

Improvements of Alignment Algorithms for Polyphonic Music Retrieval

Pierre Hanna, Matthias Robine, Pascal Ferraro, and Julien Allali

SIMBALS Project
LaBRI - University of Bordeaux 1
F-33405 Talence cedex, France
`firstname.name@labri.fr`

Abstract. Existing algorithms for estimating the music similarity generally consider monophonic music or reduced polyphonic music. Alignment algorithms are currently applied in different domains such as string matching or bioinformatics. Their applications to monophonic music leads to efficient and robust systems. In this paper, we propose to study and improve the adaptation of such alignment algorithms to polyphonic music, without information on the component voices of this music. A representation for polyphonic music, based on existing works that consider quotiented sequences, is proposed and justified with examples. Extensions for the alignment algorithm are also presented for taking into account the limiting problem of overlapped notes. Other limitations such as transposition invariance, multiple alignment problem are discussed. Experiments with symbolic performed music show the improvements induced by the adaptation proposed.

1 Symbolic Music Retrieval

Automatic estimation of music similarity is one of the major open problem in the Music Information Retrieval research area. From a computational point of view, estimation of similarity consists of determining algorithms that calculates a measure which indicates the degree of similarity between two musical segments.

There are many applications of such similarity estimation systems. Content-based music retrieval systems consist of searching a musical piece in a database given a short excerpt of music (query). This excerpt is generally assumed to be monophonic: only one note sounds at any given time. It may be perfect, for example an excerpt of a exact piece. But it may also be *inexact*, for example hummed or whistled, leading to slight time or pitch deviations. At the opposite, musical pieces of the database may be polyphonic, for example made of chords or defined by several different tracks corresponding to different instruments.

In this paper, we propose to investigate the adaptations of classical string-matching algorithms to this problem. We consider only symbolically encoded music. Symbolic music is defined by musical events, such as beginnings or endings of notes for example. Each note is then defined by a few attributes such as pitch, duration or onset time. Several algorithms have been introduced during the last few years. Geometric algorithms consider geometric representations of music and compute the distance between objects [14, 12]. Other algorithms adapted from

string matching domain and based on N-grams are proposed in [2, 13]. Another technique evaluates the best alignment between the two pieces compared [11]. This paper discusses this last approach which is described in Section 2.

Considering polyphonic music implies adaptations of these algorithms mainly dedicated to monophonic music. Existing works generally require a monophonic reduction of a polyphonic piece [13], for instance by considering the note with the highest pitch. Such approaches are thus based on the assumption that firstly the query is the main theme or melody of the musical piece searched, and secondly that it is always the highest pitches which compose the melody (it is rarely the case in an orchestra, where the polyphony could not be only reduced to the voice of the Western concert flute for example). In order to avoid such assumption, we propose here to study algorithms that take into account all the notes of a polyphonic musical piece. A way would be to consider all the distinct monophonic lines induced by the polyphony. But this naive method may imply a combinatorial explosion in the cases of large scores [7].

Adaptations of N-grams and geometric algorithms to polyphonic music have been proposed in [8, 2, 7]. Adaptations of alignment algorithms are more difficult. However, we think that it may lead to efficient and precise systems since it has been shown that alignment algorithms are very robust in the cases of queries with slight time or pitch deviations [3]. Therefore, we propose in this paper a study of the possible adaptations to polyphony and the induced problems.

In Section 2, we present the existing alignment algorithm for monophonic music and we propose improvements allowing to represent and compare polyphonic music. Then, in Section 3, a few experiments with real symbolic music are proposed to validate these adaptations. Finally, we discuss the limitations and the future works in Section 4.

2 Adaptation of Alignment Algorithm to Polyphonic Music

An algorithm for monophonic music comparison has been proposed by Mongeau and Sankoff in [9]. After a brief description of this algorithm, we discuss the limitations induced by this approach and we propose a few adaptations in order to take into account polyphonic music.

2.1 Monophonic music

Any monophonic musical score can be represented by a string (sequence of symbols). Measuring similarity between sequences is a well-known problem in computer science which has applications in many fields such as computational biology, text processing, . . . [5].

Among several methods, Smith and Waterman's approach [11] consists in detecting local conserved areas between two sequences. This is called *local alignment* or *local similarity problem*. A similarity score is calculated by considering elementary operations that transform one string into the other. Typical transformations between sequences include deletion of a symbol, insertion of a symbol between two existing symbols and substitution of a symbol by another.

Mongeau and Sankoff proposed in [9] an adaptation of the local alignment for the computation of a similarity score between monophonic musical pieces. This algorithm considers any monophonic piece as a sequence of ordered pairs with the pitch of the note as the first component and its length as the second. Several improvements have been recently proposed in [4, 3] that takes into account musical information. Evaluations during the second and third Music Information Retrieval Evaluation eXchange (MIREX) show that such systems are some of the most precise systems in the monophonic context.

2.2 Representation of Polyphonic Music

A few representations for notes in the polyphonic context have been proposed. Monophonic music is represented by sequences of notes. Each of these notes is defined by a pitch and a duration. A first possibility for representing polyphonic music is to define several monophonic lines [7]. For example, each of these lines can be associated to a perceived voice. This representation is very intuitive and effective since it allows the direct application of monophonic music algorithms. However, it requires the prior knowledge of separated voices. This knowledge implies an important limitation since automatic discrimination of voices is a very difficult problem. Other representations based on monophonic reductions have been proposed [7, 13]. But these reductions require the choice of orders for simultaneous notes, for example according to the pitch values. Here again, this choice induces the limitation of this representation.

To take into account the polyphonic nature of musical sequences, we propose to use a quotiented sequence representation introduced in [6] and to consider the note onsets. Notes that start at the same time are grouped together. Inside each of these groups, no arbitrary order is defined. However, in order to avoid a combinatorial explosion, pitches are represented modulo 12, so that only one occurrence of the same note in different octaves is preserved.

This approach leads us to consider any polyphonic sequence as a sequence of series of unordered triplets. As for monophonic sequence [9], each triplet represents a note n and is defined by the pitch of the note $p(n)$, its duration (coded in sixteenth note values) $d(n)$ and its onset $o(n)$. The pitch of the note is coded modulo an octave and is thus defined by integer values in the range $[0; 11]$.

An important problem is induced by the representation of polyphonic music when considering note durations. This problem is illustrated by Figure 1 which shows the musical score of an excerpt of *Swan Lake* by Tchaikovsky. This excerpt is polyphonic. Some notes in the upper staff (first note of the third bar for example) still sound whereas other notes start in the lower staff. The representation of the third bar thus results in six triplets. This representation is not satisfactory since, for example, the second triplet contains only one eighth note (E, lower staff) and not the half note that still sounds at the same time (A, upper staff).

In order to take into account the possibility that one note still sounds at the same time another note starts, we propose to introduce the notion of *tied notes*. For example, if a new note n_2 starts whereas another note n_1 still sounds, we



Fig. 1. Polyphonic excerpt of *Swan Lake* by Tchaikovsky.

represent a chord composed of n_2 and a tied note n'_1 , where n'_1 has onset time $o(n'_1) = o(n_1)$, pitch $p(n'_1) = p(n_1)$ but with a smaller duration $d(n'_1)$:

$$d(n'_1) = d(n_1) - [o(n_2) - o(n_1)] \quad (1)$$

The duration d_{n_1} of the note n_1 is also modified:

$$d(n_1) = o(n_2) - o(n_1) \quad (2)$$

This example considers only one tied note, but the number of consecutive tied notes is not limited. For example, Figure 2 illustrates the polyphonic representation with tied notes. The third bar is represented by six triplets, four of which are composed of tied notes A.

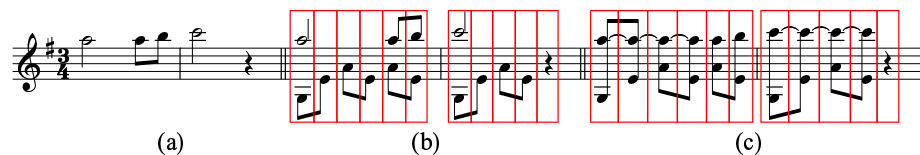


Fig. 2. Illustration of the representation of polyphonic music with tied notes: a monophonic query corresponding to one voice (a), a first polyphonic representation based on onsets (b), and the representation chosen that considers tied notes (c).

The idea of considering tied notes is somewhat similar to the concept of *notebits* introduced in [10]. However, one of the main difference is the algorithm related to these tied notes: we propose here to consider music with polyphonic tracks, or not to take into account information about tracks or voices. The problem is thus more difficult. The representation of polyphonic music with tied notes allows alignment algorithm to take into account overlapped notes by defining one specific operation. In Section 3, experiments show that the improvements induced by this approach are important.

2.3 Overlapped Notes

In the context of polyphonic musical comparison, we propose to consider the edit operation introduced by Mongeau and Sankoff [9], namely fragmentation. It

involves the substitution of one symbol by several. Mongeau and Sankoff introduced this operation in a monophonic context. We have developed an adaptation of their algorithm to the polyphonic context.

We restrain the application of the fragmentation operation to tied notes and rests. We propose to not consider the consolidation operation, since we do not need to represent tied notes in the monophonic context. The score associated to this operation is highly negative except if one note of the query matches with a sequence of tied notes. One note matches with tied notes only if the pitch of all tied notes are the same, and if the onset associated to all tied notes are also the same. These two properties impose that all the tied notes considered are associated to the same original note. If these conditions are satisfied, the score of the fragmentation operation is the substitution score associated to the note of the query and the resulting note of the tied notes, whose duration is the sum of the durations of the tied notes concerned.

From a practical point of view, the number of tied notes has to be limited in order to reduce the complexity of the retrieval algorithm. Results of experiments lead us to fix this limit to 6 notes.

2.4 Multi-track Query

When considering adaptations of polyphonic music similarity systems, it is necessary to study a few problems specific to the context of polyphony. One of these problems has been described by Pardo and Sanghi as the *multiple alignment problem* [10]. For example, a query can represent polyphonic pieces by skipping from one part to another one. For instance, in applications such as query-by-humming, users can sing a verse then imitate the guitar riff following this verse. In such cases, comparing a query to polyphonic pieces cannot be performed by comparing independently the query to each track of the pieces tested.

The representation of polyphonic music introduced previously allows the retrieval of polyphonic pieces with a query composed of fragments of different tracks, since all the notes sounding at the same time are all considered.

2.5 Transposition Invariance

A recent interesting analysis of existing approaches for retrieval of polyphonic music compares the geometric and the alignment algorithms [7]. One of the conclusion is that alignment algorithms *face severe problems in combining polyphony and transposition invariance* [7]. The problem of the transposition invariance property is induced by the representation of polyphonic music. Interval or contour representations of pitches allows retrieval systems to be transposition invariant in the monophonic context, but such representations cannot be applied in polyphonic context.

However, a new algorithm specific to this problem of transposition invariance have been proposed in [1]. The new dynamic programming algorithm proposed permits to take into account multiple local transpositions, and can perfectly be applied to representations of polyphonic music. Experiments show that applying this algorithm lead to a polyphonic music retrieval system that is transposition invariant. For now, the main disadvantage is the significant time computation added by the algorithm. But optimizations are under development.

3 Experiments

We present in this section the results of the experiments we performed to show the accuracy of the polyphonic music retrieval system presented in the previous section. During MIREX 2006, the second task of the symbolic melodic similarity contest consisted in retrieving the most similar pieces from mostly polyphonic collections given a monophonic query. Two collections was considered. The *karaoke* collection is composed of about 1000 *.kar* files (Karaoke MIDI files) with mostly Western popular music. In the following, we consider this collection.

A few monophonic queries have been defined from polyphonic musical pieces. These queries are short (less than 20 notes). Some of them are exact but some of them are interpreted and thus have slight time or pitch deviations. Some queries have been transposed in order to test the property of transposition invariance. The queries considered are excerpts of:

1. *Swan Lake* by Tchaikovsky (19 notes)
2. *The Winner Takes It All* by ABBA (18 notes, inexact)
3. *Silent Night*, popular song (11 notes, inexact)
4. *Fur Elise* by Beethoven (17 notes, transposed)
5. *Ah, vous dirai-je, Maman* by Mozart (14 notes, transposed)
6. *Angels We Have Heard On High*, popular song (14 notes)
7. *The Final Countdown* by Europe, (9 notes, inexact)

Experiments consist of computing the similarity score between one query and each piece of the collection. The similarity score is normalized, by dividing it by the maximum similarity score. This maximum score is simply obtained by computing the similarity score between the query considered and itself. In the following results, normalized scores are real values inside the range $[0; 1]$. The higher the score, the more similar to the query the piece is. The result expected is the retrieval of the piece of the collection that corresponds to the query. Table 1 thus shows the score, the rank (obtained according to the similarity score) of this corresponding piece and the score obtained by the piece that has been estimated as the most similar (first rank).

Moreover, in order to evaluate the improvements induced by the adaptations presented in the paper, we run two algorithms. The first algorithm is the basic alignment algorithm, and is denoted *normal* in tables. The other algorithm consider tied notes and is denoted *improved*. This is the only difference between the two algorithms.

During these experiments, we consider three different tracks for each piece of the database. Each track may be polyphonic, and the mix of three tracks may lead to chords composed of several notes. In the case of short query, taking into account more tracks leads to high similarity score whatever the piece tested. We focus on tracks which may lead to retrieval errors with existing approaches, due to overlapped notes.

Table 1 show the results obtained. These results clearly show the improvements induced by taking into account tied notes. Without improvement, the alignment algorithm does not allow the retrieval of the correct musical piece. Many other pieces obtain a better similarity score (> 50 indicates that the

	Normal			Improved	
Query	Score	Rank	Rank 1 Score	Score	Rank
1	0.55	> 50	0.85	1	1
2	0.52	> 50	0.83	0.93	1
3	0.68	> 50	0.95	0.98	1
4	0.69	4	0.72	1	1
5	0.59	> 50	0.86	1	1
6	0.58	> 50	0.83	1	1
7	0.79	> 50	0.87	0.98	1

Table 1. Results of polyphonic music retrieval from a monophonic query with a basic alignment algorithm and with improved algorithm. The rank and the score of the correct piece are indicated. The correct piece is retrieved by applying the improved alignment algorithm.

correct piece has not been ranked in the first 50 pieces). By introducing and considering tied notes, the results are far better. For the queries proposed, the correct piece is always estimated as the most similar piece. The similarity score is always 1, except for queries that was not an exact excerpt of the polyphonic piece. The score remains very important in these cases.

However, it is important to note that the system may be improved. Indeed, by introducing the fragmentation operation, the similarity score computed is very important, even for dissimilar pieces. We think that additional musical information has to be considered in order to reduce these scores.

4 Conclusion

In this paper we present improvements for the adaptation of alignment algorithm for symbolic polyphonic music retrieval. Experiments show that the consideration of tied notes significantly improves the quality of the retrieval system. The technique presented for representing polyphonic music with these tied notes, and for taking into account them during the alignment process without prior knowledge about voices are the major contribution of this paper. Applications of the algorithm obtained involve retrieving musical pieces from any excerpt of any instrument, not only from the main melody or theme.

The main limitation of this approach concerns the high number of simultaneous notes (or tied notes) that may sound at the same time. If this number is too important regarding the length of the query, the similarity score with any musical piece may be high. We think it is thus necessary to add some additional musical information, for example by considering more complex similarity between one note and one chord, or to avoid important intervals between successive notes. We could also apply rules induced by voice separation algorithms.

Furthermore, the algorithms presented here leads to important computation time, depending mainly on the size of the musical piece collection. Optimizations have to be proposed in order to be able to apply these algorithms to huge databases. The method presented in this paper computes an exact solution of the optimisation problem of sequence alignment in quadratic time. Actually, for our application, this exact solution is not necessary. An approximation by indexing

all subsequences of the database may be obtained in linear time and then may lead to a more efficient algorithm.

5 Acknowledgement

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

References

1. Julien Allali, Pierre Hanna, Pascal Ferraro, and Costas Iliopoulos. Local Transpositions in Alignment of Polyphonic Musical Sequences. In *Proceedings of the 14th String Processing and Information Retrieval Symposium (SPIRE)*, October 2007.
2. Shyamala Doraisamy and Stefan Rger. Robust Polyphonic Music Retrieval with N-grams. *Journal of Intelligent Information Systems*, 21(1):53–70, 2003.
3. 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.
4. Carlos Gómez, Soraya Abad-Mota, and Edna Ruckhaus. An Analysis of the Mongeau-Sankoff Algorithm for Music Information Retrieval. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 109–110, Vienna, Austria, September 2007.
5. Dan Gusfield. *Algorithms on Strings, Trees and Sequences - Computer Science and Computational Biology*. Cambridge University Press, 1997.
6. 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.
7. Kjell Lemström and Anna Pienimäki. Approaches for Content-Based Retrieval of Symbolically Encoded Polyphonic Music. In *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC)*, Bologna, Italy, 2006.
8. Anna Lubiw and Luke Tanur. Pattern Matching in Polyphonic Music as a Weighted Geometric Translation Problem. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, pages 289–296, Barcelona, Spain, October 2004.
9. Marcel Mongeau and David Sankoff. Comparison of Musical Sequences. *Computers and the Humanities*, 24(3):161–175, 1990.
10. Bryan Pardo and Manan Sanghi. Polyphonic Musical Sequence Alignment for Database Search. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 215–222, 2005.
11. T.F. Smith and M.S. Waterman. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147:195–197, 1981.
12. Rainer Typke, Remco C. Veltkamp, and Frans Wiering. Searching Notated Polyphonic Music Using Transportation Distances. In *Proceedings of the 12th ACM Multimedia Conference (MM)*, pages 128–135, New-York, USA, 2004.
13. Alexandra L. Uitdenbogerd. *Music Information Retrieval Technology*. PhD thesis, RMIT University, Melbourne, Australia, July 2002.
14. Esko Ukkonen, Kjell Lemström, and Veli Mäkinen. Geometric Algorithms for Transposition Invariant Content-Based Music Retrieval. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*, pages 193–199, Baltimore, USA, October 2003.