An XML-based Synchronization of Audio and Graphical Representations of Music Scores

Adriano Baratè, Luca A. Ludovico
Laboratorio di Informatica Musicale (LIM)
Dipartimento di Informatica e Comunicazione (DICo)
Università degli Studi di Milano
Via Comelico, 39/41 – 20135 Milan - Italy
{barate, ludovico}@dico.unimi.it

Abstract

This paper presents an overview on a future IEEE standard aimed at providing an overall description of music. This format, known as IEEE PAR1599, is based on the XML meta-language. Its purpose is taking into account the heterogeneous multimedia representations of music, such as audio tracks, video clips and graphical instances of score. The key characteristics of the format are the possibility to enjoy such heterogeneous contents in a synchronized way and the possibility to switch from a particular representation to another in real-time.

After a short description of the particular XML language we have adopted, a case study will be presented: an application installed at the exhibition "Celeste Aida" held at Teatro alla Scala in November 2006 - January 2007.

1. Introduction

This paper deals with a music application supporting content-based browsing, synchronization and indexing of images, video and audio. In the following, we will describe a software which implements an interactive and advanced enjoyment of music thanks to the underlying XML-based format.

As regards the latter aspect, at LIM[1] a standard language for symbolic music description is under development. This language, known as MX, is a meta-representation for describing and processing music information within a multilayered environment, in order to achieve integration among structural, score, interpretative, and digital sound levels of representation. Furthermore, the proposed standard should integrate music representation with already defined and accepted common standards.

The development of the MX format follows the guidelines of IEEE PAR1599.[2] This recommended practice deals with applications and representations of Symbolic Music Information using the XML language, as described in [1]. The proposed IEEE PAR1599 comes at a point in time when it is possible to draw from the experience of existing, previous efforts. After standardization process, MX will be applicable to any kind of software dealing with music information, e.g. score editing, OMR, music performance, musical databases, composition, and musicological applications.

This paper constitutes a preview of the MX format, which is currently undergoing the balloting phase of the IEEE standardization process. Besides, this work provides an example about the class of MX-based applications aimed at the integration of audio and video contents in music context.

MX has been treated in detail in many papers already published and cited among the references, so in the following we will describe only the key characteristics of the format necessary to understand the final case study.

The application based on MX technology is an evolution of other software tools implemented for research [2] or demonstrative purposes [3]. Its purpose is playing MX files containing symbolic, audio and video contents in a synchronized way. This application, besides implementing the concept of a comprehensive description of music in all its multimedia forms, has some practical implications as well. First, it supports a direct, intuitive and immediate approach towards music notation and listening, even for untrained people. Besides, it can be applied to any music genre, from any historical period and culture. For these reasons, we think that the technologies here described are aimed not only at academic research and cultural exhibitions, but also at the market of portable devices and multimedia entertainment in general. Finally, we think that such application can provide also a useful didactic instrument.

[1] Laboratorio di Informatica Musicale, Dipartimento di Informatica e Comunicazione, Università degli Studi di Milano.

[2] Project Authorization Request 1599 (PAR1599), the IEEE formal document about the definition of a commonly acceptable musical application using the XML language.

Eight International Workshop on Image Analysis for Multimedia Interactive Services(WIAMIS'07)
0-7695-2818-X/07 $20.00 © 2007 IEEE

2. MX characteristics

The concept of comprehensive description of music probably represents the main purpose of MX format. Specific encoding formats to represent peculiar music features, such as audio tracks or symbolic scores, are already commonly accepted and in use; but those formats are characterized by an intrinsic limitation: they can describe very accurately music data or metadata for score, audio tracks, computer performances of music pieces, but they are not conceived to encode all these aspects together. On the contrary, we are interested in a comprehensive description of music, in order to create a unique application able to load and keep synchronized heterogeneous descriptions of music. A comprehensive analysis of music richness and complexity highlights six different levels of music description: general, logical, structural, notational, performance, and audio layers.

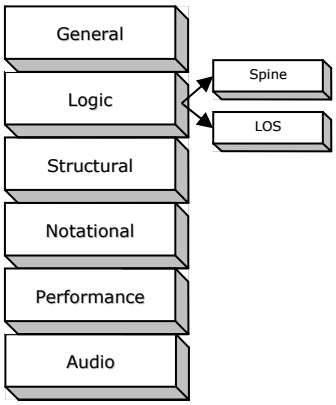

Figure 1. MX multi-layer structure.

MX strives to provide an XML-based description of the aforementioned levels, by implementing mechanisms to represent both synchronization and links towards external descriptions of multimedia contents, encoded in commonly accepted formats. These matters will be described in detail in the following subsections.

The most recent release of MX is version 1.6. The complete Document Type Definition (DTD) of MX 1.6 format, together with a number of complete examples, is available at http://www.lim.dico.unimi.it/mx/mx.zip.

As said before, this comprehensive format has to support complete synchronization among time-based and space-based descriptions of score, meaning that audio, video and graphical contents are kept synchronized with music advancing. From an applicative perspective, this should happen also when the user switches from a performance to another or from a score edition to another.

In order to achieve a comprehensive description of music and complete synchronization among both homogeneous and heterogeneous representations of music contents, MX is based on two key concepts: an XML-based multi-layer structure and a space-time construct called spine. In the following sub-sections, we will define these concepts in detail.

2.1. Multi-layer structure

The first key feature of MX is given by its multi-layer structure, where each layer is devoted to describe a different degree of abstraction in music information (see Figure 1).

The key characteristics that a comprehensive format should support can be summarized as follows:

- Richness in the multimedia descriptions related to the same music piece (e.g., graphical, audio, and video contents);
- Possibility to link to a piece a number of media objects of the same type (e.g., different performances, many score scans coming from different editions, etc.).

General layer is mainly aimed at expressing catalog information about the piece. This layer simply contains some basic alphanumeric information about the music work, including aspects such as digital rights management.

Logic layer contains information referenced by all other layers, and it represents what the composer intended to put in the piece. It is composed of two elements: i) the Spine description, used to mark music events in order to reference them from the other layers and ii) the LOS (Logically Organized Symbols) element, that describes the score from a symbolic point of view (e.g., chords, rests). Structural layer contains explicit descriptions of music objects together with their causal relationships, from both the compositional and the musicological point of view. It represents how music objects can be described as a transformation of previously described music objects. Notational layer links all possible visual instances of a music piece. Here MX references the graphical instances containing images of the score. Performance layer links parameters of notes to be played and parameters of sounds to be created by a computer performance. Finally, Audio layer describes audio information coming from recorded performances.

As regards XML structure, each layer is mapped to a sub-element of the root element. Thus, the general structure of an MX file is the one shown in Figure 2. This approach allows MX to import a number of different formats aimed at music encoding without modifying such commonly accepted encodings. For example, BMP, EPS, and TIFF formats – linked in Notational layer – can be employed to describe score graphic information, whereas common file types such as AAC, MP3, and WAV – linked in Audio layer – can be used to represent audio aspects of music.

0-7695-2818-X/07 $20.00 © 2007 IEEE

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE mx SYSTEM ...>
<mx>
   <general>
      ...
   </general>
   <logic>
      ...
   </logic>
   <structural>
      ...
   </structural>
   <notational>
      ...
   </notational>
   <performance>
      ...
   </performance>
   <audio>
      ...
   </audio>
</mx>
```

Figure 2. General structure of an MX file.

2.2 Spine

Considering music as multi-layered information, we need a means to link and synchronize the heterogeneous facets that compose such information. Accordingly, we introduced the concept of spine, namely a structure that relates time and spatial information. Spine is made of an ordered list of events, where the event definition and granularity are chosen by the author of the encoding. The events listed in spine in general correspond to all the symbols which compose the music piece. Such events are not only listed and put in order by spine, but they are also marked through a unique identifier. These identifiers, conceptually similar to unique key constraints in a database, are referred by all the instances of the corresponding event representations in other layers. Each spine event can be described:

- in 1 to *n* layers; e.g., in Logically Organized Symbols, Performance, and Audio layers;
- in 1 to *n* instances within the same layer; e.g., in three different audio clips mapped in Audio layer;
- in 1 to *n* occurrences within the same instance; e.g., the notes in a song refrain that is performed 4 times (thus the same spine events are mapped 4 times in Audio layer, at different timings).

As we have affirmed, the events listed in the spine structure can correspond to one or many instances in other layers. Figure 3 illustrates such example, applied only to Notational layer. Let a particular event listed in spine, namely event e12, be the 12th note appearing on the trumpet part of a score. By using its identifier, we can investigate the note pitch and rhythmic value, two data described in Logic layer. In this case, we discover that the considered music event corresponds to a E quaver. Now, let us assume that the considered piece has two score versions attached: as a consequence, in Notational layer there will be two points where event e12 is referred. Accordingly, if many audio tracks are present, event e12 will be described in a number of lines belonging to Audio layer.

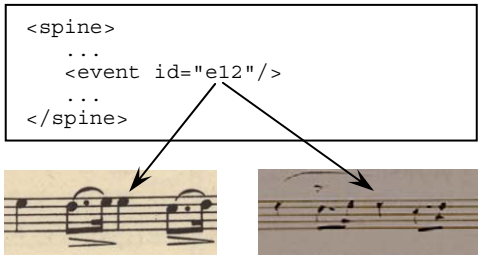

Figure 3. Event e12 graphical mappings.

Thanks to spine, MX is not a simple container for heterogeneous media descriptions related to a unique music piece; rather, those descriptions present a number of references to a common structure, and this aspect originates also a form of synchronization among layers (inter-layer synchronization) and among instances within a layer (intra-layer synchronization), as intuitively shown by vertical dotted lines in Figure 3. Through such a mapping, it is possible to fix a point in a layer instance (e.g. Notational layer) and jump to the corresponding point in another one (e.g. Audio layer). This peculiarity will be used in the application presented in the following to allow an evolved and integrated form of music enjoyment.

3. An MX Case Study

The application we are going to present allows an integrated enjoyment of different media contents related to the same music piece. This case study is based on the "Triumphal March" from Verdi's *Aida* – Act II. The choice of that piece was imposed by the leitmotiv of the exhibition "Celeste Aida", held at Teatro alla Scala for the opening of 2006/07 opera season.

Besides presenting a number of iconographic materials such as fashion plates and sketches, the software offers two versions of the score (an autographical version and a printed one), one video and three audio performances of the aforementioned Aida's piece. Everything is described in a single MX file: catalog metadata and other related material are described in General layer, the logic description of the score is presented in Logic layer, and audio/video contents are linked in Audio layer.

A screenshot of the application is presented in Figure 4. The central part of the interface is dedicated to display the score, as this is the main media in terms of visual extent and interaction. The upper left part of the window contains the controls related to audio/video interaction.

0-7695-2818-X/07 $20.00 © 2007 IEEE

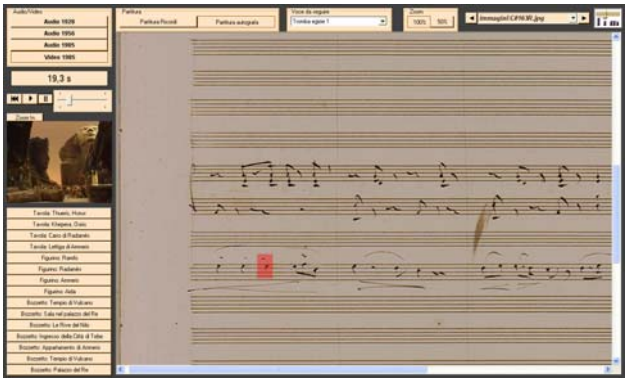

Figure 4. Screenshot of the application.

Three different clips are presented: an audio clip of a 1928 version conducted by C. Sabajno, another audio track of a 1956 version conducted by H. Von Karajan, and both an audio and a video clip of a 1984 version conducted by L. Maazel. All the executions are performed by the Orchestra of Teatro alla Scala of Milan. When one of the aforementioned versions is chosen to be played, the corresponding audio and/or video is executed.

Since there are many parts and voices in the score and they are all coded into the MX file, in the upper central part of the interface there is a list of the instruments to follow. In the central part of the application window, a red rectangle graphically shows the event currently playing in the clip.

The upper right section of the interface houses the controls devoted to managing score visualization. This MX file maps two graphical scores: the autographical version by Verdi and the version published by Ricordi editions. Two buttons are used to switch between the two graphical instances of the score. It must be noted that this switch operation is performed in real time, even when an audio/video clip is running. The red rectangle that indicates the current event is immediately repositioned to the same event in the new selected score. There are also buttons devoted to control the zoom value (50% or 100%) and the current page to display. This happens only in Pause state, when the graphical representation of the score is not synchronized with music; otherwise, the current page is automatically selected to follow the audio/video clip execution.

The left bottom section of the interface is used to open other graphical elements related to the piece, such as fashion plates or sketches. Of course, those objects are not synchronized with other multimedia contents.

This visual interface allows a number of different ways to enjoy music. First, it is possible to select a score version, an audio track, a leading instrument and simply follow the evolution of the instrumental part. This is a first way to explore a music piece, as music can be listened to and watched in a synchronized fashion. A second way to enjoy music through this application is more interesting: it consists in switching from an aural/visual representation to another. In other words, it is possible to compare in real time different versions of the score (the autographical and the printed one) or different performances. When the user decides to switch from a representation to another, the application goes on from the point previously reached. Finally, the software suggests a third way to enjoy music, that consists in altering the original sequence of music events. It is possible to jump – forward or back – from a point of the score to another, both in its visual and in its aural representations; of course, the effect will be the same as the former and the latter are synchronized.

4. Concluding remarks

The application we installed at the exhibition "Celeste Aida", based on MX language, is an interface conceived for the spread of art and music among untrained people. The application is characterized by a comprehensive multimedia description of music, synchronization among heterogeneous contents and easy user interaction. What we have described in this paper represents one of the latest technological advances in music description and interactive multimedia, and we hope that its potentiality will be employed in other projects after the IEEE standardization.

5. Acknowledgments

The authors want to acknowledge researchers and graduate students at LIM, and the members of the IEEE Standards Association Working Group on Music Application of XML (PAR1599) for their cooperation and efforts. Special acknowledgments are due to Denis Baggi and Goffredo Haus for their invaluable work as working group chair and co-chair of the IEEE Standard Association WG on MX (PAR1599).

6. References

[1] G. Haus and M. Longari, "A Multi-Layered, Time-Based Music Description Approach Based on XML", *Computer Music Journal*, vol. 29, no. 1, pp. 70–85, 2005.

[2] A. Baratè, G. Haus, L.A. Ludovico, "MX Navigator: An Application for Advanced Music Fruition", *Proceedings of AXMEDIS 2006 - 2nd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*, pp. 299-305, Leeds, UK, 2006.

[3] D. Baggi, A. Baratè, G. Haus, L.A. Ludovico, "A Computer Tool to Enjoy and Understand Music", *Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT 2005)*, pp. 213-217, London, UK, 2005.