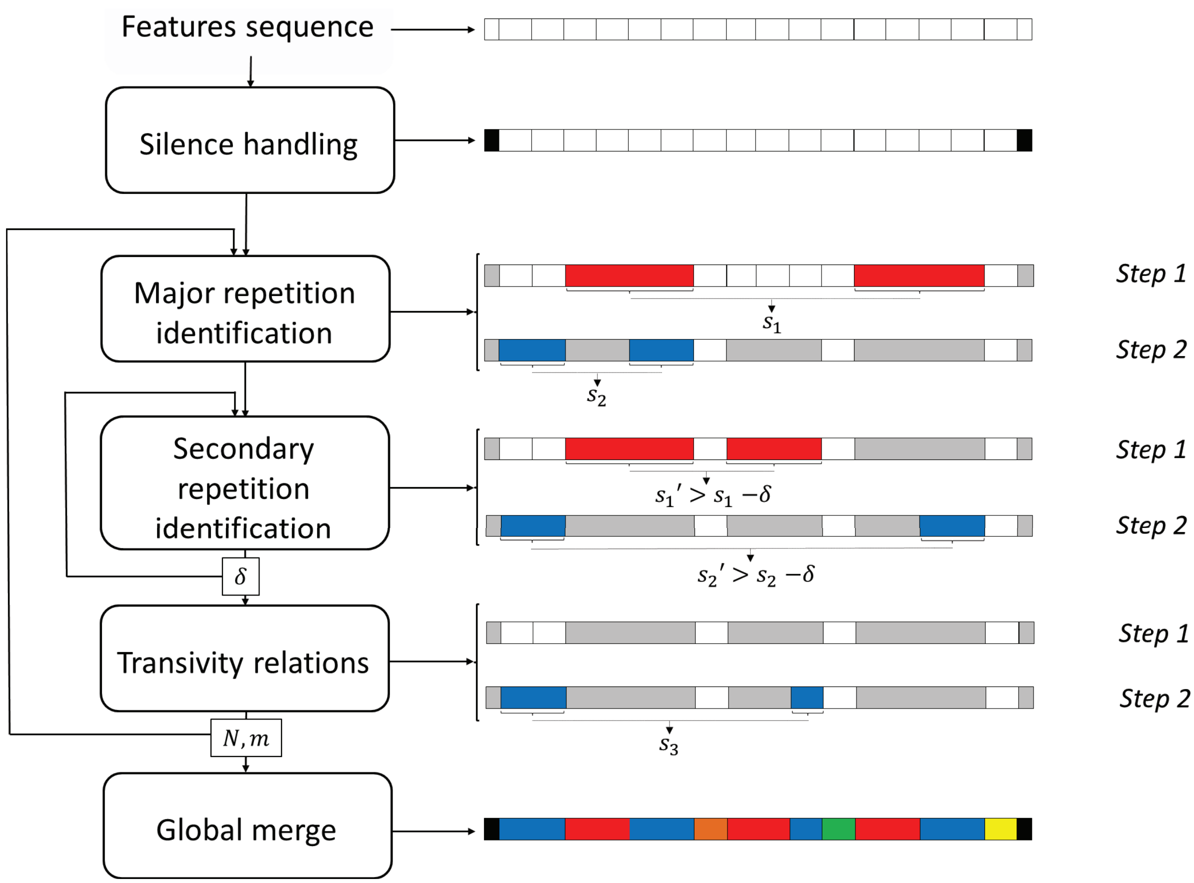

Figure 1. Overview of the system