

STRUCTURED REPRESENTATION OF HARMONY FOR MUSIC RETRIEVAL

Matthias Robine, Pierre Hanna, Thomas Rocher

LaBRI - Université de Bordeaux
F-33405 Talence cedex
France
firstname.name@labri.fr

Pascal Ferraro

Pacific Institute for the Mathematical Sciences
University of Calgary
Calgary, Canada
pascal.ferraro@labri.fr

ABSTRACT

Harmony is one of the main properties of Western music. Structured systems based on this property are generally dedicated to music analysis. In this paper we discuss the representation of the harmony of a musical piece for music retrieval. Five levels are considered, from notes to the main tonality. A tree structured representation is then proposed in order to take into account all these levels and all the links between them. The consideration of such a tree structure, instead of a list of sequences, requires techniques for tree comparison. An editing algorithm allowing the comparison between two trees is thus presented. Preliminary experiments considering two levels of the tree are proposed. They show the interest of taking into account more than one level for comparison purposes based on harmony, whereas, to our knowledge, all the existing retrieval systems consider only one level.

1. INTRODUCTION

Automatically estimating the similarity between musical pieces is one of the major open problems in the music information retrieval research field. The applications induced are numerous, from music classification to music retrieval. From a computational point of view, estimating the musical similarities consists of developing algorithms that compute a measure which is the degree of similarity between two musical pieces. There have been several proposals for such algorithms, the majority of them considering statistics on low-level audio features. Representation of music with temporal sequences allows the application of techniques from bioinformatic or text processing fields. Experiments demonstrate a significant improvement in the performances of audio similarity measures using temporal sequences of features [1]. Adaptations of alignment algorithms lead to robust systems since they can be parameterized to take into account spurious errors in the temporal sequence [3].

Music is composed from several properties that are interdependent. Among them melody, harmony, tempo, rhythm, timbre, form, or style can be cited. A global representation

for all these characteristics and their connections is a complex but very exciting challenge. We choose to deal here with the harmony component of music. In Western music, harmony may be defined as the use of different pitches simultaneously, and chords, actual or implied. It often refers to the vertical aspects of music, distinguished from the idea of melodic line, or the horizontal aspect. Although harmony is only one of the components of the music, other properties such as the melody, rhythm, or even the form of a musical piece can be addressed by an analysis of the harmony. Melody may be extracted with voice separation methods, and the form can be estimated by the chord sequence and especially by the observation of cadenzas. In addition, the main tonal properties (*i.e.* information about chord sequences or tonality) are contained in the harmony. We assume that retrieval systems may be improved by combining all the tonal properties instead of considering each of these properties independently (such as chords or notes). Therefore, an adapted representation has to be proposed, and specific algorithms for comparison have to be developed.

We propose a new global representation of tonal properties in Section 2. An algorithm for comparing such representations is detailed in Section 3 and the results of preliminary experiments are described in Section 4. Finally, perspectives are presented in Section 5.

2. REPRESENTATION OF THE HARMONY

In this section we discuss the way to represent all the information related to the harmony of a musical piece. This representation may be used for music retrieval purposes by considering different levels of the tonal parameters.

2.1. List of Sequences

A first idea is to successively consider sequences from a musical piece for each of the tonal elements presented in the previous section. This representation thus consists of a list of different sequences, *i.e.* a sequence of notes, a sequence of chords, a sequence of tonalities and a main tonality.

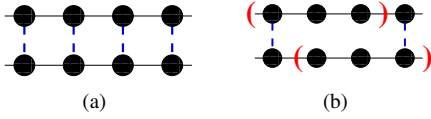


Figure 1. A sequence of notes match perfectly with another one in the case (a). However, this alignment does not take into account the harmony structure, shown in the brackets on (b). With this additional information, a better matching is computed, which respects the harmony.

These sequences may be used in different ways for music retrieval or classification purposes. It is thus possible to choose the level for comparing two music pieces, depending on the choice of the tonal properties concerned. Highest levels could also be used as filters: in a large database for example, only the pieces being similar on highest levels (with a same chord sequence for example) would be compared at the note level. Another idea is to combine together scores obtained by comparisons on different levels. Thus, some musical pieces could be similar regarding harmony and melody without matching very well on each level. Every linear combination of the scores obtained for each level is therefore possible, depending on the purpose of the analysis.

However this representation is somewhat limited: without a structured information about the harmony, some inadequate choices could be made when estimating the similarity. A typical case is illustrated by Figure 1. We then propose to use a global structure for the representation of the harmony, instead of a list of sequences.

2.2. Harmony Tree

Many works have already proposed to structure the representation of music. Thus, methods based on the theory from Schenker [7] or Lerdahl and Jackendoff [4] induce a tree, with rule-based reduction algorithms. [6] also proposes a tree representation for tonality guessing based on time reduction. However, the number of levels in the induced trees is not defined, since it depends on the musical data. It is thus difficult to use this kind of structure to compare two different musical pieces. Because two similar pieces can be represented by trees with different depths, there is no sense for a comparison using a fixed level.

We propose to represent the harmony of a musical piece with an ordered tree, according to the time, on five levels illustrated in Figure 2. A rooted tree is said *ordered* if the set of children of a given vertex are ordered. These are therefore trees for which the left-to-right order among the sibling vertices is significant. In our case this order is given by the time. Let us define the five levels of the harmony tree:

1. The main tonality of a musical piece is the root of the structure.

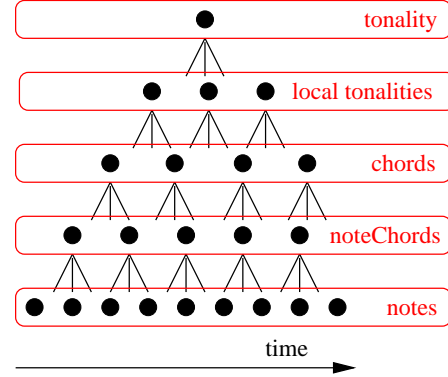


Figure 2. Ordered tree for harmony representation.

2. The second level is the sequence of local tonalities, *i.e.* the successive modulations, each of them being linked to the root.
3. The third level is the chord sequence, as the harmonic background of a jazz standard could be notated, for example. Each chord of this level contributes to a tonality and is therefore linked to a local tonality above.
4. The fourth level contains the sequence of noteChords, that is a vertical reduction of a polyphony. Several noteChords could sound during one chord, they are therefore all linked to a father chord.
5. The fifth and last level contains the sequence of the notes of a musical piece. Each note is a part of a noteChord.

3. TREE-TO-TREE EDITING PROBLEM

The comparison of trees is an important operation applied in several fields, such as molecular biology or pattern recognition. To compute similarity between trees, edit distance metrics, initially introduced for string to string comparison problem were first extended to compare ordered trees [8]. A distance between two trees is thus computed as the minimum cost of a sequence of elementary operations that convert one tree into the other. We consider here an extension of the Selkow's algorithm [8], that computes the local similarity between two trees by considering an optimal sequence of edit operation transforming these two trees [5]. In our application, edit operations are constrained such that depth and order relationships between vertices are preserved. The computation of a local similarity allows to detect local conserved areas between both trees. The solution of such a problem is based on the notion of prefix mapping between trees.

Definition 1 Let T be a tree rooted in r , any partial subtree of T rooted in r is called a prefix of T or a T -prefix. By convention, the empty tree θ is a T -prefix.

Note that a particular prefix of T rooted in r is $T[r]$ itself. Let T_1 and T_2 be two trees and let x_1 and x_2 be two vertices of T_1 and T_2 , the sets of $T_1[x_1]$ -prefixes and $T_2[x_2]$ -prefixes are respectively denoted by $\mathcal{T}_1[x_1]$ and $\mathcal{T}_2[x_2]$. A similar definition can be proposed for a forest:

Definition 2 Let F be a forest made of n trees T_1, \dots, T_n respectively rooted in r_1, r_2, \dots, r_n . A F -prefix is a sub-forest of F made of any prefixes of T_1, \dots, T_n .

The local prefix mapping problem for a given pair x_1, x_2 of vertices is to find a (possibly empty) prefix ρ_1 of $T_1[x_1]$ and a (possibly empty) prefix ρ_2 of $T_2[x_2]$ such that the score of the optimal sequence of edit operations transforming ρ_1 into ρ_2 is the maximum over all scores of sequences of edit operations between prefixes of $T_1[x_1]$ and $T_2[x_2]$. The score of the sequence solving the optimal local prefix mapping problem (called local score) for a given pair x_1, x_2 of vertices is denoted by $LS(T_1[x_1], T_2[x_2])$:

$$LS(T_1[x_1], T_2[x_2]) = \max\{S(\rho_1, \rho_2), (\rho_1, \rho_2) \in \mathcal{T}_1[x_1] \times \mathcal{T}_2[x_2]\}.$$

Note that a local prefix problem between two forests $F_1[x_1 \dots y_1]$ and $F_2[x_2 \dots y_2]$ is similarly defined as:

$$LS(F_1[x_1 \dots y_1], F_2[x_2 \dots y_2]) = \max\{S_F(\rho_1, \rho_2), (\rho_1, \rho_2) \in \mathcal{F}_1[x_1 \dots y_1] \times \mathcal{F}_2[x_2 \dots y_2]\}.$$

where $\mathcal{F}_1[x_1 \dots y_1]$ and $\mathcal{F}_2[x_2 \dots y_2]$ represent respectively the set of $F_1[x_1 \dots y_1]$ -prefixes and $F_2[x_2 \dots y_2]$ -prefixes.

Local similarity between two trees is then defined as the score of the best pair of local prefixes in trees T_1 and T_2 :

$$LS(T_1, T_2) = \max\{LS(T_1[x_1], T_2[x_2]), (x_1, x_2) \in V_1 \times V_2\}.$$

In order to evaluate local similarity, the algorithm thus needs first to find maximum similarity between prefixes of $T_1[x_1]$ and $T_2[x_2]$, for any pair of vertices (x_1, x_2) of $V_1 \times V_2$, and then to determine the best pair of vertices x_1^{Max}, x_2^{Max} of T_1 and T_2 . The local similarity is computed using a dynamic programming based algorithm using the following recursive relation:

$$S(F_1[x_1 \dots y_1], F_2[x_2 \dots y_2]) = \max \begin{cases} 0 \\ S(F_1[x_1 \dots y_1 - 1], F_2[x_2 \dots y_2 - 1]) + S(T_1[y_1], T_2[y_2]) \\ S(F_1[x_1 \dots y_1], F_2[x_2 \dots y_2 - 1]) + S(\theta, T[y_2]) \\ S(F_1[x_1 \dots y_1 - 1], F_2[x_2 \dots y_2]) + S(T[y_1], \theta) \end{cases}$$

4. EXPERIMENTS

We propose to carry out experiments on the harmony tree representation introduced in this paper for musical piece comparison. The algorithm described in the previous section

has been implemented. The *noise* collection is the Essen folksong database which contains more than 5000 musical pieces, symbolically encoded as MIDI files. We chose monophonic pieces because the comparison of polyphonic sequences of notes remains a difficult problem which requires a complete study. Therefore, we chose to focus on the improvements induced by considering tonal properties instead of melody only. We also restrained our studies to two levels of the harmony tree: the note and the chord levels. The note level is directly obtained from the MIDI files. The chord level is computed by the software *Melisma Music Analyzer* developed by Temperley and Sleator, which automatically estimates the chords according to the notes of the MIDI files [9]. Once the two levels of the harmony tree are defined, a similarity score is computed according to the algorithm presented in the previous section. This score is a positive real number. The higher the score, the more similar the musical pieces.

Concerning the queries, we considered two famous examples of music copyright infringement cases in the United States [2]. Two different musical pieces are associated to each case. These two pieces have been analyzed as very similar by a court. Each of these pieces is successively considered as the query, and is added to the noise collection. The notes and the chord sequences of each piece are known, so that the two levels harmony tree is manually defined from the chord annotations and the MIDI files. We expect the comparison algorithm not only to retrieve the query in the collection as the most similar piece, but also to retrieve the associated piece which has been judged as very similar, even if potentially harmonically and/or melodically different.

Table 1 shows the results of the different experiments. Three retrieval methods have been tested: the first method considers only the sequence of notes (corresponding to the deepest level of the harmony tree), the second one considers only the chord sequence (corresponding to the chord level), and the third one consider the harmony tree restrained to the two levels note and chord. For each query, the rank and the similarity score of the associated musical piece are given. We have chosen to represent notes and chords with interval values in order to take transpositions into account. A note is described by a pitch and a length. A chord is only described with one integer according to the *line of fifths* (similar to circle of fifths) [9]. Even if these representations are rather simple, we expect these experiments to show that the combination of two levels by comparing a tree structure improves significantly the music retrieval system.

The results of these first experiments are very promising since for each of the cases considered, the correct musical piece is estimated as the most similar by the system with the harmony tree representation. It is interesting to observe that when the system only considers the chord or the note sequence, the most similar piece is not always detected. In the *Autumn* case, the correct similarity estimation is performed

Representation	Musical piece Similarity score		
<i>Heim vs Universal (1946)</i>			
Query <i>Vagyok</i>			
Note	Vagyok 200.3	Perhaps 47.4	<i>Essen</i> Rank 1 35.6
Chord	48.7	22.9	23.3
Tree	249.1	57.7	34.8
Query <i>Perhaps</i>			
Note	Perhaps 170.9	Vagyok 47.4	<i>Essen</i> Rank 1 23.2
Chord	61.7	22.9	43.6
Tree	232.7	57.7	52.2
<i>Autumn leaves vs La Maritza</i>			
Query <i>Autumn Leaves</i>			
Note	Autumn 49.9	Maritza 11.4	<i>Essen</i> Rank 1 17.2
Chord	19.5	19.5	13.0
Tree	69.4	35.2	21.9
Query <i>Maritza</i>			
Note	Maritza 125.3	Autumn 11.4	<i>Essen</i> Rank 1 44.2
Chord	19.5	19.5	13.0
Tree	144.8	35.2	34.5

Table 1. Results of a few experiments with copyright infringement cases, by considering three different representations (notes, chords and two levels harmony tree).

by considering chord sequences, whereas in the other case, it is performed by considering the note sequence. The similarity can even be estimated as very poor, by considering only the notes with the query *Maritza* for example (which scored 11.4 whereas the best score of the Essen database is 44.2). Therefore, these first experiments clearly show that considering both chord and note levels improve the quality of the music retrieval system.

5. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new structured representation that takes into account the tonal properties of a musical piece. The application to music retrieval requires algorithms for tree alignment. We have presented an algorithm that allows the comparison of two harmony trees. We have also presented preliminary experiments. They show the interest of taking into account more than one level for comparison purposes based on harmony, whereas, to our knowledge, all the existing retrieval systems consider only one level.

The first results are promising: our structured representation succeeds on some retrieval of pieces where systems only considering note or chord sequences failed. Other experiments have to be proposed in the future. Comparison methods have to be improved for each level of the harmony tree. Then, a complete evaluation of the retrieval system based on this representation has to be proposed.

We work on the analysis methods that can provide the

parameters of the harmony tree from a musical piece. Such methods exist for each level of the tree as presented in this paper, but we plan to use the structured information to improve them. Indeed, we believe that the information on a level could be pertinent to analyze the parent level. Furthermore, we plan to use the representation proposed to describe the harmony of a musical piece and thus propose a visualization of its musical analysis.

6. ACKNOWLEDGEMENT

The authors would like to thank Laurent Soulié for fruitful discussion about the model proposed. This work is part of the SIMBALS project (JC07-188930), funded by the French National Research Agency. This research is also supported by the Aquitaine Regional Council.

7. REFERENCES

- [1] M. Casey and M. Slaney, "The Importance of Sequences in Musical Similarity," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, 2006.
- [2] *Music Plagiarism Project*, Columbia Center for New Media Teaching and Learning. [Online]. Available: <http://ccnmtl.columbia.edu/projects/law/library/>
- [3] P. Hanna, P. Ferraro, and M. Robine, "On Optimizing the Editing Algorithms for Evaluating Similarity Between Monophonic Musical Sequences," *Journal of New Music Research*, vol. 36, no. 4, pp. 267–279, 2007.
- [4] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: MIT Press, 1985.
- [5] A. Ouangraoua, P. Ferraro, L. Tichit, and S. Dulucq, "Local Similarity Between Quotiented Ordered Trees," *Journal of Discrete Algorithms*, vol. 5, no. 1, pp. 23–35, March 2007.
- [6] D. Rizo, J. Inesta, and P. Ponce de Leon, "Tree Model of Symbolic Music for Tonality Guessing," in *Proceedings of the 24th IASTED International Conference on Artificial Intelligence and Applications (ICAIA)*, Innsbruck, Austria, 2006, pp. 299–304.
- [7] H. Schenker, *Der Frei Satz*. Vienna: Universal Edition, 1935, published in English as *Free Composition*, translated and edited by E. Oster, Longman, 1979.
- [8] S. Selkow, "The Tree-to-Tree Editing Problem," *Information processing letters*, pp. 184–186, 1977.
- [9] D. Temperley, *The Cognition of Basic Musical Structures*. The MIT Press, 1999.