

VISUALISATION OF MUSICAL STRUCTURE BY APPLYING IMPROVED EDITING ALGORITHMS

Pierre Hanna, Matthias Robine and Pascal Ferraro

SIMBALS Project

LaBRI - University of Bordeaux 1

F-33405 Talence cedex, France

firstname.name@labri.fr

ABSTRACT

This paper presents a method for visualising the musical structure of symbolic monophonic music. This method is based on the computation of a similarity matrix that exhibits repetitions. Some recent techniques are very accurate when the different parts of the structure are repeated with few variations. If it is generally the case in the context of the Western popular music, musically similar segments may also exhibit significant variations of tonality (global or local transpositions), instrumentation or tempo for example. It is particularly true with Western classical music. We consider symbolic representation of music to be able to take into account more musical information. Editing algorithms compute a similarity score between symbolic musical sequences. Several improvements have been proposed during the last few years. We propose to directly apply these algorithms. The method induced computes a similarity score between overlapping segments. It enables the visualisation of the musical structure, even in the cases of local variations. Experiments on musical pieces illustrate the improvements induced by this new approach.

1. INTRODUCTION

Automatic analysis of the structure of musical pieces is one major open problem of the Music Information Retrieval research area. This analysis may be very useful for very different applications such as music comparison, classification or visualisation. Moreover, once the general structure is known, it is then easier to focus on each part to detect finer variations (local variations), or the repetitions of short patterns such as sequences of chords. Visualising the structure of a musical piece may also help the listener to understand the artistic intentions of the composer. Repetition is an important tool for composing music (hopefully often with variation and contrast), it is thus one of the main properties of music. At different levels, music can be analyzed as a self-similar structure. A general structure (for example the ABA form of a minuet) is often composed of a few parts (exposition of a theme, development, etc), which may be also composed of repetitions of patterns. The automatic extraction of these structures at different levels is a very difficult problem. It generally relies on the estimation of similarity between the different parts of a given musical piece. In this paper, we propose to

adapt existing methods for the estimation of the melodic similarity to compute similarity matrices that exhibit repetitions into a musical piece. In Section 2, existing systems for the computation of similarity matrices are presented and the originality of our approach is detailed. Then, the algorithm proposed for the computation of similarity is described in Section 3. A few experiments with music are presented in Section 4, before discussing the limitations and the future work in the last Section.

2. SIMILARITY MATRIX

One of the main methods for the visualisation of the musical structure of audio music relies on the computation of a matrix [6] denoted $M(i, j)$, which indicates the similarity between frames i and j of the same musical piece. Frames represent a small time interval of music. The similarity is generally indicated by a score s : the higher the score $s(i, j)$, the more similar the frames i and j . An image is then computed according to this matrix by defining each pixel color $p(i, j)$ to the corresponding similarity score $s(i, j)$. Since the similarity $s(i, i)$ is maximum, the diagonal $p(i, i)$ of such image is easily identifiable. In this paper, we propose to associate dark colors to high similarity scores (the opposite choice is made in [6] for example). That's why the main diagonal will appear black on a mainly white background. Such images clearly allow the visualisation of the structure of musical pieces. Dark zones indicate similar regions. In particular, dark diagonal lines (different from the main diagonal) will exhibit repeated parts such as theme, chorus, verse, etc. Figure 1 shows an example of such a self-similarity matrix. The musical structure is manually annotated and corresponds to dark diagonals.

Existing approaches essentially consider audio signals. Features are extracted directly from audio samples. The first main features applied were the mel frequency cepstral coefficients (MFCC) [6]. These coefficients are estimated from the smoothed spectrum of the audio signal and primarily describe the timbral properties of the sound. Considering only these coefficients implies limitations of structural analysis systems. Indeed, many popular music pieces are composed of one repeated chorus, played successively with different instruments : for example, the first chorus can be played by a singer with a single piano whereas the second chorus can be repeated with many

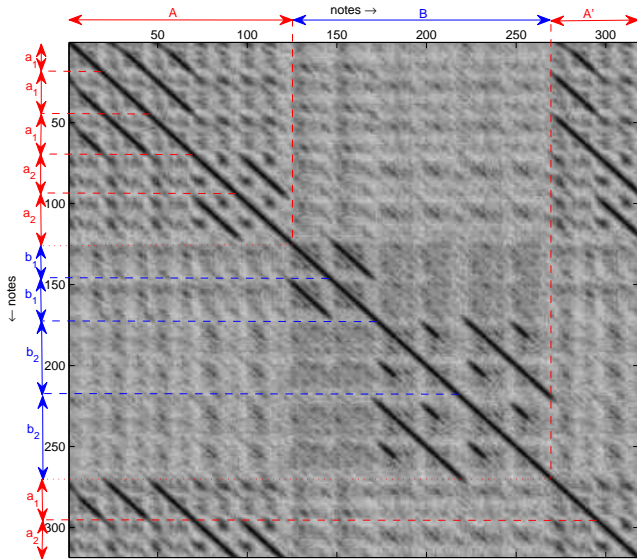


Figure 1. Visualisation of musical structure of the *minuet* part of the *Water Music Suite No.1 in F* (Haendel) by a self-similarity matrix computed with the method described in this paper. The structure has been manually annotated.

other instruments (drums, guitar, etc.). However these different parts cannot be analyzed as repetitions of the same part. This important limitation leads researchers to consider other features describing in a better way tonal information, since the harmonic information is one of the main information (tonality, chords, etc) that induces the structure of music. In [2], chroma-based representations (chromagram or pitch class) are applied for encoding musical structure. Experiments showed that musical structure analysis is improved by using such representations. Other researches investigate the possibility to take into account local musical variations such as tempo deviations or key transpositions [11].

Following this way, we propose in this paper to investigate methods for musical structure analysis that enable the detection of repetitions even with local musical variations. Some recent techniques are very accurate when the different parts of the structure are repeated with few variations. If it is generally the case in the context of the Western popular music, musically similar segments may also exhibit significant variations of tonality (global or local transpositions) or tempo for example [10]. We aim at developing a system that takes into account such variations. This investigation requires to develop systems that consider elements of musical theory, in particular information associated to harmonic content. Such systems can become very complex, as it will be explained in the next Section 3. Therefore we restrain for now our study to symbolic monophonic music. In monophonic music, no more than one note is sounded at any given time. This restriction is discussed in Section 5. The originality of the method presented in this paper relies on the consideration of symbolic melodic similarity whereas the existing methods generally consider audio features or representations that regroup several properties.

3. SYMBOLIC MELODIC SELF-SIMILARITY

Representing music with symbols is the basis of the score notation. Musical structure analysis is generally performed by music experts by considering such notation. No timbral information is thus taken into account. Then considering symbolic representation of musical pieces may help to develop efficient systems that automatically analyze musical structure. Symbolic representations regroup harmonic properties under the form of note sequences. Each note is defined by a few properties such as pitch, duration or dynamic. We propose here to study the improvements enabled by considering such symbolic representation of music.

Previous experiments of structural analysis from symbolic music have been proposed in the last few years. In [5], a self-similarity image of a Bach's piece is presented. This image has been generated from a symbolic format (MIDI file) by applying the simple following algorithm: matrix value $M(i, j)$ is colored white if the two notes i and j have the same pitch (they *match*), it is colored black otherwise. This approach is limited in the case of variations, and also because only one note at the time is considered. However, note intervals are very important for the musical analysis. We thus propose to improve such method by considering intervals instead of absolute pitch, by considering also relative note duration and by taking into account weighted matches according to the consonance.

In the approach introduced in [5], only two cases are considered for note-to-note comparison: two notes are the same or not. This induces black or white pixels. However, the relations between two notes are more complex. For instance, replacing a C note in a pattern with a G note (fifth) may slightly modify the pattern in comparison with substituting with a $C\#$ note. The relation between two notes may be correlated to the consonance interval. Therefore, we propose to assign to each pair of notes different similarity scores. This choice leads to the computation of gray images instead of black and white. For example, the fifth (7 semitones) and the third major or minor (3 or 4 semitones) are the most consonant intervals in Western music [8] and are associated with a high score, which will imply dark gray colors. Note that determining this score according to the consonance interval is totally different than determining it according to the pitch difference (in semitones for example).

Another improvement concerns local variations due to modulations, key changing or transpositions. Modulations frequently occur in classical music. In popular music, several pieces are composed of repetitions of choruses or verses that are transposed. It is important to be able to detect repetitions, even in these cases. We propose in this purpose to apply another well-known representation for the pitch. On the contrary of the *absolute* pitch representations, the *interval* representation is transposition invariant, since it is simply the number of semitones between two successive notes [13]. Using this representation, two patterns will be represented with the same sequence of symbols, even if one of this pattern is transposed.

Previous approaches do not take into account note du-

rations. When considering symbolic music, it is yet important to consider this information and to combine it with the pitch information [4]. That's why we propose to calculate a similarity score between two notes according these two factors. The process of computation of this score relies on the one presented in [9]. Furthermore, patterns may be repeated at different tempi. Using the same approach as the one described previously for the pitch and the transposition invariance, we propose then to represent durations with relative values (ratio between two successive note durations).

All these improvements have been implemented and experimented but limitations still appeared when trying to visualise structures of a musical piece. Indeed, by considering only one note-to-note comparison (more precisely two notes for relative/interval representations) for each pixel, the information is very localized whereas the analysis of high level musical structures requires more global information, and thus more successive notes. That's why we propose to compare segments of notes to other segments of notes. The length l of the segment is fixed and the segments overlap. The matrix value $M(i, j)$ represents the similarity between the segment $(i, i + 1, \dots, i + l)$ and the segment $(j, j + 1, \dots, j + l)$. All the properties of the similarity matrix described in Section 2 are preserved. Depending on the length l of the segments, the similarity matrix may exhibit low-level (with low l) or high-level (with high l) musical structure. Considering segments of notes have already been proposed in [3] but the authors conclude that important limitations still remain, in particular in the cases of local transpositions or tempo variations. The question of the similarity estimation between segments of notes is raised.

Several techniques for evaluating symbolic music similarity have been introduced during the last few years: geometric algorithms or algorithms adapted from string matching domain. Editing algorithms compute a similarity measure between two strings of symbols as the minimum score sequence of elementary operations needed to transform one of the strings into the other. This similarity measure makes use of the dynamic programming principle to achieve an algorithm with quadratic complexity. Similarity scores can be normalized and are values in the range $[0; 1]$ (1 corresponds to a perfect similarity whereas low values indicate low similarity). This approach have been applied in the monophonic musical context by Mongeau and Sankoff [9] and a few improvements have been recently proposed [4]. Edit-based systems are very flexible: they permit to take into account musical properties by adapting the computation of the scores of edit operations.

Some recently proposed improvements are very interesting in the context of the visualisation of the musical structure. In [1] a new dynamic programming algorithm is proposed that take into account multiple local transpositions. The representation of pitch is absolute and the algorithm estimates the optimal transposition. It is important to note that several transpositions are allowed inside the musical pieces being compared. Applying this algorithm allows the detection of local transpositions inside each segment. Other improvements concern local vari-

ations such as the presence of passing notes [12]. Taking into account such musical elements allows the algorithm to detect high similarity between segments that are composed of some different notes but which are musically similar.

4. EXPERIMENTS

Experiments have been performed with monophonic MIDI files. The track containing the theme has been extracted, or a track regrouping the main melody from different instrument has been created. Considering these manually sequenced MIDI files avoid errors that may be induced by an automatic transcription process. Images are computed from similarity matrices such that the maximum similarity is indicated by black pixels. The length of the segments has been fixed to 6 notes in order to be able to visualise short repetitions. Experiments show that this size also allows the visualisation of the global musical structure.

The first experiments we performed concern the *minuet* part of the *Water Music Suite No.1 in F* (HWV 348) by Haendel. Figure 1 shows the similarity matrix computed with the improved editing algorithm explained in this paper. Diagonals indicate that parts of the musical piece on the vertical axis are detected as similar than other parts of the same piece on the horizontal axis. By considering other diagonals than the main diagonal, the general structure *ABA* of a minuet is thus visible, with diagonals concentrated in big different squares. For each part, some patterns are also repeated. For example, the pattern a_1 is repeated 4 times since there are 4 little diagonals on its line (i.e. 3 times in part *A* then 1 in *A'*).

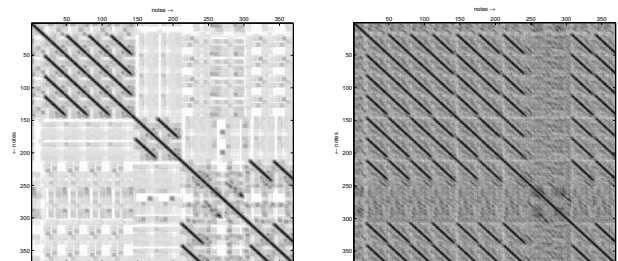


Figure 2. Self-similarity matrix of the song *In the year 2525* computed with usual editing algorithm (left) and with improved algorithm (right): repeated modulated parts becomes visible.

In order to evaluate the accuracy of our method, the visualisation has been performed on musical pieces composed of repeated parts with variations. The song *In the year 2525* by Zager and Evans is composed of a part (chorus) that is repeated 8 times [11]. Two repetitions are modulations by a minor second, two others are modulations by major second. Figure 2 (left part) shows the similarity matrix obtained by our system without considering variations such as modulations. This figure can be compared to the right part of Figure 2 obtained by taking into account local variations. All the repeated parts are visible in this image.

Another example is given by Figure 3 with the similarity matrix of the *final* of the *Firebird Suite* by Stravin-

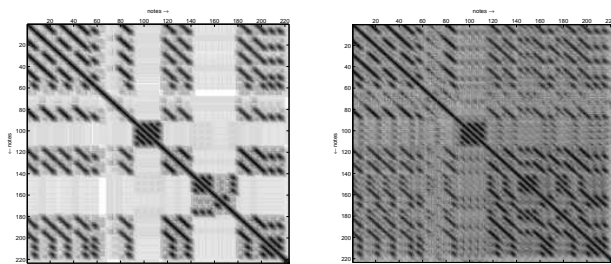


Figure 3. Self-similarity matrix of the piece *Firebird* computed with usual editing algorithm (left) and with improved algorithm (right): repeated modulated patterns become visible.

sky. Here again, several repetitions with local variations become visible with the improved algorithm.

5. CONCLUSION AND PERSPECTIVES

We have proposed to experiment the adaptation of improved editing algorithms to the visualisation of the musical structure of musical pieces. Experiments show the advantages of considering musical information in order to take into account local variations. Two main limitations still remain. The first one is the monophonic assumption. For now, the editing algorithm is still under development for considering polyphonic music. A few studies have already been proposed [7] but improvements are still necessary. When these methods will be more precise, the approach proposed in this paper may directly be applied to polyphonic music. The second limitation is that we only consider symbolic representation of music. However, several algorithms for automatic transcription have been proposed. Experiments have to be performed with such systems in order to test the robustness of the combination of these methods. Otherwise, adaptations of the algorithm proposed with chromagram may be proposed.

An important perspective is the automatic clustering of musical pieces by considering the similarity matrices showed in this paper. We aim at experimenting existing segmentation methods to obtain the general structure of any musical piece (for example *ABA* for the minuet of Figure 1).

We have shown that considering symbolic music may lead to more precise music analysis systems. The approach presented here is limited to the melodic similarity, and may be extended to other musical elements such as rhythm, harmony, etc.

6. ACKNOWLEDGEMENT

This work is part of the SIMBALS project (JC07-188930), funded by the French National Research Agency.

7. REFERENCES

[1] Julien Allali, Pierre Hanna, Pascal Ferraro, and Costas Iliopoulos. Local Transpositions in Alignment of Polyphonic Musical Sequences. In *Proceed-*

ings of the 14th String Processing and Information Retrieval Symposium (SPIRE), October 2007.

- [2] Mark A. Bartsch and Gregory H. Wakefield. To Catch a Chorus: Using Chroma-Based Representations for Audio Thumbnailing. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 15–18, New Paltz, New York, USA, 2001.
- [3] Roger B. Dannenberg and Ning Hu. Pattern Discovery Techniques for Music Audio. In *Proceedings of the third International Conference on Music Information Retrieval (ISMIR)*, Paris, France, 2002.
- [4] Pascal Ferraro and Pierre Hanna. Optimizations of Local Edition for Evaluating Similarity Between Monophonic Musical Sequences. In *Proceedings of the 8th International Conference on Information Retrieval (RIAO)*, Pittsburgh, USA, 2007.
- [5] J. Foote and M. Cooper. Visualizing Musical Structure and Rhythm via Self-Similarity. *Proceedings of International Computer Music Conference (ICMC)*, 2001.
- [6] Jonathan Foote. Visualizing Music and Audio Using Self-Similarity. In *Proceedings of the seventh ACM international conference on Multimedia (ACMMM)*, pages 77–80, Orlando, USA, 1999.
- [7] Pierre Hanna and Pascal Ferraro. Polyphonic Music Retrieval by Local Edition of Quotiented Sequences. In *Proceedings of the 5th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 61–68, Bordeaux, France, 2007.
- [8] Frederick J. Horwood. *The Basis of Music*. Gordon V. Thompson Limited, Toronto, Canada, 1944.
- [9] Marcel Mongeau and David Sankoff. Comparison of Musical Sequences. *Computers and the Humanities*, 24(3):161–175, 1990.
- [10] M. Müller and M. Clausen. Transposition-Invariant Self-Similarity Matrices. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 47–50, Vienna, Austria, September 2007.
- [11] M. Müller and F. Kurth. Towards Structural Analysis of Audio Recordings in the Presence of Musical Variations. *EURASIP Journal on Applied Signal Processing*, 2007(ID 89686):18 pages, January 2007.
- [12] Matthias Robine, Pierre Hanna, and Pascal Ferraro. Music Similarity: Improvements of Edit-based Algorithms by Considering Music Theory. In *Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*, pages 135–141, Augsburg, Germany, 2007.
- [13] Alexandra L. Uitdenbogerd. *Music Information Retrieval Technology*. PhD thesis, RMIT University, Melbourne, Australia, July 2002.