The Interactive-Music Network

DE4.3.2
Multimedia standards for music coding

Version: 2.4
Date: 01/06/2005
Responsible: IRCAM

Project Number: IST-2001-37168
Project Title: The Interactive-Music Network
Deliverable Type: PUBLIC
Visible to the Working Groups: YES
Visible to the Public: YES

Deliverable Number: DE 4.3.2
Contractual Date of Delivery: 31/05/2005
Actual Date of Delivery: 01/06/2005
Title of Deliverable: Multimedia standards for music coding
Work-Package contributing to the Deliverable: WP4
Nature of the Deliverable: Public
Working Group: Multimedia Standards
Author(s): Grégoire Carpentier, Jérôme Barthélémy

Abstract:
This document reports the uses and practices of music coding in a multimedia environment. It deals with proprietary formats or open standards dedicated or related to various forms of music content: audio, structured audio, music notation and metadata. It also addresses the issue of interoperability between heterogeneous content multimedia standards and their potential integration into a global, scalable, integrated multimedia music access framework.

Keyword List:
Music, multimedia, infotainment, edutainment, music notation, standards, music libraries, optical music recognition, music distribution, protection, accessibility, music creation, education, music archives, music publishing.

Table of Contents

1 Introduction: Rationales and scope of the document

1.1 Rationales for studying music in relation with multimedia

1.1.1 Some historical points

Music notation as known actually in Europe and America (and commonly referred to as Common Western Musical Notation or CWMN) is the product of a very long process which began in the XI century with the first musical notation by Guido d'Arezzo. It is generally considered that this form of notation has gained a remarkable stability in the XVIII and XIX centuries, and can be considered as something like the result of a process of formalization, or in today's words: a standard.

In the XX Century however, due to diverse influences like Contemporary music, ethnomusicology, and popular music, and with the emerging of industrial products such as phonograph, and more recently computer, music and its acceptation have changed. Musical practice in the preceding centuries was based on personal or collective practice on musical instruments, particularly the piano in the XIX century. This practice was not absolutely reserved to professionals, but mainly to amateurs, with for example the flourishing literature of what is known as "transcriptions", mainly done for 4 hands piano from originals conceived for orchestra, string quartet, or other forms of collective practices. The only way to distribute music was by the mean of musical scores, written in that common language known as CWMN.

With the appearance of the phonograph in the XIX century, a completely new mean for storage and distribution of music was born, which made possible the exact reproduction of any audio event – for example, a speech, a conference, a conversation – but particularly musical events. The distribution of music has been completely renewed by the new paradigm of "recording". We can consider the new P2P, Napster like, distribution mechanism as the last known avatar of this – the last but certainly not the least.

Studies in ethnomusicology were made possible by the use of these new recording devices – phonograph first, but later tape recorder. Consider for instance Bartok or Kodaly, recording songs and music from Central Europe, and introducing in their music new meter and rhythmic schemes directly. Even jazz can be considered as partially the result of the phonograph considered as a musical instrument. Electro acoustic music, where the recording device replaces completely the score, has changed the landscape in a more radical manner. Pierre Henry [1], one of the most important composers of "concrete music", or electro acoustic music, who created with Pierre Schaeffer the "Groupe de Recherches Musicales" (GRM) in 1950, is also considered by the musicians of the electronic scene as a founding father (with Karlheinz Stockhausen, John Cage and Robert Moog, Moog synthesizer's father).

In the 1960, with the appearance of the computer, a number of pioneers have tried to apply the new tool to music, with huge success. Two different domains have been explored : music production and music analysis. As a example in the domain of music production, CSound, developed by Barry Vercoe at the prestigious Massachusetts Institute of Technology (MIT), has gained a worldwide audience. In the same while, new schemes for music notation and exchange – like MIDI – have gained a worldwide audience. Even with the strong criticism made by recognized researchers from academia like Eleanor Selfridge-Field [8], MIDI is still recognized as the only one available format for music notation exchange and studies, even for academic research.

In the same while, the capabilities of the computer for repetitive tasks, and the process of digitization have made possible the generation of powerful tools for audio signal analysis – tools mainly based on the Fast Fourier Transform (FFT). This kind of tool has made become possible not only first an analysis of the harmonic characteristics of the audio signal, similar to the analogue-based spectrogram previously discovered in the 40s and applied by Emile Leipp [2] to musical studies, but also the development of compression schemes which are at the basis of the new distribution models (P2P networks as well as more traditional distribution models such as i-TV or i-Tunes).

The emergence of these new tools, together with the Internet, has particularly made possible the copy of music at a extremely low price, such a low price that it can be perfectly neglected, making it possible a mass production of music copies in a totally inexpensive manner. These issues, as well as many other related to intrusion of mass production technology in music production, have been put in light recently in an issue of the well-known Leonardo Music Journal [6], devoted to "the musical implications of grooves, pits and waves".

All these topics address the issue of new musical experiences in a growing range of domains and involving a higher and higher technology level. They are the evidence of a strong connecting point between music creation and diffusion on one hand, technologies and media and the other hand, and can therefore not be simply ignored anymore. The fail of recent project Net4Music has demonstrated that exchange of music, music concepts, music experiences based only on CWMN music sheets, is inadequate to the actual needs of the music community. This community indeed tends toward the exploration of new music schemes and concepts that can't be supported by the CWMN, for instance:

- Audio advanced concepts, such as 3D audio, with possible user interaction in a virtual 3D space where position in space interacts with audio rendering (see LISTEN project [4]).
- Fine grain synthesized music, using synthesizer tools controlling all aspects of evolution of pitch or timbre (CSound, Max/MSP).
- Interaction between scene (gesture, voice, instruments) and synthesized audio, for real-time realization of effects in live performances (see MEGA project [9]).

These emerging, growing behaviours outline the need for providing to music production and distribution actors a multimedia framework that, on one hand, copes with their current and potential future interests and on the other hand, integrates heritage from the past: number of musicians indeed – composers, instrumentalists, or even amateurs – are still attached with the concepts inherited from centuries of CWMN, and they should also not be ignored by new technologies issues. Lot of them are working in a closely related fashion with the most advanced tools in the computer domain, and are even actively participating to the evolution of the new music tools (e.g. Philippe Manoury).

1.1.2 What is multimedia music?

Multimedia music is music conveyed to the user by the mean of different medias – graphics in different flavours, vector as well as bitmaps, audio in different flavours, pulse code modulation as well as what is actually known as "structured audio", these different medias being combined, synchronized, interacting for the purpose of bringing users a richer and deeper experience of music. It's also music giving to the user the ability to interact with, from simple manipulation of volume to 3D audio interaction, selection of sources through the mean of powerful search engines, combination of sources (playlists or remix), or even direct manipulation of composition or synthesis parameters.

This definition of multimedia music infers some specific requirements which will are detailed below (See section 2).

1.2 Scope of the document

This document aims at studying the needs and requirements for inserting music and music notation in multimedia production, current solutions and uses, based either on de facto standards or on open standards. It first gives an overview of current actors, formats, standards and multimedia technologies, then focuses on current uses of music-related multimedia content and ongoing multimedia standardization projects. Last will be considered a few multimedia music future uses scenarios, related with currently existing and missing technologies.

1.2.1 The multimedia landscape evolution

The multimedia landscape is perpetually and very quickly evolving. Some years ago, in the early 90s, the CD-ROM was considered as the new mass medium, and seemed to have the strongest support from industry as a delivery medium. Some years after, the CD-ROM is not anymore considered as being so.

Due to this situation, several platform standards for CD-ROM were introduced, most of which have failed or are failing, and new solutions have been developed, the most relevant part of these have been developed in relation with Internet (web-based solutions).

1.2.2 Open standards and proprietary formats

It is generally recognized that open standards have some advantages, first interoperability. As it has been developed by several entities from industry as well as from academic research working together in a closely related fashion, an open standard must comply with requirements coming from several different vendors and several different fields of research, taking into account several use cases and users. Integration of these different requirements has been generally part of the standardization work, as it's the case in the MPEG process.

Due to this interoperability, the MPEG standard does not rely on any particular system. Particularly, MPEG is available on many different kind of devices (from computer to HDTV), by different distribution channels (from Internet to DVD), and is not dependent of systems or hardware.

Another advantage of an open standard is in its persistency. As an open standard is not dependent of a single vendor, there is generally a huge support for its implementation from those actors having elaborated it.

A good example of an open standard is the MIDI standard – which was elaborated in the 80s by leading synthesizer makers. In the meanwhile, it must be recognized that a lot of proprietary formats, acting as de facto standards have been in the past very successful – and still they are. First of all, the postscript standard, elaborated by Adobe in the 80's, but many others in the following years.

What is an "open standard"?

Generally speaking, an open standard is a standard for which the specifications are available to everyone interested in implementing it.

More specifically, open standards are generally available through the mean of an international organisation, independent from any vendor – such as ISO, W3C, or OASIS – which is devoted to the purpose of editing, publishing and reviewing standards, and this according to a specific work plan and to a predefined set of rule regarding its elaboration. Such a rule is, for example, the need of participation of different entities – industry, academy, research and so on.

As an example, an ISO/IEC standard such as MPEG must be elaborated by several countries, several industrial partners, and must be approved on the basis of a general consensus between all interested parties.

The Flash format, being presented by Macromedia as an open standard, doesn't comply with these requirements, since it has been elaborated by a single vendor. It has not been generally approved by a wide community of industrial users. Improvements, compatibility and future of the standard is not guaranteed by any independent organization. Due to these inconveniences, Flash cannot really be considered as an open standard.

1.2.3 Needs & requirements for a multimedia music standard

The evolution of information technology has recently produced changes in the usage of music notation, transforming the visual language of music from a simple written "alphabet" for music sheets to a tool for modeling music in computer programs, cooperative work on music and other multimedia integration tasks. More recently, millions of music users have discovered the multimedia experience, and thus, the traditional music notation model is likely to be replaced with something much more suitable for multimedia representation of music.

The definition and design of a multimedia music standard, in accordance to the multimedia music definition of section 1.1.2, implies the identification of two kinds of requirements:

- General requirements for the development of multimedia standards
- Specific multimedia music-related requirements

The latter ones particularly address the question of synchronization of music notation with other media. The modeling of music notation is a rather complex and composite problem. Music notation is a multi-layered piece of information, which may be used for a wide range of different purposes: from audio coding and entertainment to music sheet production, music teaching, music analysis, content query, provision of enhanced or adapted music for consumers with specific needs, etc. In the current Multimedia and Communication age many new music-related applications are strongly attracting the market attention and most of them will become more and more widespread in a short time. In order to identify a unique notation format to satisfy all these application fields and to set the basis for the new forthcoming applications, several aspects have to be considered, ranging from information modeling and format to integration and synchronization of the music notation information into other media and cross-media tools and formats.

1.2.4 Overview of current multimedia music-related standards and formats

1.2.4.1 Audio formats & audio file formats

There are many formats for audio, either non compressed or compressed ones. Compression means that the original signal is analyzed in order to detect redundancy, and that this redundancy is removed in order to reduce data size, and thus to obtain a gain of bandwidth in the case of on line distribution. Compression can be either lossy or lossless. Lossy compression means compression of audio data in a form where, when data are expanded, they can loose a part of their original data. Psychoacoustic models are used in order to decide which part of the signal is to be recorded, and which part is to be discarded. In lossless compression, the original signal and the decoded signal are bitwise identical. Lossless compression schemes are not widely used, since the compression ratio usually obtained is very far from compression ratios obtained with lossy compression schemes such as mp3 or AAC. We review in this document the most widely used audio formats, and the most prominent emerging ones.

Most of the currently used audio formats and file formats are based on lossy compression schemes, such as mp3, AAC, WMA. These compression schemes make distribution of music almost inexpensive for consumers, and are currently available within the whole range of multimedia frameworks.

As far as lossless compression schemes are concerned, we don't report all of them in this document: Will be let out of scope those devoted to archiving purposes only, and not to multimedia distribution and diffusion, as they generally achieve a somehow poor compression rate: generally from 1.5 to 2, compared to ratios of 10 and more that are available in lossy compressions.

We don't study in details in this document specific issues related to lossy compression schemes, such as artefacts introduced by compression, loss of information, and related problems for archiving purposes.

We don't elaborate comparison between different compression schemes regarding quality and/or efficiency.

1.2.4.2 Audio effects, 3D audio, multi-channel audio

The technology improvements achieved recently have been made possible for end users a number of technologies that greatly enhances their audio and music experience.

Multi-channel audio has been present in the cinema production since the late 30s-early 40s with the production and release of Disney's "Fantasia". However it was wildly complex, expensive, and had no chance of working in a home environment. Practical multi-channel digital audio began in theatres in the early 90s. These new digital audio technologies advanced quickly to allow DVD, Direct Broadcast Satellite (DBS), Digital Cable, and Terrestrial Digital Television (DTV) to deliver this same high quality multi-channel digital audio directly to the consumers home.

The concept of 3D audio has existed for sometime, but only recently has modern computing technology enabled the real-time processing needed to deliver 3D audio. Because 3D audio allows sounds to be perceived as emanating from different locations, it has been considered to be the gemstone for enabling less confusing simultaneous communications between the different actors in mission-critical applications. In such cases, speakers appear to be heard at different locations, allowing easy in-context understanding of who is speaking. These technologies have been recently applied to the cultural domain, for example in the European projects LISTEN [4], or CARROUSO [10].

The same progresses in modern computing technology have made also possible the implementation of audio effects acting in real time on the user's computer side. Such effects include delay, reverberation, remix and many other much more complicated.

Such audio technology is not widely available today in multimedia frameworks, even if widely available in the research area or in professional environment.

1.2.4.3 Structured audio formats

We review in this document structured audio formats, that is, audio formats devoted to creation, transmission and rendering of parametric sound representations (Vercoe, Gardner, Scheirer 1998 [3]). *Structured Audio* means transmitting sound by *describing* it rather than *compressing* it. Examples of structured audio formats are audio synthesis languages such as CSound, and the linear-prediction model of speech. We focus our interest on the two first categories, even if the third – linear-prediction model of speech – could be also of interest in the musical domain (lyrics). We also consider also the Musical Instrument Digital Interface (MIDI), as well as the Sound Description Interchange Format (SDIF), as being forms of structured audio.

As MIDI, in this sense, is the more widely used structured audio format, we focus our interest mainly on it, while interesting models such as CSound are not generally used in multimedia. A specific case is the case of MPEG SA, which is the only case of structured audio embedded in a multimedia framework.

1.2.4.4 Graphics

Even if less important for the musical domain, graphics are of importance for multimedia music:

- Raw graphics can be used for non traditional images of music, such as spectrograms, sonograms, or scan of images (music scores, manuscripts...), and have to be used to cover some requirements of multimedia music publication, particularly in the educational domain.
- Vector graphics are more important, since a standard for music coding is not available in any multimedia framework. Vector graphics can be used to cover this lack, and is generally used to this aim. One can refer to the review of musical publications on line reported below, to see that Flash is very often used to this aim.

1.2.4.5 Transfer protocols

Audio transfer protocols and standards such as S/PDIF, ADAT, AES/EBU are considered out of the scope of a multimedia standard review and will therefore voluntary not be reported in the present document.

- This review aims at giving a rich overview of currently existing or under development multimedia standards, focusing on content authoring, editing and rendering tools and facilities.

Any consideration on multimedia content transfer protocols or delivery technologies (cable networks, wireless networks…) will not be addressed in this document. However a few words about 3GPP and 3GPP2 will be said in the QuickTime paragraph (see section 3.6.4).

1.2.4.6 Multimedia frameworks

We end this multimedia state-of-the art with a short introduction and comparison of today's most frequently used multimedia frameworks. In this section are both reported:

- Open standards such as MPEG or SMIL, covering many aspects of multimedia.
- Commercial and specific software products such as Macromedia Flash or Director, which are not initially intended for music coding, but enable audiovisual features of strong interest regarding our definition of multimedia music (see section 1.1.2) and the development of a so-called standard.

1.2.5 Application scenarios

Application scenarios imagining potential future uses of multimedia music are provided in section 7. The overall purpose of this section is to demonstrate the benefits of integrating a music notation standard into the presently under development MPEG multimedia frameworks: MPEG-4, MPEG-7 and, though less directly, MPEG-21. The MPEG frameworks indeed are designed to supports a quite large range of multimedia content (Advanced Audio Coding, video coding, structured audio coding, 2D & 3D graphics, raw and vector graphics, scene description, user interaction…) that can be used in connection with music notation in furtherance of the Musicnetwork's vision of multimedia music.

Application scenarios are therefore useful to the following purposes:

- Exemplify to the MPEG and Music Notation communities simple cases where music notation and other multimedia object types are integrated resulting in mutual added value.
- Understand and consequently refine the definition of a music notation standards requirements in order to possibly approach a call for technology as a next step.

This last issue have been the main activity of the MPEG Ad Hoc Group (AHG) on Symbolic Music Representation (SMR) during the last year. The two first scenarios presented in section 7 (7.1 & 7.2) have been developed in the scope of the MPEG AHG on SMR, and approved by the MPEG group. They have helped the MPEG AHG on SMR designing adequate and relevant requirements for a symbolic music representation model.

For more information see the MPEG AHG on SMR Web Site.

2 Needs and requirements

2.1 General standard-related requirements

The following three general requirements have been recognized as essential ones for the elaboration of standards and formats in a multimedia context, regardless any specific application scenario.

2.1.1 Scalability

It is generally admitted that a well-designed multimedia standard is likely to support various kind of content, from a simple text to a complex structure with different forms of media (audio, video, graphics) that need to be synchronized together and compliant with user-interaction facilities. It would be however completely irrelevant to design a content-independent framework, meaning for instance great amount of data transfer only to display a simple text on a cellular phone or a portable device.

Scalability stands therefore as a way to answer a question of paramount importance in the multimedia world: What is needed in terms of resources, objects and hardware to design, deliver and render a specific kind of content with specific user-interaction features. In other words: what do you need for what you want to do in the most effective manner?

In the MPEG frameworks scalability is solved by the concept of profiles, which allow the resources to be content dependant: low profiles for poor, simple content, rich profiles for rich, complex content.

2.1.2 Interoperability

This is a typical feature of MPEG applications. Interoperability requires independence of platforms, hardware, applications, environments, programs and devices: Multimedia content should be accessible either from personal computers, i-TV, Mobile Devices, PDAs, Tablet PCs, special hardware, cars, etc.

The need for an open, interoperable, free standard for multimedia production is recognized since the beginning of multimedia in the 80s. Unfortunately, multimedia production is actually trusted by proprietary frameworks, like QuickTime in the early 90s, and later Flash or Director. There are some reasons to this situation:

- Proprietary frameworks benefits of proprietary authoring tools, generally not very expensive, easy to understand, very easy to use.
- Proprietary frameworks takes in charge compatibility problems between different platforms, at the expense of choice for end users.

At the opposite, interoperability and openness means that choice is left open for end users, and choice is left open for implementation.

Proprietary frameworks are sometimes restricted in their implementation of new technologies. This is particularly true for technologies required for music and audio, which are not supported in the most widely used multimedia proprietary frameworks.

2.1.3 Extensibility

Extensibility is also generally considered as a criteria of paramount importance in the design of a multimedia standard. Extensibility means for a standard never to be frozen to its current form at a given time, in terms of content authoring, rendering, interaction between different contents and user interaction. On the contrary an extensible standard should be designed in order to make possible an easy, quick and coherent adjunction of content and features.

In other words, extensibility means the ability to foresee the standard's future evolutions and to minimize the constraints fixed by the core development. For instance it should be easily possible to design content data structures of a various scale of complexity out of a core library of simple elements. The MPEG-7 framework – with its content description aims – is a good example of such a standard, where extensibility relies on its XML structure (see section 3.4.7.3).

2.2 Specific music-related requirements

The following three general requirements have been recognized of high importance for a multimedia music standard which match with our definition of multimedia music (see section 1.1.2).

2.2.1 High quality rendering

High quality of rendering for audio means that the audio quality must be quasi undistinguishable from CD audio quality. This is now accessible with the most recent compressions schemes, at rates of 160 kbps, but this not always true for early compression schemes, of whom latest evolutions are still at use in most cases.

Lossless compression schemes are needed as well for commonly used stereo audio files and for surround, multichannel sound files. Though mp3 Surround tools have been recently developed by Fraunhofer IIS, support for surround sound, multichannel audio, and 3D spatialization is still not very frequent.

A high quality of rendering for graphics is also needed, and possibly the ability to adapt to the end-user's rendering devices. A vector graphics-based rendering engine must be supported in order to render music scores.

2.2.2 Synchronization between different media

This is a straightforward requirement for the design of a multimedia music standard. Multimedia music means indeed various kind of contents: Audio, video, graphics, images, music notation, text, structured audio, etc. Synchronization between these media is for instance needed for:

- Processing audio files of different formats
- Scores or/and lyrics real-time displaying
- Real-time subtitles displaying
- Real-time score following
- Animation for educational purposes
- Structured audio score rendering
- Real-time DSP
- Real-time events or metaevents triggering

Synchronization can be achieved on the basis of a timeline. As in most of sound sequencers the time representation should be both temporal and metric, and conversions from the temporal domain into the metric domain and vice versa should be straightforward. As a consequence, any shrinking and stretching operation in the music execution time should be replicated in the execution rate of the other media and vice versa. Algorithms of shrinking and stretching for audio, video, etc., are already available.

2.2.3 Multi-layer rendering

The heterogeneity of multimedia music contents requires not only synchronization between these media, but also the ability to design and develop suited layers for each kind of content. As simple audio digitally distributed on CDs has been extended to video, audio, lyrics/subtitles, and basic interaction possibilities on current DVDs, it is likely desirable to add different layers for a complete distribution of music-related content, providing more advanced interaction features.

2.2.4 Annotation, indexation, metadata

Annotation of music scores is a traditional activity. Teachers, musicians (professionals or not), pupils, all musicians using scores put currently annotations on these at least to help them playing music (fingering, cue notes, attention marks, breath marks…), or to play it in a slightly different way, what is used to be called "interpretation" (adding bowing, dynamics, tempo…).

By using multimedia systems, annotation systems can be enlarged to other kind of annotations: video or audio (for example showing details of execution), still images (for example manuscripts) and also graphical representation of music scores or handwritten music manuscripts. A lot of interesting new applications could be considered in the field of music education. Several examples of tools present music notation annotated with audiovisual content. Those applications are not flexible enough. For example, they are not supported by any authoring tools and thus the content for the lessons is limited to that provided by the tool producer. Distribution of traditional scores can also be enhanced by annotations.

Powerful indexing mechanisms must also be available, not only based on traditional metadata such as author, title and so on, but also based on description of content (timbre, mood, tempo, meter, rhythm, melody and so on). These descriptions must be made available with a fine-grained level of granularity, that is with the ability of the system to address segments or sections of the media. Such features are needed as well for information retrieval and for musical analysis encoding.

2.2.5 Integration in a global multimedia content delivery framework

It seems more relevant to design a multimedia music standard as an extension of currently existing or under development multimedia standard rather than aiming at a global standalone framework which satisfies at once all the requirements. An appropriate strategy would therefore by to identify, design and develop the technologies that are presently missing in multimedia frameworks such as MPEG-4 or MPEG-7. This development should of course be driven by integration facilities into a host framework.

Another important requirement of multimedia distribution is its easy integration with digital asset management systems, and with databases. At the opposite of authoring tools like Director, Flash or Premiere where media content is directly inserted in multimedia flow, loosing relationship with the original content, more recent approaches have shown the interest of a non-destructive approach, particularly, XML based approaches such as SMIL, where the multimedia framework defines only temporal and spatial organisation of media.

2.2.6 Content-based analysis and retrieval

This requirement is related to query-by-content by using combination of traditional query (melody, rhythm, key tonality, ambitious, instrumentation) with audio-visual content description (low-level and high-level description). Evaluation of similarity, based on statistical approaches and audio-visual descriptors, combined with traditional query could be achieved this way, thus enhancing the traditional cataloguing paradigm. Description of music notation content should be extensible to traditional cataloguing schemes, in order to encompass current practices in libraries, and interoperable with audio and visual descriptions.

2.2.7 Real-time features

Openness to real-time processing functions means that some processing tools must be available, either on the client side or even on the authoring side. For example, for karaoke systems, or "intelligent accompaniment systems" which could be implemented on the client side to automatically follow the execution of an end user on a MIDI instrument, or an instrumentalist playing in front of a microphone. But more simply, mixing tools can also be made available to the user.

2.2.8 End-user interactivity features

Openness to end-users interactivity means that some interactivity could be implemented on the user's side, and that this kind of interactivity can be implemented in a general manner, not only by the mean of authoring on the production side. For example, transposition is a general need of end users. A good framework for multimedia music should implement this functionality in a standard manner, without having to implement it at the authoring step.

3 Music & multimedia: state of the art

We review here some of the most important actors in the multimedia domain, together with a list of products in relation with multimedia production, authoring or diffusion. We don't speculate about their future, nor about their possible evolutions. The landscape is likely to evolve quickly, some of the main actors will probably disappear, and possibly new actors will appear. The relative weight of the actors also is likely to evolve, and we don't take here any position on this point.

3.1 Audio & Media Industry actors

3.1.1 Macromedia

Macromedia is the most widely known actor in the multimedia domain in the Internet, with its leading products Flash and Director (Shockwave). Located in San Francisco, Macromedia is working on authoring software since 1984. Macromedia develops several products, from authoring tools to free viewers and players. In a short future Macromedia is to be acquired by Adobe and merged in a combined company, Adobe Systems Incorporated (see below, section 3.1.2).

Macromedia authoring tools are very widely used, this being probably due to their particular approach of the multimedia production, this approach being generally based on the concept of a "timeline", particularly convenient for time-based multimedia content, such as scenarios, story-boards, but also for music, since music scores are conforming to a time line (the five-line staff), where events are placed by the composer.

Multimedia authoring tools: Director, Flash, Authorware

- Director and Flash are detailed below (see sections 3.6.1 and 3.6.2).
- Authorware is oriented towards e-learning applications. It is compliant with standards from the AICC (Aviation Industry Computer-Based Training Committee) or with the ADL Shareable Courseware Object Reference Model (SCORM), and includes functions for automatic tracking of students results, functions for managing accessibility such as text-to-speech. Applications developed with Authorware can be delivered on corporate networks, CD/DVD, and the web.

Content production: Freehand (vector graphics), Fireworks (graphics), Dreamweaver (web sites), SoundEdit (audio)

- Freehand is a design software designed for tight integration with Flash and Director products. It is based onto vector-based tools for designing print layouts, Macromedia Flash MX animations, or application interfaces.
- Fireworks is also a design software oriented towards bitmap-based graphics such as photo editing. It provides also a tight integration with other Macromedia tools.
- Dreamweaver is a web site editor, with support for CSS (cascading style sheet), XML and web services, as well as support for dynamic web pages technologies such as PHP (Hypertext pre-processor), ASP (Microsoft's Active Server Page), or JSP (Java Server Page).
- SoundEdit is an application for editing audio, with support for many different formats such as WAV, AIFF or AU. It provides tools for visual analysis of sounds, spectral view or Fast Fourier Transform.

http://www.macromedia.com/

3.1.2 Adobe

While mainly oriented towards business-oriented electronic publishing, Adobe develop several products for multimedia content production and diffusion.

The main Adobe's product is the Acrobat framework, based on the PDF format (Portable Document Format), itself based on postscript, a standard for printed documents. This framework is mainly oriented towards business-oriented electronic publishing, electronic diffusion of business documents, but is also sometimes used in multimedia publishing.

Apart of content in PDF distributed with the Acrobat framework, the content produced with Adobe can generally be distributed in Windows Media formats, in the QuickTime format, or in the MPEG-2 (DVD) format.

Authoring & production tools: Premiere (authoring), After Effects (animation, visual effects), Encore DVD (DVD authoring), Audition (Digital audio).

- Premiere is a mostly a video editing tool. It provides support for many video formats such as MPEG-1, MPEG-2, DV, AVI (Microsoft), Windows Media 9 Series, and QuickTime. For audio, Premiere provides support for WAV, Windows Media Audio, mp3, and AIFF as well as audio-only AVI and QuickTime formats. Content produced with Premiere can be directly exported to DVD (MPEG-2).
- After Effects is oriented towards video effects, compositing, animation and visual effects.
- Encore DVD is oriented towards DVD authoring, with text tools, menu creation, and support of interactivity.
- Audition is an audio editing environment. It provides digital signal processing (DSP) tools and effects, mastering and analysis tools, and audio restoration features. It provides support for WAV, AIFF, mp3 and WMA formats.

http://www.adobe.com

Adobe to acquire Macromedia

Recently Adobe has officially announced a definitive agreement to acquire Macromedia in an all-stock transaction. The future company resulting from this acquisition is to be named Adobe Systems Incorporated. Today there is no clear position from Adobe about the future of both Macromedia and

Adobe products. "We do not anticipate any changes with our ongoing business plan", Adobe says. It seems however that over time, Macromedia product will transition to the Adobe brand.

The core tenets behind the success of Adobe and Macromedia main free software, FlashPlayer and Adobe Reader will apparently remain the same, but it makes no doubt that Adobe Systems will encourage the development of an industry-defining, cross-media, rich-client technology platform across multiple operating systems and devices, through the complementary of Macromedia Flash and Adobe PDF. The planned release of Macromedia Studio MX this year (2005) will not be affected by this operation.

Both Adobe and Macromedia have until now encouraged and supported the definition and development of SVG. Even if Adobe's position on the future support of SVG is still not clear ("*The combined company will continue to work with customers and partners to define a future roadmap for our products*"), one can wonder whether potential changes in field of open standards for vector graphics are likely to be expected in the future.

http://www.adobe.com/aboutadobe/invrelations/adobeandmacromedia.html

3.1.3 Microsoft

Microsoft is active in the domain of multimedia, even if this activity is more recent than other actors. Microsoft is actually present mainly with its Windows Media technology, but also with Internet Explorer, and even with its development tools such as Visual Basic, .net, and C#. These tools can be used to integrate Windows Media technologies in custom products, or to enhance or to extend capabilities of Windows products : decoders, user interfaces, production tools and so on.

Microsoft is also active in the domain of media coding, and produces its own technologies for streaming media, audio and video compression. Windows Media integrates a proprietary technology for Digital Rights Management, which has made of this tool the tool of choice for online music distribution before the i-Tunes success (Universal Music).

Moreover, the next Windows exploitation system, forthcoming in 2006, will probably integrate much more capabilities than those actually available. This new exploitation system, whose current name is "Longhorn", will integrate a new presentation subsystem, code-named "Avalon", which integrates many functionalities, among these being powerful tools for vector graphics and animation ("Sparkle"), a new mark-up language for user interfaces, named "XAML" for eXtensible Application Markup Language, to be used for presentation and navigation in information. The "Sparkle" component for animation in Avalon has been qualified of "Flash-killer", by reference to the Macromedia Flash product.

http://www.microsoft.com
http://longhorn.msdn.microsoft.com

3.1.4 Apple

Apple QuickTime is Apple's multiplatform multimedia technology for handling video, sound, animation, graphics, text, interactivity, and music. As a cross-platform technology, QuickTime can deliver content on Mac OS X, as well as all major versions of Microsoft Windows.

Main products: QuickTime Player, QuickTime Pro (Media authoring), QuickTime Streaming Server (Broadcast streaming Video), QuickTime Broadcast (Encoding software for live events), DVD Studio Pro (DVD authoring).

- QuickTime technology, QuickTime Player, QuickTime Pro and QuickTime Broadcast are detailed below (see section 3.6.4).

- QuickTime Streaming Server is based on RTP/RTSP (Real-Time Transport Protocol/Real-Time Streaming Protocol). It provides support for streaming QuickTime, MPEG-4 and 3GPP files. It also provides support for streaming mp3 content using Icecast-compatible protocols over HTTP.
- DVD Studio Pro is a DVD authoring tool. It provides support for interactivity, menus, scripting control, remote control interactivity, and DVD-ROM content. It provides MPEG-2 encoding for DVD and MPEG-4 for web streaming., and Dolby digital AC-3 compression and 5.1-channel surround sound.

http://www.apple.com/software/

3.1.5 Fraunhofer IIS

The Fraunhofer Institute for Integrated Circuits IIS develops software, microelectronic circuits, devices and systems up to the scale of complete industrial installations for IT and communications applications. Research activities focus on: audio encoding up to international standardization, low-power circuits for battery-powered terminals, high-frequency circuits, integrated digital systems for telecommunications, satellite navigation, quality assurance by means of automatic image recognition, ultrafine-focus x-ray systems, image sensor technology and high-speed camera systems and digital cinema.

Fraunhofer IIS is the leading international research lab in the field of high quality low bit rate audio coding. In its "Audio & Multimedia (AMM)" department cluster a team of 80 engineers focuses on the development of audiovisual solutions. Fraunhofer IIS has been the main developer of the most advanced audio coding schemes, like MPEG Layer-3 (mp3) and MPEG AAC (Advanced Audio Coding). Fraunhofer IIS plays a major role in the MPEG-4 Audio standardization process and contributes to many other standardization efforts as well, like 3GPP, AES, DMDA, DRM, DVB, DVD, EBU, ISMA, MPEG-7, MPEG-21, ITU-R WP6A and ITU-R WP6Q.

Technology: MPEG-2, MPEG-4, DRM.

- MPEG-2: MPEG Layer 3 (Full name of mp3, see section **Errore. L'origine riferimento non è stata trovata.**), MPEG-2 AAC (Advanced Audio Coding, see section 3.4.1.3), ENSONIDO® (Surround sound over stereo headphones).
- MPEG-4: MPEG-4 Audio features & profiles and MPEG-4 Audio Lossless Coding (ALS) (see section 3.4.1.7), MPEG-4 Video (high quality low bit rate video coding, thanks to a block-based predictive differential video coding scheme), MPEG-4 Systems (Integration and interaction for natural and synthetic multimedia, see section 3.6.5), Integrated Error Robust Solutions (Transmission of MPEG audio data over error prone channels).
- DRM: LWDRM® (A DRM system convincing through usability and security), Watermarking (Watermarking used for Digital Rights Management), Audio Scrambler (Making Compressed Digital Audio Secure).
- Further Audio technology: AudioID (Automatic audio identification and fingerprinting, in connection with the MPEG-7 framework, see section 3.4.7.3), Watermarking (Embedding data into music by imperceptibly changing the audio signal), Core Design Kits (Implementing MPEG audio codecs on fixed point processors).

Software: mp3 Surround (mp3 Surround encoder/decoder/player demoware), mp3 VBR-Header SDK (for mp3 developers). The Fraunhofer IIS MPEG-4 player and the Fraunhofer IIS MPEG-4 encoder and streaming server are no longer available to the public, since the evaluation period has expired on December 31, 2004. This software demonstrated MPEG-4 technologies available at Fraunhofer IIS.

http://www.iis.fraunhofer.de/index.html

3.1.6 RealNetworks

RealNetworks was the first company able to distribute streaming media content over the Internet. Due to the strong competition however, and particularly from Microsoft, the position of RealNetworks is not more a leading position. RealNetworks is actually (01/2004) suing Microsoft charging that the company has abused its monopoly to gain an unfair advantage in the digital media market.

Main products: Helix Producer (production tool), RealOne player.

- Helix Producer is a software for digital media encoding and broadcasting. It provides encoding in RealMedia proprietary format from input formats such as AU, AVI (Microsoft's Audio Video Interlaced format), Quicktime, WAV, mp3, MPEG-1, and AIFF.
- The RealOne player is the player for viewing the RealMedia proprietary format, detailed below (see section 3.6.7). The Real player supports also SMIL content (see section 3.6.6).

http://www.realnetworks.com/

3.2 Music notation industry actors

Though it is not the purpose here we say a few words about music notation actors, as music notation coding still remains a missing technology in integrated multimedia. For more information please refer to the MusicNetwork's deliverables 4.1.1 and 4.1.2 (Music Notation Coding) available from the MusicNetwork's web page.

Today's music notation market is actually driven by two main actors: Finale and Sibelius. Both companies develop technologies for editing, scanning (optical recognition), and printing high quality music scores. Theses technologies have been recognized by the music community to achieve a high degree of interest for professional or amateur musicians, instrumentalists, composers, students and teachers purposes. Their wide success have even encouraged music universities and music schools to set up training courses devoted to music notation software.

However such high quality music notation facilities are still neglected by current multimedia frameworks and standards, and the music community is still affected with the lack of open standards (which have built the success of the Internet) capable of embedding current music notation technology in a multimedia context. In particular, neither Finale nor Sibelius use an interchange format as their default format. They have rather designed their own, binary and undocumented formats, which keep them out of the multimedia area, although the conversion from Finale and Sibelius formats to interchange formats and vice-versa is still possible. The best export formats at the moment are NIFF, MusicXML, and SCORE. Others are possible, but not MIDI, that loses almost all visual information (see section 3.5).

3.3 Standardization bodies

3.3.1 MPEG

The Moving Picture Expert Group (MPEG – http://www.chiariglione.org/MPEG) is a working group of ISO/IEC Joint Technical Committee 1 subcommittee 29, devoted to the development of standards for coded representation of digital audio and video. Established in 1988, the group has produced MPEG-1, the standard on which such applications as Video CD are based, MPEG-2, the standard on which such applications as Digital Television set top boxes and DVD are based, MPEG-4, the standard for multimedia for the fixed and mobile web and MPEG-7, the standard for description and search of audio and visual content. Work on the new standard MPEG-21 "Multimedia Framework" has started in June 2000. So far 3 Technical Reports and 8 standards have been produced and 8 more parts of the standard are at different stages of development. Many Calls for Proposals have already been issued.

The MPEG Industry Forum (MPEGIF - http://www.mpegif.org/) is a not-for-profit organization devoted to promotion of MPEG technologies. The forum organizes events and exhibitions, carries interoperability tests and develops an MPEG-4 Certification program.

3.3.2 W3C

The World Wide Web consortium (http://www.w3.org/) is devoted to the development and evolution of the Web. As such, the W3C is strongly involved in all standardization activities around the Internet and Web tools, and first with HTML and its evolutions. But in addition to this core activity, the W3C has developed several open standards for the web, for different purposes. Some of these standards are related to multimedia, such as SMIL and SVG. Some of these are also related to metadata, such as RDF.

The W3C has also developed XML, a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. XML is a method for structuring data, which defines rules or guidelines for designing text-based file format for structured data. XML, as HTML, uses tags and attributes (of the form name="value"). XML doesn't define by itself any tag, but derived languages (XHTML, SMIL...) does.

XML is now a rich family of technologies, composed of many XML-based languages:

- XSL is an XML-based language for expressing style sheets, and XSLT is an XML-based transformation language for adding, removing or rearranging tags and attributes in an XML file.
- RDF Site Summary (RSS – also called Really Simple Syndication) is a lightweight multipurpose extensible metadata description and syndication format. RSS is an XML application, conforms to the W3C's RDF Specification and is extensible via XML-namespace and/or RDF based modularization.

There are several XML-based vertical applications defined outside of the W3C by several organizations, such as OASIS (http://www.oasis-open.org).

3.3.3 Web3D Consortium

The Web3D Consortium was formed to provide a forum for the creation of open standards for Web3D specifications, and to accelerate the worldwide demand for products based on these standards through the sponsorship of market and user education programs. Web3D applications have been actively pursued by many organizations for quite some time. This community has spearheaded the development of the VRML 1.0 and 2.0 specifications (see section 3.4.4.1), which provide the basis for the development of associated applications. The organizations involved in this effort felt that the creation of an open consortium focused exclusively on Web3D would provide the structure necessary to stabilize, standardize, and nurture the technology for the entire community.

Today, the Web3D Consortium is utilizing its broad-based industry support to develop the X3D (see section 3.4.4.2) specification, for communicating 3D on the web, between applications and across distributed networks and web services. Through the well-coordinated efforts with the ISO and W3C, the Web3D Consortium is maintaining and extending its standardization activities.

http://www.web3d.org/

3.4 Audio and media coding

3.4.1 Audio formats

We review below only the most widely used audio formats, eventually coming with their associate file format, when this latter is unique.

3.4.1.1 PCM audio

PCM (Pulse Code Modulation) is a common method of storing and transmitting uncompressed digital audio. Since it is a generic format, it can be read by most audio applications—similar to the way a plain text file can be read by any word-processing program. PCM is used by Audio CDs and digital audio tapes (DATs). PCM is also a very common format for AIFF and WAV files.

PCM audio can be encoded in various resolutions, sampling rates, and number of channels, but is often recorded with 16 bits resolution, 44.1 kHz sampling, and stereo. With stereo recording, it is 1536 kbps, the rate of an audio CD.

Recently, different attempts have been made to improve the quality of audio on physical devices, and have led to the development of two new digital music formats: SACD and DVD-Audio.

The SuperAudio CD (SACD) uses a new process of sound recording and reproduction called Direct Stream Digital™ (DSD). DSD enables a much more direct signal path than the Pulse Code Modulation (PCM) format of original CD, which requires a number of interpolation and over-sampling filters during recording and playback.

DVD-Audio uses completely different technology to achieve improvement of performance. DVD-Audio discs take advantage of higher sampling rates — up to 192 kHz, compared to 44.1 kHz for standard CDs. DVD-Audio discs use the Meridian Lossless Packing (MLP), a lossless compression scheme that allows discs to hold up more information than standard PCM CDs.

3.4.1.2 mp3

mp3 stands for MPEG 1 audio layer III, and is a widely used and known compressed audio format. In 1987, the Fraunhofer IIS started to work on perceptual audio coding in the framework of the EUREKA project EU147, Digital Audio Broadcasting (DAB). In a joint cooperation with the University of Erlangen, the Fraunhofer IIS finally devised a very powerful algorithm that is standardized as ISO-MPEG Audio Layer-3 (IS 11172-3 and IS 13818-3).

From an historical point of view, three different schemes for compressing audio were available in MPEG 1, layer I, layer II, layer III. The difference for these three compression schemes were in the compression rate, and in the complexity of the decoder. But very soon after the publication of the MPEG standard, decoders were made available on PCs due to the increase in power of the computers.

mp3 achieves data compression by using perceptual coding techniques addressing the perception of sound waves by the human ear. For example, joint stereo coding takes advantage of the fact that both channels of a stereo channel pair contain far the same information. These stereophonic irrelevancies and redundancies are exploited to reduce the total bitrate. The so-called masking effects, either temporal or auditory, are also used to reduce the amount of data.

mp3 is integrated in the whole range of proprietary as well as non-proprietary multimedia solutions available in the world, including QuickTime and Macromedia Flash.

http://www.iis.fraunhofer.de/amm/techinf/layer3/index.html

3.4.1.3 AAC

AAC (Advanced Audio Coding), like mp3, is based onto a psychoacoustic model. AAC provides significantly better quality at lower bit-rates than mp3. AAC was developed under MPEG-2 and is integrated in MPEG-4. AAC supports a wider range of sampling rates (from 8 kHz to 96 kHz) and up to 48 audio channels, plus up to 15 auxiliary low frequency enhancement channels and up to 15 embedded data streams. AAC works at bit rates from 8 kbps for mono speech and up to in excess of 320 kbps for high-quality audio. Three profiles of AAC provide varying levels of complexity and scalability.

MPEG-2 AAC is the consequent continuation of the truly successful coding method ISO/MPEG Audio Layer-3. Like all perceptual coding schemes, MPEG-2 AAC basically makes use of the signal masking properties of the human ear in order to reduce the amount of data. Doing so, the quantization noise is distributed to frequency bands in such a way that it is masked by the total signal, i.e. it remains inaudible. Even though the basic structure of this coding method hardly differs from the ones of its predecessors, a closer look reveals some new aspects worth paying attention to. The crucial differences between MPEG-2 AAC and its predecessor ISO/MPEG Audio Layer-3 are:

- Filter bank: in contrast to the hybrid filter bank of ISO/MPEG Audio Layer-3 - chosen for reasons of compatibility but displaying certain structural weaknesses - MPEG-2 AAC uses a plain Modified Discrete Cosine Transform (MDCT). Together with the increased window length (1024 instead of 576 spectral lines per transform) the MDCT outperforms the filter banks of previous coding methods.
- Temporal Noise Shaping (TNS): A true novelty in the area of time/frequency coding schemes. It shapes the distribution of quantization noise in time by prediction in the frequency domain. In particular voice signals experience considerable improvement through TNS.
- Prediction: A technique commonly established in the area of speech coding systems. It benefits from the fact that certain types of audio signals are easy to predict.
- Quantization: by allowing finer control of quantization resolution, the given bit rate can be used more efficiently.
- Bit-stream format: the information to be transmitted undergoes entropy coding in order to keep redundancy as low as possible. The optimization of these coding methods together with a flexible bit-stream structure has made further improvement of the coding efficiency possible.

http://www.mpeg.org/MPEG/aac.html
http://www.iis.fraunhofer.de/amm/techinf/aac/index.html

3.4.1.4 WMA

Microsoft's Windows Media Audio (WMA) format is a relatively late entry into the field of proprietary audio formats. WMA performs very good at lower bit-rates and is reported to produce quality indistinguishable from the original CD at 128 kbps. WMA is supported by most full-featured player programs and by many portable players. WMA is royalty-free when incorporated into software that runs on the Windows platform.

http://www.microsoft.com/windows/windowsmedia/music/default.aspx

3.4.1.5 RealAudio

RealAudio was the first widely used system for streaming audio over the Internet. It is a proprietary format, but it is used by many online music stores for sample clips of songs.

RealAudio permits encoding of audio from a very low bit rate, for speech encoding, to high bit rate for quality audio.

http://www.real.com

3.4.1.6 Ogg Vorbis

Ogg Vorbis is a recent audio encoding and streaming technology. The Ogg Vorbis specification is in the public domain, free for commercial and non commercial use. Decoders are available for integration in Windows Media Player or in SMIL (RealOne player).

Like mp3 or AAC, the Ogg Vorbis format is based on a psychoacoustic model.

http://www.vorbis.com/

3.4.1.7 MPEG-4 General Audio

In trying to cover a broad range of application scenarios, the MPEG-4 audio coder includes coding tools from several different coding paradigms, such as parametric audio coding, synthetic audio, speech coding and subband/transform coding. Within this comprehensive "tool box" the high-quality part of the MPEG-4 audio functions will be covered by the so-called "General Audio (GA)" coders. In a GA coder, the input signal is first decomposed into a time/frequency (t/f) spectral representation by means of an analysis filterbank, which is then subsequently quantized and coded.

The core part of the MPEG-4 audio GA coder is based on MPEG-2 AAC technology which is complemented by a number of additional coding tools. In this way, both specific MPEG-4 functionalities (like scalability) are added and further enhancement in coding performance is achieved. The Fraunhofer IIS has been the key contributor in the progression of MPEG-2 AAC technology towards the MPEG-4 system.

Just as in MPEG-1 and MPEG-2, Audio compression is the core functionality of MPEG-4. However, the range of bitrates covered by MPEG-4 Audio and the compression ratio and sound quality have been improved much over previous systems and can be considered state of the art today. Besides this MPEG-4 has a lot additional new functionalities. Fraunhofer IIS provides technology supporting the following functionalities:

- Bitrate scalability
- Bandwidth scalability
- Encoder complexity scalability
- Decoder complexity scalability
- Error robustness tools
- Low Delay Audio Coding

MPEG-4 Audio provides several so-called profiles to allow the optimal use of MPEG-4 in different applications. At the same time the number of profiles is kept as low as possible in order to maintain maximum interoperability. MPEG-4 offers the following profiles:

- Speech Audio Profile
- Synthesis Audio Profile
- Scalable Audio Profile
- Main Audio Profile
- High Quality Audio Profile
- Low Delay Audio Profile
- Natural Audio Profile
- Mobile Audio Internetworking Profile

http://www.iis.fraunhofer.de/amm/techinf/mpeg4/audio.html

3.4.1.8 MPEG-4 Audio Lossless Coding

Lossless coding is to become the latest extension of the MPEG-4 audio standard. The MPEG audio subgroup is currently working on the standardization of lossless coding techniques for high-definition audio signals. As an extension to MPEG-4 audio, the amendment "ISO/IEC 14496-3:2001/AMD 4, Audio Lossless Coding (ALS) defines methods for lossless coding. The basic technology for MPEG-4 ALS was developed by the NUe Group at Technical University of Berlin.

- MPEG-4 ALS defines efficient and fast lossless audio compression techniques for both professional and consumer applications. It offers many features not included in other lossless compression schemes.
- General support for virtually any uncompressed digital audio format (including wav, aiff, au, bwf, raw).
- Support for PCM resolutions of up to 32-bit at arbitrary sampling rate (including 16/44.1, 16/48, 24/48, 24/96, 24/192).
- Multi-channel / multi-track support for up to 256 channels (including 5.1 surround).
- Support for 32-bit IEEE floating point audio data.
- Fast random access to any part of the encoded data.
- Optional storage in MP4 file format (allows multiplex with video).

Besides these outstanding features, a global MPEG standard for lossless audio coding will facilitate interoperability between different hardware and software platforms, and will thus promote long-lasting multivendor support. MPEG-4 ALS is expected to be finalized within year 2005.

http://www.nue.tu-berlin.de/forschung/projekte/lossless/mpeg4als.html

3.4.2 PCM-based Audio file formats

We review here the most used, PCM-based audio file formats: WAV, AIFF, AU, SDII.

3.4.2.1 WAV

WAV is the default format for digital audio on Windows PCs. WAV files are usually coded in PCM format, which means they are uncompressed and take up a lot of space. The WAV format comes from the RIFF (Resource Interchange File Format) format, which was created by Microsoft. WAV files can support many different sampling frequencies, resolutions, multiple channels, and a number of compression algorithms, but the most frequently used is raw PCM data.

WAV is the standard audio file type on computer using the Windows system, but can be used on Macintosh computers as well as on other computers (Linux). Compression can create compatibility problems on these platforms.

3.4.2.2 AIFF- AU

AIFF and the later AIFF-C is the default audio format for the Macintosh, and AU is the default format for SUN systems. Both of these formats are supported on most other platforms and by most audio applications. Each of these formats can be compressed, but compression sometimes creates compatibility problems with other platforms. The two formats are generally PCM based.

AIFF files support only PCM data. They can specify any resolution from 1 to 32, and any sample rate.

AIFF-C supports compression schemes in data chunks.

AIFF, and AIFF-C, is supported by QuickTime.

The AU format supports many resolutions, from 8 bits linear PCM to 64 bits IEEE floating point, and many different sampling rates, from 11.025 kHz to 48 kHz. AU isn't widely supported outside the UNIX community.

3.4.2.3 Sound Designer II

SDII (Sound Designer II, sometimes seen abbreviated as SD2) is a monophonic/stereophonic audio file format, originally developed by Digidesign for their Macintosh-based recording/editing products. It is the successor to the original monophonic Sound Designer I audio file format.

An SDII file can be monophonic or stereophonic. When stereo is used, the tracks are interleaved (sample-001-left, sample-001-right, sample-002-left, sample-002-right, etc.) Files also store sample rate and bit depth information.

The SDII file has become a widely accepted standard for transferring audio files between editing applications. Most Mac CD-ROM writer software, for example, specifies SDII or Audio Interchange File Format as the file format needed when making audio CDs.

The SDII file has also become accepted among PC audio application developers. This makes transferring audio from Mac to PC platforms much easier. When used on a PC, the file must use the extension of ".sd2".

Globally speaking, Sound Designer II is same as AIFF with added proprietary information such as markers and regions.

3.4.3 Multichannel audio

Concerning 3D sound a clear distinction should be made between *channel-oriented* formats like 5.1 surround sound and *object-oriented* formats like BIFS. The former is passive, platform and speaker-configuration dependent. The latter allows dynamic interaction at the object level, it is platform and speaker-configuration independent.
Multichannel audio formats like WMA, mp3 Surround and MPEG Multichannel belong to the first category, and are reported in this section. Object-oriented formats are addressed in the next section.

3.4.3.1 MPEG multichannel audio

MPEG 2 provides a backwards-compatible multichannel extension to MPEG-1; up to 5 main channels plus a low frequency enhancement (LFE) channel can be coded; the bit rate range is extended up to about 1 Mbit/s.

http://www.mpeg.org/MPEG/audio.html

3.4.3.2 mp3 Surround

mp3 Surround supports high-quality multi-channel sound at bit rates comparable to those currently used to encode stereo mp3 material, resulting in files half the size of common compressed surround formats. At the same time, the new format offers complete backward compatibility to any existing mp3 software and hardware devices.

Attracting attention at several recent events and fairs, the new mp3 Surround format has become eagerly anticipated by the vast mp3 user community. The web site www.mp3surround-format.com now provides users immediate download access to free mp3 Surround evaluation software, demo samples and detailed technology information

The evaluation encoder enables the creation of mp3 Surround material out of five or six channel ".wav" files. The Fraunhofer IIS mp3 Surround player is capable of decoding and playing back the surround format's files as well as stereo mp3 material. This software-only solution runs on any standard PC with multi-channel audio capabilities.

mp3 Surround was developed by Fraunhofer IIS in collaboration with Agere Systems. Using a psychoacoustic technique called binaural cue coding, mp3 Surround captures the spatial image information of multi-channel sound. This method is critical in achieving the compact file size that mp3 users expect.

http://www.mp3surround-format.com
http://www.iis.fraunhofer.de/amm/download/mp3surround/pressrelease.html

3.4.3.3 Windows Media Audio

Windows Media Audio 9 provides support for surround sound playback in six (5.1 audio) or eight (7.1 audio) channels.

http://www.microsoft.com/windows/windowsmedia/music/default.aspx#surroundsound

3.4.4 Object-oriented audio formats

There is no wide support for audio effects in the multimedia standards world, either in de facto standards or in open standards. The only support can be found in VRML (Virtual Reality Markup Language), an open standard dedicated to description of 3D scenes, and in MPEG (MPEG-2 and MPEG-4). There is also a support for multichannel audio in WMA. QuickTime offers also a support for multichannel audio, as it supports MPEG-4.

3.4.4.1 VRML

VRML is a standard developed by the Wed3D consortium for virtual reality, VRML standing for Virtual Reality Markup Language.

It's a SGML-based language which permits to describe 3D scenes, and contains support for MIDI and Wav audio.

VRML uses a simple model for placing sound sources in a 3D space. Sounds can be attached to objects, and to their position in a 3D space. Attenuation of sources is based on a simple, elliptical model. There is no support for reverberation, delay , filters or other types of audio effects.

Recently VRML was enhanced to the XML-based X3D format (see below).

http://www.w3.org/MarkUp/VRML/

3.4.4.2 X3D

X3D is an Open Standards XML-enabled 3D file format to enable real-time communication of 3D data across all applications and network applications. It has a rich set of features for use in engineering and scientific visualization, CAD and Architecture, Medical visualization, Training and simulation, multimedia, entertainment, educational, and more

X3D belong to object-oriented formats for the coding of 3D audiovisual scenes. It supports spatialized audio and video through audiovisual sources mapped onto geometry in the scene.

http://www.web3d.org/

3.4.4.3 MPEG 4 AudioBIFS

The AudioBIFS system is part of the Binary Format for Scene Description in the MPEG-4 International Standard. AudioBIFS allows the flexible construction of sound scenes using streaming audio, interactive presentation, 3-D spatialization and auralization, and dynamic download of custom signal processing routines. MPEG-4 sound scenes are based on a model which is a superset of the model in VRML 2.0.

In addition to the possibility to describe the acoustics of a room, AudioBIFS enables the positioning and description of virtual acoustic sources and the mixing of audio objects. Like in BIFS, composition takes place during play back according to the scene description. It is also possible to use the same audio objects in several scenes.

MPEG-4 AudioBIFS also considers the position of the user as well as the position of the acoustic sources in order to enable acoustic room simulation, which is based on information about geometry of scenes, on material characteristics of the 3D objects, as well as on room acoustics defined by perceptual parameters.

Here is a list, with short descriptions, of each of the eight BIFS nodes that comprise the MPEG-4 Version 1 AudioBIFS toolset:

- AudioSource: Enables the insertion of a sound source (connection to an elementary audio stream).
- AudioMix: Mix N channels of sound to produce M channels of sound.
- AudioDelay: Delay a sound for a short amount of time relative to the rest of the audio subgraph.
- AudioSwitch: Select N channels of sound out of a set of M channels.
- AudioBuffer: Cache sound for use in interactive playback.
- AudioFX: Execute parametric sound-effects processing given as SAOL (MPEG SA) code.
- Sound: Attach the sound created with an audio subgraph into a 3D world.
- Sound2D: Attach the sound created with an audio subgraph into a 2D scene.

In addition, a few of the general-purpose BIS nodes have associated sound behaviour:

- Group: Group multiple nodes together for hierarchical transformation.
- ListeningPoint: Specify location of virtual listener in scene.
- TermCap: Query terminal for available playback resources.

MPEG-4 Version 2 extends the simple virtual-reality model to include two rich and robust techniques for creating virtual audio environments:

- AcousticScene: Group sounds together in an auralization process.
- AcousticalMaterial: Specify reflection and transmission impulse responses for an object in a scene.
- DirectiveSound: Specify frequency-dependant directivity modelling for a sound.
- PerceptualParameters: Enable the creation and the modification of environmental acoustic effects separately for each sound source, adjusted to characterize the perceptual quality of the source and the environment in a 3D space.

http://web.media.mit.edu/~eds/mpeg4/

3.4.5 Structured audio

Structured audio is poorly integrated in multimedia frameworks. MIDI itself, while 20 years old, is not integrated in multimedia proprietary frameworks like Director or Flash. A support for MIDI is integrated in Windows Media or in QuickTime. The RealOne player supports also MIDI integrated in SMIL animations.

3.4.5.1 MIDI

MIDI (Musical Instruments Digital Interface) was developed in the early 80s, answering to the need of the growing industry of electronic musical instruments. The National Association of Music Merchandisers (NAMM) proposed in 1982 the adoption of an universal standard for transmitting and receiving musical performance information between all types of electronic instruments, which was first called UMI, for Universal Musical Interface, and finally become MIDI in 1983.

Support for MIDI is integrated in Microsoft Windows operating system and in Apple Mac OS, so MIDI is integrated in products like Windows Media Player or QuickTime.

http://www.midi.org/

3.4.5.2 Csound

Created in 1985 by Barry Vercoe, Csound is one of the most widely used software for sound synthesis. It supports several sound synthesis methods, analysis and resynthesis, support for room simulation and 3D modelling, and physical and mathematical instruments modelling. Unfortunately, Csound is not implemented in any multimedia framework. However an evolution of Csound (MPEG-SA, see below) has been integrated into MPEG.

http://www.csounds.com/

3.4.5.3 MPEG-SA

MPEG SA comes from the idea that that using CSound would be a good way to put high-quality audio on a WWW page. It was developed at the Machine Listening Group in the MIT, and was integrated into MPEG in 1997.

MPEG SA is based on the same concepts as CSound, with an orchestra and a score language, and enables a complete description of audio by mean of parametric coding rather than compression.

http://web.media.mit.edu/~eds/mpeg4/

3.4.5.4 SDIF

SDIF stands for Sound Description Interchange Format. The general idea of SDIF is to store information related to signal processing and specifically of sound, in files, according to a common format to all data types. Thus, it is possible to store results or parameters of analysis and synthesis.

SDIF is an established standard for the well-defined and extensible interchange of a variety of sound descriptions including spectral, sinusoidal, time-domain, and higher-level models. SDIF consists of a basic data format framework and an extensible set of standard sound descriptions. The SDIF standard has been created in collaboration by IRCAM, CNMAT, and IUA-UPF.

http://recherche.ircam.fr/equipes/analyse-synthese/sdif/
http://www.cnmat.berkeley.edu/SDIF/

3.4.6 Vector graphics

3.4.6.1 Postscript, PDF

Postscript is a language for description of a printed page. Developed by Adobe in 1985, it has become an industry standard for printing and imaging.

The PDF (Portable Document Format) is based on Postscript, and on the ability of almost all software on major operating systems such as Windows or MacOS to generate postscript using their widely available Postscript printing device driver. PDF content can be created by using current softwares in conjunction with Adobe's Acrobat Distiller, and viewed by using Adobe's free Acrobat Reader.

http://www.adobe.com/products/postscript/main.html
http://www.adobe.com/products/acrobat/

3.4.6.2 SVG

SVG (for Scalable Vector Graphics) is a standard (a recommendation) of the World Wide Web Consortium. SVG is a language for describing two-dimensional graphics and graphical applications in XML.

SVG contains support for all kind of 2D graphics, including b-splines, fonts, time-line based animation, and interactivity. Interactivity is supported by the mean of a scripting language, the ECMAScript [7], a standard developed by the European Computer's Manufacturer's Association. This scripting language allows a very complete interactivity with the SVG content, to which script can access by the mean of a DOM interface (the DOM interface is the standard interface developed by the W3C to XML documents).

SVG content can be produced directly from Adobe's products such as Illustrator, but also from Postscript content, by using Adobe's Illustrator as a converter. It can also be produced by using several conversion tools which are available from the w3c (http://www.w3.org) pages.

http://www.w3.org/Graphics/SVG/

3.4.6.3 OpenGL

OpenGL is an environment for developing portable, interactive 2D and 3D graphics applications. Since its introduction in 1992, OpenGL has become the industry's most widely used and supported 2D and 3D graphics application programming interface (API), bringing lots of applications to a wide variety of computer platforms. Though OpenGL is rather a 3D visualization and rendering standard than a content description one, we report it here as it could be a nice 2D interactive way to render music scores with minimum of animation.

The OpenGL® API (Application Programming Interface) began as an initiative to create a single, vendor-independent API for the development of 2D and 3D graphics applications. Prior to the introduction of OpenGL, many hardware vendors had different graphics libraries. This situation made it expensive for software developers to support versions of their applications on multiple hardware platforms, and it made porting of applications from one hardware platform to another very time-consuming and difficult. The lack of a standard graphics API was seen as an inhibitor to the growth of the 3D marketplace and prompted the creation of such a standard.

The result of this work was the OpenGL API, which was largely based on earlier work on the SGI® (Silicon Graphics Inc.) IRIS GL™ library. The OpenGL API began as a specification, then SGI produced a sample implementation that hardware vendors could use to develop OpenGL drivers for their hardware. The sample implementation has been released under an open source license (see http://oss.sgi.com).

All OpenGL applications produce consistent visual display results on any OpenGL API-compliant hardware, regardless of operating system or windowing system. OpenGL allows new hardware innovations to be accessible through the API via the OpenGL extension mechanism. In this way, innovations appear in the API in a timely fashion, letting application developers and hardware vendors incorporate new features into their normal product release cycles. OpenGL API-based applications can run on systems ranging from consumer electronics to PCs, workstations, and supercomputers. As a result, applications can scale to any class of machine that the developer chooses to target.

Modifications to the OpenGL API are made through the OpenGL Architecture Review Board, an industry group that contains founding, permanent, and auxiliary members. The current version of the OpenGL API is 1.4.

Software developers do not need to license OpenGL to use it in their applications. They can simply link to a library provided by a hardware vendor. Hardware vendors do need to have a license to create an OpenGL implementation for their hardware.

http://www.opengl.org/

3.4.6.4 Flash

The Flash format, developed by Macromedia, is mainly based on a vector graphics format, similar in functionalities to the Freehand format of the same vendor.

It is a 2D vector graphics format, comprising shapes such as circles, lines, curves (b-splines), text, and so on. Objects can be animated by using a time-line based animation, or by using scripting with the proprietary ActionScript language.

Flash content can be generated from Postscript or PDF, or by using the Freehand format.

http://www.macromedia.com/software/flash/

3.4.6.5 MPEG BIFS

MPEG Binary Format for Scenes Description makes possible to define so-called ”scenes“ consisting of several audiovisual objects which can be part of complex interactive multimedia scenarios. The individual objects are encoded and transmitted separately in a scene which is then composed after decoding of individual objects.

BIFS describes a scene as a hierarchical structure, a graph, which is composed of several nodes. More than 100 different types of nodes are defined, from media nodes such as AudioClip or MovieTexture, to nodes such as Text, or shapes such as Circles, rectangle and son on. Appearance and behaviour of nodes can be controlled by the mean of their exposed set of parameters. Certain types of nodes called sensors, such as TimeSensor or TouchSensor, can interact with users and generate appropriate triggers, causing changes in the scene appearance and behaviour. Moreover, MPEG-4 scenes interactivity can be defined by using script nodes, by using the syntax of the ECMAScript language.

MPEG BIFS can be generated by using the Envivio's 4Mation authoring environment, or generated from SVG or SMIL by using the XMT tools (see below in section 3.6.5).

3.4.7 Metadata

3.4.7.1 RDF

The Resource Description Framework (RDF), developed by the w3c, integrates a variety of applications from library catalogues and worldwide directories to syndication and aggregation of news, software, and content like personal collections of music, photos, and events using XML as an interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web.

Resource Description Framework is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web. RDF emphasizes facilities to enable automated processing of Web resources. RDF can be used in a variety of application areas; for example: in resource discovery to provide better search engine capabilities, in cataloguing for describing the content and content relationships available at a particular Web site, page, or digital library, by intelligent software agents to facilitate knowledge sharing and exchange, in content rating, in describing collections of pages that represent a single logical "document", for describing intellectual property rights of Web pages, and for expressing the privacy preferences of a user as well as the privacy policies of a Web site.

The development of RDF has been motivated by the following uses, among others:

- Web metadata: providing information about Web resources and the systems that use them (e.g. content rating, capability descriptions, privacy preferences, etc.)
- Applications that require open rather than constrained information models (e.g. scheduling activities, describing organizational processes, annotation of Web resources, etc.)
- To do for machine processable information (application data) what the World Wide Web has done for hypertext: to allow data to be processed outside the particular environment in which it was created, in a fashion that can work at Internet scale.
- Interworking among applications: combining data from several applications to achieve new information.
- Automated processing of Web information by software agents: the Web is moving from having just human-readable information to being a world-wide network of cooperating processes. RDF provides a world-wide lingua franca for these processes.

RDF is designed to represent information in a minimally constraining, flexible way. It can be used in isolated applications, where individually designed formats might be more direct and easily understood, but RDF generality offers greater value from sharing. The value of information thus increases as it becomes accessible to more applications across the entire Internet.

RDF has a simple data model that is easy for applications to process and manipulate, and independent of any specific serialization (data format) syntax. Any expression in RDF is based on an underlying structure defined as a collection of triples, each consisting of a subject, a predicate and an object. A set of such triples is called an RDF graph.

RDF has formal semantics that provides a dependable basis for reasoning about the meaning of an RDF expression. In particular, it supports rigorously defined notions of entailment, which provide a basis for defining reliable rules of inference in RDF data. As a consequence, inference rules are capable of dynamically generating RDF statements based on existing statements. For example, a simple rule can be defined for inferring that the type of a resource is a type of all of its base classes.

The vocabulary is fully extensible, being based on URIs (Universal Resource Identifiers) with optional fragment identifiers (URI references, or URIrefs). URI references are used for naming all kinds of things in RDF. Literal can also appear in RDF.

RDF has a recommended XML serialization form, which can be used to encode the data model for exchange of information among applications.

RDF can use values represented according to XML schema datatypes, thus assisting the exchange of information between RDF and other XML applications.

RDF is an open-world framework that allows anyone to make statements about any resource. In general, it is not assumed in RDF that complete information about any resource is available. RDF does not prevent anyone from making assertions that are nonsensical or inconsistent with other statements. Designers of applications that use RDF should be aware of this and may design their applications to tolerate incomplete or inconsistent sources of information.

RDF is a W3C Proposed Recommendation at the date of 15 December 2003.

http://www.w3.org/RDF/

3.4.7.2 Dublin Core

The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices.

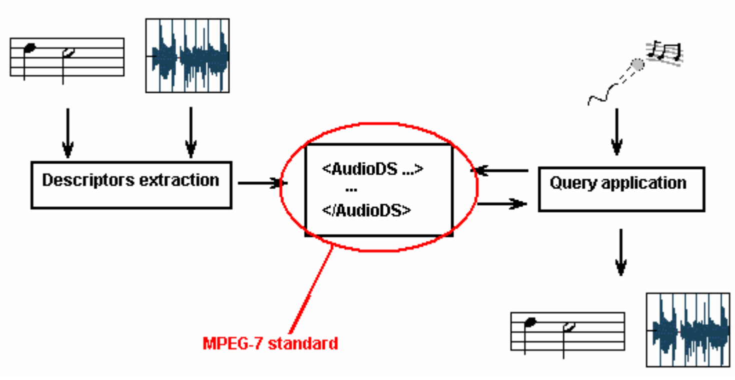
http://dublincore.org/

3.4.7.3 MPEG-7

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the standards known as MPEG-1 and MPEG-2, and the MPEG-4 standard. MPEG-7, formally named "Multimedia Content Description Interface", is a standard for describing the multimedia content data that supports some degree of interpretation of the information's meaning, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes aims to support as a broad range of applications as possible. MPEG-7 is not restricted to database retrieval applications such as digital libraries, but extended to areas like broadcast channel selection, multimedia edition and multimedia directory services.

MPEG-7 is a standard for describing multimedia content, so that users can search for that content as effectively as they actually use text based search engines. MPEG-7 standardizes the description itself, but doesn't standardize neither the method for extraction of this description nor the search engines or other applications that uses that description.

The following schema illustrates the scope of the MPEG-7 standard:



Typical applications of MPEG-7 in the audio domain are:

- Query by humming: search for a song by humming or whistling a tune. An example of query by humming is available at http://www.musicline.de/de/melodiesuche/input.
- Query by example: given a sound excerpt, search for sounds in a database having similar characteristics.

MPEG-7 doesn't standardize the process of descriptors extraction from media content, so that process can be either a manual or an automatic process.

MPEG-7 descriptions can be located with the associated material, or located elsewhere on the network. In this case, a mechanism linking description with the content is needed. MPEG-7 can also be used independently of other MPEG standards, for example for describing CD audio content or even analog content.

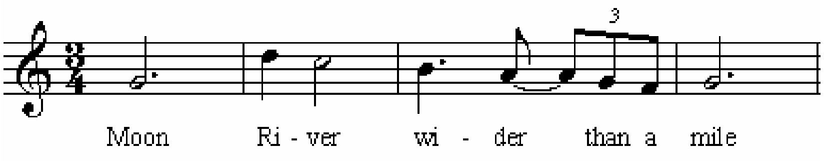
Description Definition Language, Description Schemes and Descriptors

MPEG-7 is XML based. It defines the Description Definition Language (DDL), which is an XML-based language which allows the creation of new MPEG-7 Description Schemes and descriptors. A particular DDL specifies the constraints that a valid MPEG-7 description should respect.

Description Schemes specify structure and semantics of relationships between components of a particular Description Scheme, which can be either Descriptors or descriptions Schemes. Descriptors describe the syntax and semantics of a particular feature.

For example, the MelodyDS Description Scheme describes melody by describing features such as meter, key, and scale, which are optional, and describes melody itself by using one of two options: the first, named MelodyContour, describes melody by intervals, using only five levels of information, and the second, named MelodySequence, describes the melody by precise intervals.

Here is an example of a melody description:



```
<AudioDS xsi:type="MelodyType">
  <MelodyMeter>
    <Numerator>3</Numerator>
    <Denominator>4</Denominator>
  </MelodyMeter>
  <MelodyScale>1 2 3 4 5 6 7 8 9 10 11 12</MelodyScale>
  <MelodyKey mode="otherMode">
    <KeyNote display="sol">G</KeyNote>
  </MelodyKey>
  <Contour>
    <ContourData>2 -1 -1 -1 -1 -1 1</ContourData>
  </Contour>
  <MelodySequence>
    <StartingNote>
      <StartingFrequency>391.995</StartingFrequency>
      <StartingPitch height="4">
        <PitchNote display="sol">G</PitchNote>
      </StartingPitch>
    </StartingNote>
    <NoteArray>
      <Note>
        <Interval>7</Interval>
        <NoteRelDuration>2.3219</NoteRelDuration>
        <Lyric>Moon</Lyric>
        <PhoneNGram>m u: n</PhoneNGram>
      </Note>
      <Note>
        <Interval>-2</Interval>
        <NoteRelDuration>-1.5850</NoteRelDuration>
        <Lyric>Ri-</Lyric>
      </Note>
      <!— Remaining notes are elided -->
    </NoteArray>
  </MelodySequence>
</AudioDS>
```

MPEG-7 textual descriptions tend to become very large in size and thus inefficient. For this reason, a binary format (BiM) has been defined which makes possible compression of these descriptions.

http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

3.5 Symbolic music coding

Music notation codes are so many that a complete, exhaustive review of such formats would be impossible. As music has been one among the first early applications of computers, the evolution of music codes have been driven by the joint evolution of platforms, systems, software, etc. Today music formats mainly differs depending on their use case: Music notation, real-time performance, content description, control of synthesis, etc.

A naïve, intuitive approach for classifying today's music formats would be the following: Binary formats, ASCII formats and XML-based formats. We review here nowadays most commonly used music codes. For a more complete study please refer to E. Selfridge-Field's *Beyond MIDI* [8] or visit http://www.music-notation.info.

- Binary formats: MIDI, RMTF, Finale, Sibelius, XMF, NIFF, Capella, SASL, SCORE, RMF, SSS.
- ASCII formats: PDF, GUIDO, abc, MusiXTeX, MusicTeX, MuseData, OMNL, CHARM.
- XML-based formats: SMDL, MNML, MML, XScore, MusicXML, GUIDO XML, WEDELMUSIC, MDL, SVG, SMIL.

As it is not the main purpose here, we will not go further in details for these formats. Some of them are finely detailed in MusicNetwork's deliverables DE4.1.1 and DE4.1.2, devoted to music notation coding, and available from the MusicNetwork's web page. Are reviewed in these deliverables the formats that bear some significant interest in the scope of our multimedia music definition (see section 1.1.2), in terms of integration and synchronization with other media.

3.6 Multimedia frameworks

We review here the main multimedia frameworks, in their actual state. We don't speculate about their future, nor about their possible evolutions. The landscape is likely to evolve quickly, some of the main actors will probably disappear, and possibly new actors will appear. The relative weight of the actors also is likely to evolve, and we don't take here any position on this point.

3.6.1 Flash

Flash is a proprietary system developed by Macromedia, Inc. It is composed of an authoring tool, and a viewer available for free on the most widely available platforms (Macintosh and Windows based). The viewer can be integrated in an HTML page, so Flash content can be easily integrated in Web content.

Flash is based on vector graphics. In conjunction with scripting (Actionscript), and animation (timeline based as well as scripting based), this has made of Flash the most widely used multimedia framework for the Web. The very low bit rate induced by the use of vector graphics, scripting, and timeline based animation has made of Flash a very convenient format for Web-based animation, storyboards, high quality graphics, and a suitable alternative to HTML for the development of Internet web sites.

Following a survey conducted by NPD Online in June 2003, 97.4% of Web users can experience Flash content, having the Flash player already installed.

But even if Flash is considered as the most widely used multimedia framework for the Web, its poor support of audio and music makes it not very well suited for multimedia music applications. As compressed audio format, Flash supports only the mp3 format, which is now more than 10 years old, and which suffers from actually known disadvantages, such as audible artefacts, and a compression ratio which is not more at the level of the actual state of the art. Flash doesn't support MIDI, and doesn't support any kind of structured audio, audio effects or 3D audio.

Audio and music support

The support of audio can be considered as being poor in Flash. The only compression scheme supported is mp3, together with a proprietary compression scheme known as “Nelly Mosser” for which no information is available to our knowledge. Flash supports also uncompressed schemes, PCM based, such as WAV or AIFF.

For structured audio, no format – even MIDI - is supported.

Authoring and production

Flash benefits from a proprietary authoring tool developed by Macromedia.

Graphics - Scores

The ability of Flash to import encapsulated postscript (EPS) and Freehand format makes Flash a possible useable technology for scores diffusion, in a vector graphic format.

Interactivity, animation

Interactivity and animation can be implemented in Flash by using ActionScript, a proprietary scripting language

Openness, extensibility

Flash is in principle a closed system on the client’s side. No extensions can be developed, no decoders can be added, and no interactivity other than interactivity defined on the authoring side with the ActionScript scripting language can be defined to enhance the standard viewer (this is not to be confused with the extensibility functions available in the new Flash MX 2004, which are available in the Flash authoring application).

Flash is open to XML, and able to exploit XML data in a client-server architecture, via http-based protocol, or via XML socket based, real-time exploitation of data. With this functionality, it’s for example possible to imagine a Flash client application exploiting XML data available on line, for example XML-based metadata such as RDF, or Dublin Core, or even MPEG-7 metadata in their XML format.

The Flash file format is itself now open, as well as some parts of the source code, and many developers are developing new Flash based solutions. For example, the NorthCode company (www.northcode.com) has developed SWFStudio (http://www.northcode.com/swfstudio/), a software which makes possible to build stand-alone executables from Flash content. In this configuration, it becomes possible to build plugins to Flash executables. The same society has developed a plugin development kit in order for other developers to build their own extensions to Flash (with the restriction that this works only with Flash stand alone applications – it’s always impossible with the Flash standard client).

http://www.macromedia.com/software/flash/

3.6.2 Director

Director, developed by Macromedia, is an authoring tool for publishing interactive multimedia content. It is based on the concept of a timeline where actors (cast members) are placed and can be animated, and interact with user. Director use a scripting language (Lingo) to develop user’s interaction and animation.

Director makes it possible to build executables which are directly playable on the user’s device. Director can also be used as an authoring tool for other types of content, like Apple QuickTime.

Director is mainly used to develop CD-Rom based multimedia products, but is also sometimes used to develop Web-based products.

Audio and music support

The support of audio can be considered as being poor in Director as well as in Flash. The only compression scheme supported is mp3. For structured audio, no format – even MIDI - is supported. A support for mixing in real time of multiple audio is available. Director supports also uncompressed schemes, PCM based, such as WAV or AIFF.

Authoring and production

Director benefits from a proprietary authoring tool developed by Macromedia.

Graphics - Scores

Director supports embedding of Flash objects. The ability of Flash to import encapsulated postscript (EPS) and Freehand format makes Flash a possible useable technology for scores diffusion, in a vector graphic format.

Interactivity, animation

Interactivity and animation can be implemented in Director by using the Lingo proprietary scripting language.

Openness, extensibility

Director provides a Software Developement Kit for developing external functions on the user's side as well as on the server's side (plugins). This SDK let the developer with a working knowledge of the C language, and with a working knowledge of computers systems, develop extensions called Xtras to Director in the following areas: Sprites, Transitions, Lingo, Tool and Multiuser Xtras . Sprites provide a way to add new media data types. Transitions provide a way to extend the list of available transitions. Script (Lingo) Xtras provide a way to add new commands to the Scripting language. Tool Xtras are used to extend the functionality of the authoring environment. Multiuser Xtras extend the functionality of the Multiuser server.

Xtras developed for Director are also compatible with Authorware

There is a wide community of developers developing Xtras for Director. Some of these are available through the Macromedia web site.

http://www.macromedia.com/software/director/

3.6.3 Windows Media

Being preinstalled with every version of Microsoft Windows sold, Windows Media Player is becoming increasingly widespread on the web.

Audio and music support

Windows Media Player supports proprietary compressed audio (Windows Media Audio), and supports also MIDI. WMP supports multichannel audio, in multiple configurations (5.1, 7.1).

Authoring and production

Production of Windows Media content can be done in multiple ways: by the mean of Windows Media Encoder, or by the mean of the toolkits provided by Microsoft for this purpose. These toolkit can be accessed by the mean of the C++ language, the Visual Basic language, or even by the mean of an HTML interface.

Graphics - Scores

There is no support for vector graphics in Windows Media, making it not very suitable for diffusion of content with music scores

Interactivity, animation

No support of interactivity – scripting, controls – is directly available in Windows Media.

Openness, extensibility

Customization of the Windows Media Player is possible by using the Software development Kit provided to this end by Microsoft. By using the SDK, it's possible to develop a customized end-user interface driving the Windows Media content, in any language supported by the Windows Media SDK (C++, Visual Basic, HTML, .net with C#…).

http://www.microsoft.com/windows/windowsmedia/

3.6.4 QuickTime

Apple doesn't provide authoring tools, but software like Adobe's Premiere or Macromedia Director are able to produce QuickTime content, generally by the mean of a plug-in. Apple's QuickTime-related products (QuickTime Pro, QuickTime MPEG-2 Playback and QuickTime Broadcaster) are not literally authoring tools, but provide however a few creating, encoding and editing functionalities.

Recent releases of QuickTime are able to support the MPEG-4 format.

Apple's products.

- QuickTime 7 Pro enables H.264 video creation, audio and video capture, multi-channel audio creation and multiple files export. It is an easy-to-use tool for creating AAC audio files and 3G files for mobile viewing, editing videos and exporting movies.
- QuickTime MPEG-2 Playback Component provides QuickTime users with the ability to import and play back MPEG-2 content, including both multiplexed and non-multiplexed streams. It is suited for content creators with projects such as Professional content production and transcoding video content (from MPEG-2 video to MPEG-4 for example).
- QuickTime Broadcaster is a tool for producing live broadcast events.

Audio and music support

QuickTime 7 Player supports a wide-range of industry-standard audio formats, including AIFF, WAV, MOV, mp3, MP4 (AAC only), CAF and AAC/ADTS. For structured audio, QuickTime supports MIDI. There is no support for audio effects or 3D audio. Multichannel audio is supported by QuickTime 7 up to 24 audio channels, enabling standard surround formats (2.1, 5.1 and 7.1).

Authoring and production

Adobe's Premiere or Macromedia Director can generate QuickTime content. There are also a number of production tools available, like FinalCut Pro for instance.

Scores

QuickTime supports a native vector graphic format, but the lack of authoring tool for this format makes it not very suitable for multimedia authoring. QuickTime can integrate Flash content, saved from the Flash application directly as a QuickTime movie, either as a movie with a video (bitmapped) track or as movie with a Flash track The Flash track retains its native format, this means that Flash vectors are not converted into bitmaps or QuickTime vectors. Bitmapped graphics embedded in a Flash .swf remain bitmaps after import into QuickTime.

Interactivity, animation

No scripting language is available for defining interactivity, but interactivity can be defined by using the QuickTime Software Development Kit provided by Apple.

Openness, extensibility

Extensions to QuickTime can be defined on the user's side by using the QuickTime Software Development Kit provided by Apple. It provides interfaces in C or Java QuickTime content can be embedded in a web page, but only a restricted set of functions are available from scripting languages such as JavaScript, making QuickTime not very well suitable for development of interactive content on the Web.

Timeline-based, raw graphics animation is provided by authoring tools such as Adobe Premiere or Macromedia Director.

QuickTime & 3GPP

3GPP (and 3GPP2) stand for Third Generation Partnership Project. It s a collaborative third generation (3G) telecommunications specifications-setting project, comprising North American and Asian interests developing global specifications for ANSI/TIA/EIA-41 Cellular Radiotelecommunication Intersystem Operations network evolution to 3G, and global specifications for the radio transmission technologies (RTTs) supported by ANSI/TIA/EIA-41.

3GPP2 was born out of the International Telecommunication Union's (ITU) International Mobile Telecommunications "IMT-2000" initiative, covering high speed, broadband, and Internet Protocol (IP)-based mobile systems featuring network-to-network interconnection, feature/service transparency, global roaming and seamless services independent of location. IMT-2000 is intended to bring high-quality mobile multimedia telecommunications to a worldwide mass market by achieving the goals of increasing the speed and ease of wireless communications, responding to the problems faced by the increased demand to pass data via telecommunications, and providing "anytime, anywhere" services.

3GPP and 3GPP2 have become worldwide standards for the creation, delivery and playback of multimedia over 3rd generation, high-speed wireless networks. These standards seek to provide uniform delivery of rich multimedia over newly evolved, broadband mobile networks (3rd generation networks) to the latest multimedia-enabled cell phones and QuickTime has adopted 3GPP & 3GPP2 into the core architecture giving QuickTime users mobile multimedia capabilities. The QuickTime family of standards-based products therefore now provides solutions for the creation, delivery and playback of mobile multimedia over 3G networks.

http://www.apple.com/quicktime/
http://www.3gpp2.org/

3.6.5 MPEG

MPEG has been developed since 1988 by the Moving Picture Expert Group, a working group of ISO. This working group is devoted to the development of standards for coded representation of digital audio and video. Established in 1988, the group has produced MPEG-1, the standard on which such products as Video CD and mp3 are based, MPEG-2, the standard on which such products as Digital Television set top boxes and DVD are based, MPEG-4, the standard for multimedia for the fixed and mobile web and MPEG-7, the standard for description and search of audio and visual content. Work on the new standard MPEG-21 "Multimedia Framework" has started in June 2000.

We refer here mainly to the MPEG-4 standard, but parts of the MPEG-7 standard are of interest for our purpose, as well as parts of the early MPEG-1 and MPEG-2 standards.

MPEG-4 systems

MPEG-4 offers an extensive systems part supporting scene descriptors, object descriptors and all related features, such as Intellectual Property Management and Protection (IPMP) provisions. This system allows for simple (old style) single source audio coding as well as complex, object oriented audio scenes. In order to make full use of all of the MPEG-4 features, a transport mechanism must support the MPEG-4 systems requirements.

In addition MPEG-4 defines the following transport formats, although other transmission systems with different application specific transport formats are possible as well:

- MPEG-4 Flexmux: This is not really an transport format but rather an interface description between MPEG-4 and an arbitrary transport format. During the development of MPEG-4 it served with some additions as a simple file format.
- MPEG-4 File Format: This is the file format for MPEG-4 content. It is derived from Apple’s QuickTime media format and supports e.g. editing and streaming of MPEG-4 content.
- LOAS (Low overhead audio stream): This is a audio-only transport format for applications where an MPEG-4 audio object needs to be transmitted and additional transport overhead is an issue.
- LATM (Low overhead Audio Transport Multiplex): This is the multiplex part of the LOAS described above. Can be used if only a multiplex is needed.
- MPEG-4 over RTP (Real Time Protocol): This is protocol defined by the IETF (Internet Engineering Task Force), but there are several ways how to transmit MPEG-4 content over IP networks (e.g. the Internet) using RTP. One possibility to transmit MPEG-4 Audio is to use LATM and RTP.

Audio and music support

MPEG supports many compressed audio schemes, the most relevant being mp3 and more recently AAC. AAC provides a quality indistinguishable from CD quality at range of 64 kbits/sec (for mono channel audio.

MPEG supports structured audio (MPEG SA) and MIDI (integrated in MPEG SA). It supports also sound effects, 3d audio and multichannel audio.

Authoring and production:

There are a lot of authoring tools dedicated to MPEG-4 content: MPEG-4 Studio (Kyungpook National University, 2D), MPEG-4 Toolbox (Institute for Research and Technology Hellas, 3D), Maxpeg Author (Digimax & National Taiwan University, 3D), 4Mation (Envivio, 2D), Studio Author (iVast, 2D), V4Studio (GPAC, 2D), XMTEdit (IBM, 2D), Harmonia (ENST).

XMT

XMT is a framework for representing MPEG-4 scene description using a textual syntax. XMT allows content authors to exchange their content with other authors, tools or service providers, and facilitates interoperability with both the X3D, developed by the Web3D consortium, and the Synchronized Multimedia Integration Language (SMIL) from the W3C consortium. Automatic transcription tools from SMIL to XMT, and then from XMT to BIFS are available.

Scores

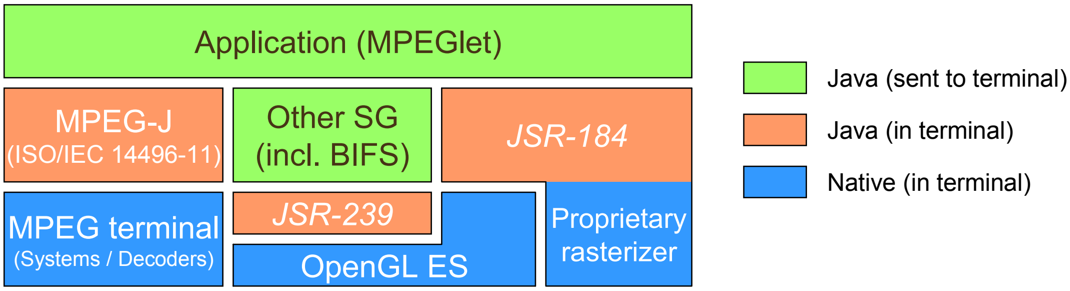
MPEG-4 BIFS offers a range of 2d graphics primitives suitable to represent vector graphics content such as music scores.

Interactivity, animation:

MPEG has recently started a new activity aiming at the definition of a new Graphics API as an extension of MPEG-J[1]. The current status of the technical issues and requirements is reported in the WD 1.0 of this new specification (MPEG-4 Part 21).

The graphics API provides low-level access to rendering methods so that applications can create their own scenes, special effects, etc. in a faster and more scalable way than it is possible today with the existing MPEG-J APIs.

To enable as many applications as possible, MPEG-4 Part 21 proposes the use of the well-known industry standard OpenGL and in particular the ES version targeted at embedded systems. The binding from Java to OpenGL is defined by the JSR-239 expert group. However, some applications may prefer using a simple scene graph and possibly a proprietary rasterizer optimized for such scenes. In this case, the specification defined by JSR-184 expert group is recommended. The figure below, taken from Part 21 WD 1.0 (w6549), depicts the block organization of systems and APIs in an MPEG-4 terminal using the proposed specification.



The java MPEGlets (delivered with the content) can be used:

- to directly interact with the graphic sub system (OpenGL API);
- to implement functionality similar to that of the basic graphic BIFS Nodes;
- to directly affect the Decoder (such as MPEG-J API of MPEG-4 Part 11).

Metadata

MPEG-7 implements a very rich framework for metadata. MPEG-7, formally called "Multimedia Content Description Interface" standardize a set of description schemes and descriptors, a language to specify description schemes (the Description Definition Language DDL), and a scheme for coding these

[1] MPEG-J is a java-based API built into the MPEG-4 format. Developers can look at this to extend the functionality of MPEG-4. An MPEG-J “MPEGlet” delivered to the end-user might be used to control the media objects in the MPEG-4 streams and enhance the interactivity of the media experience.

descriptions. It enables the needed effective and efficient access (search, filtering and browsing) to multimedia content. It permits fine-grained description and access to segments and sections of medias.

Openness, extensibility

In principle, extensions to the MPEG framework can only be made by participating directly to MPEG works, and by working in relationship with all the members of the WG. This ensures that a strong interoperability can be maintained in all parts of the MPEG framework, and that the standard can be maintained.

The domains actually covered by MPEG, in the audio domain as well in the scene composition, interactivity and all domains identified above for multimedia music, makes actually from MPEG the most suitable framework for that purpose.

http://www.chiariglione.org/mpeg/
http://www.mpeg.org/MPEG/index.html
http://www.iis.fraunhofer.de/amm/techinf/mpeg4/systems.html

3.6.6 SMIL

The Synchronized Multimedia Integration Language (SMIL, pronounced "smile"), developed by the World Wide Web Consortium (W3C), enables simple authoring of interactive audiovisual presentations. SMIL is typically used for "rich media"/multimedia presentations which integrate streaming audio and video with images, text or any other media type. SMIL is an easy-to-learn HTML-like language, and many SMIL presentations are written using a simple text-editor.

SMIL doesn’t define any particular type of media (such as vector or raster images, videos, text, or audio data). Instead of media content, SMIL describes media composition, that is the layout of the different elements on the screen, as well as their time attributes. SMIL describes temporal and spatial organisation of media while not defining the content itself. Every type of media – even Flash content – can be part of a SMIL animation.

The SMIL language is based on XML, and thus is text based, making it very easy to generate, even from a database and a middleware.

Versions

The first version of SMIL (SMIL 1.0) has been published in November 1997, and a second version (SMIL 2.0) has been published in August 2001. Some players which were compatible with version 1.0 are not compatible with version 2.0 of SMIL (QuickTime).

The World Wide Web consortium has also defined a specification, named XHTML + SMIL, which integrates a subset of the SMIL 2.0 specification with XHTML. It includes SMIL 2.0 modules providing support for animation, content control, media objects, timing and synchronization, and transition effects. The SMIL 2.0 features are integrated directly with XHTML and CSS, and can be used to manipulate XHTML and CSS features. The profile is designed for Web clients that support XHTML+SMIL markup. Internet Explorer 6.0 supports XHTML + SMIL.

Viewers

The only complete player available is the RealPlayer from RealNetworks. There is also a player available (not for free) from Oratrix, but the player is to be considered more as a “reference software”, that is, a software to be used by implementers or developers to test compatibility with their own products, than an end-user product.

Internet Explorer 6.0 implements a support for XHTML+SMIL profile.

Audio and music support

Wav, AIFF, or mp3 are generally supported in viewers. Proprietary viewers such as RealPlayer or QuickTime supports also proprietary audio media types. This way, MIDI content can be integrated in a SMIL animation, but will be rendered correctly in players supporting this content type.

Authoring and production

SMIL being an open standard, there are a lot of production and authoring tools available. An interesting environment is the Gr*i*NS pro Editor for SMIL 2.0. This environment creates also presentations for RealOne, HTML+TIME or 3GPP/Mobile.

In addition, SMIL being XML-based, there are numerous ways to generate SMIL from a database and a middleware.

Scores

Being not a media content descriptor, SMIL doesn't support by itself vector-graphics content. SVG content can be integrated in a SMIL animation.

Interactivity, animation

Interactivity in SMIL is very poor, due to the lack of a scripting language.

A very rich framework for animation has been developed. SMIL being XML-based, it's composed of elements which have attributes such as position, transparency. Almost every attribute of every element can be animated this way. SMIL provides also support for video transition effects (fade, push…).

Metadata

The earlier SMIL 1.0 specification allowed authors to describe documents with a very basic vocabulary using the "meta" element. In the version 2.0 of SMIL, the same element is supported, but new capabilities have been introduced for describing metadata using the Resource Description Framework Model and Syntax.

Openness, extensibility

SMIL is theoretically the most widely open multimedia framework. In principle, almost every type of media can be integrated in SMIL, provided that a specific decoder is available in the viewer on the client side for that specific media type.

In practice, this means that the inclusion of media type and their compatibility is dependent on the viewer used on the client side. We are reviewing below the Real Player implementation from the point of view of inclusion of media type of particular interest for our purpose : audio and vector graphics.

http://www.w3.org/AudioVideo/

3.6.7 RealMedia

RealNetworks media are limited to audio, from speech, monochannel to surround, 5.1 channel music, and video. There is no native support for interactivity, vector graphics, but the Real Player supports the W3C's standard for synchronized multimedia SMIL (see section 3.6.6), and thus interactivity, animation and support of vector graphics can be integrated this way.

Audio and Music support

Real audio supports compression of audio from 16 Kb/sec (monophonic, very low quality) to 352 Kb/sec (stereophonic, very high quality).

RealAudio Surround codecs preserve the matrixed multi-channel surround audio in conventional "surround sound" audio. Surround audio can consist of four sound channels (left, right, left surround, and right surround) or 5.1 channels (additional subwoofer and center).

Authoring and production

The Helix producer enables encoding of streaming media (audio, video), in the native Real formats, with different bit rates. The Helix producer cannot generate SMIL animations – but can generate the included media.

Scores

Real Media doesn’t include any native support for vector graphics, but vector graphics can be integrated in SMIL animations by integration of Flash or SVG content.

Interactivity, animation

Real Media doesn’t include any native support for interactivity or animation, but these functionalities can be integrated in SMIL animations by integration of Flash or SVG content.

Metadata

There is no support for metadata in Real Media.

Openness and extensibility

Real Media doesn’t include any native support for scripting or extensibility, but the SMIL support for these features must be taken in account.

RealPlayer

- Audio: The Real Player includes support for its audio proprietary format as well as for mp3 format. It’s also possible to include MIDI content.
- Vector graphics, animation: Real Player supports the integration of Flash content (only Flash 3 and Flash 4), with some restrictions such as for audio content, which must be integrated using another channel (mp3, or rm). Interaction with Flash content is also supported, enabling in this manner capabilities of interactions with timeline from the user . It’s for example possible to develop a simple user interface in Flash, composed of some buttons for playing, stopping, fast reviewing or forwarding an audio track, but this kind of interaction will be limited to interaction with the timeline.
- Real Player supports also integration of SVG.

http://www.realnetworks.com/info/real10_platform/

3.6.8 Comparison of multimedia frameworks

Here are reviewed the most important features of multimedia frameworks, from the specific point of view of music requirements as expressed in section 2. Video features are not covered here, since they are outside of the scope of this document.

	Flash	Director	Windows Media	QuickTime	SMIL (RealPlayer)	MPEG 4
Production						
Authoring tools	Proprietary (Macromedia)	Proprietary (Macromedia)			GriNS (Oratrix)	Envivio 4Mation
Encoding			Windows Media Encoder (Microsoft), Adobe Premiere (plugin)	Macromedia Director, Adobe Premiere FinalCut Pro		XMT
Viewers						
Computer	Flash (Macromedia)	Shockwave (Macromedia)	Windows Media Player (Microsoft)	QuickTime Player (Apple)	Realplayer One, Internet Explorer 5.5, Internet Explorer 6 (XHTML + SMIL), QuickTime 4	QuickTime, Envivio
Mobile				3GPP, 3GPP2	GriNS (SMIL Basic, SMIL 2.0)	
Support						
Audio	mp3, PCM	mp3, PCM	mp3, PCM, WMA	mp3, AAC	mp3, wav, AIFF	mp3, AAC
Audio effects						AudioBIFS
3D audio						AudioBIFS
Multi channel audio			5.1, 7.1	5.1, 7.1	5.1 (RealPlayer)	5.1
Structured audio			MIDI	MIDI	MIDI	MIDI, MPEG SA
Vector graphics	Flash (Free Hand)	Flash			Flash, SVG (RealPlayer)	BIFS, 2D graphics
Interactivity, scripting	ActionScript	Lingo	Visual Basic, .net...			JBIFS
Metadata						
					RDF	MPEG 7
Extensibility						
		Plugins (C)	C, C++, C#, Visual Basic	C, Java		

4 Uses: Review of on-line musical publications

This section aims at giving the evidence of emerging needs for multimedia support in music content distribution. A quick glance at some music-related French websites gives a fair overview of the massive, increasing use of technology for providing music content through multimedia platforms.

4.1 Hypermedia: A new platform for musicology

The MINT department (Musicology, Computer Science & New Technologies) of OMF (Observatoire Musical Français) was created in 1997 to study how musicology could benefit from multimedia technologies (electronic publications, document browsing, navigation systems, training courses, etc). A web site is devoted to reference available Internet resources for musical analysis, and review more than 200 different on-line publications related to music and musical heritage, making use of multimedia technologies. Special thanks must be given to Pierre Couprie and OMF for this review of hypermedia analysis on the Web.

http://www.omf.paris4.sorbonne.fr/omf-articles.php3?id_rubrique=20&id_article=30
http://www.omf.paris4.sorbonne.fr/ANAnews.php3

4.2 On-line papers on the way to multimedia music

The publications listed below have been reviewed for the presence of audio, score, or other musical representation like sonogram (spectral analysis of sound). We have also analysed the different publications searching for a form of synchronisation between audio and representation of music like score or sonogram. The presence of a synchronisation is summarized below in the column entitled “visual following”.

These publications have been analysed as regarding interactivity. All publications have a minimum form of interactivity, which mainly consists of hyperlinks. But a number of publications have more advanced forms of interactivity, like listening games:

http://www.ac-dijon.fr/pedago/music/bac2002/risset/index.html

When these more advanced forms of interactivity have been detected, they are summarized below in the column “Interactivity”.

Most frequently used media and multimedia features:

- Audio
- Vector & raw graphics
- Synchronisation
- Interactivity

Most frequently used music representations:

- Scores
- Sonograms

Most frequently used technologies:

- Audio: QuickTime, Real, mp3, MIDI, AVI and Cubase.
- Scores: Finale or PDF in few cases.

	Audio	Score	Sonagram	Visual following	Interactivity	Shockwave	Flash	QuickTime	Real	mp3	MIDI	AVI	CD Link	Finale	PDF	Cubase
http://atelier.feuillantine.free.fr/analyse/bach/CBT1/BWV8	X	X		X		X		X								

71/1-intro.html																
http://atelier.feuillantine.free.fr/analyse/messiaen/rouss1.html	X	X		X		X										
http://fboffard.free.fr/	X	X			X		X									
http://jan.ucc.nau.edu/~tas3/bachindex.html	X	X											X			
http://musique.baroque.free.fr/constantes.html	X	X							X							
http://patriciagray.net/musichtmls/flash/mozart.html	X	X		X	X		X									
http://perso.club-internet.fr/phillal/PAGES/ARCANA/arcana.html		X														
http://perso.club-internet.fr/phillal/PAGES/deserts.html										X						
http://perso.club-internet.fr/phillal/PAGES/HPPM/Guigue/guigue.html		X														
http://perso.club-internet.fr/phillal/PAGES/HPPM/Jodlowski/jodlo1.html		X														
http://perso.club-internet.fr/phillal/PAGES/integrales.html		X														
http://perso.club-internet.fr/phillal/PAGES/IONISA/ionisation.html		X														
http://perso.wanadoo.fr/josquin.desprez/	X	X									X				X	
http://perso.wanadoo.fr/mhenninger/artic/beeth/beeth.htm		X														
http://perso.wanadoo.fr/mhenninger/artic/passion/passion.htm		X														
http://www.cite-musique.fr/gamelan/	X			X	X	X										
http://www.ac-bordeaux.fr/Pedagogie/Musique/aria3sui.htm		X									X					
http://www.ac-bordeaux.fr/Pedagogie/Musique/grisey.htm																X
http://www.ac-bordeaux.fr/Pedagogie/Musique/symp1.htm		X														
http://www.ac-bordeaux.fr/Pedagogie/Musique/pagwop10.htm		X									X					
http://webpublic.ac-dijon.fr/pedago/music/bac2002/risset/pages/risset.html	X		X	X	X		X									
http://www.ac-grenoble.fr/Partiels/	X		X	X	X	X										
http://www.aeiou.at/bt133.htm	X	X						X		X		X				
http://www.aeiou.at/bt-ero.htm	X	X						X		X		X				
http://www.aeiou.at/bt-sym5.htm	X	X						X		X		X				
http://www.aeiou.at/bt-sym6.htm	X	X						X		X		X				
http://www.aeiou.at/bt-sym7.htm	X	X						X		X		X				
http://www.aeiou.at/bt-sym9.htm	X	X						X		X		X				
http://www.colleges.org/~music/modules/op11/	X	X		X			X									
http://www.colleges.org/~music/modules/pierrot/	X	X		X			X									
http://www.colleges.org/~music/modules/vox.html	X	X					X	X								
http://www.ethnomus.org/ecoute/animations/badong/badong.html	X						X									
http://www.ethnomus.org/ecoute/animations/diphonique/hai1.html	X		X	X	X		X									
http://www.ethnomus.org/ecoute/animations/nzakara/nzakara.html	X	X		X	X		X									
http://www.ethnomus.org/ecoute/animations/quintina/seq1.html	X	X	X	X	X		X									
http://www.geocities.com/Athens/Agora/1985/analys.html	X								X							
http://www.ina.fr/grm/acousmaline/polychromes/ferrari/index_fer.html	X	X	X	X	X	X										
http://www.ina.fr/grm/acousmaline/polychromes/parmegiani/index.html	X	X	X	X	X		X									
http://www.ina.fr/grm/acousmaline/polychromes/racot/inde	X	X	X	X	X		X									

x.html																
http://www.ina.fr/grm/acousmaline/polychromes/sud/index_sud.html	X	X	X	X	X	X										
http://www.ommadawn.dk/mou/omm/analysis.html		X														
http://www.patriciagray.net/musichtmls/Flash/fille.html	X	X		X			X									
http://www.teoria.com/articulos/analysis/BWV772/index.htm	X	X									X			X	X	
http://www.teoria.com/articulos/analysis/chopin4/index.htm	X	X									X			X		
http://www.teoria.com/articulos/analysis/chopin8/index.htm	X	X									X			X		
http://www.teoria.com/articulos/analysis/debussy8/index.htm	X	X									X			X		
http://www.teoria.com/articulos/analysis/BWV861/index.htm	X	X									X			X	X	
http://www.teoria.com/articulos/analysis/BWV846/index.htm	X	X									X			X	X	
http://www.uwsp.edu/music/jsobaski/b-intro.htm		X														
http://www.zappa-analysis.com/	X										X					

5 Music notation standardization projects

Are reported in this section ongoing or past standardization projects for music notation, that claim a strong interest for bringing music content into the multimedia age, i.e. aiming at defining an open standard that can be easily synchronized with other media, and easily integrated in a global multimedia framework.

5.1 WedelMusic

WEDELMUSIC is an innovative idea to allow the distribution and sharing of interactive music via Internet totally respecting the publisher rights and protecting them from copyright violation. WEDELMUSIC allows content distributor (publishers, archives, etc.), corporate consumers (theatres, orchestras, music schools, libraries, music shops), and users (students, musicians, etc.) to manage interactive multimedia music in WEDELMUSIC XML format. WEDELMUSIC is a project, a format, a technology and a set of tools altogether.

Project

WEDELMUSIC stands for WEb DELivery of MUSIC and is a European IST project for the delivery of musical scores on the Internet. It aims at interactive music delivery through Internet with respect to copyrights and DRM. It also addresses special issues like accessibility for impaired people. The WEDELMUSIC is now terminated since 2002.

Format

A WEDELMUSIC object contains data for the delivery of all kind of music contetn (audio, symbolic, image). It has an identification number and an archive backup, and authorizations are defined for its manipulation. WEDELMUSIC objects may include:

- Multilingual cataloguing information,
- Music notation scores,

- Audio files (e.g., WAVE, mp3, MIDI),
- Image of music scores (PNG, TIF, PDF, etc.),
- Multilingual lyric (XML)
- Video files (MPEG, AVI, MOV, QT, etc.)
- Documents (DOC, HTML, XLS, PDF, PS, etc.)
- Pictorial images (TIFF, GIF, PNG, JPEG, BMP, etc.)
- Animations and sliding show (FLASH, PPT, etc.)
- Synchronizations between all media
- Printing preferences

Technology

WEDELMUSIC solution presents:

- Reliable mechanisms for protecting multimedia content and Digital Right Management
- Tools for automatic building, converting, and storing, digital musical content
- A unified XML-based format for modelling music including audio, symbolic, image, document, tools for distributing, sharing digital content according to several business and transaction models

WEDEL tools

- WEDELMUSIC Editor
- Music formats converter
- Analysis and symbolic music processing GUIs
- Impaired people GUIs
- Viewers and players
- WEDEL format content delivery servers
- Local providers for management of WEDEL format content and authorizations mechanisms on the place of delivery (orchestras, theatres, libraries, music schools, music shops, etc.)

http://www.wedelmusic.org/

5.2 The Interactive Music Network

Funded by the European Commission and established to help bringing music into the interactive multimedia era. The MUSICNETWORK is a Centre of Excellence to bring the music content providers, cultural institutions, industry, and research institutions together. The MUSICNETWORK draws on the assets and mutual interests of these actors to exploit the potentials of multimedia music content with the new technologies, tools, products, formats and models.

http://www.interactivemusicnetwork.org/

5.3 Music XML

Music XML (different from Recordare’s MusicXML format) is a European IMS project aiming at defining an IEEE commonly acceptable standard for musical application using XML. This project is to develop an XML application defining a standard language for symbolic music representation. The language will be 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 repre-sentation with already defined and accepted common standards. The standard is to 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.

http://www.computer.org/standards/1599/par.htm

6 Patent issues

6.1 W3C and patents

As a general rule, open standards developed by the W3C, such as HTML, VRML, or SMIL, are royalty-free, and free of patents. The W3C Patent Policy governs the handling of patents in the process of producing Web standards. The goal of this policy is to assure that Recommendations produced under this policy can be implemented on a Royalty-Free (RF) basis [5].

This policy is a guarantee for developers implementing W3C standards: they can in principle implement the W3C recommendations without having to cope with patent problems, and they don’t have to pay royalties to patent owners.

Sometimes however, the case can arise where some patent owner claims the ownership on a technology that is tightly integrated in Web standards, such as in the well-known case of Eolas. Eolas, a spin-off of the University of California, claims the ownership of the technology which permits embedding plug-ins in web pages, such as what is defined by the HTML standardized tag <OBJECT>, referring to the patent US 5,838,906, owned by the University of California.

For this reason, Eolas has sued Microsoft for using its technology for making it possible to embed ActiveX objects in a web page, and a court recently ruled in favour of Eolas. It has become clear that in this case, all uses of the <OBJECT> tag, as well as uses of the <EMBED> tag (which is not part of the HTML standard however), is covered by the patent (the patent is actually in re-examination, due to a requirement of W3C’s Director Tim Berners-Lee).

6.2 MPEG-4 and patents

Most of the technologies used in MPEG are covered by patents. For example, the well-known mp3 compression scheme is covered by a patent owned by Thomson Multimedia and Fraunhofer IIS.

Simple users doesn’t have to take care of licensing for these patents, for using mp3 for encoding or for decoding mp3 streams, but implementers – developers or industries - implementing software for encoding or decoding patented technologies have to beware of these patents.

“Licensing MPEG-4 is an important issue, and also one that is cause for much confusion. First, one should understand the roles of the different organizations involved in getting MPEG-4 deployed. Below is a short clarification of the role of some of the main players.

ISO/IEC MPEG is the group that makes MPEG standards. MPEG does not (and cannot, under ISO rules) deal with patents and licensing, other than requiring companies whose technologies are adopted into the standard to sign a statement that they will license their patents on Reasonable and Non-Discriminatory Terms (also called RAND terms) to all parties that wish to create a standards-compliant device, *(hardware or software) or create a standards-compliant bitstream. 'Non-Discriminatory' means that the patent needs to be licensed to al parties that wish to implement MPEG-4 on the same terms. 'Reasonable' is not further defined anywhere.*

In developing MPEG-4 part 10 / H.26L, the ITU/MPEG Joint Video Team is now attempting to establish a royalty-free baseline coder, and to this end it also asks of proponents to submit a statement that specifies whether they would want to make available any necessary patents on a royalty-free basis, if (and only if) others are prepared to license under the same terms.

The MPEG-4 Industry Forum, M4IF has in its statutes that it shall not license patents or determine licensing fees, but has nonetheless played an important role in driving the availability of licenses for the patents needed to implement MPEG-4. M4IF can discuss issues pertaining to licensing, and has acted as a catalyst, in the very literal sense, in getting patent pools going. In the years 1999 and 2000, M4IF adopted a series of resolutions recommending on ways that a joint license ('patent pool') might be established; it also mentioned names of parties/people that could play a role in this process. The process was detailed for the Systems, Visual and Audio parts of the standard, and involved an independent evaluator and a neutral administrator in all cases. Even though it will not actively pursue these, M4IF encourages alternative patent pools to be created, the more the better, and if possible even competing ones. (Competition is good, also in licensing, for the same reasons as why technology competition is good). It should be noted that no-one is forced to do business with any patent pool; one can also go straight to all the individual licensors (e.g., at least 18 in MPEG-4 Visual) as the licenses are always non-exclusive. However, doing so is cumbersome and there is a risk that negotiating 18 individual licenses turn out more costly than doing business with a one-stop joint licensing scheme.

M4IF has recommended that licensors create patent joint licensing schemes for specific profiles, as profiles are the interoperability points of MPEG-4, and only when one implements a profile there is a standards-compliant implementation. Also, M4IF held a poll among its members to determine for which profile there was the most interest.

There are the 'patent pools' (joint licensing schemes) and their administrators. It is the licensors that determine who will be their licensing agent(s), and they alone. Other parties (including ISO/IEC MPEG, M4IF or the licensees) have no say in this choice. MPEG LA is an example of a licensing administrator, licensing MPEG-2 Video and MPEG-2 Systems. MPEG LA also announced licensing for MPEG-4 Visual. Dolby licenses MPEG-2 AAC, and has announced a joint licensing scheme for some of the patents needed to implement MPEG-4 AAC. Thomson licenses mp3 (MPEG-1 Layer III Audio). Patent holders determine the fees; MPEG LA, Dolby, Thomson (and there are others) collect on behalf of the patent owners and distribute the proceeds.

Patent pools usually (always?) allow new patents to enter the pool when they are found essential. Sometimes this is because patents were submitted later, sometimes because they issued only at a later date. A license pool is never (at least in the case of MPEG standards, and as far as the author knows) 'closed' after a certain date. In MPEG-2, many patents were added after the start of licensing.

It should be clear – yet cannot be stressed enough – that neither M4IF nor MPEG receives even one cent of the collected royalties, nor does do they want to. M4IF has among its members licensors, licensees and entities that are neither of those. Collectively, the members have an interest in fair and reasonable licensing, because the standard will fail without it. " – Quoted from MPEG-4 Overview by Rob Koenen (The full paper is available from the MusicNetwork's web page.)

7 MPEG-integrated application scenarios

This section is devoted to application scenarios imagining potential future uses of multimedia music. The overall purpose of this section is to demonstrate the benefits of integrating a music notation standard into the presently under development MPEG multimedia frameworks: MPEG-4, MPEG-7 and, though less directly, MPEG-21. The MPEG frameworks indeed are designed to supports a large range of multimedia content (Advanced Audio Coding, video coding, structured audio coding, 2D & 3D graphics, raw and vector graphics, scene description, user interaction…) that can be used in connection with music notation in furtherance of the Musicnetwork's vision of multimedia music. These application scenarios are useful first to exemplify to the MPEG and Music Notation communities simple cases where Music Notation and other multimedia object types are integrated resulting in mutual added value; secondly, these scenarios can be useful to better understand and consequently refine the definition of the requirements in order to possibly approach a call for technology as a next step.

Each scenario is more or less built on the same unchanged template, made out of three main parts:

- Media streams (interleaved, multiplexed, etc…)
- Multimedia scene construction (MPEG-4 BIFs…)
- User interaction (Java, Query…)

The two first scenarios (Enhanced karaoke & Interactive music tutor) are straightforward applications that introduce Music Notation technology integrated with existing MPEG-4 and possibly MPEG-7 technology. They have been developed in the scope of the MPEG Ad-hoc group on Music Notation, and has been approved by the MPEG group.

The other scenarios (Improvisation training, Musical analysis, Pattern discovery, Interactive music course, Conducting course) are to highlight the need of technology required by somehow more advanced applications. Of course many other application scenarios could be presented, following the template of this document.

7.1 Scenario 1: Enhanced karaoke

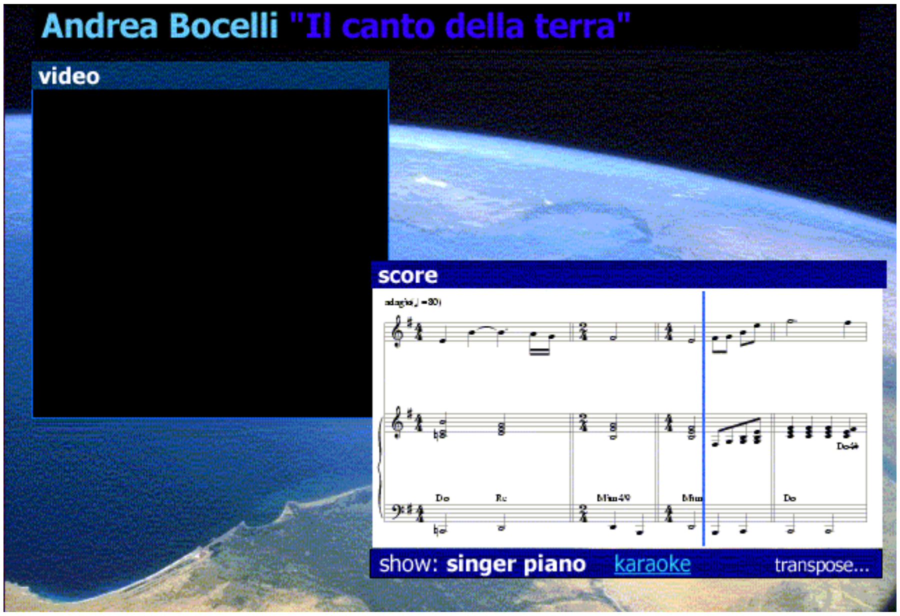
In this scenario, the purpose is conveying to the user a set of multimedia objects so that he may be allowed to interact with this content by selecting or stopping part of it and replace the stopped components by local performances. Since karaoke is a successful application only dealing with audio and lyrics, we call this application "enhanced karaoke", since it also involves musical instruments other than voice, and it also involves more interaction with the end-users.

Involved objects and content

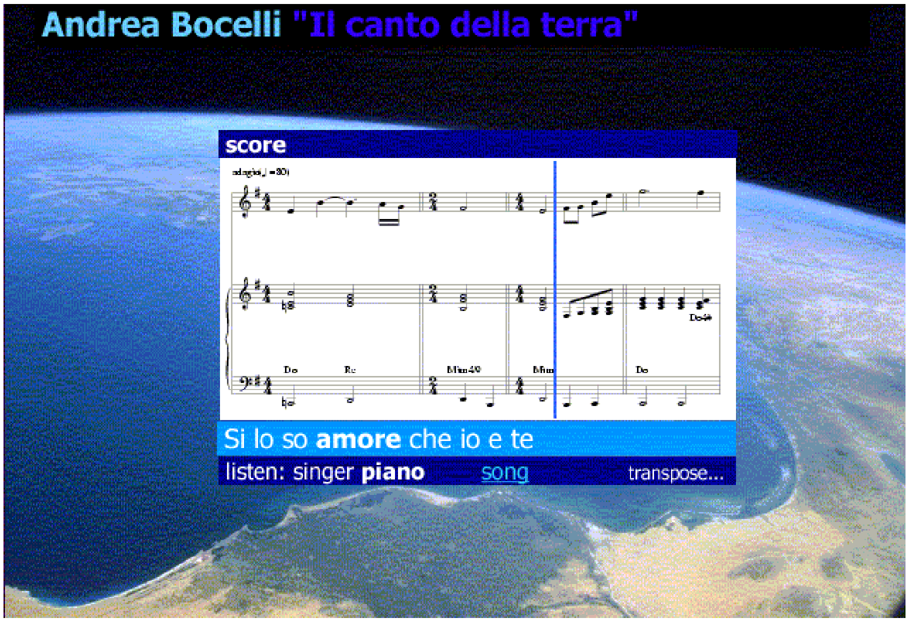
In this scenario, several objects are involved in relationship to one song. For what concerns audio, three stereo AAC objects may be used to encode the singer's voice, a guitar and piano; a fourth object, e.g. an SA object, is used to synthesize in real-time the bass through access units carrying non time-stamped SASL commands (so the decoding timestamp of the access unit is used to synchronize the events). Four other main objects are present, a video accompanying the song (the video may report the scene of the opera or the simple clip of the song), a text containing lyrics, music notation content and a scene description including graphic shapes acting as selection buttons and interaction sensors and routings.

Scene description and interaction

The scene description allows the display of the accompanying video (e.g. a singer), and it contains some icons to be used for the selection of the different instruments and voices and of the text. By default all the AAC and SA objects are active and the text display is not active. Finally, the music notation decoder is active and displays a score with all the parts. If the user does not click on one or more of the icons, a line moves over the visualized score in synchronization with the musical content. A scene mock-up with just voice an one audio track is shown in the following picture.



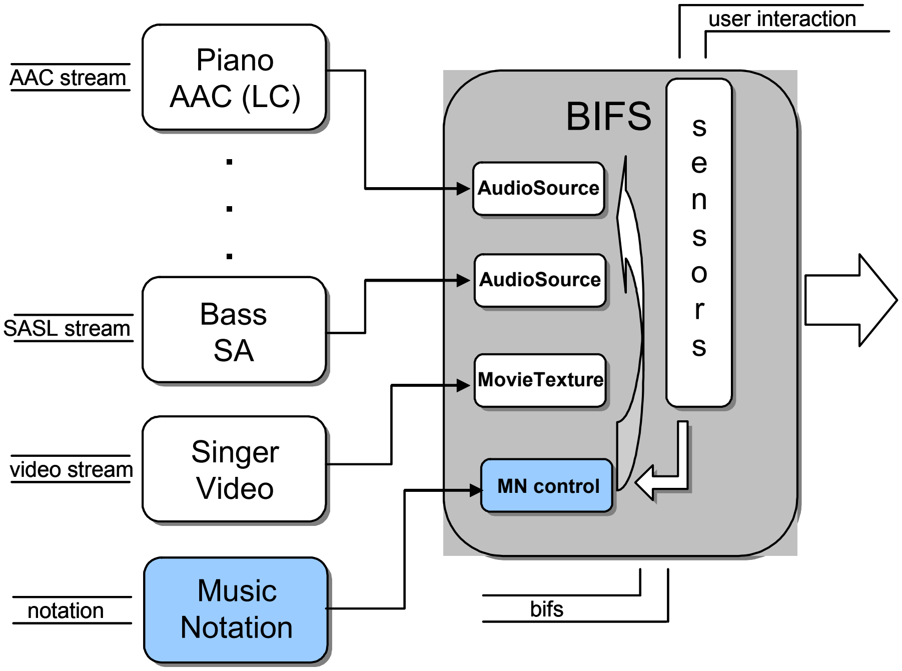
If the user clicks on the voice icon (“singer”, in the picture), the video is minimized and text is displayed synchronized with the music, so that a normal karaoke application is enabled. If the user clicks on one or more of the instrument icons, whatever is the state of the text display, that instrument is muted and the music notation decoder highlights (either by changing colour or by a new window) the part that has been muted, always with a line /cursor moving on it synchronously with the rest of the sounds to highlight which music notation symbol has to be played. If two or more sound parts are muted a similar behaviour occurs for all of them. Whenever the user clicks again on the corresponding icon the previous situation is restored in relationship with that particular part or text. The following picture is another mock-up of the same application scenario.



Si lo so **amore** che io e te

In addition, the user has a button allowing him to transpose (music transposition), since users have not always the same voice as the original singer, or dispose of an instrument slightly different from the original, for example a tenor saxophone instead of an alto saxophone, in which case he has also to see the score part transposed. To this purpose, all music objects must be transposed (not difficult if those objects are SA ones, some processing may be required for AAC in AudioFX or different tracks may be available), and the MN object must be transposed too.

The main blocks for this scenario are summarized in the following block diagram.



Main required MPEG tools

The tools already available in MPEG required for this simple application scenario are:

- MPEG-4 AAC (LC for instance)
- MPEG-4 SA
- MPEG-4 Video (e.g. Simple Profile)
- MPEG-4 BIFS (any profile supporting 2D graphics, timed text and multiple sounds)

Main requirements for Music Notation

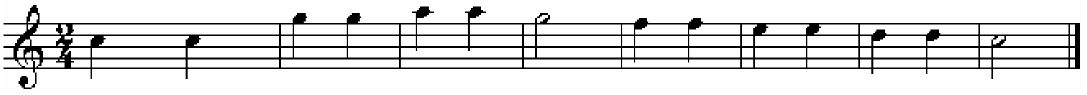
First of all it is necessary to have a proper music notation format with its normative decoding process. This means having a format carrying the music notation and in addition a different chunk offering the possibility to describe proper synchronization between score “events” and times (score alignment with live performance). The music notation format must support all the necessary functionality to correctly display music notation information (different fonts, different justification as a function of note duration and constant, different size, colours, etc.. as in the text), particularly in synchronization with other media in the scene. Further, in music it is needed to add also justification parameters since the spacing among symbols has musical meaning.

The MN object must be able to represent in a synthetic manner music objects. Music objects are essentially notes, but also more synthetic objects such as trills, arpeggios, portandos, and so on which should not be represented as the notes actually played, but as single objects.

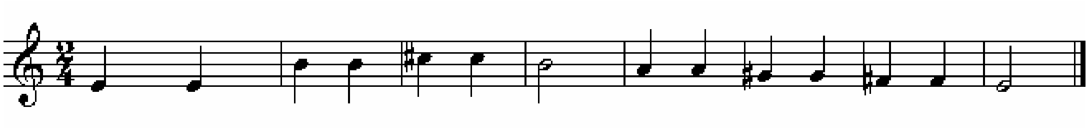
Further, interaction is necessary between the user and the media. This means having the music notation decoder interfaced to the scene with one (or maybe two) nodes with suitable fields able to receive necessary information to drive the decoder and at the same time delivering information from the decoder to other fields of relevance. In this particular example a field is required containing on/off state for each of the parts (to be possibly routed to AudioSource nodes or to an AudioSwitch node for the audio object switching).

As mentioned above, in some cases, the user may further wish to use a different instrument, thus the music notation need to be transposed. This will change the visual representation. Transposition MN node shall be able to transpose correctly a part, or an instrument part, according to the rules currently in use in music notation. For example, a correct transposition of a very simple extract.

Original displayed score:



Transposed displayed score:



7.2 Scenario 2: Interactive music tutor

In this scenario the purpose is having the user look for a training category in an archive of courseware and subsequently download multimedia interactive content matching the search criteria. In this case the user has the possibility to access multimedia sequences containing a required feature and interactively work with this content to learn and compare his/her ability by this content. The way in which an eventual live performance may be compared or measured against the downloaded interactive presentation (e.g. scoring) is outside the scope of the standardization and it is related to any individual application that may automate the evaluation process based on the available content. Nevertheless, a suitable model must be available to describe musical notation in a way to allow with the required precision this comparison.

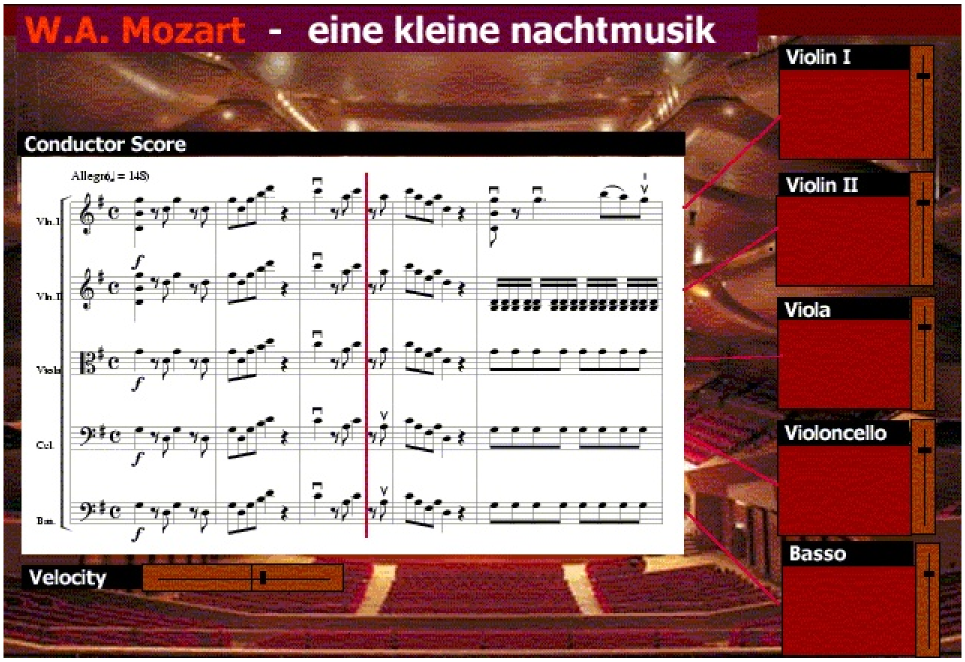
Involved objects and content

In this scenario the user tool has access to a possibly wide library containing performances of music pieces for educational purposes. All the available material is annotated by suitable descriptors according to the MPEG-7 standard with additional features related to notation. The available material is encoded in multimedia files composed by several objects each. Concerning audio, each instrument that is supposed to have a main role in the performance is encoded as an independent audio object (e.g. AAC LC). Each of these instruments also has a close-up video recording. Audio is passing through a processing node offering the possibility to slow or accelerate the performance (factor 0.5 to 1.5) without altering the pitch. Finally music notation is available, and a scene description for content composition and user interaction is provided.

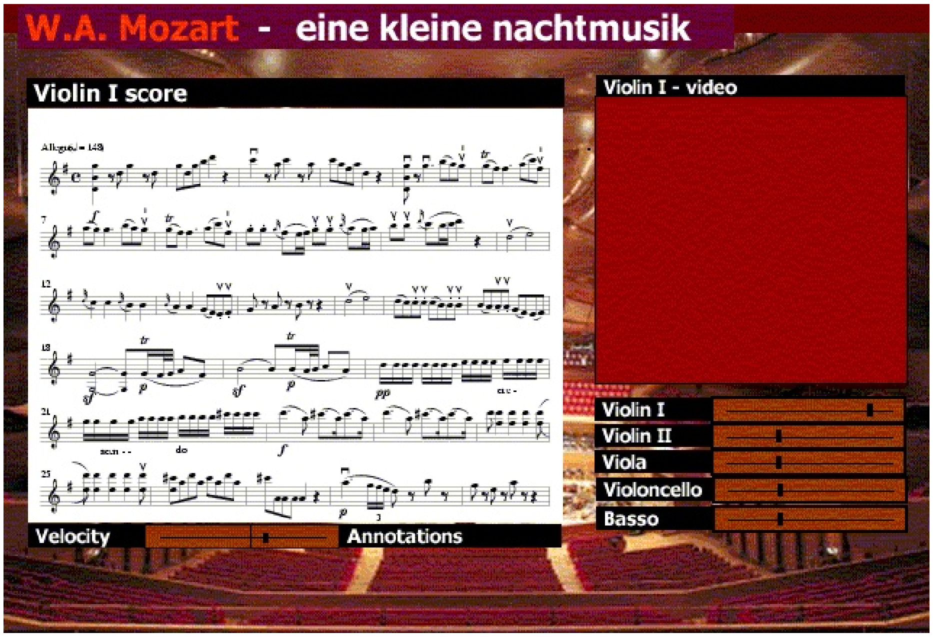
Query, scene description and interaction

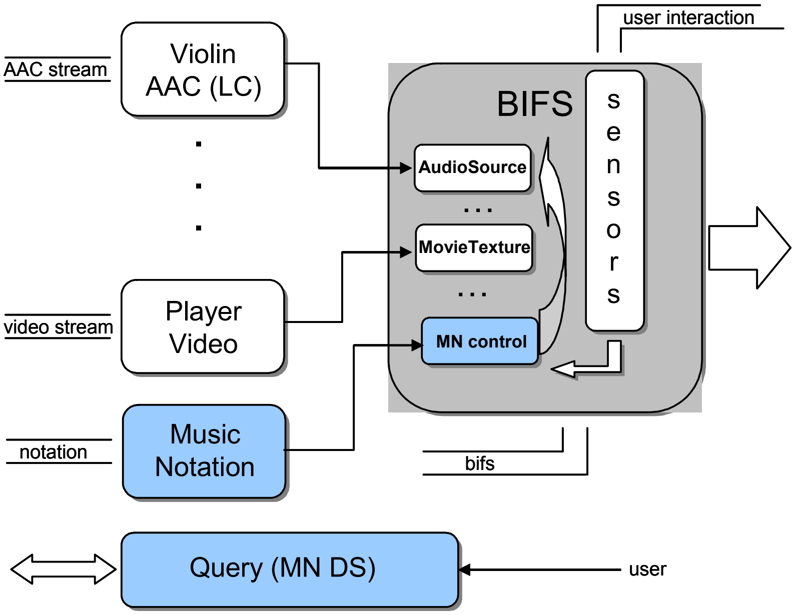
The user has the possibility to query the database for the particular skill to exercise he/she is looking for. For instance chromatic scales on the violin, or staccatos on the piano, and so on. The search will provide him links to material available for download and view examples and possibly exercise the desired features.

Each scene description allows the display of the accompanying videos (close-up of the instruments), and it contains different control icons to be used -- e.g. to affect the speed of the performance or completely mute parts in the performance. By default all the AAC objects are active and the different close-up videos are available as small resolution movies (the video may show the movements of the hands of a reference player, or the gesture of the conductor to be followed, etc.). Finally, the music notation decoder is active and displays a score with all the parts. If the user does not click on one or more of the video pictures, a line moves over the visualized score in synchronization with the musical content, like in the following picture.



If the user clicks on the picture of the instrument he/she is interested in, the video is magnified for that instrument, the music notation is reformatted to present only the selected part and not the main score with all the parts anymore (always with a line moving on it synchronously with the rest of the sounds), the sound of that instrument is enhanced in intensity over other instruments. Other parts may also be muted or reduced in volume. The user also has the possibility to control the execution speed of the performance through suitable control icons (like sliders). In this case the sound is slowed down by an AudioFX node implementing a speed change effect and the music notation tool behaves accordingly maintaining synchronization with the audio. The user is also usually interested in repeating some sections, marking them and restarting from the marked point several times (sound can be buffered, but this is a feature possibly related to a non normative use of the normative file; indeed a precise synchronization between score and other media, especially sound, is a strong requirement). Whenever the user clicks again on the corresponding video the previous situation is restored in relationship with that particular part and instrument. The following pictures show another view (magnified instrument) of this scenario and a block diagram summarizing the main blocks involved.





Main required MPEG tools

The tools already available in MPEG required for this simple application scenario are:

- MPEG-4 AAC (LC for instance)
- MPEG-4 Video (scalable, FGS)
- MPEG-4 BIFS (2D visual profile)
- MPEG-4 AudioBIFS (AudioFX node for processing)
- MPEG-7 DSs

Main requirements for Music Notation

As in the previous example, first of all it is necessary to have a flexible music notation format with its normative decoding process. This format must support all the necessary functionality to correctly display music notation information, particularly in synchronization with other media in the scene. This means, as said earlier, having a format carrying the music notation and in addition a different chunk offering the possibility to describe proper synchronization between score "events" and times. More than this, a suitable "subset" of the music notation functionality should be "visible" at the MPEG-7 description layer, in order to allow a query on relevant aspects of a score that may be worth searching for. The main requirements are:

- Production of main score and parts from the same synchronized music notation model
- Definition of sections
- Stop and play
- Accelerate and decelerate the execution rate
- Score alignment with live performance (similar to the first case)
- The MN object must be able to represent in a synthetic manner music objects. Music objects are essentially notes, but also more synthetic objects such as trills, arpeggios, portandos, and so on which should not be represented as the notes actually played, but as single objects.
- Description of musical content: it shall include all elements needed to describe music notation at a high level, including details of execution such as dynamics (staccato, pizzicato, legato, slurs, fingering, bowing…), rhythmic and meter details (tempo, rhythm, time signature…).
- Query by example: it shall be possible to select a segment of music notation to search for similar music, at the notation level.

Interaction is necessary between the user and the downloaded media. This means having the music notation decoder interfaced to the scene with one or more nodes with suitable fields able to receive necessary information to drive the decoder and at the same time delivering information from the decoder to other fields of relevance. In this second example a field is required containing on/off state for each of the parts (to be possibly routed to AudioSource nodes or to an AudioMix node for the audio object enhancement). In addition a field is necessary to control the speed of the score line display. To summarize the main interaction requirements:

- Showing selected single part with needed visualization parameters.
- Showing main score with required visualization parameters.
- Transposing the selected parts to be played with a different instrument
- Selecting parts to be muted or reduced in volume
- Accelerating and decelerating the execution rate for the music notation
- Adding some execution annotations such as fingering, bowing etc. that are typically added to the music notation during the rehearsal and during music studying.

7.3 Scenario 3: Improvisation training – The virtual band

In this scenario the user practices jazz improvisation or accompaniment through an enhanced "Band-in-a-Box"-like environment and is therefore to download various kind of multimedia content, from music notation to structured audio, audio and video samples. The user has the possibility to access multimedia sequences containing required features and interactively work with this content: changing tempo, style or key according to his will. Eventually scenario 2 could be embedded in this latter use case, assuming the user likely to reproduce his/her favourites jazz musicians solos and to evaluate his/her own performance with a client-side scoring application.

Involved objects and content

In this scenario the user tool has access to a wide library of jazz standards in a symbolic music notation format, which content is similar to the Real Books music sheets, eventually completed by hyperlinks to discographies or basic music theory: Scales for improvisation, piano or guitar voicing examples, etc… For each jazz tune, music notation content is related either to structured audio streams (MIDI or MPEG-SA) for the simulation of a virtual band (accompanying the user when he/she practices improvisation and completing the user's performance when he/she practices accompaniment), or compressed multichannel audio data streams routed to an AudioMix node for an "a la carte" rhythm section.

Query, scene description and interaction

The user has the possibility to query the database for the particular jazz tune to practice. The search will provide him/her links to accompaniments (either MIDI-like or audio-like) available for download and links to commercial recordings containing versions of the song he/she is like to perform. For each of these references, audio content as well as transcriptions of solos are to be downloadable and rendered in an "Enhanced Karaoke" manner (see section 7.1). Eventually, each scene description should also allow, synchronously to the rendering of famous musicians' solos and their notated transcriptions, the display of accompanying videos (close-up of the instruments) and images (fingering).

Each scene description contains different control to affect the speed or the key of the performance, or completely mute tracks of the virtual band. By default the rhythm section is realized by independent AAC objects that can be mixed or muted according to the user's preferences. If the user wishes to change the tempo or the key of the performance, he can choose either to keep on working with the AAC streams and performs elementary DSP (time stretching & pitch shifting) through an AudioFX node, or to switch from compressed audio streams to SA streams, containing MIDI data or SASL commands. In the latter case, SA streams should carry non time-stamped SASL commands, so the decoding timestamp of the SA stream's access unit is used to synchronize the events. If the user wishes to modify the style of the accompaniment (swing, Latin jazz, jazz waltz, bebop, half-time feel, double-time feel, etc…), structured audio is also to be preferred.

Main required MPEG tools

The tools already available in MPEG required for this simple application scenario are:

- MPEG-4 AAC (LC for instance)
- MPEG-SA
- MPEG-4 Video (e.g. Simple Profile)
- MPEG-4 BIFS (2D visual profile)
- MPEG-4 AudioBIFS (AudioSource for scene construction, AudioSwitch for channel selection, AudioFX node for processing, AudioMix for mixing rhythm section channels)

It is straightforward here that if he user wishes a minimum of interactivity with multimedia content symbolic data is to be preferred for scores (SMR streams) and audio (structured audio streams, MIDI or MPEG-SA). It makes indeed no doubt that if the user want to perform a tune at a twice lower tempo than the original one, the use of structured audio appears as a much better solution than DSP through an AudioFX node.

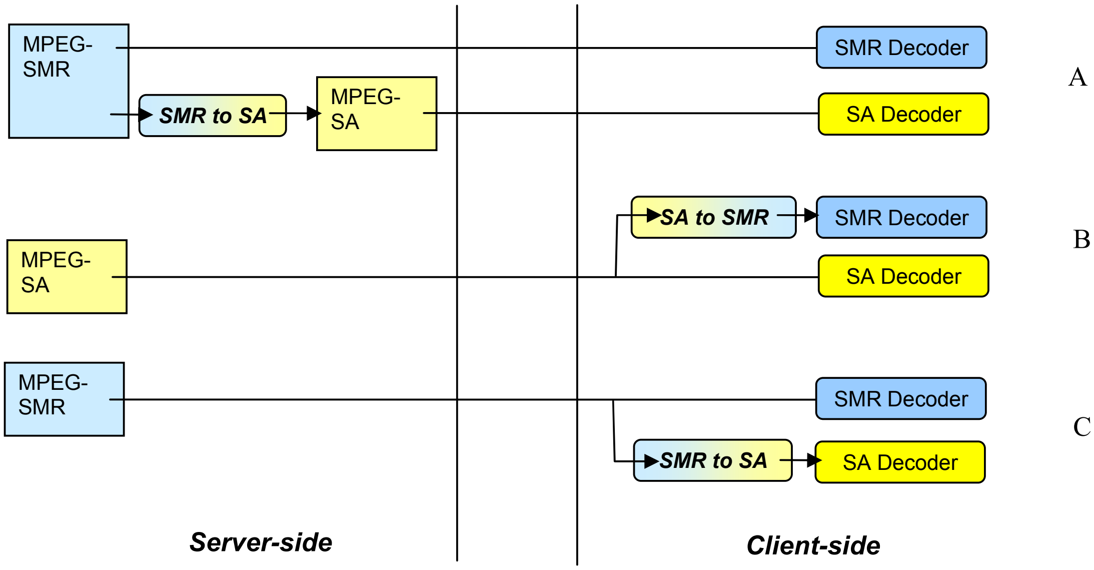
MPEG-4 Structured Audio (SA) specifies different tools for the description of sound synthesis and sound processing algorithms (Structured Audio Orchestra Language, SAOL) and for related control (Structured Audio Score Language, SASL). In addition, MPEG-SA supports the inclusion of a standard MIDI file and sound fonts. Now assume a content publisher who has symbolic music information/notation files in a proprietary format (e.g., Finale, Sibelius, Capella, etc.), and wants to publish them for MPEG-4 compliant devices so that end-users can see the music notation, and hear the corresponding music (synthesized sound). The following steps must be followed in order to perform this task using MPEG-4:

- use a converter to convert the original symbolic music representation/notation format to the MPEG-SMR format;
- inside the authoring tool a specific tool may generate the MPEG-SA files as well as the appropriate synchronization information from the MPEG-SMR file;
- the MPEG-SMR and MPEG-SA files are converted into binary and sent to the client;
- on the client the two streams are decoded by the two decoders and sent to the compositor.

On the client side some manipulation may be done even if the two binary streams are separate. Again, a song may be transposed and/or its speed of execution may be changed. In these cases, the manipulation has to be done on both nodes, the SMR Node and the AudioSource Node.

Another situation could be that a publisher wants to send some MIDI files (or SASL, both supported by MPEG-4 SA) to MPEG-4 compliant devices. Some of these devices may support SMR visualization, and thus enable the MIDI files to be automatically converted (through some specific algorithm) into SMR on the client side and rendered. In a similar way only the SMR may be delivered. This time, the MIDI information (or SASL, if SA is available) can be locally generated on the client side.

The figure bellows summarize these different possibilities:



Main requirements for music notation

The general requirements for music notation in terms of rendering and synchronization with other media remain the same as those exposed in the previous scenarios. Specific requirements for this scenario are:

- Production of the condensed, Real Books-like, jazz tune score and, if available, single parts of accompanying instruments.
- Section handling, allowing user-defined repetitions (head, choruses, coda…)
- Tempo variations: the user should have the possibility to perform the tune at a different tempo than the original one (natural compressed audio or structured audio should be modified in consequence).

- Key changes: the user should have the possibility to perform the tune in a different key than the original one (audio and music notation should both be modified in a proper manner).

7.4 Scenario 4: Musical analysis – Enhanced navigation in databases

In this scenario the user navigates through on-line music pieces and databases, thanks to on-line musical analysis tools. He/she is likely to access a wide range of music-related multimedia content, from simple audio files and music sheets to complex, interactive and hyperlinked music analysis objects. Music analysis frameworks enclose at-the-once internal links to a single music piece and external links that provide the user searching and navigating facilities and thanks to whom he/she is able to browse large music databases and sort information on specific criteria.

Involved objects and content

The user tool has access to a wide library of music pieces, all related to their appropriate analysis, mainly depending on their genre: harmonic analysis for chord sequences, pattern-oriented analysis for motives, themes, counterpoints and fugues, schenkerian analysis, etc… Audio content is mainly delivered with AAC streams, eventually through multiple channels, according to the needs and complexity of the related musical analysis. This latter object is delivered with MN streams and rendered in a enhanced, Flash-like manner, which provides the user the opportunity to interact with the score: Play, stop, repeat, jump to selection, etc… Structured audio streams could eventually complete this core content, in case the analysis purposes require some selection to be rendered very slowly, or abstraction the timbre of instruments, as in the case of a piano reduction for instance.

In addition, each music piece is completed by a set of metadata descriptors that embed information about instrumentation, scales, modes and temperaments, orchestra, conductor, performers, etc., in the scope of the MPEG-7 content description standard.

Query, scene description and interaction

The user starts with downloading the global framework of the music piece he/she is interested in. This framework basically corresponds to the table of chapters in a DVD, and allows as well linear listening than elementary navigation in the piece: overtures, expositions, acts, scenes, movements, coda, finales, etc… This first content should therefore be regarded as a very simple analysis of the music piece (section 4.2 itemizes a significantly high number of Web pages that provide such “active listening”).

Once the user has selected the segment to be rendered, the corresponding score, completed with musical analysis annotations, is streamed to his/her MPEG-4 terminal. The user can then choose to hear the whole segment (with synchronized score following) or only relevant parts or voices suggested by the analysis. Depending on the case, AAC or structured audio will be preferred.

The user also have the possibility to select regions of the score and ask the navigation tools to recover all the occurrences of similar regions, either within the analysed music piece, or, in a larger scope, within a composer’s complete works or within a given musical period. The segment selection is to be possible as well at the whole score level than at single/grouped parts level, in order to cope both with harmony and melody/motives features. It should therefore be possible to search in databases for music pieces containing similar harmonic progressions, melodies with similar contour (the MPEG-7 MelodyDS can be used for this purpose), or using similar scales or modes.

On the metadata side, a MPEG-7 layer is used to render additional information about the composer, performers, instrumentation and temperaments. The user has the opportunity to query the database with combined analysis and metadata criteria. For instance he/she may want the database-browsing tool to find all classical works for piano and a soloist string instrument, or all the 20th century pieces for violoncello solo that mainly use the pentatonic scale. Query functionalities should be available from both the analysed score layer and the MPEG-7 layer.

Main required MPEG tools

The tools already available in MPEG required for this simple application scenario are:

- MPEG-4 AAC (LC for instance)
- MPEG-SA
- MPEG-4 BIFS (2D visual profile)
- MPEG-4 AudioBIFS (AudioSource for scene construction, AudioSwitch for channel selection, AudioFX node for processing, AudioMix for mixing rhythm section channels)
- MPEG-J (for interaction with the user)
- MPEG-7 DSs

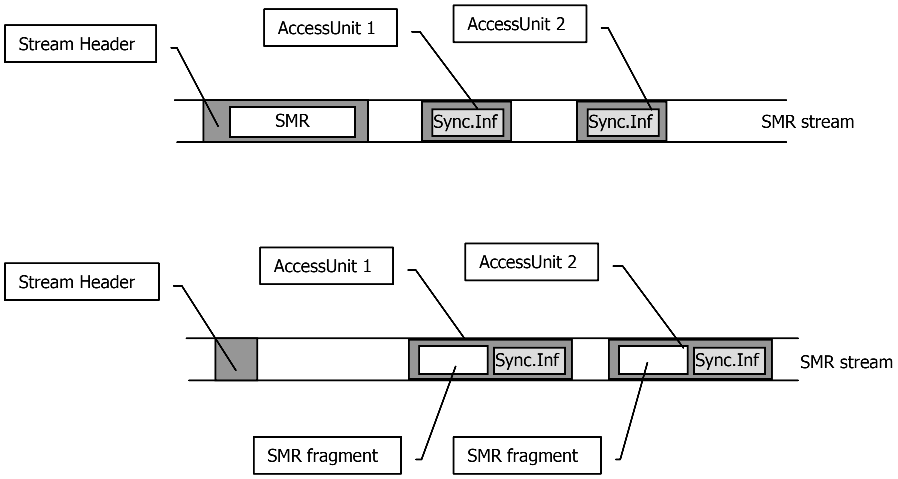
This last scenario addresses two major kinds of issues in multimedia music delivery. The first ones are related to the encoding of the content in the bitstream, in order to make any content being streamed in an efficient manner, anticipating the user's behaviour. The second ones cope with synchronization between different media and user interaction.

Regarding the encoding of SMR into the bitstream, something similar to the SASL/MIDI case might be defined. SASL information can be encoded in MPEG-4 in three different ways:

- part (or all) of the SASL file can be included in the header (ObjectTypeSpecificInfo) and SASL instructions shall be time-stamped;
- SASL lines can be inserted in Access Units with timestamps (so it is possible to deliver certain Access Units with content intended for a later use);
- SASL lines can be inserted in Access Units without timestamps, in this case the decoding time stamp of the AU is the reference time and SASL instructions are scheduled immediately.

Each of the above three possibilities can be associated with different use cases. For SMR information the same, or similar, approach may be followed. Regarding time and synchronization information with other media, SMR might be a different chunk of information in the header (if it is known in advance) or it might be a separated type of data in the Access Units, as mentioned above.

These three cases identify situations in which content is more or less available (reasonably) before its playback (everything in the header) from those situations when this is not possible/reasonable (use of Aus). Furthermore, if all the content (e.g., the SMR information of a symphony) is ready for use and broadcasted with the live event, it is not advisable to put everything in the header. Users may actually at any time switch on a different part of the music piece, or query the database for enhanced browsing before switch back to the initial stream at the very point he/she left it. In this latter case the server will be required to periodically retransmit at reasonable intervals (carousel) all that chunk of information, which does not sound really efficient.



SMR stream and Access Units (AUs)

A complete solution including AUs implies defining how to break content down into AUs in meaningful ways. This is an additional issue related to the definition of the SMR format. In SASL, each AU has associated timing information (the Decoding Time Stamp/Composition Time Stamp). In some cases, the AU may contain additional time stamping for the SA scheduler so that, for example, all the events from 10 to 13 can be included in one AU at time 10. SMR, and its synchronization with other media, can be seen as a very similar case. A graphical representation of the two cases discussed above is shown in the following figure; in the first case only synchronization information is encoded in AUs (SMR in the header) whereas in the second case SMR fragments are also included in it.

As far as synchronization between different media and user interaction are now concerned, if the application involving SMR only makes use of a graphics API, then it must do its own composition. MPEG-J Extensions may help this process. The synchronization among different media is done by MPEG-4 Systems and may be driven by the application itself (from an interactivity point of view).

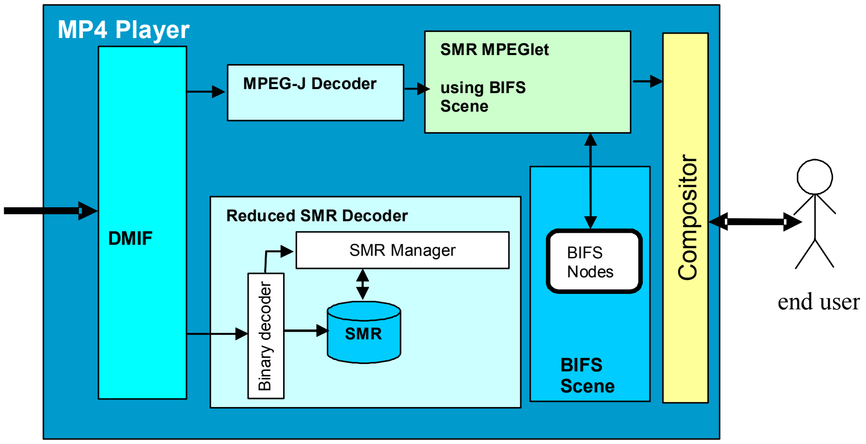
It is important to have a clear understanding of how the graphics API and all the other MPEG-4 media can be synchronized together, in order to permit the best integration of the SMR functionality. In fact, MPEG-J (MPEG-4 Part 11) already allows an application to access the decoders (to start/stop etc.) but it does not allow accessing the content of composition buffers, which may be useful for an application doing its own composition. MPEG-J extensions with OpenGL may allow a much more configurable visual-decoder-like implementation, by MPEGlets controlling visual rendering and composition. Timing information is in any case provided by the Systems stream, not by MPEG-J in general or by new extensions. Therefore, this should still be used to synchronize SMR, synthetic animations and other media[2].

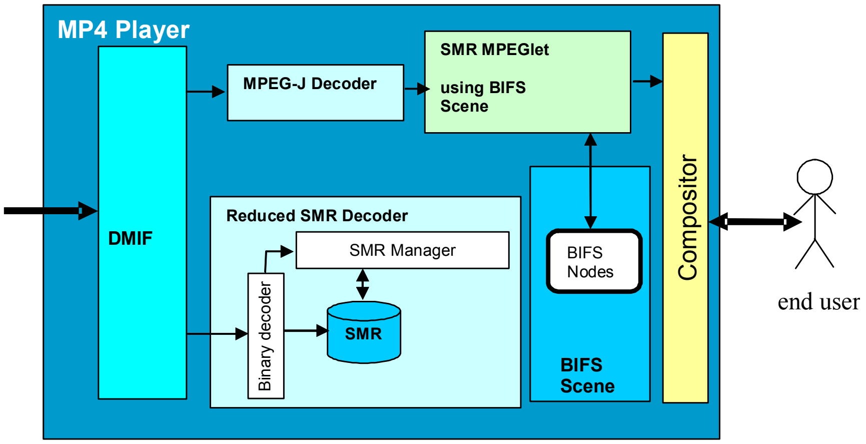
Two possible solutions to use MPEG-J as formatting engine are shown on the following pictures:

[2] It can also be noticed that sound support is orthogonal to the new extensions, which are defined at the moment only for video compositing. This means that two solutions are possible to include audio content in the multimedia scenes:

1. the already defined Audio BIFS and Advanced Audio BIFS nodes can be used together with other BIFS functionality and with the MPEG-J extensions, providing a more advanced graphic behavior
2. design something specific to audio, like for instance proposing a similar simple API for audio and symbolic music representation in the same spirit as the graphics API: low-level tools that can be combined together to create higher-level features such as Audio BIFS.

At a first glance, the first approach seems to be much easier to implement, unless some really very specific functionality will be required for some important application scenario that might be investigated in MPEG. BIFS is alternative to MPEG-J extensions; using BIFS is easier from a content developer point of view. But for an application developer, it may not be the case. Both approaches have their advantage.





The first one uses the MPEG-J Extensions for rendering whereas the second one uses the standard scene to render a music score representation in BIFS. In both cases:

- The *Reduced SMR decoder* decodes the incoming stream and it stores the SMR content, but it does not produce directly visual composition buffers.
- The *SMR MPEGlet* interacts with the *SMR Decoder* using a specific API interface (derived from the general decoder interface) allowing the MPEGlet to access the SMR data structures and also to be notified when SMR events (i.e. synchronization events) happen. Thus the SMR MPEGlet can produce the visual representation of the score by using the MPEG-J Extension (in the first case) or by using the scene graph API (in the second case).
- For synchronization, the *SMR Manager* in the decoder may produce events sent to the MPEGlet through an event notification interface allowing the MPEGlet to properly display the desired music notation synchronized with other media.

It should be noted that the SMR Decoder API has to be normatively defined (in the two cases above a specific SMR BIFS node is not necessary). These two solutions have the benefit of flexibility (the MPEG-J application can manage the score information in a customized way) but on the other side rendering of complex scores could be critical for performance.

However, different solutions where:

- the score is generated by MPEGlets using BIFS
- the score is generated by MPEGlets using MPEG-J Extensions
- the score is generated directly by the decoder

may even coexist, in the sense that a player of a given Profile may implement the last solution when performance is needed (this solution in any case does not make use of any available MPEG graphic tool available, it has to completely define its own graphic layer); another player may use the reduced decoder with an MPEGlet exploiting BIFS or MPEG-J Extensions for rendering when flexibility is needed (like in educational applications).

Main requirements for music notation

The general requirements for music notation in terms of rendering and synchronization with other media remain the same as those exposed in two first scenarios. Specific requirements for this scenario are:

- Production of music scores completed with rich analysis annotations
- Section handling, allowing user-defined selection and repetitions
- Production of “hyperlinked” scores to allow easy enhanced navigation in music pieces and databases
- Synchronization with MPEG-7 layer for combined query: SMR content and metadata criteria

8 Documents, white papers, tutorials and presentations

These links have been accessed on May 2005.

8.1 Macromedia Flash

A white paper presenting the Flash framework:

http://www.macromedia.com/software/flash/survey/whitepaper_jul03.pdf

8.2 QuickTime

An overview:

http://developer.apple.com/documentation/QuickTime/RM/Fundamentals/QTOverview/QTOverview.pdf

The QuickTime 7 user’s guide:

http://images.apple.com/quicktime/pdf/QuickTime_7_User_Guide.pdf

8.3 Windows Media

An overview of Windows Media 9:

http://www.microsoft.com/windows/windowsmedia/technologies/overview.aspx

An overview of Windows Media Encoder:

http://www.microsoft.com/windows/windowsmedia/wm7/encoder/whitepaper.aspx

An overview of Windows Media Server 2003:

http://www.msfn.org/modules.php?modid=2&action=show&id=82

8.4 SMIL

A tutorial on using SMIL, by the Boston University:
http://www.bu.edu/webcentral/learning/smil1/

An overview of SMIL 2.0, focusing on concepts and structure:
http://www.computer.org/multimedia/mu2001/pdf/u4082.pdf
Support slides for a tutorial on SMIL:
http://homepages.cwi.nl/~media/SMIL/Tutorial/SMIL-4hr.pdf

8.5 MPEG-4

An overview of MPEG-4 technologies, applications, benefits of MPEG 4:
http://www.iis.fraunhofer.de/amm/techinf/mpeg4/mp4_out_20027.pdf
Some overviews of MPEG-4 BIFS:
http://www.iis.fraunhofer.de/amm/download/bifs_en.pdf
http://www-artemis.int-evry.fr/Publications/library/preda/Tran-ICME2003.pdf
An on-line tutorial on BIFS:
http://gpac.sourceforge.net/tutorial/bifs_intro.htm
A document about MPEG-4 and Quicktime 6:
http://a320.g.akamai.net/7/320/51/cad38f4a7f9b46/www.apple.com/mpeg4/pdf/MPEG4_v3.pdf
A white paper presenting an MPEG-4 webcasting solution from Envivio and Cisco:
http://www.envivio.com/images/products/031217_wp_4forum.pdf
A paper on Audio BIFS:
http://www.mcl.ie.cuhk.edu.hk/scheirer_tom99.pdf
Some papers on structured audio in MPEG-4:
http://web.media.mit.edu/~eds/mpeg4-old/saolc.pdf
http://web.media.mit.edu/~eds/CMJ1999.pdf
http://web.media.mit.edu/~eds/MMSys-99.pdf
http://web.media.mit.edu/~eds/papers/icassp98.pdf

8.6 RealMedia:

Documentation on producing multimedia documents with RealNetworks digital media delivery platform:
http://service.real.com/help/library/guides/realone/IntroGuide/PDF/ProductionIntro.pdf

9 References

[1] Chion, Michel, *Pierre Henry*, Paris, Fayard, 1980.

[2] Emile Leipp *Acoustique et musique*. Paris, Masson, 1951.

[3] Vercoe, B. L., Gardner, W. G., Scheirer, E. D. 1998. *Structured audio: The creation, transmission, and rendering of parametric sound representations*. Proceedings of the IEEE, 85, 5, pp. 922-940.

[4] LISTEN project http://listen.gmd.de/.

[5] W3C patent policy http://www.w3.org/Consortium/Patent-Policy-20030520.html.

[6] *Leonardo Music Journal*, Volume 13, GROOVE, PIT AND WAVE, MIT Press.

[7] ISO/IEC DIS 16262 Information technology - *ECMAScript: A general purpose, cross-platform programming language*.

[8] Selfridge-Field, Eleanor, *Beyond MIDI: The Handbook of Musical Codes*. MIT Press, 1997.

[9] MEGA project : http://www.megaproject.org/

[10] CARROUSO project :
http://www.ircam.fr/projets_europeens.html?&L=0&tx_ircam_pi1[showUid]=4&cHash=37e44f4456

[11] Pereira, Fernando & Ebrahimi, Touradj, *The MPEG Book*, Prentice Hall Imsc Press Multimedia Series, July 2002.

10 Glossary of Acronyms

3GPP	
AAC	Advanced Audio Coding, an audio compression scheme developed by the MPEG group (Dolby, Fraunhofer, AT&T, Sony, and Nokia).
AC3	Dolby Audio Coding, an audio compression scheme developed by Dolby.
AHG	Ad Hoc Group
AICC	Aviation Industry Computer-Based Training Committee
AIFF	Audio Interchange File Format
ASP	Microsoft's Active Server Page
AU	Default Sun systems audio format
AUs	MPEG Access Units (ISO/IEC 14496-1)
AVI	Audio Video Interlaced, a video and audio format developed by Microsoft.
BIFs	BInary Scenes for Files
CSS	Cascading Style Sheet
CWMN	Common Western Musical Notation
DAB	Digital Audio Broadcasting
DAT	Digital Audio Tape
DBS	Direct Broadcast Satellite
DCMI	Dublin Core Metadata Initiative
DDL	Description Definition Language
DS	Description Scheme
DSD	Direct Stream Digital
DSP	Digital Signal Processing
DTT	Digital Terrestrial Television
DVD	Digital Video Disk
FFT	Fast Fourier Transform
GRM	Groupe de Recherches Musciales
HDTV	High Definition Television
HTML	HyperText Markup Language
IEEE	Institute of Electrical & Electronic Engineers

ISO	International Standards Organisation
I-TV	Interactive TV
JSP	Java Server Pages
MIDI	Musical Instrument Digital Interface
MIT	Massachusetts Institute of Technology
MLP	Meridian Lossless Packing
mp3	A coding standard for compression of audio data: MPEG-1 Layer 3
M4IF	MPEG-4 Industry Forum
MPEG	Moving Picture Experts Group
MPEGIF	MPEG Industry Forum
MPEG-GA	MPEG General Audio
MPEG-SA	MPEG Structured Audio
MPEG-J	MPEG Java
NAMM	National Association of Music Merchandisers
OASIS	Organization for the Advancement of Structured Information Standards
OMF	Observatoire Musical Français
P2P	Peer to Peer
PCM	Pulse Code Modulation
PHP	Hypertext Preprocessor
RDF	Resource Description Framework
RTP	Real-Time Transport Protocol
RTSP	Real-Time Streaming Protocol
SACD	Super Audio CD
SCORM	Shareable Courseware Object Reference Model
SD2	Sound Designer II
SDIF	Sound Description Interchange Format
SGML	Standard Generalized Markup Language
SMIL	Synchronized Media Integration Language, a language for synchronisation of multiple media developed by the W3C.
SMR	Symbolic Music Representation
S/PDIF	Sony Philips Digital Interface
SVG	Scalable Vector Graphics
UMI	Universal Music interface
URI	Universal Resource Identifiers
VRML	Virtual Reality Modelling Language
W3C	World Wide Web Consortium
WAV	A sound format developed by Microsoft standing for wave /waveform
WMA	Windows Media Audio
XAML	eXtensible Application Markup Language
XHTML	eXtensible HyperText Markup Language
XML	eXtensible Markup Language
XSL	eXtensible Stylesheet Language
XSLT	A language for transforming XML documents