IEEE 1599: a Multi-layer Approach to Music Description

Luca A. Ludovico
Laboratorio di Informatica Musicale (LIM)
Dipartimento di Informatica e Comunicazione
Università degli Studi di Milano
Via Comelico 39/41 – I-20135 Milano, Italy
luca.ludovico@dico.unimi.it

***Abstract*—IEEE 1599 is a new XML-based format to describe heterogeneous music contents comprehensively. In a single file, music symbols, printed scores, audio tracks, computer-driven performances, catalogue metadata, text and graphic contents related to a single music piece are linked and mutually synchronized within the same framework. Heterogeneous contents are organized in a multilayered structure that supports different encoding formats and a number of digital objects for each layer. The article aims at describing this approach for the description of heterogeneous music-related contents, and providing guidelines to understand key concepts of IEEE 1599 standard.**

***Index Terms*—IEEE 1559, XML, music, multi-layer representation, synchronization, music standards.**

I. Introduction

Music can be described according to many points of view. For example, a performer is usually interested in scores, a musicologist in autograph documents or score analysis, a dee-jay in audio recordings, and a listener in live or recorded performances. When talking about music, one may mean any of these aspects.

Music information is intrinsically heterogeneous, since it involves a number of different media, as well as a number of different descriptions. Consider for instance the case study of an opera house [1] or a music publisher [2], the materials and documents of which can include: scores and symbolic representations of music, audio and video recordings, fliers, playbills, posters, photos, sketches, fashion plates, costumes and related accessories, stage tools, maps and equipment, and other text documents commonly used in the evening's programs, such as bibliography, discography, libretto, short descriptions, and reviews of music works. This list does not claim completeness, but is sufficiently wide to illustrate the heterogeneity of data and metadata belonging to a single work. All documents give a contribution to the overall description of the music piece, and there is considerable interest – from a cultural, scientific and commercial point of view – to provide access to such information, as shown by the growing amount of digitalization projects and music-oriented multimedia databases.

This article illustrates the guidelines of a format explicitly designed to represent and manage music-related information.

II. Representing Heterogeneous Music-Related Information

In the digital domain, there is a class of problems related to computer-based representation of music information. When the original information belongs to the physical world, it must be translated into a binary representation in order to be managed, organized, preserved and made available. This is the typical case of autograph scores, iconographic material, and physical objects. Digitalization campaigns have the purpose of transforming real-world objects in computer files. While music sheets, printed photos, and sketches can be scanned, 3D objects can be acquired from different perspectives thanks to digital cameras. Other materials automatically come in digital format, such as computer-made texts and scores, in which case digitalization is not required.

The problem of digitizing information is not the focus of this writing, even though the results of such processes are relevant for its purposes, since data and metadata must be represented and saved in a suitable format. As stated before, a comprehensive music description capable of dealing with heterogeneous music-related information is desired, and it must include all symbolic documents (i.e. scores and texts from a semantic point of view), still graphics (i.e. iconographic material) as well as audio and video (i.e. performance recordings).

Needless to say, many file formats are available to represent either symbolic or multimedia information. For example, AAC, MP3 and PCM are commonly used to encode audio recordings; Csound, MIDI and SASL/SAOL represent well-known standards for computer-driven performance; GIF, JPEG and TIFF files can be used to represent music scores; DARMS, NIFF and MusicXML are used for score typing and publishing.

Thus, specific encoding formats to represent particular music features are already known and in use, however such formats are characterized by an intrinsic limitation. In fact, while they can describe music data or metadata for score, audio tracks, computer performances of music

pieces, they are not intended to encode all these aspects together. Nevertheless, commonly accepted standards cannot be ignored, even if they describe music from a limited perspective. There are at least two reasons to continue supporting available standards, namely, the characteristics of each format in function of its specific application field, and the availability of huge collections of documents already encoded with such formats.

A unique but comprehensive representation of music is highly desirable, to satisfy the needs of musicologists as well as of performers, of music students as well as of untrained people who are simply interested in music. If a format is available to catch all the different aspects listed above in a single document, the result would be considerably rich in information and capable of targeting a huge audience.

In conclusion, the goal of IEEE 1599 standard consists of providing an overall description of music and its contents. Hence, a new format to represent all music-related data and metadata has been proposed. The key characteristics of this format are:

- *Richness in multimedia descriptions* for the same music piece. Symbolic, logic, graphic, audio, and video contents can be encoded within, or linked by the same document;
- For each type of multimedia description, possibility of *linking a number of digital objects*. For instance, many performances of the same piece, or many score scans from different editions are related to a single file;
- *Full support for synchronization among time-based contents*. In a dedicated player, audio and video contents can be synchronized as the score advances while music is being played, even when switching from a particular performance to another, or from a score edition to another;
- *User-friendly interaction* with the music contents. This format offers all the characteristics to implement software applications that make interaction with music contents possible and easy. In an IEEE 1599 browser, the user can click any region of the score and jump to that point, while the audio does likewise. Hence, it is possible to navigate the audio track while highlighting the related part of the score.

III. IEEE 1599, A NEW STANDARD FOR MUSIC

A. Definitions and Brief History

IEEE 1599 is the official name of MX, an acronym standing for *Musical application using XML* and used in scientific articles and conferences during the IEEE standardization process. The development of the format followed the guidelines of IEEE 1599, *Recommended Practice Dealing With Applications and Representations of Symbolic Music Information Using the XML Language*. This project proposed to represent music symbolically in a comprehensive way, opening up new ways to make both music and music-related information available to musicologists and performers on one hand, and to non-practitioners on the other. Its ultimate goal is to provide a highly integrated representation of music, where score, audio, video, and graphical contents can be appreciated together.

IEEE 1599 is the result of research efforts at the *Laboratorio di Informatica Musicale*, or LIM, of the *Università degli Studi di Milano*. This IEEE standard is sponsored by the Computer Society Standards Activity Board and was launched by the Technical Committee on Computer Generated Music (IEEE CS TC on CGM) [3]. The work has been endorsed by the international research program *Intelligent Manufacturing Systems* and financially supported by the *Commission for Technological Innovation* of the Swiss Federal Government through the *University for Applied Sciences of Southern Switzerland*, SUPSI.

An early phase of the project, around year 2000, consisted of evaluating the features of existing formats for music description, in particular SMDL, enhanced NIFF and XML. The IEEE standardization project began in 2001, and contacts were made with other research groups in the academic and commercial world. In the second phase in 2002, a prototypal version of the format was released, originally known as *Musical Application using XML*, or MAX [4]. This format was discussed at *MAX 2002*, the first international conference on musical application using XML, organised by IEEE CS TC on CGM, analyzed and evaluated with other research groups, and eventually validated by the implementation of software applications. In particular, tools for *music visualization* [5], *content-based retrieval* [6], and *automatic segmentation* [7] were developed.

The IEEE final evaluation process, known as balloting, ended in July 2008 with the result of making IEEE 1599 an international standard.

B. XML and Music

IEEE 1599 is an XML-based format. There are many advantages in choosing XML to describe information in general, and music information in particular, as the large number of other music-oriented XML languages demonstrates. This subject has been treated in many texts and scientific papers, but in this context the interest lies mainly in the reason why XML and music make such a good couple. Most of these topics have been treated in [8].

First, XML is a formal meta-language suitable for declarative representations. It is capable of describing entities – such as music objects – as they are, without unnecessary overload of information. The representation is modular and hierarchical; nevertheless, it allows explicit interaction among modules. Internal relationships are formal and explicit. All these aspects have an important counterpart in music languages, which are also formal and hierarchical.

XML modularity offers advantages for a well-organized, comprehensive description. Each part constitutes a separate entity of the overall description, while still maintaining its own identity. However, interdependency is allowed, and made explicit by the XML-based format.

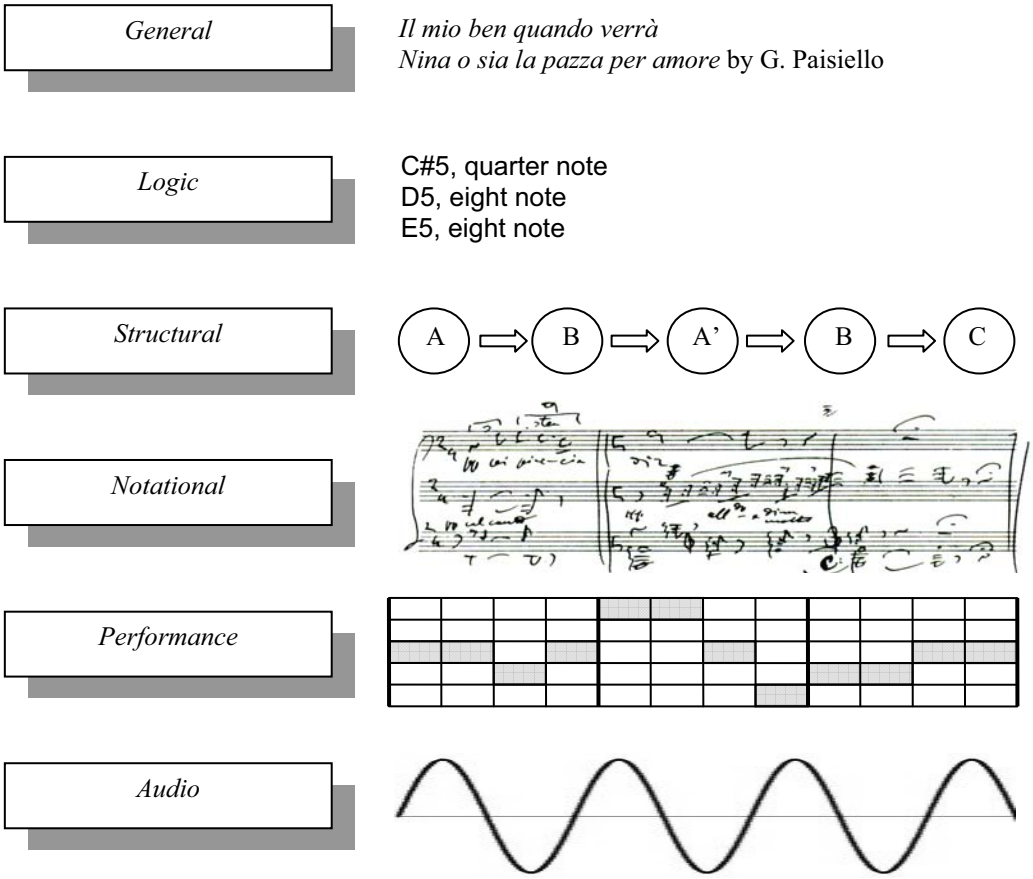

Figure 1. The characteristic multi-layered structure of IEEE 1599. In the right part of the figure, intuitive graphical examples are provided to illustrate the purpose of the layers.

Languages derived from XML are extensible, as they support extensions and external entities. This aspect is fundamental for further developments of music language and unforeseen uses of the description format.

Finally, XML is open to user contributions, easy to read, decode, edit, and understand. In principle, anyone can give suggestions and implement specific parts of the format. Even though it is a lot easier to use than binary formats, XML is not meant to be managed directly, and actually it should be read and written with computer-based systems. For instance, music XML cannot be printed in its text form in order to be played at sight by a human performer, special software is needed to decode the format and represent it as a sequence of music symbols. Fortunately, applications to edit XML files can be easily found on the marketplace, and usually basic editors are freely available. Besides, software to work on a particular XML-based format can be implemented without royalties or licenses, since most formats are free.

In conclusion, an XML-based language allows *inherent readability*, *extensibility* and *durability*. A detailed coverage of this matter is presented in [9]. These are the main reasons why the format proposed here relies on XML to represent and organize music information in a comprehensive way.

As stated above, IEEE 1599 is an XML-based format, thus it inherits all the features of XML. It is open, free, and easy to read by humans and computers, and can be edited by common software applications. Moreover, it is strongly structured, it can be extended to support new notations and new music symbols, and it can thus become a means of interchange for music with software applications and over the Net.

IV. Key Features of IEEE 1599

The advantages of the format and possible applications of a comprehensive description of music have been mentioned above. Section 2 has provided definitions and examples of richness in music communication, and mentioned formats for music. Finally, Section 3 has dealt with the advantages of an XML application to music. In this section, a proposal of an XML-based format for the representation of music in all its aspects is described.

In order to integrate heterogeneous musical information – expressed in any suitable format – within a single description, a new XML-based encoding has been developed. The following discussion will introduce its key features, which are different from other music-oriented XML-based languages such as MusicXML. In addition, this brief overview on IEEE 1599's most interesting characteristics should allow understanding of the advanced applications show by other articles in this publication.

A. Multi-layer Structure

As stated before, a comprehensive description of music must support heterogeneous materials. Thanks to the intrinsic capability of XML to strongly provide structures

for information, such representations can be organized in an effective and efficient way. IEEE 1599 employs *six different layers* to represent information, as explained in [10] and shown in Fig. 1:

- *General* – music-related metadata, i.e. catalog information about the piece
- *Logic* – the logical description of score symbols
- *Structural* – identification of music objects and their mutual relationships
- *Notational* – graphical representations of the score
- *Performance* – computer-based descriptions and executions of music according to performance languages
- *Audio* – digital or digitized recordings of the piece.

Consequently, a generic IEEE 1599 document presents an XML structure similar to the one shown in Fig. 2.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ieee1599 SYSTEM
      "http://www.mx.dico.unimi.it/ieee1599.dtd">
<ieee1599>
   <general>...</general>
   <logic>...</logic>
   <structural>...</structural>
   <notational>...</notational>
   <performance>...</performance>
   <audio>...</audio>
</ieee1599>
```

Figure 2. The XML stub corresponding to the IEEE 1599 multi-layer structure.

Needless to say, not all layers must, or can, be present for a given music piece. Of course, the higher their number, the richer the musical description.

Richness has been mentioned in regard to the number of heterogeneous types of media description, namely symbolic, logic, audio, graphic, etc. But the philosophy of the IEEE 1599 standard allows one extra step, namely, that each layer can contain many digital instances. For example, the *Audio* layer could link to several audio tracks, and the *Structural* layer could provide many different analyses for the same piece. The concept of multi-layered description (i.e. as many different types of descriptions as possible, all correlated and synchronized) together with the concept of multi-instance support (i.e. as many different media objects as possible for each layer) provide rich and flexible means for encoding music in all its aspects.

This approach allows adopting of *ad hoc* encoding to represent information. In fact, while a comprehensive format to represent music is not available, popular existing standards must be taken into account. This is a not a contradiction because of the two-sided approach of IEEE 1599 to music representation, which is: keep intrinsic music descriptions inside of the IEEE 1599 file and media objects outside of the XML document, in their original format. Fig. 3 shows the relationship between the group constituted by the *General*, the *Structural*, and the *Logic* layers and the one including the *Notational* (e.g., GIF, JPEG, TIFF for a score), the *Audio* (e.g., AAC, MP3, WAV), and the *Performance* (e.g., Csound, MIDI, MPEG) layers. Intrinsic music descriptions, typically catalogue metadata and logical representations of music events clearly reside inside the XML file (see the upper block in Fig. 3), whereas media files are extern but they are linked from the corresponding IEEE 1599 layers (see the lower block in Fig. 3).

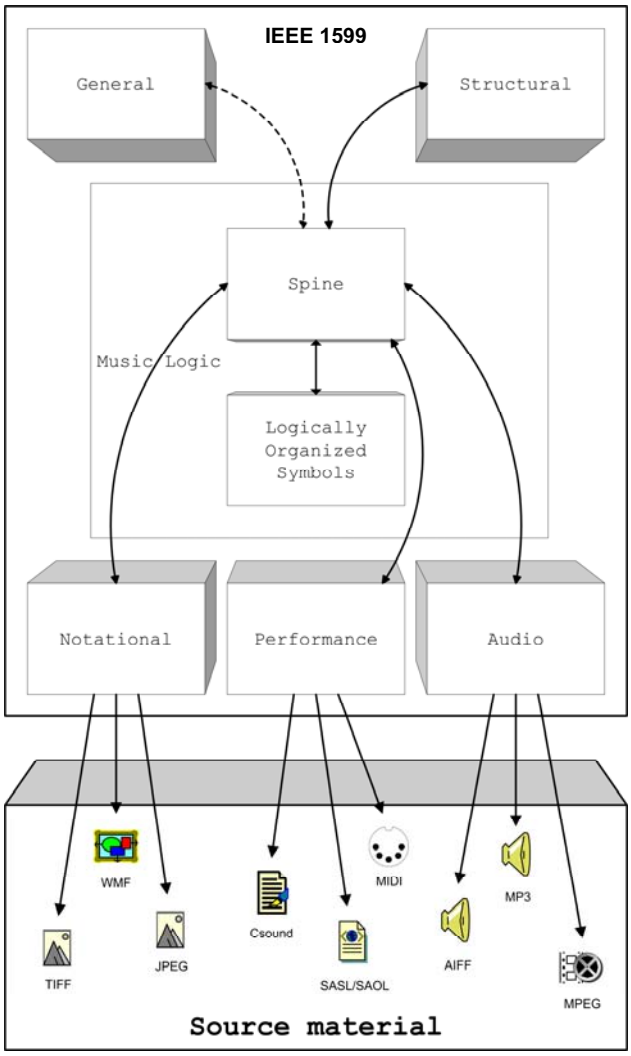

Figure 3. Contents encoded inside the IEEE 1599 document and external media objects.

Consider the following examples. The symbols that belong to the score, such as chords and notes, are described in XML, in the *Logic Layer*. On the contrary, MP3 files and other audio descriptions are not translated into XML format, rather they are linked and mapped inside the corresponding layer, namely the *Audio Layer*.

It should be clear that the description provided by an IEEE 1599 file is flexible and rich, both in regard to the number and to the type of media involved. In fact, thanks to this approach, a single file can contain one or more descriptions of the same music piece in each layer. For example, in the case of an operatic aria, the file could house: the catalogue metadata about the piece, its author(s) and genre; the corresponding portion of the libretto; scans of the original manuscript and of a number of printed scores; several audio files containing different performances; related iconographic contents, such as

sketches, on-stage photographs, and playbills. Thanks to the rich information provided by a single document, software applications based on such a format allow an integrated fruition of music in all its aspects.

B. The Spine

The *spine* is a sort of glue needed in a multi-layer framework. In this approach, heterogeneous descriptions of the same music piece are not simply linked together, but a further level of detail is provided: whenever possible, media information is related to single music events. The concept of music event is left intentionally vague, since the format is flexible and suited to many purposes. A music event can be defined as the occurrence within the score, or its abstraction, of something that is considered important by the author of the encoding. For instance, in a conventional sense, notes and rests of a score can be interpreted as music events. From a more general standpoint, all symbols in a score could be music events, ranging from clefs to articulation signs. Also, for particular purposes, other interpretations are allowed. For instance, in jazz music the event list could correspond to the harmonic grid. In dodecaphonic music, an interesting event could be the occurrence of a series. In segmentation, music events could be the starting and ending points of aggregated music objects. In a particular framework for music analysis, events could refer to the highest and lowest notes within an episode of a given vocal or instrumental part. And so on.

As explained later, the spine constitutes a list of music events that are sorted and labeled in order to allow references from other layers. Against common sense, the score does not correspond necessarily to the list of events referred to by other layers. Because if it were, music works with no notation, such as pieces where the performance is improvised and a true score does not exist, or music for which the score is unknown, could not be supported by IEEE 1599, and the user would be forced to encode the reconstructed score entirely. In this way, a traditional annotated score or a complete encoding of the piece is not required to produce a valid IEEE 1599 document.

Thanks to these premises, the second key concept of the format can be introduced, namely the *spine*. It consists of a sorted list of events, where the definition and granularity of events can be chosen by the author of the encoding.

The spine has a fundamental theoretical importance within the format. It provides both a glue function among layers and an abstraction level, as the events identified in it do not have to correspond to score symbols, or audio samples, or anything else: it is the author who can decide, from time to time, what goes under the definition of music event, according to the needs.

Since the spine simply lists events to provide a unique label for them, the mere presence of an event in the spine has no semantic meaning. As a consequence, what is listed in the spine structure must have a counterpart in some layer, otherwise the event would not be defined and its presence in the list (and in the XML file) would be absolutely useless. For example, in a piece made of *n* music events, the spine would list *n* entries without defining them from any point of view, as shown in Fig. 4. Say that a given event (e.g. `e1`) corresponds to a note: It is contained in many annotated scores and it is played in a number of audio tracks, and its meaning, presence and behavior cannot be understood by analyzing the spine structure. These aspects are treated in the *Logic*, *Notational*, and *Audio* layers respectively.

```
<ieee1599>
  <logic>
    <spine>
      <event id="e1" timing="0" hpos="0"/>
      <event id="e2" timing="2" hpos="2"/>
      <event id="e3" timing="2" hpos="2"/>
      <event id="e4" timing="1" hpos="1"/>
      <event id="e5" timing="1" hpos="1"/>
      ...
    </spine>
    <los>...</los>
  </logic>
</ieee1599>
```

Figure 4. A short example of spine. The listed events can be semantically defined inside the `LOS` sub-element and referenced by all the other layers.

In regard the other meaning of the spine, namely the glue function, once again Fig. 3 helps understand its central role in the format. Music events are not only listed in the spine, but also marked by a unique identifier. These are referred to by all instances of the corresponding event representations in other layers.

Please note that each spine event can be described:

- in 1 to *n* layers, e.g. in the *Logic*, *Notational*, and *Audio* layers. For example, a music symbol has a logic definition (a C-pitched eighth note), a graphical representation and an audio rendering
- in 1 to *n* instances within the same layer, e.g., in three different 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).

The events listed in the spine structure can correspond to one or to many instances in other layers. For example, Fig. 5 illustrates a common situation where music events are notes and rests.

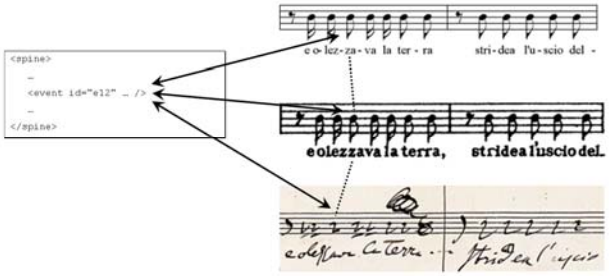

Figure 5. Spine event e12 mapped in three different score versions, corresponding to three graphic files.

Let a particular event listed in spine, namely event `e12`, be the 12th note appearing in the tenor part of an aria. By

using its identifier, one can investigate note pitch and rhythmic value, two data described in the *Logic Layer*. In this case, one discovers that the considered music event corresponds to a B-pitched eighth note. Now, assume that the considered piece has 3 score versions attached. As a consequence, in the *Notational Layer* there will be 3 entries where event `e12` is referred. More complex examples could involve also heterogeneous contents.

Thanks to the spine, IEEE 1599 is not a simple container for heterogeneous media descriptions related to a unique music piece. It shows instead that those descriptions can also present a number of references to a common structure. As intuitively shown by the vertical lines in Fig. 5, this aspect creates synchronization among instances within a layer (*intra-layer synchronization*), and – when applied to a complex file – also synchronization among contents disposed in many layers (*inter-layer synchronization*).

V. Conclusions

The format proposed here has been designed to achieve a comprehensive description of music, content interoperability, and deliverability. On one side, this way of encoding music pays special attention to on-line accessibility, digitalization of analogue material and preservation of artifacts from the past, independently of the cultural origin and language. On the other, the format is meant to provide an integrated and evolved fruition of music, thus representing a new approach to music delivery and enjoyment.

It is hoped that such an effort will open the way for a large number of new applications with increased power and flexibility, as well as new markets for these kinds of applications. The repercussions may have a wide effect on music education, media entertainment, enjoyment and development of music as a whole.

References

[1] G. Haus, “Rescuing La Scala’s Music Archives”, in *Computer*, vol. 31, no. 3, 1998, pp. 88–89.

[2] G. Haus and L.A. Ludovico, “The Digital Opera House: an Architecture for Multimedia Databases”, in *Journal of Cultural Heritage*, vol. 7, no. 2, 2005, pp. 92–97.

[3] D.L. Baggi, “Technical Committee on Computer-Generated Music”, in *Computer*, vol. 28, no. 11, 1995, pp. 91–92.

[4] G. Haus and M. Longari, *Proceedings of the First International IEEE Conference on Musical Application using XML (MAX2002),* IEEE Computer Society, 2002.

[5] D.L. Baggi, A. Baratè, G. Haus and L.A. Ludovico, “A computer tool to enjoy and understand music”, in *Proceedings of EWIMT 2005 – Integration of Knowledge, Semantics and Digital Media Technology*, 2005, pp. 213–217.

[6] A. Baratè, G. Haus and L.A. Ludovico, “An XML-Based Format for Advanced Music Fruition”, in *Proceedings of SMC 2006 – Sound and Music Computing 2006*, 2006.

[7] G. Haus and L.A. Ludovico, “Music Segmentation: An XML-oriented Approach”, in *Lecture Notes in Computer Science*, vol. 3310/2005, 2005, pp. 330–346.

[8] P. Roland, “The Music Encoding Initiative (MEI)”, in *Proceedings of the first IEEE International Conference MAX 2002 – Musical Application using XML (2002)*, IEEE Computer Society, 2002, pp. 55–59.

[9] J. Steyn, “Framework for a music markup language”, in *Proceeding of the First International IEEE Conference on Musical Application using XML (MAX2002)*, IEEE Computer Society, 2002, pp. 22–29.

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

Luca A. Ludovico (Milan, 1977) received his bachelor degree in Computer Engineering from the Politecnico di Milano (Milan, Italy) in 2003 and his Doctorate in Computer Science from Università degli Studi di Milano (Milan, Italy) in 2006. His major field of study is Music Informatics.

Since 2005, he has been Researcher and Assistant Professor at the Department of Informatics and Communication - Università degli Studi di Milano. At the moment, he teaches a course of Fundaments of Computer Science and a laboratory on music applications.

Prof. Ludovico is the Italian coordinator of the activities related to the IEEE Technical Committee on Computer Generated Music. Due to his role, he had a prominent part in the standardization process of IEEE 1599 format.