MX STRUCTURAL METADATA AS MIR TOOLS

Goffredo Haus, Alberto Pinto
LIM-DICo
Università degli studi di Milano
{haus, pinto}@dico.unimi.it

ABSTRACT

In this paper two main topics are covered: a new model for melodic fragments or indexes and its embedding in MX, the new XML-based format currently undergoing the IEEE standardization process, in order to take advantage of its structural layer in a MIR-oriented perspective. Our goal is to show how music information retrieval methods can be improved by the use of particular melodic invariants coming from graph theoretic tools applied to melodies in order to catch more musical transformations than other methods as, for example, permutations of subfragments. This approach to the melodic similarity problem leads to associate a *musical graph* to each melodic *segment* in order to extract MIR-useful invariant quantities.

Keywords: MX, XML metadata, music information retrieval, structural layer, invariants, melodic similarity, musical graphs.

1. INTRODUCTION

The importance of music information retrieval in large musical databases increases as increase the number and dimension of multimedia databases.

Music can take advantage from XML languages because of its intrisically layered structure (cfr. [17]), from audio to structural information. Unfortunatly there is currently no defined, independent standard for representing music information that can describe and process all the different layers which characterize music information. For each layer of music information, there is one or more accepted standards (e.g. MIDI for performances, NIFF for notation and so on) and/or one or more proprietary formats. None of them can be suitably applied to other layers (cfr. [18]).

XML provides an effective way to represent multimedia information at different levels of abstraction and XML metadata provide useful tools for the retrieval processes. The IEEE "Definition of a Commonly Acceptable Musical Application Using the XML Language" project has been developing an XML application defining a standard language for symbolic music representation. The language is a meta-representation of music information for describing and processing said music infromation within a multilayered environment, for achieving integration among structural, score, MIDI, and digital sound levels of representation.

Furthermore, the proposed standard should integrate music representation with already defined and accepted common standards.

The standard will be accepted by any kind of software dealing with music information, e.g. score editing, OMR systems, music performance, musical databases, and composition and musicological applications.

1.1. The MX structure

In MX, the new XML-based format MX currently undergoing the IEEE standardization process, music information is represented according to a multi-layered structure and to the concept of space-time construct. Infact, music information can be structured by a layer subdivision model, as shown in Fig. 1.
Each layer is specific to a different degree of abstraction in music information: General, Structural, Music Logic, Notational, Performance and Audio. [1] gives an exhaustive description of this format and the issue of the integration between MX and other formats is covered in [3].
The main advantage of MX is the richness of its descriptive features, which are based on other commonly accepted encodings aimed at more specific descriptions.

The multi-layered music information structure is kept together by the concept of *spine*. Spine is a structure that relates time and spatial information (cfr. Fig. 2), where measurement units are expressed in relative format. Through such a mapping, it is possible to fix a point in a layer instance (e.g. Notational) and investigate the corresponding point in another one (e.g. Performance or Audio).

The *Structural* layer contains explicit descriptions of music objects together with their causal relationships, from both the compositional and musicological point of view, i.e. how music objects can be described as transformation of previously described music objects.

A particular structural object is *Theme* which represent exactly the concept of musical theme of the particular piece (or part of it) under consideration. Theme objects may be whether the output of an automatic segmentation process or the result of a musicological analysis.

Therefore a

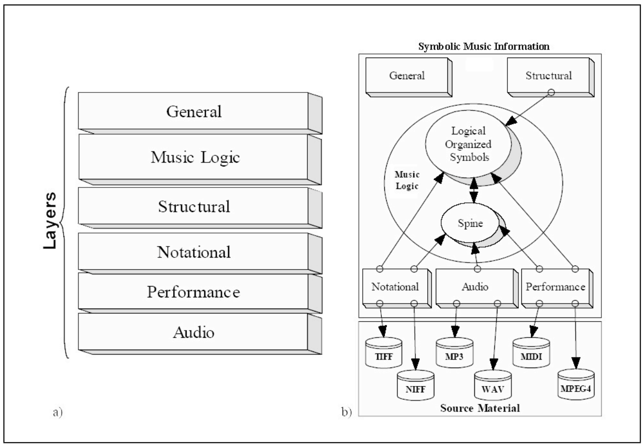


Figure 1. (a) Music information layers and (b) relations among them

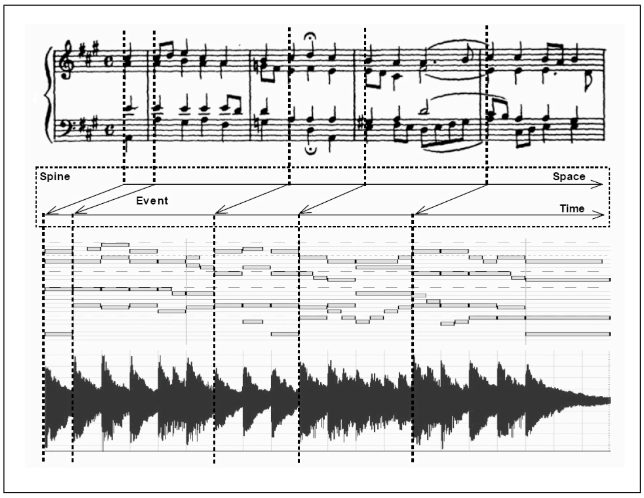


Figure 2. Spine: relationships between Notational, Performance and Audio layer

content-based information retrieval method cannot leave aside the possibility of explot the presence of new specific metadata specifically oriented to retrieval in order to improve processes and algorithms.

We'll return to this subject in the section 5 after the introduction of the graph model of thematic fragments (TFs) for music information retrieval processes.

1.2. Music information retrieval

The increase of files size and number makes the teatment of musical themes like pure sequences of bytes quite impossible (cfr. [5] and [6]). On the contrary, it is necessary to consider the peculiar characteristic of the musical context, in other words we are in need of a *model*.

We are first going to analyse two approaches known in literature: Lerdahl & Jackendoff's grammars and Tenney & Polansky's metrics. Then we will propose a new approach based on graph theory. This allows a more precise and fast retrieval and facilitates the recognition of some particular transformation a musical theme could have undergone, like traspositions or inversions.

2. RELATED WORK

XML is an effective way to describe music information. Nowadays, there is a number of good dialects to encode music by means of XML, such as MusicXML, MusiXML, MusiCat, MEI, MDL (cfr. [3] for a thorough discussion). In particular, we have at least two good reasons to mention MusicXML. MusicXML is a comprehensive way to represent symbolic information. As a consequence, MusicXML was integrated in a number of commercial programs. Among them, it's worthwhile to cite one of the leading applications for music notation: Coda Music Finale. One of the key advantages of MusicXML over other XML-based formats is represented by its popularity in the field of music software. However, all the encoding formats we listed before are not interested in semantic descriptions of metadata. In MPEG-7 context, currently there are initiatives to integrate OWL ontologies in a framework opportunely developed for the support of ontology-based semantic indexing and retrieval of audiovisual content. This initiative follows the Semantic Level of MPEG-7 MDS (Multimedia Description Schemas), and TV-Anytime standard specifications for metadata descriptions. Despite of MX Semantic Layer, MPEG-7 Semantic Level describes music information from the real world perspective, giving the emphasis on Events, Objects, Concepts, Places, Time in narrative worlds and Abstraction. Therefore, MPEG-7 ontology is only aimed at the description of music performance and not of score information, as in the case of MX. You can find a complete discussion of these topics in [4].

Genre, an intrinsic property of music, is probably one of the most important descriptor used to classify music archives. Traditionally, genre classification has been performed manually but many automatical approaches are provided by the state of art. In [2], three different categories of genre classification are proposed: i) manual approch based on human knowledgment and culture; ii) automatic approch based on automatic extraction of audio features; iii) automatic approch based on objective simliraty measures. Taxonomy use is the main difference beetwen the two automatic approaches: in the first a given taxonomy is necessary, in the second is not required. In MX, we have tried to classify genres by an OWL ontology, in order to get a taxonomy as flexible as possible and capturing the complexity of real world genre classifications.

Although important in music information retrieval, genre classification is not the main topics of this paper. Here we are interested in the more specific retrieval of musical themes independently from their possible classification in one or another genre.

However the treatment of musical themes cannot leave aside semantic considerations [5] [6] about the particular context in which we operate. Themes aren't pure sequences of bytes or symbols; on the contrary, they represent relationships among scale degrees.

Most music information retrieval methods adopt a functional approach, that is to say they treats musical themes like functions and therfore they base similarity function on metrics in functional spaces. Unfortunatly, doing so we loose the underlying relalationships between consecutive notes, wich could be very important from a musicological point of view to find affine similarities between themes [15]. Functional models bases more or less on metrics like Tenney & Polansky's [9] [8]. They founded on the concept of *morphology*, substantially a finite sequence of comparable elements.

In this context melodies are finite sequences, i.e. functions defined on a finite subset of $\mathbf{N}$ and whose values are in mono ($\mathbf{Q}$) or n-dimensional ($\mathbf{Q}^n$) spaces, according to the characterizing parameters. Thus, distance concepts are those erhedited by metrics defined on (discrete) functional spaces.

The measure of those distances involves the creation, the recognition and the analysis of variations and trasformation of morphological parameters like pitches, onsets, harmonic relations, sequences of timbre related values and, more generally, any kind of observable related to melody.

As pointed out by Tenney [8], we have to distinguish between statistical and morphological properties. Generally, statistical are *global* and *time independent*, like the mean value or standard deviation of a parameter, while morphological characteristics are described by the *'profile'* of parameters and depend on elements ordering. It is also possible to use statistic measures of parametric profile as parameters at a higher level (hierarchy of profiles); in this way we can analyse the melodic profile at different levels by the application of different metrics.

Definition 1 *A morphology is an ordered set of elements belonging to the same ordered set M. The elements in M are identified by M_i , $i = 1, \cdots, L$; $L = |M|$.*

Some examples of morphologies are pitch sequences, rhythm sequences, harmonic sequences, etc. (cfr. [8] [10]).

Morphological metrics are metrics on *morphologies* (metrics on ordered sets).

Definition 2 *Given a set S, a function*

$$d : S \times S \longrightarrow \mathbf{R} \tag{1}$$

is a distance function *or a* metric *if* $\forall a, b, c \in S$ *holds:*

1. $d(a, b) \geq 0$
2. $d(a, b) = 0$ *iff* $a = b$
3. $d(a, b) = d(b, a)$
4. $d(a, b) \leq d(a, c) + d(c, b)$.

(S, d) *is called a* metric space.

Metrics on spaces of real valued functions are useful models for morphological metrics. For example, given two continuous real valued functions $f(t)$ e $g(t)$, defined on $[m, n]$, there are two intuitive amplitude metrics:

$$d(f, g) = \sup\{|f(t) - g(t)|\} \qquad \text{(sup)} \tag{2}$$

and

$$d(f, g) = \int_n^m |f(t) - g(t)| dt \qquad \text{(amplitude)} \tag{3}$$

Working in discrete spaces, the integrals will be replaced by sums.

By replacing $f(t)$ and $g(t)$ with their derivatives of any order, Tenney and Polansky obtained metrics by measuring the mean amplitude difference of the corresponding rate of changement of the two functions. For discrete functions the derivative is substituted by the *difference function* of first (second, third, $\cdots$) order. Given two morphologies M, N, of lenght L, the amplitude metrics is:

$$\sum_{i=1}^{L} |M_i - N_i| \tag{4}$$

or, normalized,

$$\sum_{i=1}^{L} \frac{|M_i - N_i|}{L} \tag{5}$$

Generalizing the concept, it is possible to make an average of the higher order derivatives, like in the Sobolev metric:

$$d(f, g) = \sum_{i=0}^{n} \left[\sqrt{\int (f(t)^i - g(t)^i)^2} \right] \tag{6}$$

where i is the order of derivative. An analogous metric is the L_1-version of the previous metric, normalized respect to the number of derivatives:

$$d(f,g) = \sum_{i=0}^{n} \frac{\left[\int |f(t)^i - g(t)^i|\right]}{n} \qquad (7)$$

then, weighting difference functions, we obtain the (linear ordered) metric:

$$d(M,N) = \frac{1}{\sum \alpha(i)} \sum_{i=0}^{n} \alpha(i) \cdot \frac{\sum_{j=1}^{L-i} |M_j^i - N_j^i|}{L-i} \qquad (8)$$

where i is the order of difference function on M and N. Obviously the lenght decreases step by step.

Starting from $i = 1$ implicates the exclusion of the elements of M and N, so melody tranposition will have zero distance.

$\alpha(i)$ represents a weight function indexed by the order of difference function $n < L - 1$, where L is the lenght of M and N. The weight function extabilishes the importance of each order of derivation.

Equation 8 inspired our similarity function by a suitable replacement of derivatives with graph powers. As we'll show in this process we'll loose simmetry, so the resulting similarity function will not be a metric. We'll describe it later in detail after the introduction of some graph theory concepts.

For the sake of completeness, we must cite before the various reductionistic approaches utilized to mitigate the obvious mismatches typical of this approach. This methods lead to reduce musical information into some "primitive types" and then to compare the reduced fragments with functional metrics.

A very interesting reductionistic approach refers to Fred Lerdahl e Ray Jackendoff's studies. Lerdahl and Jakendoff [11] published their researches oriented towards a *formal description of the musical intuitions of a listener who is experienced in a musical idiom*. Their work wasn't directly related to MIR, their purpose was the development of a formal grammar which could be used to analyze any tonal composition. However, in case of tematic fragments, it would be possible to reduce the TFs into "primitive types", showing formal similarities according to the defined grammar.

The aim is to describe, in a simplified manner, the *analitic system of the listener*, i.e. rules which allow the listener to segmentates and organizes a hierarchy of musical events. On this basis, *score reductions* are applied, gradually deleting the less significant events. In this way we can obtain a simplification of the score and in the meantime we can preserve sufficient information which allow to mantain recognizability.

The study of these mechanisms allows the construction of a *formal grammar* able to describe the fundamental rules followed by human mind in the recognition of the underlying structures of a musical piece. Unfortunatly the grammar construction is very complex and is not free from ambiguities.

3. BACKGROUND

In this section we'll give some graph-theoretic definitions and results which will be used later. We'll suppose the reader knows the elementary notions of graph theory; for reference texts cfr. for example [12] and [13].

In this paper we'll consider quite only eulerian connected oriented multigraph defined on finite sets. To fix the notation, now we give some definition.

Definition 3 *A graph G is* included *in a graph G' if $V_G \subseteq V_{G'}$ and exists a partition $\mathcal{P}(A_{G'})$ of the arrows set of G', $A_{G'}$, in trails such that the diagram*

$$\begin{array}{ccc} A_G & \xrightarrow{\ i_A\ } & \mathcal{P}(A_{G'}) \\ \downarrow{\scriptstyle \partial_0, \partial_1} & & \downarrow{\scriptstyle \partial'_0, \partial'_1} \\ V_G & \xrightarrow[\ i_V\ ]{} & V_{G'} \end{array} \qquad (9)$$

commute; where $i = (i_A, i_V)$ is the usual graph inclusion.

Remark 4 *If the partition is* the finest *one, i.e. all classes are singleton, the definition collapse to the standard inclusion.*

Remark 5 *The inclusion defined above works with labelled graphs. It may be useful enlarging this concept to the situation where V_G is 'quite' included in $V_{G'}$. In fact in musical graphs is not relevant to preserve the labels because we study objects invariant by musical tranformations (i.e. permutation of labels preserving the metric structure defined on V_G).*

So we will give the next definition.

Definition 6 *A graph G is* weakly included *in a graph H iff exists a subgraph G' of H isomorphic to G.*

$$G \cong G' \subseteq H \qquad (10)$$

3.1. Graph complexity

Definition 7 *The* complexity $k(G)$ *of a graph G is the number of equally oriented spanning trees of G.*

The following three propositions evidence the importance of graph complexity in our model. Their importance will be more clear in the next section.

Proposition 8 *Let H be the graph obtained from a graph G substituing an arrow $a : v_i \to v_j$ with a couple of arrows $a_1 : v_i \to v_k$ and $a_2 : v_k \to v_j$, with $v_i, v_j, v_k \in V_G$. Then*

$$k(H) \geq k(G). \tag{11}$$

Proof. Let T be a spanning tree which contains the arrow $a_1 : v_i \to v_k$. If we replace a with the trail a_1a_2, the result from T with the substitution of a by a_1 or by a_2 is yet a spanning tree; so the complexity of G can only increase. ■

Proposition 9 *Let H be the graph obtained from a graph G of order $n(G)$ adding a vertex v_{n+1} to V_G and replacing an arrow $a : v_i \to v_j$ with the couple of arrows $a_1 : v_i \to v_{n+1}$ and $a_2 : v_{n+1} \to v_j$. Then*

$$k(H) \geq k(G). \tag{12}$$

Proof. We can divide the spanning trees of G in two classes: X={spanning trees containing a} and Y={spanning trees which do not contain a}. Obviously we have $k(G) = |X| + |Y|$. If $\alpha \in Y$ then $\alpha \vee a_2$ is a spanning tree of H. Thus the complexity of H is at least $|Y|$. Let's consider a tree $\beta \in X$. Replacing a by the trail a_1a_2 we obtain yet spanning tree. Finally we have $k(H) \geq |Y| + |X| = k(G)$. ■

Proposition 10 *Graph complexity is a monotonic function respect of the order relation of graph inclusion previously defined.*

Proof. Let G be strongly contained in H. Then, if $V_G \subset V_H$, the corollary 8 implies that $k(G) \leq k(H)$ and if $V_G \leq V_H$, the corollary 9 implies $k(G) = k(H)$. Now let G be included weakly in H. The invariance of complexity under isomorphism and the previous case proves the theorem. ■

4. THE GRAPH MODEL

Now we are going to describe how to build a graph model of a TF. We consider a structured set of TF (a *database*) and a TF (the *query*) which has to be compared with every set-element. This is how to proceed:

1. build a representative graph for every TF
2. work with graphs instead of TF.

In this way, we can recognise a greater number of relevant musical similarities and we can reduce the number of TF wich should be compared. Moreover, TFs of the archive which have the same representative graph can be identified yet.

Let M be a TF of length $m = |M|$ and consider three characterizing sequences of observable: pitches $\{h_s\}_{s\in\mathbf{I}}$, lengths $\{d_s\}_{s\in\mathbf{I}}$ and accents $\{b_s\}_{s\in\mathbf{I}}$, $\{I = 1, ..., m\}$. Then let (V, d) be a metric space on a finite set of element (V). V and d depends upon the musical system we are considering.

Now, let's consider the linear graph G_l obtained by associating a vertex labelled by h_s to every element $h_s \in V$ and an oriented arrow $a_s : h_s \to h_{s+1}$ to every couple (h_s, h_{s+1}), so that $\partial_0 a = h_s$ and $\partial_1 a = h_{s+1}$.

We can then define a weight function $p : A_{G_l} \to \mathbf{Q}\times\mathbf{Q}$ by:

$$p : a_s \to (d_s, b_s) \in \mathbf{Q}^+ \times \mathbf{Q}^+ \ \ \forall s = 1, ..., m-1 \tag{13}$$

where $a_m : h_m \to h_1$ and $p(a_m) = (d_1, b_1)$.

Then we quotient the vertex set by identifying the vertex with the same label.

Definition 11 *Let M be a TF, we call* musical graph representing M *(and we write $G(M)$) the graph obtained by the process described above.*

Now we are going to analyze the properties of the graph G(M) representing a TF M.

Proposition 12 *Musical graphs are eulerian, connected, oriented multigraphs.*

Proof. The proof is trivial if one considers the construction described above. In fact we have sent every interval of the original TF in an equally oriented arrow. The melody is a sequence of intervals, so it is clear that such a sequence represents an oriented trail in the graph which uses every edge once and once only. The closure of the trail is imposed by the definition, because we suppose the last interval being the last note-first note one. ■

4.0.1. Hamiltonian graphs

Serial music is an important part of contemporary music and also in diatonic and tonal contexts.

So an important class of TFs are those which correspond to series: they have an interesting interpretation in our model. In fact, series correspond to *hamiltonian graphs*.

The next proposition characterizes the TFs which contains a *series*.

Proposition 13 *A TF M contains a series iff its representative graph $G(M)$ is hamiltonian.*

Proof. In fact, given a series M the resulting graph is necessarily hamiltonian, because it contains all the intervals of the series. Viceversa an hamiltonian path in G represents obviously a series M. ■

Example 14 *Let's consider a dodecaphonic series (cfr. [19])*

Its representative graph is clearly C_{12}, which is hamiltonian.

Remark 15 *Cyclic graphs C_n are a trivial example of series.*

4.0.2. Equivalences

One of the reasons why we needed to enlarge the concept of similarity was the recognition of new kinds of transformations; particularly we were interested in the permutation of melodic subfragments. These tranformations are important not only for musicologic research purposes in contemporary music but also for investigations in canonic-imitative music.

Now, let's analyse two concepts of equivalence of TFs that will be central in the model.

4.0.3. Euler-equivalence of TFs

An equivalence concept which comes out from the representative graph construction is the one concerning the different trails in the graph.

Definition 16 *We say that two TFs are* Euler-equivalent *iff they have he same representative graph.*

Let's try to better understand what this could mean from a musical point of view by some propositions.

Proposition 17 *A graph is Eulerian iff it is decomposable in edge-disjoint cycles.*

Musically the proposition means that in an equivalence class there are TFs which admits a common cycle decomposition.

Example 18 *Now, consider the TF (cfr. [20]):*

and let's permute the cycles with the evidenced start and stop points (B,A,B) e (B,A,G,A,B):

The two TFs have exactly the same representative graph.

Therefore when we consider a particular musical graph we are really considering all the TFs corresponding to all the different eulerian circuits of the graph (with fixed starting point)

Musically this means that we *identify TFs obtained by particular permutations of subfragments*. We want to point out that these aren't arbitrary permutations. Otherwise there would be no advantages in respect of Tenney & Polansky's non ordered interval metrics ([8], [9]).

Is it possible to compute exactly the number of euler-equivalent TF by the next result.

Proposition 19 *Every class of euler-equivalent TFs has cardinality given by*

$$c \cdot \prod_{i=1}^{n} (d_i^+ - 1)! \qquad (14)$$

where c is the number of equioriented spanning trees of the same representative graph.

Proof. The proposition follows from the Cayley theorem. ■

4.0.4. Equivalence of TFs

Now we give a more general notion of equivalence, which includes also the standard tranformations of music theory.

Definition 20 *Two TFs are* equivalent *iff their representative graphs are isomorphic.*

Remark 21 *Equivalence implies euler-equivalence.*

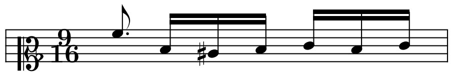
Remark 22 *Standard melodic transformations are included into the isomorphism definition. In fact isomorphism implies an isometry between the metric spaces of vertices; therefore if we consider for example the standard equally tempered metric space (S^1) it is evident that tranformations like* transposition *and l'*inversion *are isometries $i : V \to V$.*

The retrogradation *consists in the inversion of the orientation, so we have just to consider the opposite graph.*

4.1. Subgraphs and inclusions

Besides euler-equivalent TFs there is another interesting class of TF that can be obtained from a given one. The *eulerian subgraphs*. In fact it is possible to choice vertex and arc subsets such that the resulting graph remains eulerian.

Example 23 *Consider the two TFs (cfr. [20])*



Of course the seconf TF is an eulerian subgraph of the first one.

This is a quite particular case, which points out how we can recognise TF inclusions by a simple graph difference.

The necessary condition in this case is that the embellishments must be closed circuits *of the graph.*

Thus we have the following proposition.

Proposition 24 *Let M be a TF and M' another TF obtained from M by the addition of closed embellishment to the notes of M. Then $G(M)$ is an (eulerian) subgraph of $G(M')$.*

Now we will define a notion of inclusion using the "weak inclusion" defined before.

Consider the musical theme (cfr. [20]):

It is clear, musically speaking, that the first TF is a variation of the second one.

Now we will try to formalize this concept by the next definition.

Definition 25 *We say that a TF A is included in a TF B if the representative graph of A is weakly included in the representative graph of B $G(A) \subseteq G(B)$ 6 such that $i : V_G(A) \to V_G(B)$ is an isometry.*

4.2. Necessary conditions

The problem of deciding the inclusion of a TF into another one, using the "large" concept of inclusion, moved from the TFs to their representative graphs.

Teorically it may also be possible to use *string matching* tecniques, but there are at least three facts which lead to exclude those methods.

First of all, we should compare all the TFs undisctintively, and this would be particulary heavy from a computational point of view, expecially real-time.

Another handicap of the sequential approach is the fact that it's rigorously *note-dipendent*, i.e. it depends, in our graphic language, from vertex labelling. The transformation of TFs should be not contempled.

At the end, the natural equivalence class of TFs should be that concerning the isometries between the vertex set only, excluding permutations of sub-TFs.

Large part of our work consisted in searching for graph *invariants monotonic* respect the partial order relation of *inclusion* in the graph set defined in 6. Using graph invariants we could consider a TF with all its transformations by isometries. This process really does not depend on vertex labelling and so we can consider a TF toghether with all its transformed ones.

To fasten the inclusion test however it would be necessary to find out a set of good *necessary conditions*, which can reduct the set of TFs to compare. Now we will explore the condition so far identified.

4.2.1. Order and size

The first invariant we are going to consider are graph order and graph size.

Definition 26 *The* order *and the* size *of a TF M are the order $n(G)$ and the size $m(G)$ of the graph G representing M.*

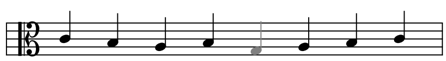
Of course we can observe that:

Remark 27 *If $G \subseteq H$ then*

$$n(G) \leq n(H) \qquad m(G) \leq m(H), \tag{15}$$

i.e. the number of vertex and the number of arrows in G have to be less than or equal to their respective in H.

From a musical point of view the condition concerning the arrows is evident (the fragment $M(H)$ must have more intervals than $M(G)$). Vertex condition is less trivial. In fact it says that the total number of distinct notes must increase or remain stationary in an inclusion, so we could never have a situation like this:



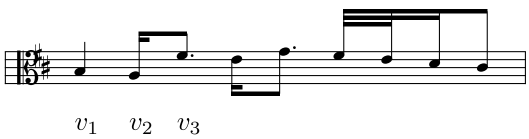
The second fragment includes G, but the first doesn't. These fragments aren't comparable. Moreover our inclusion relation induces a *partial* order on the musical graph set.

4.2.2. The complexity

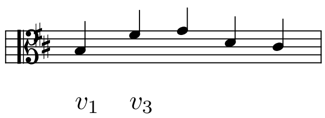
Another significative invariant is the number of spanning trees of the graph representing the TF. At the graph level, an embellishment is obtained by the substitution of an arrow with an equioriented trail of the same total weight, with the same beginning and end vertex.

Consider the simplest variation: the insertion of one note.

Example 28 *In this fragment*



the first note of the alla lombarda *rhythm, corresponding to vertex* v_2 *is the embellishment of the line*



So, in the correspondent graph we have:

- *first fragment: the arrows* $a_1 = (v_1, v_2)$ *and* $a_2 = (v_2, v_3)$
- *second fragment: the arrow* $b_1 = (v_1, v_3)$

where $|b_1| = |a_1| + |a_2|$.

In this example, vertex v_2 *has minimum degree (*$v^+ = v^- = 1$*), i.e. there are no other trails passing trough it, because we wanted to point out the non decreasing property of graph complexity in the substitution of* M_2 *with* M_1.

In fact the second fragment (M_2*) has* $k(G(M_2))$ *spanning tree and if we replace the arrow* b_1 *with the trail* $a_1 a_2$ *such trees continue to span the graph (*M_1*).*

This fact pointed out by the example helds also with added vertex belonging to the first graph. Let's better formalize this fact with the following definition.

Definition 29 *The* complexity $k(M)$ *of a TF* M *is the complexity of its representative graph* $g(M)$ *(cfr. def. 7).*

We would remind that

Remark 30 *The complexity of a graph is equal to the number of spanning trees of the graph*

Proposition 31 *Let* M *be a TF and* M' *a variation of* M *obtained by the insertion of notes. So we have* $k(M) \leq k(M')$.

Proof. At the representative-graph level, the process which transforms M in M' consists of a substitution of oriented arrows by oriented trails. In this way the proposition follows from the 8 and 9. ■

4.2.3. The degree sequence

Of course the inclusion of a TF M in another TF M' implies that the number of corresponding notes can only increase.

Definition 32 *The degree sequence of a TF* M *is the valence sequence of its representative graph* $G(M)$.

Proposition 33 *If a TF* M *is included into another TF* M' *then the respective valence sequences are such that*

$$d_i(M) \leq d_i(M') \quad , \quad \forall i \in I \tag{16}$$

Proof. The proposition follows by the definition of inclusion of TFs. ■

4.3. The metric on V_G

By definition, given a musical graph G, the set V_G is a metric space.

It's really important considering the note set only as a metric space as we are interested in invariant objects of all musical transformations. The lonely really important facts are the distance ratios between the notes.

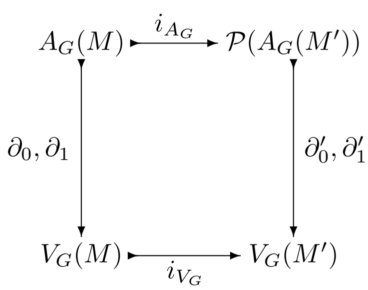
Certainly, the most general possible transformation in our context is an *arbitrary permutation* of nodes; however this may cause an outgoing of the isometry set. Then, even a learned listener should run into serious difficulties in recognising such a transformation. Similarly, if we think of changing the rhythm, the probabilities of recognise the TF should tend to zero.

We would remind at this point the musical transformation that a TF can undergo:

1. transposition;
2. specular inversions;
3. retrogradations.

Thus it is possible to give a necessary condition for the inclusion that will have a great relevance in recognising those transformations:

Proposition 34 *If a TF* M *is contained into another TF* M' *then the inclusion function* $i_{V_G(M)}$



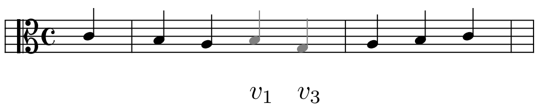
(17)

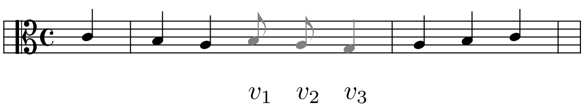
is an isometry.

Proof. The proposition is obvious using the definition of (weak) inclusion and the definition of musical tranformations. ■

4.4. The power graph

Let's consider the TFs (cfr. [20])





They differ by a passing note (closure of the third). At the graph level we can observe that the arrow $a_1 : v_1 \to v_3$, wich is present in the first graph, is replaced by two arrows $a_1 : v_1 \to v_2$ and $a_2 : v_2 \to v_3$ in the second one, so a trail of lenght 2 replaced an arrow.

Let's remember the *transitive closure* operation.

Definition 35 *Let G be a graph. We call* transitive closure *of G the graph $\overline{G}$ such that $V_{\overline{G}} = V_G$ and $A_{\overline{G}}$ contains all and only the arrows such that:*

$$a_i = b_i b_j \quad \wedge \quad \partial_1 b_i = \partial_0 b_j \ , \ b_i, b_j \in A_G \tag{18}$$

Remark 36 *The transitive closure is an internal unary operation on the set of eulerian graphs; in fact we can observe that:*

Proposition 37 *If G is an eulerian graph of size m, then $\overline{G}$ is also eulerian of size $2m$.*

Proof. Given an eulerian cycle in G, the arrows of $\overline{G}$ are the arrows of G plus a number of arrows equal to the number of couple of adjacent arrows of G. Thus is obvious that $|\overline{G}| = 2m$. By construction, the graph resulting from a closure operation on an eulerian graph is obviously eulerian. ■

From a musical point of view the operation consists of an union of two TFs: $M(G)$ and the TF obtained from $M(G)$ taking all the notes of $M(G)$ but proceeding by jumps.

Iterating the process we can obtain the *k-th powers* of a graph.

Definition 38 *We call k-th power M^k of a TF $M(G)$ the TF whose representative graph is obtained from the graph $G(M)$ iterating (k-1) times the transitive closure operation.*

The interesting result which comes out from those definitions is the next proposition.

Proposition 39 *Let M and M' be two TFs. M' contains a variation of M by adding single notes if and only if exists an isometry $i : V_{G(M)} \hookrightarrow V_{H(M')}$ such that the equation*

$$G \setminus H \setminus (H \setminus G)^2 = \emptyset \tag{19}$$

is satisfied.

Proof. If M' contains a variation of M which is obtained by adding no more than one note between two consecutive notes of M so we will therefore have a graph inclusion $i : G(M) \hookrightarrow H(M')$ such that the set A_H will be partitionable in classes formen by trails η such that:

$$\eta = i_A(a) = h, \ h \in A_H \tag{20}$$

or

$$\eta = i_A(a) = h_1 h_2 \ , \tag{21}$$

$$with\ \partial_1 h_1 = \partial_0 h_2 \ \ h_i \in (A_H \setminus A_{i(G)})$$

In the first case we have $\eta \in A_H \cap A_{i(G)}$ though in the second case $\eta \in A_{(H \setminus i(G))^2}$. So the 19 is satisfied.

Vice versa, if exists an isometry $i : V_{G(M)} \hookrightarrow V_{H(M')}$ such that it satisfies the 19, we'll have $G \setminus H \subseteq (H \setminus G)^2$; hence $\forall a \in i(A_G)$ will be:

$$a \in A_H \cap A_{i(G)} \quad \vee \quad a = h_1 h_2 \ , \tag{22}$$

$$with\ \partial_1 h_1 = \partial_0 h_2 \ \ h_i \in (A_H \setminus A_{i(G)})$$

Hence we can partition A_H following the definition. ■

Finally we've obtained an operative sufficient and necessary condition for the inclusion which can be usefully implemented into an algorithm for the recognition of the inclusions.

4.5. Similarity function

Now let's give the notion of similarity function between graphs which will be useful to estimate the similarity between two TFs.

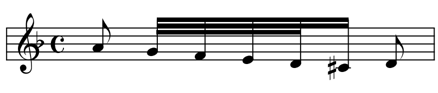
Definition 40 *Let M and M', $M \leq M'$, be two TFs with representative graphs $G = G(M)$ and $H = H(M)$. Then, given $r \in \mathbf{N}$, we call* r similarity function ordine r between M and M' *the function:*

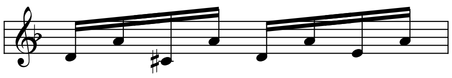
$$\sigma(M, M') = \sigma(G, H) =$$
$$= \max_{\phi} \sum_{i=1}^{r} \alpha_i \frac{|G| - |G \setminus H^i \setminus H^{i-1}|}{|G|} \tag{23}$$

where $H^0 = \emptyset$, α_i are positive coefficients which depend upon the weight assigned to the different trail lenghts and ϕ varies among all the possible isometries from V_G to V_H.

Remark 41 *The function $1/\sigma$ isn't a metric on the set of TFs, because $\sigma(M, M') \neq \sigma(M', M)$.*

Example 42 *Let's consider the two fragments (cfr. [20]):*





The first TF (A) is clearly contained in the second (B) and the function which realizes the max is the identity function. Let's calculate the similarity function with $r = 1$ and $\alpha_1 = \alpha_2 = 1$. We have:

$$\sigma(A, B) = \frac{1}{7} + \frac{6}{7} = 1 \qquad (24)$$

5. MODEL IMPLEMENTATION IN MX

5.1. MX layers

XML organizes information in a hierarchic structure, so MX represents each layer as a secondary branch of the source element. We are not interested here in describing all MX layers, for a complete treatment of the subject, see for example [1]. We'll limit our description to the *structural* layer, after a short overview of the general layer.

```
<!ELEMENT mx (general, structural?,
              logic, notational?,
              performance?, audio?)
>
```

5.1.1. General

In this layer musical information is described as a whole. It provides a general description of a musical piece and contains information about possible connected instances. This is its definition:

```
<!ELEMENT general (description, casting?,
          related_files?, analog_media?,
          notes?, rights?)
>
<!ELEMENT description (work_title?,
          work_number?, movement_title,
          movement_number?, genre?,
          author*)
>
<!ELEMENT genre (genre_spec+)>
<!ELEMENT genre_spec EMPTY>
<!ATTLIST genre_spec
          name CDATA #REQUIRED
          description CDATA #IMPLIED
          weight CDATA #IMPLIED
>
<!ATTLIST author
type CDATA #IMPLIED
>
<!ELEMENT author (#PCDATA)>
<!ELEMENT work_title (#PCDATA)>
<!ELEMENT work_number (#PCDATA)>
<!ELEMENT movement_title (#PCDATA)>
<!ELEMENT movement_number (#PCDATA)>
<!-- description of the music event (genre,
                   date, place); -->
<!ELEMENT casting EMPTY>
<!-- casting information;-->
<!ELEMENT related_files (related_file+)>
<!-- the table of related music data files,
referring to all layers, with one or more
files for the summarization of each layer;
-->
<!-- the table of related multimedia data
files, such as images, videos, and the like;
-->
<!ATTLIST related_file
file_name CDATA #REQUIRED
file_format \%formats; #REQUIRED
encoding_format \%formats; #REQUIRED
file_size_byte CDATA #IMPLIED
>
<!ELEMENT related_file EMPTY>
<!ELEMENT analog_media EMPTY>
<!-- the table of related analog media;-->
<!-- technical  information about related
import/export/restoring/cataloguing/
other operations;-->
<!ELEMENT notes EMPTY>
<!-- general notes.-->
```

5.1.2. Structural

The *structural level* of information was developed to contain the explicit description of musical objects and their causal relationships, both from musicological and compositional points of view. That is to say musical objects can be described as transformations of previously described musical objects. Those objects are usually the results of segmentation processes made by diferent musicologists toghether with their own different musical points of wiew, or also by an automatic score segmenter like *Scoresegmenter* (cfr [7]).

This is a kind of description that shows causality connections instead of temporal connections. The information contained in this layer doesn't refer to temporal ordering and absolute time instances; on the contrary it describes the causal relationships by transformations and displacements of musical object in the score, as they come out from the analysis/synthesis framework.

At the moment there isn't a definitive standard for this layer and our efforts are expecially directed towards the definition of a common acceptable standard. In this framework we belive that the introduction in this MX layer of MIR-oriented metadata, for example attributes of `Theme` could score an important goal in MIR problem. The main topic of this article deals with the introduction of melodic invariant quantities. Infact we belive that the graph model of musical fragments would be very useful in MIR processes because it provides necessary conditions for the inclusion of thematic fragments. It would be certainly possible to introduce more attributes corresponding to different formal approaches, like statistical ones, not considered by us.
Now we list the part of DTD related to the structural layer, which implements also the concept of melodic theme.

```
<!-- Structural Layer -->
```

```
<!ELEMENT structural (analysis*, PN*)>

<!-- Melodic themes -->
<!ELEMENT analysis (theme*, segment*,
                    transformation*,
                    relationship*)
>
<!ATTLIST theme
          id ID #REQUIRED
          ordinal CDATA #IMPLIED
          desc CDATA #IMPLIED
>
<!-- Desc attribute provides a textual
description of the theme.
Ordinal attribute describes the possible
numeric characterization of the theme.
It should be encoded in roman numbers:
I, II, III, IV, V, etc (e.g. I and II
themes in a Sonata.) -->

<!ELEMENT theme (occurrence+)>
<!ELEMENT occurrence (thm_desc?,
          thm_spine_ref+,
          (transposition | inversion
          | retrogradation)*)
>
<!ATTLIST occurrence
          id ID #REQUIRED
>

<!ELEMENT thm_spine_ref EMPTY>
<!-- This element is needed for
generalization of theme representation,
since there could be themes split in
different sequences of notes belonging
to the same part or even to different
parts.-->
<!ATTLIST thm_spine_ref
          spine_start_ref IDREF #REQUIRED
          spine_end_ref IDREF #REQUIRED
          part_ref IDREF #REQUIRED
          voice_ref IDREF #REQUIRED
>
<!ELEMENT thm_desc (#PCDATA)>
<!ELEMENT transposition EMPTY>
<!-- Interval is an integer number,
indicates the interval of transposition
and its iterpretation is related to the
type attribute. When type is real interval
indicates the distance in semitones,
otherwise it indicates distance
in tonal scale.
-->
<!ATTLIST transposition
          type (real | tonal) #REQUIRED
          interval CDATA #REQUIRED
>
<!ELEMENT inversion EMPTY>
<!-- Staffstep attribute must have the
same interpretation as the staff_step
attribute of noteheads. -->
<!ATTLIST inversion
          type (real | tonal) #REQUIRED
          staff_step CDATA #REQUIRED
>
<!ELEMENT retrogradation EMPTY>

<-- Invariants are quantities that
refer to a theme and do not vary
even if the theme is transformed
by canonical transformations -->

<!ELEMENT theme (invariants?)>
<!ELEMENT invariants (order, size, complexity)>
<!ELEMENT order (#PCDATA)>
<!ELEMENT size (#PCDATA)>
<!ELEMENT complexity (#PCDATA)>

<!-- Petri Nets -->
<!ELEMENT PN EMPTY>
<!ATTLIST PN
          file_name CDATA #REQUIRED
>
<!-- Segments -->
<!ATTLIST relationship
          id ID #REQUIRED
          segmentAref IDREF #REQUIRED
          segmentBref IDREF #REQUIRED
          transformationref IDREF #REQUIRED
>
<!ELEMENT relationship EMPTY>
<!ATTLIST segment
id ID #REQUIRED
>
<!ELEMENT segment (segment_event+)>
<!ELEMENT transformation EMPTY>
<!ATTLIST transformation
          id ID #REQUIRED
          description CDATA #REQUIRED
          gis CDATA #IMPLIED
>
<!ATTLIST segment_event
          id_ref IDREF #REQUIRED
>
```

5.1.3. Invariant representation

As we described in the analysis the model, there are various necessary conditions for the inclusion of TFs.

The inclusion-monotone invariants play an important role in the retrieval process and are necessary in order to delete any comparison which could surely fail.

Those quantities, which preserve their values even if the fragment changes by the application of isometries, are extremely useful to deal with themes toghether with all their transformations. This is the reason why we propose their embedding into the structural layer of MX.

In the XML formalism, an invariant should have the form of a subelement of a `Theme`: this mainly for visibility and better retrieval reasons.

6. SUMMARY AND CONCLUSION

Invariants reveal themselves as useful tools in MIR processes in order to reduce improductive comparisons, expecially in the exact match case. MX seems to be the ideal framework in which invariants should be embedded because of the presence of a structural layer. This permits to associate to a thematic fragment all its invariant quantities in a very natural way.

In particular, the graph model presented here enlarge the similarity class of tematic fragments in respect of other models known in literature. Moreover, it is clear that the more a TF presents variety in melodic and interval construction the smaller becomes its euler-equivalence class. Viceversa, TFs with repetitions tends to be more similar, increasing the cardinality of their eulerian class, so the graph model results coherent with the common musical intuition.

7. FUTURE WORK

The approach to MIR by invariants can be developed towards the enrichment of the invariant family derived from graphs or from other models. In particular, the graphical approach can be developed towards numerous directions, first of all the increasing of the number and the power of necessary conditions which are musically significant. Efforts have to be made in the direction of the integration of the rhythm and accentual dimensions into the model.

8. ACKNOWLEDGMENTS

The authors wish to acknowledge the partial support of this project by Italian MIUR (FIRB "Web-Minds" project N. RBNE01WEJT_005) and the Italian National Research Council, in the framework of the research program "Methodologies, techniques, and computer tools for the preservation, the structural organization, and the intelligent query of musical audio archives stored on heterogeneous magnetic media", Finalized Project "Cultural Heritage", Subproject 3, Topic 3.2, Subtopic 3.2.2, Target 3.2.1. We also want to acknowledge the members of the MAX WG for their cooperation and interest in our work.

This work has been made possible by the efforts of researchers and graduate students of LIM.

9. REFERENCES

[1] Goffredo Haus, Maurizio Longari. *A Multi-Layered Time-Based Music Description Approach based on XML*. Computer Music Journal, 2005, Spring Issue, MIT Press.

[2] Jean-Julien Aucouturier and Francois Pachet. *Respresenting Musial Genre: A state of the art*. Journal of New Music Research, 32, 83-93, 2003.

[3] Maurizio Longari. *Formal and Software Tools for a Commonly Acceptable Musical Application Using the XML Language*. PhD thesis, Dipartimento di Informatica e Comunicazione, Università degli Studi di Milano, 2003.

[4] Chrisa Tsinaraki, Panagiotis Polydoros and Stavros Christodoulakis *Integration of OWL Ontologies in MPEG-7 and TV-Anytime Compliant Semantic Indexing*. In Proceedings of HDMS - Hellenic Data Management Symposium, 2004.

[5] Walter B. Hewlett and Eleanor Selfridge-Field. *Music Query*. MIT Press, 2005.

[6] Walter B. Hewlett and Eleanor Selfridge-Field. *Melodic Similarity*. MIT Press, 2000.

[7] Goffredo Haus, Giovanni Sametti. *A Score Analysis/Resynthesis Environment of the Intelligent Music Workstation*. Journal of New Music Research, 24,3 (Amsterdam, 1995) Swets & Zeitlinger B.V., 230-246.

[8] James Tenney e Larry Polansky. *Temporal gestalt perception of music: a metric space model*. Journal of Music Theory, 24, 205-241, 1980.

[9] Larry Polansky. *Morphological metrics*. Journal of New Music Research, 25, 289-368, 1996.

[10] Larry Polansky. *More on morphological mutation functions: Recent techniques and developements*. Proceedings of the International Computer Music Conference, 50-60, 1992.

[11] Fred Lerdahl e Ray Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1996.

[12] Frank Harary. *Graph Theory*. Addison-Wesley, 1969

[13] Béla Bollobás. *Modern Graph Theory*. Springer, 1998.

[14] Chris Godsil, Gordon Royle. *Algebraic Graph Theory*. Springer, 2001.

[15] Luigi Verdi. *Organizzazione delle altezze nello spazio temperato*. Ensemble '900, 1998.

[16] M.Baroni, R.Dalmonte, C.Jacobini. *Le regole della musica*. EDT, 1999.

[17] Goffredo Haus. *Elementi di informatica musicale*. Jackson, 1984.

[18] Curtis Roads. *The computer music tutorial*. MIT Press, 1996.

[19] Arnold Schönberg *Harmonielehre*. Leipzig-Wien, 1911.

[20] Johann Sebastian Bach. *Orgelwerke*. ed.Peters, 1967.