Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

MUSICAL SIMILARITY IN A POLYPHONIC CONTEXT : A MODEL OUTSIDE TIME

Benoit Meudic

Musical Representation Team
Ircam, Paris
`meudic@ircam.fr`

ABSTRACT

In the context of pattern extraction from polyphonic music, we challenge an approach outside time for computing the similarity between two musical sequences which neither modelizes temporal context nor expectancy. If theses notions might play a role in our perception of musical patterns, we propose in a first step to investigate the limits of a system that ignores them. Our approach relies on a new representation of the polyphonic musical sequence which is quantized in equally-spaced beat-segments and on a new definition of the notion of similarity in a polyphonic context. In agreement with ([1], [2]), we think that text-matching methods, or pure mathematical algorithms are not directly convenient for music analysis. We think that the similarity relationships between musical sequences are the result of a cognitive process that implies to evaluate the algorithms in terms of their cognitive relevance. As few experiments have been made on people's cognitive criteria for similarity measuring, we base our criteria on heuristics that were inspired from some musical issues. Three different sets of features have been considered: pitches, pitch contours and rhythm. For each set, a similarity measure is computed. The global similarity value results from the linear combination of the three values. The algorithm was tested on several pieces of music, and interesting results were found. At the same time, new questions were raised on the notion of similarity (this research is part of the European project Cuidado).

1. BACKGROUND

The interest for music similarity has been growing for a few years. Several techniques associated with different musical representation formats have been proposed for performing similarity measurement. Most of the techniques are context-independent: the computation of the similarity value between two musical sequences does not depend on the events that have occured before or between the two sequences. Against the rule, an interesting proposition for an inductive model can be found in [3]. Very few approaches propose to analyse polyphonic data. When they do, polyphony is not considered as a specific musical issue: most of the papers aim to transform polyphony in a monophonic approximation, while others consider polyphony as a part of a more general multidimensional mathematical issue [4]. In the last case, musical results are not provided and similarity is only based on exact repetition.

Among the following approaches, the issue of polyphony is not tackled, but the proposed strategies are interesting because they illustrate numerous different ways to compute the similarity.

An approach using dynamic programming can be found in [5]. The similarity between two sequences relies on an edit distance that measures the number of basic operations (adding, deleting, moving a note…) which transform one sequence into the other. The difficulty here is to determine both the basic operations and their "cognitive cost".

Another interesting approach [1] uses statistical information about pitches and durations together with a contour representation extracted from scores in order to obtain feature maps which are formed by an unsupervised learning algorithm. Unfortunately, it is implicitly assumed that the similarity between the different melodies is a transitive relation (distances in the two dimensions super-map are euclidean), whereas this is hardly the case in music. Moreover, the temporal succession of the events is not considered in the features (except in the contour) and maybe other features should be taken into account. Also, note that this approach is not context independent as a learning process is required for initialization.

Another approach is described in [2]. The similarity is based on the length of the vector of the differences between two sequences of melotones (pitches representation) or cronota (durations representation). However, this measure has limitations. For instance, intervals between not contiguous notes that would be common to the two compared sequences are not being considered in the measure. Moreover, the cronota sequence is represented by multiples of the 1/16 note, which is not compatible with ternary rhythmic values. Choosing the common denominator of all the durations as a basic unit would imply a far too complex analysis, as most of the very small durations do not play a role in the similarity measure. We will propose another representation format in paragraph 3.

Lastly, [6] considers a set of several features such as pitches (or duration) profiles, intervals and contour as binary features. The similarity between two sequences increases with the number of shared features. This global approach does not allow for local variations. For instance, with this approach, contours can be identical or different, but not similar.

2. QUESTIONS ON SIMILARITY

The interest for music similarity has raised with new commercial applications, such as query by humming, which have emerged from the growth of Internet.

However, the notion of similarity remains very difficult to define. Usually, similarity is not considered differently according to the different objectives for which it is computed. However, we think that it would be a first step in order to make this notion clear.

For instance, query by humming and pattern extraction are very different tasks. In one case, the goal is to match two sequences that should be the same but are different because they have been interpreted from the original score. In the other case, the goal is to match two sequences which are perceived as similar but which have initially been written differently by the composer. If the first matching do not specially requires cognitive considerations (two sequences are similar because they should physically be the same), the second one often needs them: two sequences can be heard as similar while being physically different.

Another issue arises when dealing with polyphonic sequences which happen very often in music. The term polyphony doesn't automatically mean several pitches at the same time, but several voices at the same time. For instance, a "Suite" of J.S Bach for violoncelle can be polyphonic whereas only one event is played at the same time (it is a monodic polyphony). An answer to the issue of polyphony would be the automatic separation of the different voices, but no algorithm can currently perform that. Thus, we think that polyphony must be taken into account in a similarity measure, which raises new issues that will be tackled later.

Another problem arises from temporal considerations: should a similarity algorithm be independent of the context? This is hardly the case when considering that our culture, education and memory influence our perception of similarity. However, one can still wonder is there exists some universals.

3. AIMS

The method we present in this article is a new model for computing the similarity value between two non-identical polyphonic sequences.

In a first step, we have tried to challenge the notion of context. For instance, we do not introduce knowledge about tonality neither we modelize induction, memorization or learning processes. We know that this draws the limits of the system, and we know that a further step would be to integrate theses notions. But we are still very interested in exploring the limits of a system "outside time". Besides, we will not try to extract *all* but *a set* of significant patterns from a polyphonic music. Thus, it may happen that two musical sequences that are very similar perceptually are not recognised as such by our algorithm. In a first step, our limited goal is that sequences recognised as very similar by the algorithm should indeed be very similar perceptually.

4. THE SIMILARITY MODEL

4.1. Musical representation : the b.s

Performing similarity measures on not quantized music is rather difficult as soon as identical sequences with different tempi will physically appear as different whereas they are cognitively heard as similar. Our model integrates this cognitive aspect by representing each pattern as a quantified sequence of beat segments (b.s). Each b.s is a polyphonic sequence of pitches, onsets and durations in the MIDI format (see figure 1). The beat-tracking algorithm we use is described in [7].

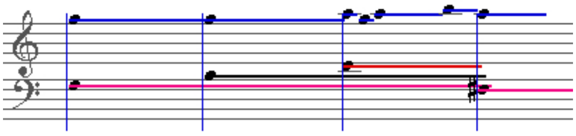

Figure 1: Beginning of the "variations Goldberg" from Bach. The vertical lines delimit the beat segments. Horizontal lines are the durations of each event.

4.2. Similarity in a polyphonic context

We assume that the notion of similarity between two polyphonic musical sequences makes sense. No information is available on the different voices of each sequence. Computing the intersection between the two sequences would appear as an intuitive way to measure what is common between the two sequences. However, the intersection could be empty while the two sequences would be perceived as similar.

Thus, we state that a sequence x is similar to a sequence x' if x is approximatively included in x'. For instance, when listening to music, we try to associate one sequence already heard with the current sequence we are hearing. We do not intersect the two sequences, but we evaluate the similarity between one sequence and a reference one. Thus, in our model, we understand similarity between two sequences x and x' as the distance from x to a certain sequence sub(x') included in x'. Note that this measure is not symetric (see equation 2).

4.3. Introduction to some cognitive aspects of the model

An important cognitive aspect of our model is that a musical sequence of b.s is considered as a whole entity (it may contain an abstract cognitive structure), and not solely as the concatenation of smaller entities. We think that several relations between non-adjacent events emerge from the whole entity. Theses relations play a role in the cognitive processes for recognizing the similarity between two sequences. To integrate this aspect, the similarity value between two sequences will not be computed from the addition of the similarity values between the smaller components:

$$S(x, x') + S(y, y') \neq S(xy, x'y') \qquad (1)$$

Where S(x, x') designs the similarity value between sequences x and x', and xy designs the concatenation of sequence x and y.

Another cognitive aspect (see 4.2) is that our similarity measure is not symmetric in a polyphonic context:

$$S(x, x') \neq S(x', x) \qquad (2)$$

If x is approximatively included in x', x will be very similar to x'. But x' will not automatically be approximatively included in x.

Last, according to cognitive aspects in [8], a similarity measure is not transitive:

$$S(x, y) + S(y, z) \leq \text{ or } \geq S(x,z) \qquad (3)$$

Our similarity computation provides a real value between 0 and 1 which state how similar are the sequences (1 is for identical). In our pattern extraction project, we use a similarity matrix for the representation of the results, and we perform clustering, but this will not be presented in this paper.

We will now describe our model for similarity. Because of lack of space, we have chosen to provide a general overview of our algorithm rather than a detailed description of a part of it.

In our model, sequences of b.s are of same length (length is expressed in number of b.s), so that each position of b.s in a given sequence can be matched with the same position of b.s in another sequence. We compute three different similarity values by considering three different sets of features: pitches (chords, pitch intervals etc…), pitch contours (contour at the top and at the bottom of the polyphony) and rhythm.

4.4. Similarity measure for pitches

We consider here the chords and the pitch intervals features. A similarity value is computed from two b.s sequences seq1 and seq2 of same length.

The only events falling on the downbeats are considered. This may be arguable, but two reasons have conducted this choice:

- Considering all the polyphonic events would require too much running time.
- The downbeats are often perceived as salient temporal position. Two sequences whose pitches coincide on the downbeat but differ elsewhere are often recognised as very similar (this has been confirmed in our experiments).

Usually, a downbeat event (dwb.event) is a chord, but it can also be a note or a rest.

In order to consider all the possible relations (see figure2) between non-adjacent dwb.event, the global similarity results from the computation of a similarity value between all the pairs of dwb.event in seq1 (dwb.event-seq1(i), dwb.event-seq1(j)) and their corresponding pairs in seq2 (dwb.event-seq2(i), dwb.event-seq2(j)). i and j (> i) are indexes of the considered b.s.

The similarity value for a pair depends on:

- the length of the different combinations of intersections between the four dwb.events considered as chords (harmonic similarity)
- the length of the intersection between all the intervals between dwb.event-seq1(i) and dwb.event-seq1(j) and the intervals between dwb.event-seq2(i) and dwb.event-seq2(j) (melodic similarity).

The similarity values between pitches or intervals are :

- 1 for equal pitches or equal intervals
- 0.5 for transposed chords
- 0 otherwise

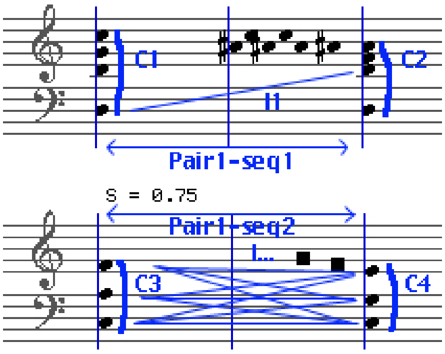

Figure 2 (see part 5 for details on the symbols): Two similar patterns in Sonat8Am-mvt2 from Mozart. All vertical and horizontal intervals between dwb.events C1 and C2 of Pair1-seq1 and dwb.events C3 and C4 of Pair1-seq2 are compared.

4.5. Similarity measure for contours

As reported in [1], contour plays an important role in the perception of melodic similarity.

Our model compares the upper and lower contours of two b.s sequences seq1 and seq2 of same length.

As above, the only events falling on a downbeat (dwb.events) are considered. An up (down) contour is the sequence of the intervals between the upper (lower) pitches of the dwb.events. Each contour of each sequence is compared with the two contours of the other sequence. Contours are very similar (see figure3) if the intervals from one sequence are similar to the corresponding intervals from the other sequence (two intervals are similar if their difference is less than 5 half tones).

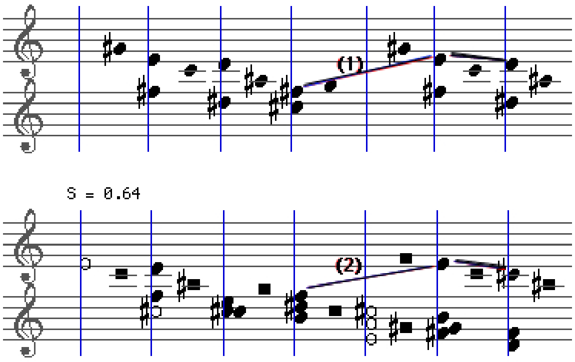

Figure 3 (see part 5 for details on the symbols): Two similar patterns in Pierrot Lunaire from Schoenberg. Lines (1) and (2) show similar contours.

4.6. Similarity measure for rhythm

We believe that rhythm is a component of the cognitive criteria we use in the recognizing of similarity. Our model compares the rhythmic structure of two sequences of b.s seq1 and seq2 of same length. In a first step, seq1 and seq2 are normalized so that the total duration of the b.s will be the same for seq1 and seq2. Then, for each b.s, onsets (temporal positions) in seq1 are associated to the corresponding onsets in seq2. Two onsets of two b.s form a pair if they share similar temporal positions in the b.s. If an onset of one sequence do not form a pair with an onset of the other sequence, then it is deleted. The similarity between two sequences of b.s is the mean of the similarity between each corresponding b.s. (as seq1 and seq2 have same length, each b.s of seq1 correspond to one b.s of seq2) The similarity between two corresponding b.s is the mean of the similarity between each pair of corresponding onsets (here, relations between non-adjacent events are not considered). Corresponding onsets are already similar because they share the same temporal position in the b.s. The similarity increases with the length of the intersection of the durations of the events corresponding to the onsets of a pair (an approximation value is considered for the intersection).

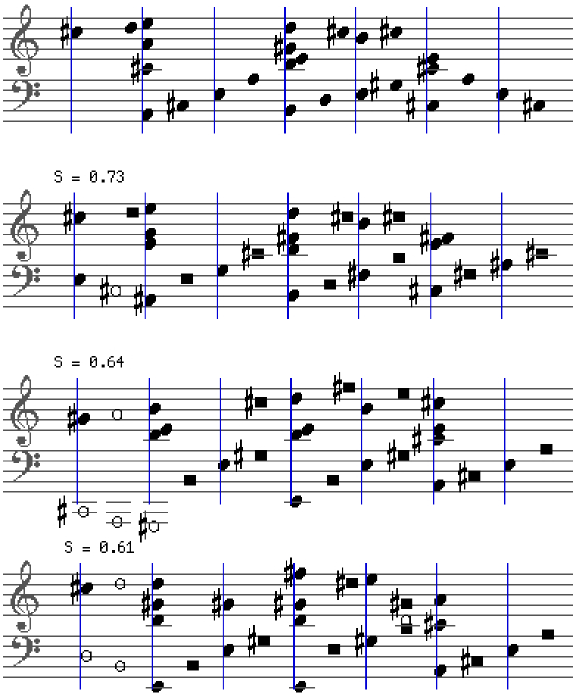

Figure 4 (see part 5 for details on the symbols): Three patterns similar to the first pattern in Sonat AM d664 op121, 1rst Mvt from Schubert.

4.7. Overall similarity measure

Each of the three above models (pitches, contour and rhythm) computes a similarity value. The three values are then linearly combined to form a global similarity measure. The similarity for pitches is twice more weighted than the two others.

Due to the non-symetric relation (equation 2), the similarity value between two non-ordered sequences seq1 and seq2 is componed of two different values: S(seq1, seq2) and S(seq2, seq1). If only one value is needed, our model considers the greater one.

5. MUSICAL EXEMPLES

Each of the above musical exemples (Figure2, Figure3 and Figure4) shows a reference pattern (the above one) together with other similar patterns. Durations are not represented. In figure4, the similar patterns are sorted according to their decreasing similarity value S. The events that determined the similarity between the patterns (for pitches, contours and rhythm) are represented in black. The square symbols only determined the similarity for rhythm.

6. CONCLUSION

We have presented the general lines for a polyphonic similarity model that integrates some cognitive principles. Interesting musical results have been presented which show that patterns can be extracted without considerations on the temporal context. We think that further investigation should be done in this direction. For instance, one could introduce a dissimilarity measure in order to consider events which contribute to the perceptual differentiation of sequences. Other features, such as statistical features, or the dynamics of the notes, should also be considered.

7. REFERENCES

[1] Petri Toivainen & Tuomas Eerola ``A computational model of melodic similarity based on multiple representations and self-organizing maps", Proceedings of the 7th ICMPC Sydney 2002.

[2] Hofman-Engl, L "Melodic transformations and similarity - a theoretical and empirical approach", Phd Thesis, Keele University, 2002.

[3] Lartillot O, "Generalized Musical Pattern Discovery by Analogy from Local Viewpoints", in Discovery Science, LNCS 2534, Springer-Verlag, 2002

[4] Wiggins, Lemstrom, Meredith "SIA(M)ESE: An algorithm for transposition invariant, polyphonic content-based music retrieval", Proceedings of the 3rd ISMIR, Ircam, 2002

[5] M. Mongeau and D. Sankoff, "Comparison of musical sequences", Computers and the Humanities 24:161-175, 1990.

[6] Cambouropoulos and Widmer, "Melodic clustering : Motivic analysis of Schuman's Träumerei", Proceedings of the 3rd Jim, Bordeaux, 2002

[7] Meudic B, "A causal algorithm for beat tracking", 2nd conference on understanding and creating music, Caserta, 2002

[8] Tversky A, "Features of similarity", journal of Psychological Review, p327-352 1977