The Interactive-Music Network

DE4.3.1
Multimedia standards for music coding

Version: 1.4
Date:06/02/2004
Responsible: IRCAM

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

Deliverable Number: DE 4.3.1
Contractual Date of Delivery: 31/01/2004
Actual Date of Delivery: 06/02/2004
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): Jerome Barthelemy, David Fuschi, P. Bellini, P. Nesi

Abstract:
This document reports the uses and practices in the domain of music coding in a multimedia environment. It reports about de facto or open standards which are in use in domains related to music, in its audio form as well as in its written form (scores), or even in structured audio forms (such as MIDI).
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 Content

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 itself the product of a very long process which began in the XIst 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 XVIIIth and XIXth centuries, and can be considered as something like the result of a normative process, a standard.
In the XXth 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 acception has changed. Musical practice in the preceding centuries was based on personal or collective practice on musical instruments, particularly the piano in the XIXth. 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 appearing of phonograph in the XIXth 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 ethno musicology was made possible by the use of these new recording devices – phonograph first, but later tape recorder. We can think of 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. We can think that Pierre Henry [1], one of the most important composers of "concrete music", or electro acoustic music, which 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 appearing 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. In the domain of music production, CSound, for example, which has been invented at the prestigious Massachusetts Institute of technology, has gained a worldwide audience.
In the same while, the capabilities of the computer for repetitive tasks, and the process of digitisation have made possible the generation of powerful tools for audio signal analysis – tools mainly based on the process known as Fast Fourier Transform. This kind of tool has made possible first an analysis of the harmonic characteristics of the audio signal, similar to the analogue-based spectrogram invented previously in the 40, and applied by by Emile Leipp [2] to musical studies, but moreover has made possible 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 emergence of the Internet, has particularly made possible the copy of music at a ridiculously low price, a so low price that it can absolutely be 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 industrial 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 points cannot simply be ignored. Exchange of music, exchanges of music concepts, exchange of music experiences or new collective music experiences based only on music sheets, their selves based on

CWMN, which ignores totally new music schemes, have been tried recently (Net4Music). By their failure, they have proved their inadequacies to the needs of the music community.
In the same while, new schemes for music notation – like MIDI – have gained a worldwide audience. Even with the strong criticism made by recognized researchers from academia (Selfridge Field)[8], MIDI is recognized as the only one available format for music notation exchange and studies, even for academic research.
A number of musicians – composers, instrumentists, or even amateurs - are attached with the concepts inherited from centuries of common western musical notation, and they also cannot simply be ignored. Some of these are working in a closely related fashion with the most advanced tools in the computer domain (Manoury), and are even actively participating to the evolution of the new music tools.
Music now frequently make uses of audio advanced concepts, such as 3D audio, in order to make possible interaction of users in a virtual 3D space where position in space interacts with audio rendering (LISTEN project)[4].
Music also uses finely grained synthesized music, using synthesizer tools controlling all aspects of evolution of pitch or timbre (CSound, MAX/MSP).
Music also uses interaction between scene (gesture, voice, instruments) and synthesized audio, for real-time realization of effects in live performances (MEGA)[9].

1.1.2 Definition of 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 tools.
This definition of multimedia music infers some specific requirments which are listed below.

1.2 Scope of the document

This document is aimed to study 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.

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 important part of these have been developed in relation with Internet, web-based solutions.

1.2.2 Open standards and proprietary standards

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 doesn’t 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, de facto standards have been in the past recognized as a success – and are already recognized. First of all, the postscript standard, elaborated by Adobe in the 80's, but many others.

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.
But, 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 workplan 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. The Flash format 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 Audio formats

We review in this document the most widely used audio formats, and the most proeminent emerging ones.
Most of the currently used audio 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.
We don't study here lossless compression schemes, which would be of interest for archiving purposes, but which are actually not used at all in multimedia distribution and diffusion, this being mainly due to the poor compression rate they achieve : 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 Audio effects, 3D audio, multichannel audio

The technology progresses achieved recently have been made possible for end users a number of technologies that greatly enhances his audio and music experience.
Multichannel audio has been present in film since the late 1930's-early 1940's 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 multichannel digital audio began in theatres in the early 1990's. 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 multichannel digital audio directly to the consumers home.
Practical multichannel digital audio began in theatres in the early 1990's. 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 multichannel 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 realtime processing needed to deliver 3Daudio. 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.

1.2.5 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). But we consider also the musical-instrument digital interface (MIDI) as being a form 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.

1.2.6 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, sonagrams, 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 encoding 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.

2 Needs and requirements

2.1 General requirements: openness and interoperability

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. For example, support of audio is

2.2 Specific requirements

Requirements as needed for distribution of multimedia music are:

- High quality of rendering for audio, support of surround sound, multichannel audio, 3D spatialization
- High quality of rendering for graphics (vector graphics)
- Easy and accurate synchronization between different media
- Powerful indexing mechanism, high quality metadata
- Annotation capabilities
- Content-based analysis and retrieval
- Easy integration in existing multimedia authoring environments
- Openness to real-time processing functions
- Openness to end-users interactivity

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 128 kbps, but this not always true for early compression schemes. Support for surround sound, multichannel audio, and 3D spatialization is 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.
An efficient tool must be available to realize accurate synchronization between different medias. This seems to be quite evident, but it's not always the case!
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 instrumentist playing in front of a microphone. But more simply, mixing tools can also be made available to the user.
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.
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.
An 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.
It would finally interesting to study in details existing formats for music scores, but since no standard format for music scores is really today widely used, and no format for music scores is actually included in any multimedia format, this topic is not adressed here. We recommend to study the document about music notation formats available on the MusicNetwork site.

2.3 Application Scenarios for Music Notation integrated in MPEG

This chapter has been developed in the scope of the MPEG Ad-hoc group on Music Notation, and has been approved by the MPEG group.
It present two straightforward application scenarios that introduce Music Notation technology integrated with existing MPEG-4 and possibly MPEG-7 technology. These two application scenarios are useful first to exemplify to the MPEG and Music Notation communities two 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.
The first section of this document introduces a scenario related with interactive multimedia content distribution for an "enhanced karaoke" application; the second section introduces a slightly more complex example (at least in terms of music notation functionality) related with music education and interactive courseware. Of course many other application scenarios could be presented, following the template of this document.

2.3.1 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 time stamp 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.

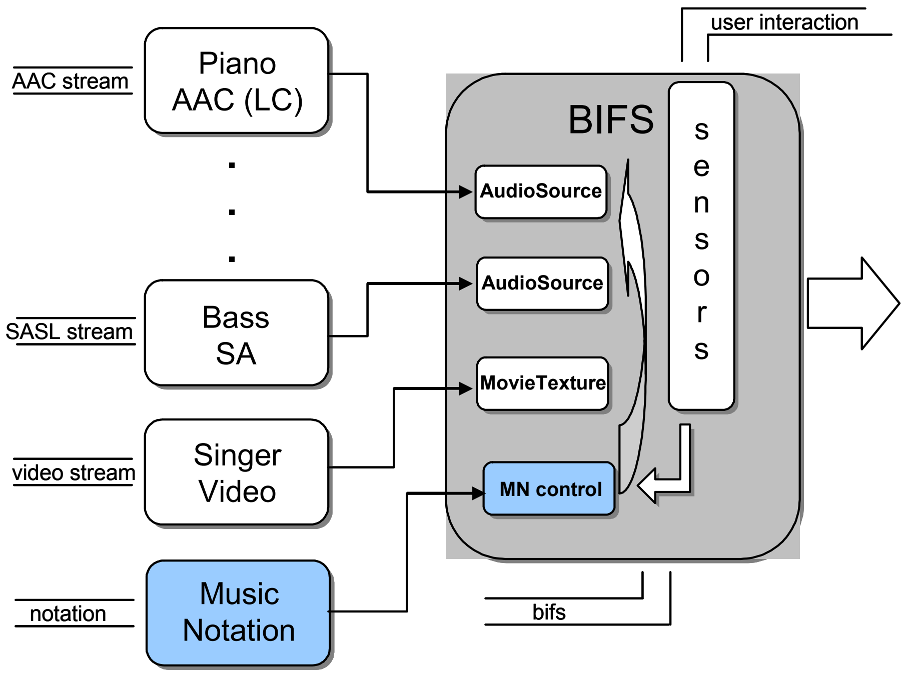
Scene description and interaction

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 color 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 behavior 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.

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:

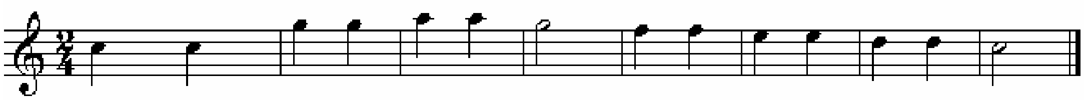
- • MPEG-4 AAC (LC for instance)
- • MPEG-4 SA
- • MPEG-4 Video (e.g. Simple Profile)
- • MPEG-4 BIFS (any profile supporting 2D graphics, timed text and multiple sounds)

Main requirements for Music Notation

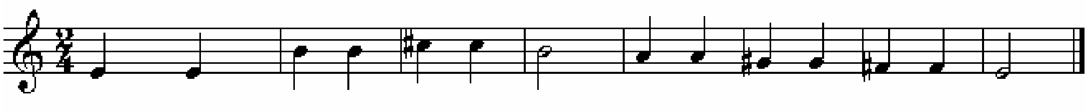
First of all it is necessary to have a proper music notation format with its normative decoding process. This means having a format carrying the music notation and in addition a different chunk offering the possibility to describe proper synchronization between score "events" and times (score alignment 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, colors, 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:



2.3.2 Scenario 2. Interactive music "tutor"

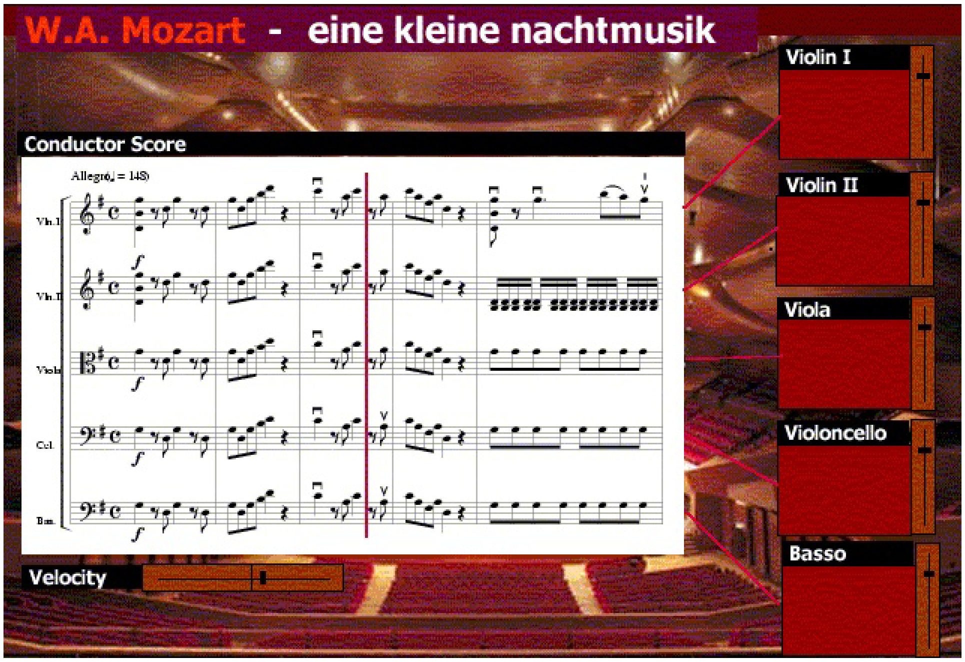
In this scenario the purpose is having the user look for a training category in an archive of courseware and subsequently download multimedia interactive content matching the search criteria. In this case the user has the possibility to access multimedia sequences containing a required feature and interactively 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

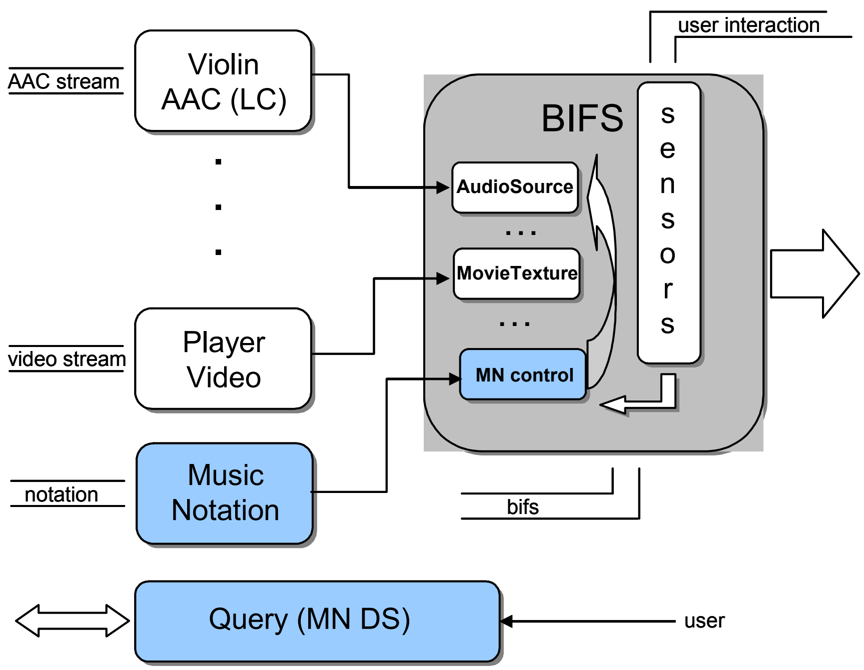
In this scenario the user tool has access to a possibly wide library containing performances of music pieces for educational purposes. All the available material is annotated by suitable descriptors according to the MPEG-7 standard with additional features related to notation. The available material 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.

3 Actors

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 Industry

3.1.1 Macromedia

http://www.macromedia.com/

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.

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 §5.1 and §5.2).
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 (Hypetext preprocessor) , ASP (Microsoft's Active Server Pages), 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.

3.1.2 Adobe

http://www.adobe.com

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 soometimes 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.

3.1.3 Microsoft

http://www.microsoft.com/

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 2005, will probably integrate much more capabilities than those actually available. This new exploitation system, for which the 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 markup 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. A very rough draft of preliminary documentation is available at 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.

Apple's main products:

- Quicktime player, Quicktime streaming server, DVDStudio (DVD production)
- Quicktime is detailed below (see §5.4).
- Quicktime streaming server is based on RTP/RTSP (Real-Time Transport Protocol/Real-Time Streaming Protocol). It provides support for streaming support for streaming Quicktime, MPEG-4 and 3GPP files. It also provides support for streaming MP3 content using Icecast-compatible protocols over HTTP.
- DVDStudio is a DVD video 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.

3.1.5 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.

Real Networks 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 §5.7). The Real player supports also SMIL content (see § 5.6)

3.2 Standardization bodies

3.2.1 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 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. So far a Technical Report and two standards have been produced and three more parts of the standard are at different stages of development. Several 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.

3.2.2 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, SVG, VRML. 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).

4 Music and media coding

4.1 Audio

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. Psycho acoustic 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 below only the most widely used formats.

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.

4.1.2 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.

4.1.3 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.

4.1.4 mp3

mp3 stands for MPEG 1 audio layer III, and is a widely used and known compressed audio format.
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.

4.1.5 AAC

AAC (Advanced Audio Coding), like mp3, is based onto a pycho-acoustic 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.

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

4.1.6 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

4.1.7 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.

4.1.8 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 format is based on a psycho-acoustic model.

http://www.vorbis.com/

4.2 Audio effects, 3D audio, multichannel audio

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 4). There is also a support for multichannel audio in WMA. Quicktime offers also a support for multichannel audio, as it supports MPEG 4.

4.2.1 VRML

VRML is a standard developed by the W3C 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.

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

4.2.2 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

4.2.3 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.
AudioBIFS :

- 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
- AudioClip : Save a short "snippet" of sound for use with interactions or loops
- AudioFX : Execute parametric sound-effects processing given as SAOL (MPEG SA) code.
- Sound (and Sound2D) : Attach the sound created with an audio subgraph into a 3D world (or 2D scene, for Sound2D).

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

4.2.4 Windows Media Audio

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

4.3 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.

4.3.1 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/

4.3.2 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, apart of the evolution of Csound has been integrated in MPEG, and known as MPEG SA, see just below.

4.3.3 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/

4.4 Vector graphics

4.4.1 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/

4.4.2 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/

4.4.3 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/

4.4.4 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 §5.5).

4.5 Metadata

4.5.1 RDF

The Resource Description Framework (RDF), developed by the w3c, integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to 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 arrive at 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's 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 a formal semantics which 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/

4.5.2 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.

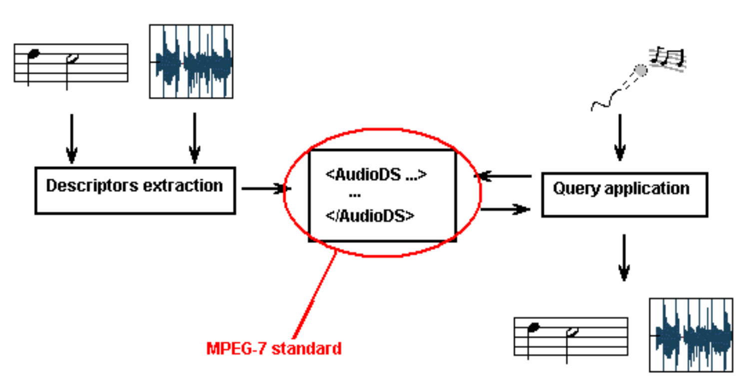
http://dublincore.org/

4.5.3 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.
- Query by example: given a sound excerpt, search for sounds in a database having similar characteristics.

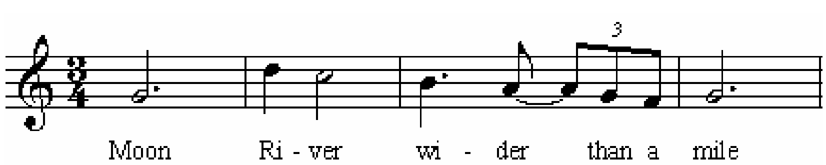
For an example of query by humming, the following resource is available (on 31/12/2003):

- http://www.musicline.de/de/melodiesuche/input

MPEG 7 doesn't standardize the process of 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 describes 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

5 Multimedia frameworks

We review here the main multimedia frameworks, in their actual state at the date of January 2004. 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.

5.1 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/

5.2 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 encapsuled 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 siide (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/

5.3 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/

5.4 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 plugin. Recent releases of Quicktime are able to support the MPEG 4 format.

Audio and music support:
Quicktime supports mp3 and AAC compressed formats, as well as proprietary formats. For structured audio, Quicktime supports MIDI. There is no support for audio effects or 3d audio.

Authoring and production:
Adobe's Premiere or Macromedia Director can generate Quicktime content. There are also a number of production tools available.

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 as provided by authoring tools such as Adobe Premiere or Macromedia Director.

http://www.apple.com/quicktime/

5.5 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, and parts of the early MPEG 1 and MPEG 2 standards are of interest as well.

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:

- Envivio 4Mation is an authoring tool dedicated to MPEG 4 content.
- 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-J enables scripting for end-users. MPEG-J is a programmatic system which specifies an API for interaction of Java code present as part of the media content with MPEG-4 media players.

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 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

5.6 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 implementors 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.

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.w3.org/AudioVideo/

5.7 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 5.6), 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.

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

Comparison

Here are reviewed the most important features of multimedia frameworks, from the specific point of view of music requirements as expressed in §2. Video features are not covered heree, 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		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					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 (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		

6 Uses : review of musical publications on-line.

Here are reviewed 55 different publications on-line related to music and musical heritage, making use of multimedia technologies
Special thanks must be given to Pierre Couprie and OMF (Observatoire Musical Français) and his review of hypermedia analysis on the Web : http://www.omf.paris4.sorbonne.fr/EARS-MINT/ANA/

6.1 Media and technologies

Media : audio, score, sonogram, synchronization and interactivity.

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 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".

Technologies :

Quicktime, Real, MP3, MIDI, AVI, Cubase are used for audio.
Finale or PDF are used in a few cases for the score.

	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/BWV871/1-intro.html	X	X		X		X		X								
http://atelier.feuillantine.free.fr/analyse/messiaen/rouss1.html	X	X		X		X										
http://citd.scar.utoronto.ca/VPAC80/Projects/Kay/BoS1.html	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://smt.ucsb.edu/mto/issues/mto.01.7.3/mto.01.7.3.wannamaker.html		X							X							
http://smt.ucsb.edu/mto/issues/mto.97.3.3/mto.97.3.3.taylor_essay.html		X									X					
http://webware.cite-musique.fr/www/gamelan/	X			X	X	X										
http://www.ac-bordeaux.fr/Pedagogie/Musique/ariabach.htm#analyse		X									X					
http://www.ac-bordeaux.fr/Pedagogie/Musique/grisey.htm																X
http://www.ac-bordeaux.fr/Pedagogie/Musique/op80bee0.htm		X														
http://www.ac-bordeaux.fr/Pedagogie/Musique/pagwop10.htm		X									X					
http://www.ac-dijon.fr/pedago/music/bac2002/risset/index.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/badong/badong.html	X						X									
http://www.ethnomus.org/ecoute/diphonique/hai1.html	X		X	X	X		X									
http://www.ethnomus.org/ecoute/nzakara/nzakara.html	X	X		X	X		X									
http://www.ethnomus.org/ecoute/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/index.html	X	X	X	X	X		X									
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.s-line.de/homepages/jscheytt/debussy/klavierwerk/laplusquelent.html	X	X								X	X					
http://www.teoria.com/articles/BWV772/index.htm	X	X									X			X	X	
http://www.teoria.com/articles/chopin4/index.htm	X	X									X			X		
http://www.teoria.com/articles/chopin8/index.htm	X	X									X			X		
http://www.teoria.com/articles/debussy8/index.htm	X	X									X			X		
http://www.teoria.com/articles/fugaBWV861/index.htm	X	X									X			X	X	
http://www.teoria.com/articles/II-V/indice.htm	X	X									X					
http://www.teoria.com/articles/iso/index.htm		X														
http://www.uwsp.edu/music/jsobaski/b-intro.htm		X														
http://www.zappa-analysis.com/	X										X					

7 Patent issues

7.1 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 recomendations 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 spinoff of the University of California, claims the ownership of the technology which permits embedding plugins in web pages, such as what is defined by the HTML standardized tag <OBJECT>, refering to the patent US 5,838,906, owned by the University of California, and for which Eolas
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 favor of Eolas.It has becomed 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 reexamination, due to a requirement of W3C's Director Tim Berners-Lee).

7.2 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 implementors – developers or industries - implementing software for encoding or decoding patented technologies have to beware of these patents.

From MPEG-4 Overview by Rob Koenen[1]:
"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

[1] full paper is available in the Standards documents on the MusicNetwork web site.

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 proposers 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. “

8 Documents, white papers, tutorials and presentations

These links have been accessed on 05/01/2004.

8.1 Macromedia Flash

A white paper presenting the Flash framework:
http://www.macromedia.com/software/flash/survey/whitepaper_jul03.pdf

8.2 Macromedia Shockwave

A white paper on Shockwave:
http://www.macromedia.com/tw/software/shockwaveplayer/whitepaper/whitepaper_nov00.pdf

8.3 Quicktime

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

The Quicktime 6.4 user's guide:
http://a1552.g.akamai.net/7/1552/51/3c2da90b3554b5/www.apple.com/quicktime/products/qt/pdf/QT6.4_UserGuide.pdf

8.4 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:
http://www.microsoft.com/windowsserver2003/docs/WMS_TechOverview.doc

8.5 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

8.6 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

8.7 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] Eleanor Selfridge-Field, editor. Beyond MIDI: The Handbook of Musical Codes. MIT Press, 1997.
- [9] MEGA project : http://www.megaproject.org/
- [10] CARROUSO project : http://www.ircam.fr/produits/technologies/CARROUSO.html

10 Glossary of Acronims

CWMN – Common Western Musical Notation
P2P – Peer to Peer
FFT – Fast Fourier Transform
i-TV – Interactive TV
MIDI – Musical Instrument Digital Interface
MPEG – Moving Picture Experts Group
ISO – International Standards Organisation
W3C – World Wide Web Consortium
OASIS – Organization for the Advancement of Structured Information Standards
DVD – Digital video disk
HDTV – High Definition Television
MP3 – A coding standard for compression of audio data: MPEG-1 Layer 3
AAC – Advanced Audio Coding, an audio compression scheme developed by the MPEG group (Dolby, Fraunhofer, AT&T, Sony, and Nokia).
AVI – Audio Video Interlaced, a video and audio format developed by Microsoft.
AC3 – Dolby Audio Coding, an audio compression scheme developed by Dolby.
WMA – Windows Media Audio
SMIL – Synchronized Media Integration Language, a language for synchronisation of multiple media developed by the W3C.
SVG – Scalable Vector Graphics
VRML – Virtual Reality Modeling Language
XML – eXtensible Markup Language
SGML – Standard Generalized Markup Language
RDF – Resource Description Framework
PCM – Pulse Code Modulation
WAV – A sound format developed by Microsoft standing for wave /waveform
AIFF – Audio Interchange File Format
OMF – Observatoire Musical Français