NINA – Navigating and Interacting with Notation and Audio

Denis Baggi*, Adriano Baratè†, Goffredo Haus†, Luca A. Ludovico†
* *Scuola Universitaria Professionale della Svizzera Italiana, SUPSI*
CH-6928 Manno, Switzerland
denis.baggi@supsi.ch
† *Laboratorio di Informatica Musicale (LIM) - Dipartimento di Informatica e Comunicazione*
Università degli Studi di Milano
Via Comelico 39/41 – I-20135 Milan, Italy
{barate, haus, ludovico}@dico.unimi.it

Abstract

This article deals with an XML-based format that allows an integrated representation of music, as well as with a working software demo that demonstrates the power of the format. The format itself is the basis of an ***IEEE international standard*** *for a comprehensive description of music contents and their synchronisation. After a theoretical introduction, a music browser based on such concepts will be reviewed. The application, called NINA (for Navigating and Interacting with Notation and Audio), has been recently presented at an exhibition about Neapolitan music.*

1. Introduction

The problem of music description and representation is complex. Needless to say, music is neither just symbols nor just sounds. Different representations can be provided for a given music piece: different symbolic scores (transcriptions, reductions, variations), various graphic objects (manuscripts, printed versions, full scores vs. parts), and human and computer-based performances (audio tracks, videos). A detailed coverage on this matter is provided in [1] and [2].

In regard of digital representation of music data and metadata, a number of well-known and commonly accepted standards are in use to express the multimedia variety of music contents.

The idea of representing music through symbols is not new, and the problem of music notation goes back several centuries. Proposals for standards for symbolic music have been around for over thirty years in Information Technology. Early computer symbolic notations appeared several decades ago, such as the *Plaine-And-Easie Code* [3] and *DARMS* [4]. Closer to us, attempts have been made to use the technology of the Standard Generalised Markup Language (SGML), a subset of which has been defined for music, namely, the Standard Music Description Language (SMDL) [5]. Though well defined, it failed to attract much attention because of lack of applications. Other XML-based languages have inspired our proposal, such as *MusicXML* and the *Music Encoding Initiative* [6].

Symbolic formats are of fundamental importance for the representation of music, but many other kinds of description can be used. For instance, for graphical representations of music-related contents, many standards are commonly in use, such as. GIF, JPEG, TIFF. Similarly, standards such as AIFF, MP3, WAV exist for audio, and Csound, MIDI, and SASL/SAOL for computer-driven performances. The variety of music-related contents will be also shown by the application of Section 4.

While popular standards for symbolic scores, graphics, audio, video exist, their use for a comprehensive description of music presents some disadvantages, as follows:

1. Data and metadata can not be synchronised across different formats, as they were not conceived for this purpose. For instance, if TIFF is selected to encode the pages of a score and AIFF is used for the corresponding performances, there is no way to keep information synchronised, unless a third format is used as a wrapper. This problem is even more evident when the user also wants to synchronise structural and musicological analyses, computer-driven performances and textual information such as lyrics.

0-7695-3040-0/07 $25.00 © 2007 IEEE
DOI 10.1109/SMAP.2007.41

2. Often music contents are encoded in binary form, unreadable for humans, proprietary and strongly dependent on the current technology.

These are key motivations for developing a brand new format to solve the mentioned problems. The general approach and the characteristics of this format are described in the next section.

2. An Overview of IEEE P1599

The format code-named MX is the core of a new IEEE standard for a multimedia and multimodal representation of music-related contents. It is based on the XML language, thus it is readable, open, extensible and durable. Its purpose is to represent, in a comprehensive way, all aspects of music, namely, the musical experience, the adventure of entering a new world, understanding a narration and recognising images, with the possibility of investigating how the whole is built.

MX is sponsored by the IEEE CS Standards Activity Board (SAB) and supported by the global research fund Intelligent Manufacturing Systems (*http://www.ims.org*). The format represents the realisation of IEEE *Project Authorisation Request* 1599 (PAR 1599) – "Definition of a Commonly Acceptable Musical Application Using the XML Language". The activity of the IEEE Working Group aims to realise a universally accepted standard for the encoding of sound, music, music notation, of other symbolic representation and the like, thus satisfying an acute need for Web and network distribution, CD-ROMs, DVDs, and the like. Moreover, the standard will serve to describe and process different *layers* of music information. P1599 has been proposed to the IEEE Standard Activity Board, approved in September 2001, and the official draft is currently undergoing editing for the balloting process.

In-depth information is provided in MX Web site *http://www.mx.dico.unimi.it*.

3. Description and Synchronisation of Music Contents

The key characteristics that a comprehensive format must support can be summarised as follows:

- *Richness* in the multimedia descriptions of the music piece (e.g. graphical, audio, video contents, and the like)
- Possibility of linking, to a piece, *several media objects* of the same type (e.g. different performances, many score scans from different editions, etc.).

MX views music structures as consisting of different *layers*, each one representing a different aspect of the piece. Figure 1 shows the relationship among the constituents of *Symbolic Music Information*, or SMI, and between the group formed by *general*, *structural*, and *logic* layers on one side, and the one consisting of layers for *notation* (graphics, e.g. TIFF, JPEG, GIF for a score or a diagram), *audio* (e.g. MP3, WAV), and *performance* (e.g. MIDI). This means that the standard selected to represent the multimedia contents of music is irrelevant, since the *same logical symbol file* is used. Depending on the kind of music, not all layers may be present, e.g. a harmonic grid makes no sense for Gregorian chant, while a score has little meaning in a jazz piece coming from a jam session.

A comprehensive analysis of music richness and complexity [7] indicates the need for *six different levels* of music description: layers for *general, logical, structural, notational, performance,* and *audio*.

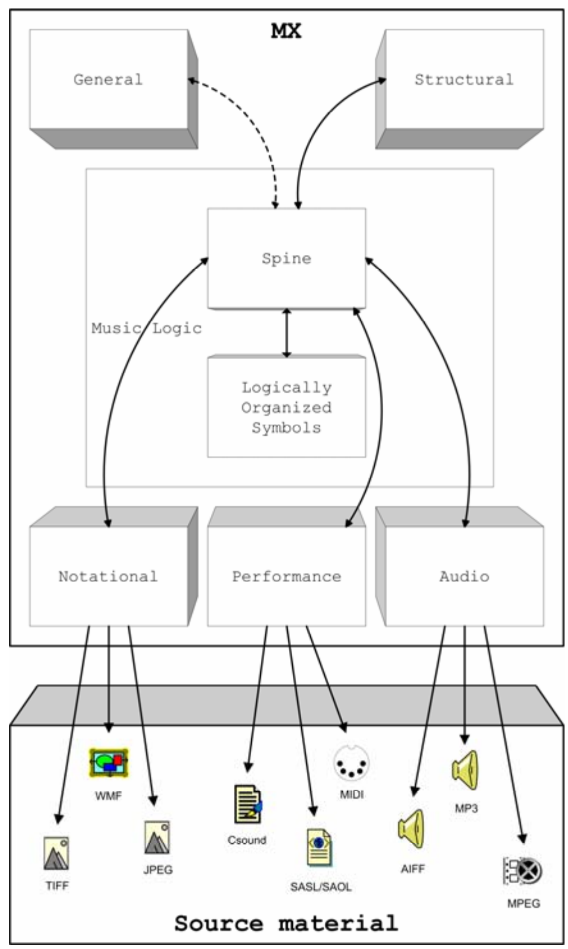

Figure 1. Layers in MX and their relationship.

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

The most recent release of MX is *Release Candidate 1* (RC1). The complete Document Type Definition (DTD) of the format, together with a number of complete examples, is available at *http://www.mx.dico.unimi.it*.

In order to obtain a comprehensive description of music and a complete synchronisation 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*. These concepts are described in detail in the following paragraphs.

3.1. Multi-layer structure

The first key feature of MX is its *multi-layer structure*, in which each layer describes a different degree of abstraction in music information.

The *General layer* contains catalogue information on the piece. It simply lists some basic alphanumeric information about the music work, including catalogue (title, author, etc.) and cast (number and names of performers) information.

The *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: the *Spine*, used to mark music events referenced from other layers, and the *Logically Organised Symbols* or *LOS*, a set that describes the score from a symbolic point of view (e.g., chords and rests), of which Figure 3 shows an example.

The *Structural* layer contains explicit descriptions of music objects, together with their causal relationships, from both the compositional and the musicological point of view. It provides descriptions of music objects and of their manipulations. Thanks to this layer, it is possible to perform different kinds of harmony-based analysis, to identify music themes, to encode subject- countersubject-answer relationships and so on.

The *Notational* layer links all possible visual instances of a music piece. In this layer, MX references the graphical instances containing images of the score, as shown in Figure 3.

The *Performance* layer links parameters of the notes to be played, and parameters of the sounds to be created by a computer performance. In this layer, existing formats such as Csound, MIDI and SASL/SAOL can be referenced.

Finally, the *Audio* layer describes audio information obtained from recorded performances and coded in common audio formats.

As regards to the 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 for music encoding among those that are commonly accepted. For example, BMP, EPS, and TIFF formats – linked in the *Notational* layer – can be used to describe score graphic information, whereas common file types such as AAC, MP3, and WAV – linked in the *Audio* layer – can represent audio aspects of music.

```
<?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.

3.2. Spine

Since in this standard music is viewed as *multi-layered information*, a method is needed to link and synchronise heterogeneous aspects of such information. To this end, the concept of *spine* has been defined, a structure that relates time and spatial information. The spine consists of an ordered *list of events*, for which the event definition and granularity are chosen by the author of the encoding. The events listed in spine correspond, in general, to the symbols which constitute the music piece, and such events are not only listed and ordered in the spine, they are also marked with a unique identifier. Such identifiers, conceptually similar to unique key-constraints in a database, are referred to by all the instances of the corresponding event representations in other layers. Each spine event can be referred to in the following ways:

- in 1 to *n* layers; e.g., in the *Logically Organised Symbols*, *Performance*, and in the *Audio* layers;
- in 1 to *n* instances within the same layer; e.g., in three audio clips mapped in the *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 the *Audio* layer, at different timings).

```
<spine>
 ...
 <event id="e12"/>
 ...
</spine>
<los>
 ...
 <chord event_ref="e12">
  <duration num="1" den="4"/>
  <notehead>
   <pitch step="E" octave="5"/>
  </notehead>
 </chord>
</los>
```

Figure 3. Event e12 description and graphical mappings.

As stated, the events listed in the spine can correspond to one or many instances in other layers. Figure 3 illustrates such an example, applied only to the *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, it is possible to investigate its note pitch and rhythmic value, two data described in the *Logic* layer. In this case, it can be seen that the considered music event corresponds to quarter note E. Now, let us assume that the considered piece has two score versions attached: as a consequence, in the *Notational* layer there will be two entries where event e12 is referred. Accordingly, if many audio tracks are present, event e12 will be described in several lines referring to the *Audio* layer.

4. A Proposal for a New Music Browser

In the previous sections, the problem of music representation has been addressed. In respect to the development of applications, MX presents characteristics that enable highly integrated and synchronised enjoyment of multimedia contents.

NINA, standing for Navigating and Interacting with Notation and Audio, is an application built at the Laboratory for Musical Informatics (LIM) of the State University of Milan, Italy, to illustrate how the standard for symbolic music works, and to show its power. This application is the evolution of a number of earlier software demos and working applications developed by the LIM staff, as documented in [8], [9], and [10]. NINA was designed and implemented for the exhibition “Napoli, nel nobil core della musica” held in May 2007 at *ResidenzGalerie* in Salzburg, Austria. One of the purposes of the exhibition was that of making music tangible and visible, bringing together all five senses, beyond hearing. The music piece chosen for this demonstration is the operatic aria “Il mio ben quando verrà”, from Giovanni Paisiello’s *Nina, o sia la pazza per amore*.

The core of the application is an XML encoding in MX format, which contains logical information about the piece, and synchronisation among the linked multimedia objects. Overall synchronisation is provided among graphic objects representing scores, audio and video clips containing human performances, and with the libretto.

In the context of that exhibition, a rich but simple user interface has been designed, conceived for laymen, to let them listen to a track with various interpreted scores, and look at different score versions simultaneously.

The screenshot of Figure 4 illustrates in general terms the user interface of NINA. Music browsing is based on windows containing different representations of multimedia contents that operate in synchronism while the music is being played.

The user is allowed to use a number of selection windows. In the lower window there are four choices among scores that can be loaded: full autograph core, manuscript copy, printed score for piano, and libretto. In the left window, the user chooses either an audio track or a video clip to listen to. The movie is launched in a dedicated player. The upper window allows selection of the instrumental part to be followed in real time on the score previously chosen.

The main part of the interface contains key graphical contents, namely, the score of the aria in one of its versions.

Figure 4. The interface of NINA.

Upon that selection, several synchronised activities start and execute in real time. The music starts playing, while on the score the running cursor indicates what is being played, here the beginning of the 2nd bar. In addition, the interface make it possible to follow every instrument (as in the examples below), as well as to view a whole vertical line (as in common music notation software). The user can move the red cursor with the mouse and initiate playing from another point in the score, while the other real time windows adjust synchronously, and of course the audio/video player cursor changes its current position accordingly.

As an implementation choice, the user is allowed to follow the evolution of any single voice. Hence, another selection window is provided in the upper part of the interface, where instruments are listed.

During the performance, users can change the instrument to be followed, as well as the audio being played and the score. This is controlled by clicking on any point of the current graphical score (the synchronization is driven by spatial co-ordinates), by dragging the slider of the audio/video player (the synchronization is driven by time co-ordinates), and even by selecting syllables from the libretto (navigation by text contents). Of course, overall synchronization is always immediately reconstructed. Unfortunately, in this article, only the graphical effects of such operation can be shown: Figures 5.a (autograph full score & audio), 5.b (manuscript copy & audio) and 5.c (printed libretto & video) show a practical example.

In conclusion, NINA is a browser that represents music with readable symbols that can be accessed and manipulated even by non-musicians. A significant result is that only one single XML file is needed for several – synchronised – renditions of the same piece, thus proving that the format used for audio and graphical contents is irrelevant. Of course, it is necessary to distinguish between how synchronisation information is attained and how such information can be stored and appropriately used. In this work, only the latter point has been treated, whereas the former has been already covered elsewhere as in [11] and [12].

5. Applicability of the MX Standard

These examples list a number of fields where MX characteristics and potentialities can be fully exploited:

- *An opera.* A DVD of an opera allows the user to: see the play on the screen, hear the music, see the score; read the libretto; choose excerpts of alternative renditions;
- *A piece by a jazz Big Band*. The harmonic grid is displayed and the name of the soloist pops up at the beginning of each solo, and alternate structures can be examined;
- *A fugue.* The theme is highlighted aurally and graphically as it gets passed among the different voices;
- *Music with a "program" or story*. For example, Vivaldi's Four Seasons come with poems that specifically refer to segments of the music;
- *A piece of Indian classical music.* The scale of the raga is shown and the melodic development is highlighted with a diagram, to follow the unfolding story the musician is telling;
- *A piece with several drums*, as in African Drumming. The user sees a graph illustrating playing and accents of five drums, to show that the various hits do not fall exactly on the beat;

Figure 5. Switching score and audio in real time.

- *Preservation of the music heritage from the past.* Symbolic encoding allows storage of documents in any media;
- *Musicological study.* Ease of queries to find music pieces with given characteristics, notes, or other – for example, all pieces utilising the lowest note of a grand piano.

New applications can be created, particularly for different cultures and kinds of music, such as those studied by ethnomusicology, where there has been little investigation and methodological research.

6. Conclusions and Future Works

The natural way to represent music is with *symbols*, a tradition that goes back several centuries in all cultures. It is high time to progress from *closed* binary audio standards to *open* symbolic representations for music. IEEE P1599 is the development of an enabling technology that serves as the basis for the realisation of unending new applications.

While music can be appreciated by just listening to it, the attentive listener senses that there is a whole world beyond the sound, just waiting to be explored. While this world can be discovered thanks to several years of formal, disciplined study, applications of this standard make its discovery available to every listener, thus taking the listener from music enjoyment to music education. This is an example of how technology can allow both users and specialists to discover the deep meaning of music, and generate connections to all aspects of culture, from art enjoyment to contemporary life and awareness.

As regards future works, P1599 is currently undergoing the balloting process by IEEE, and it is hoped that the MX format will soon become an international standard for the representation of music contents. About the integrated player, our efforts are aiming at a generalisation of the software application, in order to open any MX file (classical music, operas, pop, rock, jazz, etc.). Other features to add involve multi-platform support and on-line operability.

7. References

[1] Vinet, H., "The Representation Levels of Music Information", *Computer Music Modeling and Retrieval: International Symposium, CMMR 2003*, Berlin, Springer verlag, 2003, pp. 193-209.

[2] Moore, F. R., *Elements of Computer Music*, Prentice Hall, 1990

[3] Brook, B.S., "The Plaine and Easie Code", *Musicology and the Computer*, City University of New York Press, 1970, pp. 53-56.

[4] Erickson, R.F., "The Darms Project: A Status Report", *Computers and the Humanities* 9(6), 1975, pp. 291-298.

[5] Newcomb, S.R., "Standard Music Description Language complies with hypermedia standard", *IEEE COMPUTER*, July 1991, pp. 76-79.

[6] Roland, P., "The Music Encoding Initiative (MEI)", *Proceeding of the First International IEEE Conference on Musical Application using XML (MAX2002)*, 2002, pp. 55-59.

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

[8] Baratè, A., Haus, G., and Ludovico, L.A., "An XML-Based Format for Advanced Music Fruition", *Proceedings of the Sound and Music Computing Conference 2006 (SMC06)*, Marseille, France, 2006.

[9] Baratè, A., Haus, G., and Ludovico, L.A., "MX Navigator: An Application for Advanced Music Fruition", *Proceedings of the 2nd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS2006)*, Leeds, UK, 2006, pp. 299-305.

[10] Baratè, A., and Ludovico, L.A., "An XML-based Synchronization of Audio and Graphical Representations of Music Scores", *Proceedings of the 8th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2007)*, Santorini, Greece, 2007.

[11] Baratè, A., Haus, G., Ludovico, L.A., and Vercellesi, G., "MXDemo: a Case Study about Audio, Video, and Score Synchronization", *Proceedings of the 1st International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS2005)*, Firenze, Italy, 2005, pp. 45-52.

[12] Baratè, A., Haus, G., and Ludovico, L.A., "Music representation of score, sound, MIDI, structure and metadata all integrated in a single multilayer environment based on XML", *Intelligent Music Information Systems: Tools and Methodologies*, Idea Group Reference, 2007