

## MPEG-4: A Multimedia Standard for the Third Millennium, Part 2

**Stefano Battista**  
*bSoft*

**Franco Casalino**  
*Ernst and Young Consultants*

**Claudio Lande**  
*CSELT*

MPEG-4 defines a multimedia system for interoperable communication of complex scenes containing audio, video, synthetic audio, and graphics material. In part 1 of this two-part article (in the previous issue) we provided a comprehensive overview of the technical elements. Here in part 2 we describe an application scenario based on digital satellite television broadcasting, discuss the standard's envisaged evolution, and compare it to other activities in forums addressing multimedia specifications.

The first MPEG-4 demo over satellite

In April 1997 the Systems subgroup founded an implementation ad hoc group (IM1) with the mandate of validating the concepts of the MPEG-4 standard and promoting it by developing demonstration applications. This group also contributes to part 5 of the standard, the reference software. A system architecture was developed and successfully demonstrated at the 45th MPEG-4 meeting, where the world's first real-time display

of MPEG-4 content (multicast over satellite) was shown to the MPEG community and the world press. Here we give a concise description of the architecture and demo mock-up.

MPEG-4 broadcast scenario

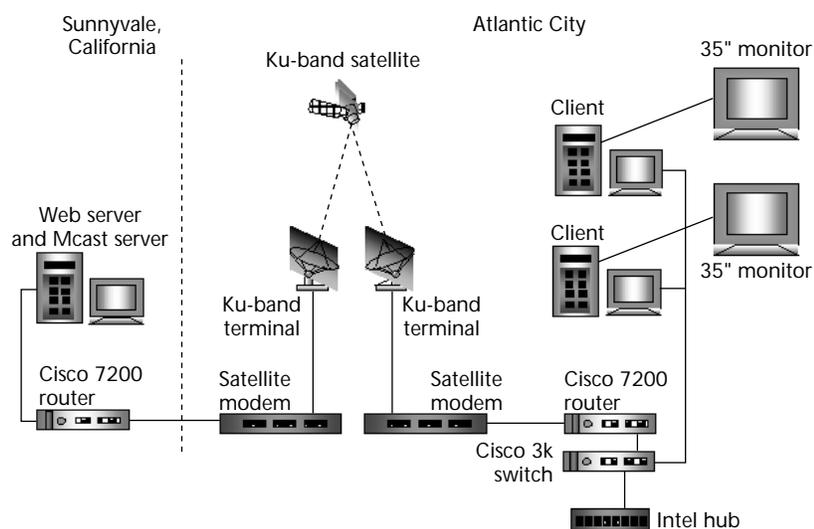
The scenario is an electronic programming guide (EPG) application where the user can choose among channels providing CD-quality audio, films, and news. The content is premultiplexed and then streamed. The target band for the overall application is 5 megabits per second. The overall system is constructed of a single (logical) origination point, a real-time, unidirectional communication channel, and a large number of end-user receiver/decoder terminals. It is a one-to-many, or possibly a few-to-many, system. The asymmetry of the architecture leads to an emphasis on reducing complexity and cost at the receiving side even if this implies increasing the complexity and cost at the transmitting side.

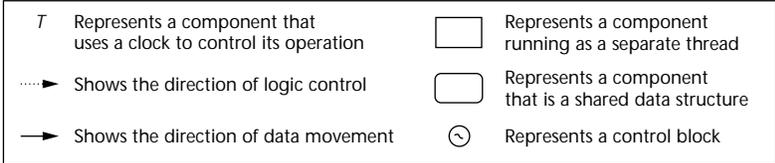
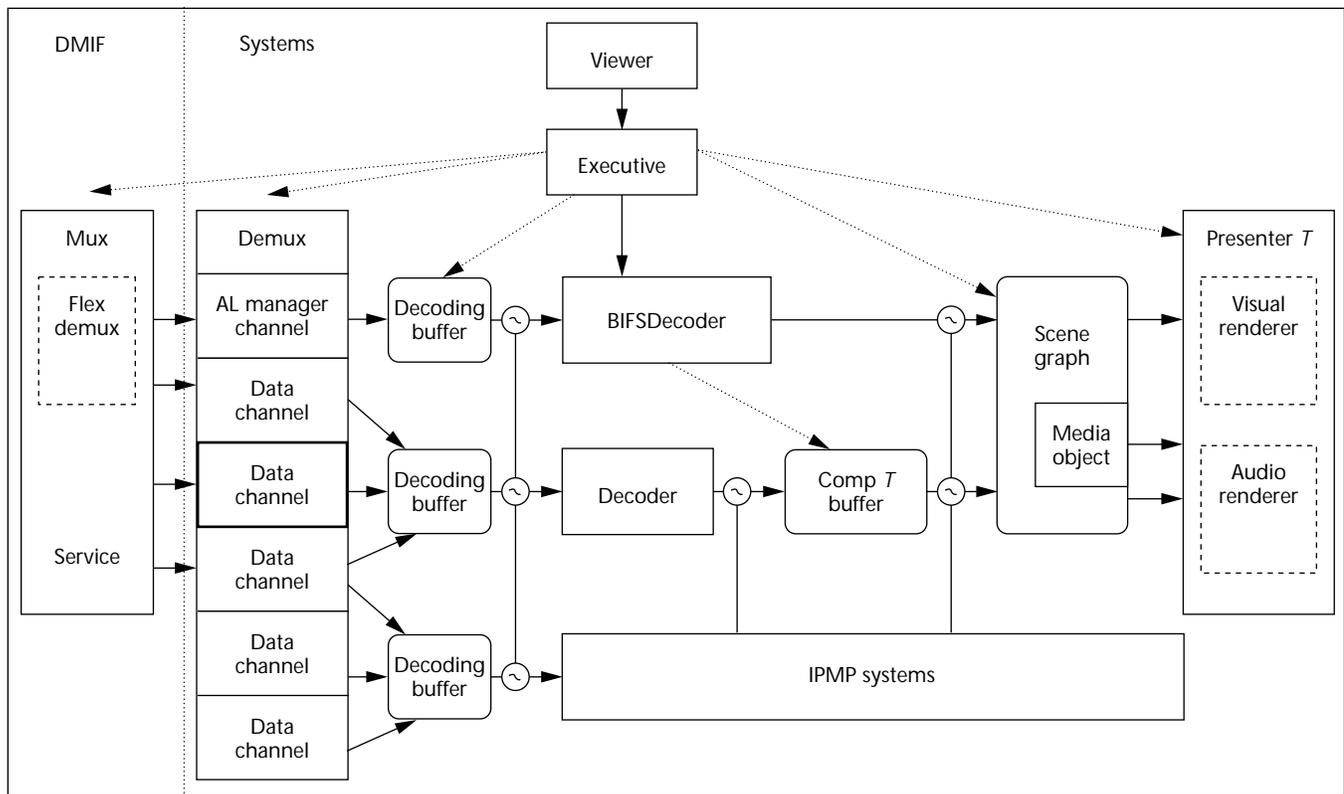
The architecture is schematically depicted in Figure 1.

Figure 1.

The demonstration consisted of MPEG-4 server components located in Sunnyvale, California linked via a hybrid satellite network to MPEG-4 clients in Atlantic City, New Jersey. MPEG-4 takes advantage of the latest compression technologies, providing rendering of multiple objects within one viewing window. The network uses the Internet Group Multicast Protocol (IGMP) to route

**Figure 1. MPEG-4 client-server setup.**





multimedia streams to only the locations where they are requested by clients, thus keeping network use to a minimum.

Now let's look more closely at the major components of the architecture, server side, client side, and content developed.

**MPEG-4 player**

The main components of an MPEG-4 terminal<sup>1</sup> include the network layer (Digital Media Integration Framework, or DMIF); the core, which includes demultiplexing, synchronization, and BIFS decoder tools; the audio and video decoders; and the presenter, which renders the scene.

**Player core.** The MPEG-4 player used in the demonstration implements a terminal compliant with the International Standard version of MPEG-4. This terminal mainly features the capability of playing an MPEG-4 multiplexed stream containing audio visual and synthetic information. The multiplexed stream consists of

- one initial object descriptor stream containing the information about how to locate the scene description and the associate streams,
- one scene description stream containing the 2D representation of the scene being played,
- one object descriptor stream containing the association between logical audio or video objects as they are referred to in the scene and their physical address, and
- one or many audiovisual objects coded in different ways.

Each stream also carries the information necessary for synchronization (time stamps, object clock references).

The architecture of an MPEG-4 terminal appears in Figure 2. The network layer part of this architecture, called DMIF, is shown for clarity—we will explain it later. The main components are

**Figure 2. MPEG-4 terminal architecture.**

the demultiplexer, the executive, the decoders, and the presenter, which includes the visual and audio renderer.

Note that the MPEG-4 system described by this block diagram operates within an application whose developer completely determines its operation. The application provides the graphical user interface to select the MPEG-4 scene to retrieve. It then creates an executive, which takes control of the application's execution. The demultiplexer sends each part of the bitstream to the appropriate component, all under the control of the main executive. The executive also creates the correct number and types of decoders, along with setting up the data paths between the components. The demultiplexer accesses the input stream and during startup creates a data channel for each elementary stream. During the initialization phase decoding buffers are also created. The BifsDecoder creates the scene graph while decoding the BIFS stream. After this step the presenter starts visiting the graph every 5 milliseconds and calls the visual and audio renderers to display the scene.

The compositor (scene graph) can employ user input events received by the presenter to change the composition information. The requirements for an MPEG-4 player include the ability for more than one decoder to run at any moment, each providing to the presenter the media object (audio or video) needed. This architecture provides that parallelism.

Each decoder runs in its own thread. Basically, it fetches elementary stream units called access units from its input buffer and produces composition units in the composition buffer. The decoder thread's behavior doesn't depend on the specific nature of the decoder, but it can be specified if necessary.

**Audio/video capability.** The platform can support different types of audio/video decoders. The video decoders integrated in the platform are JPEG, H263, and MPEG-4 Video. The latter also supports the shape format. In the same way the video renderer can support RGB, YUV, and YUVA RGBA formats. Video rendering uses the Microsoft DirectX library. Audio capabilities include G723, AAC mono, AAC stereo decoders, and MPEG-4 Audio.

The graphics engine also supports simple 2D graphics primitive display, such as rectangles, circles, spline curves, and text. Rendering was optimized to minimize the number of objects redrawn for each scene redrawing.

**Multicast server.** The Mcast\_server is a program that allows delivery of MPEG-4 content over IP multicast. Its main component, the Media Pump module, reads a file in an appropriate format and pumps it with the correct timing on a multicast user datagram protocol/Internet protocol (UDP/IP) socket. The main body of the program instantiates Media Pumps based on the information read on a particular startup file. The Media Pump module opens the indicated MPEG-4 stream file as well as its companion hint file (containing the timing information for correct delivery of the content), creates a UDP socket, and binds the socket to the indicated IP multicast address and port number.

Following complete delivery of an MPEG-4 stream file, the Media Pump resets its internal timeline and retransmits the file from the beginning. This loop continues forever.

The formats of the MPEG-4 stream file and the hint file are standardized in Version 2 of the specification. Adoption of the MPEG-4 file format is the target for Version 2.

**Network.** The network can manage Ethernet connections at 10 or 100 Mbps linked via a wide area network containing two Cisco routers and a point-to-point satellite link. The routers, linked by a full duplex 5-Mbps satellite channel, had the IGMP protocol enabled in order to propagate membership in multicast groups throughout the network. A Network Associates analyzer displayed utilization of network bandwidth.

**Satellite link.** The satellite link was provided by the GE-3 Satellite, manufactured by Lockheed Martin and operated by GE American Communications. The satellite follows a geostationary orbit at 87 degrees west, approximately 36,000 kilometers above the equator and uses the Ku band (11.7 to 14.5 GHz) for transmission. The Sunnyvale terminal employs a 1.8-meter dish antenna. A portable terminal, located on the Sheraton rooftop in Atlantic City, employs a 1.2-meter dish. Transmit power is a maximum of 16 watts at each location. Quadrature phase shift keying (QPSK) modulation with 3/4 rate Reed-Soloman forward-error correction coding provides a bit error rate typically below  $1 \times 10^{-13}$ . Propagation delay is 257 ms in each direction.

In the future, satellites will contain switching payloads, allowing multicast routing within the satellite component of the network. Other means are available today for multicast over satellite, including low-speed terrestrial return channels for initiation of group membership.

**MPEG-4 application.** The MPEG-4 application demonstrated an EPG application. The main scene consisted of six rectangles, each providing a service (Figure 3).

The first row displays three Quarter Common Interchange Format (QCIF)—that is,  $175 \times 144$  pixels—MPEG-4 videos intended as a video preview, running at 3 frames per second. The peak bit rate of the main scene reaches 506 Kbps, while the mean bit rate is 441 Kbps. Clicking on one video changes the scene, and the respective video is displayed at full screen with its associated audio. The full-screen video is again an MPEG-4 video, coded at 1,000 Kbps (peak rate) and 467 Kbps (mean rate). Although the dimension is Common Interchange Format (CIF)—that is,  $352 \times 288$  pixels—it is displayed scaled by a factor of two in both the horizontal and vertical directions. The associate audio is coded in the AAC mono at the bit rate of 64 Kbps. A “back” button lets the user return to the main scene.

The second row offers stereo music channels, all coded at 96 Kbps using the AAC encoder. This yields CD-like quality.

In the third row an application called “Tour de France” can be launched, providing a multiview scene of the famous bicycle race. The user can obtain ancillary information (such as additional information on the athletes or the location of the race) by simply clicking on the pictures.

Finally, this application includes a news program and a video channel showing an MPEG-4 Video with a girl dancing in front of a waterfall (Figure 4).

Technically this application consists of one main BIFS scene and nine channels “appended” and deleted using the Virtual Reality Modeling Language (VRML) construct Inline. This enables good modularity and facilitates testing of the overall application.

The IM1 group created the main mosaic scene in ASCII format using a first version of an MPEG-4 authoring tool, then converted it in the BIFS format



**Figure 3.** Main scene snapshot for the electronic programming guide application.



**Figure 4.** MPEG-4 Video showing a girl dancing in front of a waterfall. The background image is in JPEG format, while the girl is an MPEG-4 video object encoded from YUV and alpha raw stream.

using an encoder to reduce the size from 56 to 6 Kbytes. The same process compressed the simple audio-video channel description scene from 877 to 90 bytes. This clearly shows the importance of a binary compressed format for the scene description.

#### MPEG-4 Version 2: The next step

In October 1998, the first set of MPEG-4 standards was frozen. Work continues for Version 2, which will add tools to the MPEG-4 standard. Existing tools and profiles from Version 1 won't be replaced in Version 2; instead, technology will be added to MPEG-4 in the form of new profiles. MPEG-4 Version 2 is a backward compatible

extension of Version 1. The Systems layer of Version 2 is backward compatible with Version 1, while in the area of Audio and Visual, Version 2 will add profiles to Version 1.

### Systems

Version 2 of the MPEG-4 systems extends Version 1 to cover issues like multiuser interaction and Java (MPEG-J) support. Version 2 also specifies a file format to store MPEG-4 content.

Multiuser functionality lets several users access the same scene and interact with its content. Multiuser interactions are considered an extension of single-user systems rather than targeting a specific application domain.

The MP4 file format contains the media information of an MPEG-4 presentation in a flexible, extensible format that facilitates interchange, management, editing, and presentation of the media. This presentation may be local to the system containing the presentation, or it may be distributed via a network or other stream delivery mechanism (a TransMux). The file format is designed to be independent of any particular TransMux while enabling efficient support for TransMuxes in general. The design is based on the QuickTime format from Apple Computer.

MPEG-J is a programmatic system (as opposed to the parametric system offered by MPEG-4 Version 1) that specifies the application programmers interface (API) for interoperation of MPEG-4 media players with Java code. By combining MPEG-4 media and safe executable code, content creators may embed complex control and data processing mechanisms with their media data to intelligently manage the operation of the audiovisual session.

### Visual

The Visual part of the MPEG-4 standard will be extended with tools in the following areas:

- increased flexibility in object-based scalable coding;
- improved coding efficiency, especially for very low bit-rate applications requiring simple decoders;
- improved error robustness; and
- coding of multiple views.

On the final point, intermediate views or stereoscopic views will be supported based on the effi-

cient coding of multiple images or video sequences. A specific example is the coding of stereoscopic images or video by redundancy reduction of information contained between the images of different views.

### Audio

The following additional functionalities will be supported by MPEG-4 Audio Version 2:

- error resilience methods;
- environmental spatialization;
- low delay general audio coding;
- syntax for a backchannel, for adaptive coding and play-out of audio objects;
- small step scalability; and
- parametric audio coding.

### Synthetic/Natural Hybrid Coding

Now we'll briefly describe MPEG-4 Synthetic/Natural Hybrid Coding (SNHC) Version 2's new tools.

**Full compliance with VRML.** MPEG-4 Version 1 isn't completely compatible with VRML, hence not all the contents are exchangeable between the two standards. The missing parts are the PROTOs and EXTERNPROTOs, syntactic rules that let content creators define new nodes and use them as if they were built-in VRML nodes. PROTOs are defined through information available in the local terminal. EXTERNPROTOs are defined on remote locations, and in VRML they can also be implemented in languages different from VRML. PROTOs and EXTERNPROTOs will be added to MPEG-4 Version 2.

**Body animation (BA).** Instancing a body object in an MPEG-4 scene causes rendering of a synthetic human body in the terminal. Body objects, represented by 3D meshes, should already be present at the receiver side. (It's not necessary to transmit the body model.) To have an MPEG-4 animated body model, only animation parameters can be sent.

BA defines two different sets of data: Body Animation Parameters (BAPs) and Body Definition Parameters (BDPs). Their semantics are the analog

of facial animation parameters (FAPs) and facial definition parameters (FDPs), respectively.

BAPs represent limb motion as joint rotations from the neutral position (standing body, parallel feet, arms along the sides with the palms of the hands facing inward). BAPs have been defined so that they can produce reasonably similar animations and postures with different terminal body models. Like FAPs, BAPs' values are unconstrained. This means that the resulting animations may range from realistic human motion to exaggerated cartoon-like characters suited, for example, to video-games.

When the encoder wants stricter control over the receiving terminal body model, it must send more information in the form of BDPs. BDPs define the receiving terminal geometric body model to use for rendering. A BDP can either be used to adapt the terminal model (for example, by customizing body surface, dimensions, and texture) or to transmit a completely new one. When the encoder sends a new body model to the receiver terminal, it also sends some information called Body Definition Tables that describe how to animate the model with BAPs received from the input stream.

Body Interpolation Tables instead define how the receiving terminal can interpolate missing BAPs from the ones contained in the animation stream. This is especially useful in reducing the number of parameters to transmit to the receiving terminal.

The MPEG-4 group joined with the VRML Humanoid Animation working group to create a common architecture for human body modeling.

Putting together face and body animation, the content creator can insert complete human avatars in an MPEG-4 scene. The use of avatars is likely to become common in many different MPEG-4 applications, including virtual meetings (avatars representing real people "meet" in a virtual shared room), model-based videoconferences (instead of transmitting video information, two or more terminals send each other tiny streams of animation parameters; this allows a much higher frame rate than, for instance, traditional videotelephone), realistic networked video games, and so on.

**3D mesh encoding.** When complex 3D objects must be transferred to an MPEG-4 terminal, it's often impractical or simply impossible to transmit the entire mesh (as an IndexedFaceSet BIFS node). The need arises for a more efficient way to encode meshes. Sometimes it's not necessary to transmit the mesh's exact data because it's not required (for instance, whenever the mesh

represents an object whose shape is not well defined, like a terrain, a cloud, a rock, and so on). In such cases a lossy encoding can prove much more efficient in terms of mesh compression. MPEG-4 supports both lossy and lossless encoding. The tools developed by MPEG-4 for mesh encoding involve compression of structure or connectivity, geometry or shape, and properties such as shading normals, colors, and texture coordinates of 3D models.

Meshes can be static or animated by means of animation streams. The latter can be used to create interesting contents with morphing objects.

Mesh transmission/representation may also benefit from the Level of Detail (LOD) node, which can specify different detail levels of a 3D object depending on the viewer distance. A 3D model can have a coarser geometry when the user viewpoint lies far from the object and a finer one when the viewpoint lies close to the object.

Another important tool provided by MPEG-4 is the progressive transmission of geometry. This consists of enhancing 3D mesh details through progressive transmission of differential geometry and structure (similar in functionality to progressive texture). This also allows spatial scalability according to available bandwidth and terminal rendering capabilities.

Efficient encoding of meshes can be useful in the context of face and body animation, where 3D models must be transmitted to an MPEG-4 terminal for use as the avatars for animation and rendering.

**Advanced Audio BIFS (spatialized audio).** Advanced Audio BIFS nodes provide a parametric description of the environment in which a set of audio sources is inserted. Among the new elements taken into account are how physical objects interact with (reflect/transmit) sound, how distance and air attenuate sound, how sound reverberates in a physical space, and so forth.

These tools can model precisely how sound behaves and is perceived by people in physical space. This model can provide very realistic sound effects. Among the possible applications are highly immersive virtual reality applications (driving and flight simulators), realistic 3D video games, fine modeling of concert halls, and so on.

Delivery Multimedia Integration Framework (DMIF)

The delivery efforts for MPEG-4 Version 2 will consider the following issues:

- Add mobile operation with DMIF.
- Extend DMIF V1 QoS to Access Unit Loss and Delay parameters at the DMIF Application Interface (DAI).
- Invoke session and resource management (SRM) on demand after an initial session has been established (with the tools present in DMIF v.1).
- Allow heterogeneous connections with an end-to-end agreed-upon QoS level.
- Integrate with Internet Engineering Task Force (IETF) specified network servers.
- Provide fully symmetric consumer and producer operations within a single device.
- Enable end-to-end sessions across multiple network provider implementations.

#### Other “multimedia” standards

A few activities under development in the multimedia industry may to some extent overlap with the scope of MPEG-4, specifically in providing a standardized format for the integration of media of different nature. We’ll briefly describe the following specifications and compare them to MPEG-4: Extensible 3D (X3D), Synchronized Multimedia Integration Language (SMIL), and Broadcast HyperText Markup Language (BHTML).

#### X3D

X3D is the current specification effort of the Web3D (formerly the Virtual Reality Modeling Language, or VRML) consortium. While the specification remains work in progress, it seems more like an evolution of VRML 97 toward more “compact” implementations and interoperability with other APIs than aiming for an integrated system for continuous media.

Several efforts from the VRML community define extensions to the VRML 97 specification related to streaming of continuous media. While most implementations of VRML browsers assume that audio, video, or audiovisual clips are downloaded first to the client terminal and then played, MPEG-4 integrates streaming of scenes and continuous media embedded in the scenes. Several activities in the VRML consortium directly address the issue of media integration, like the VRML Streams working group and the VRML MPEG-4

working group. They may affect the new release from the VRML community.

MPEG-4 borrowed most of its architecture for scene graph representation from VRML 97. On the other side, the VRML community is following a different path of evolution, possibly more suitable for typical client-server applications.

#### SMIL

The W3 consortium’s SMIL specifies a format for integrating independent multimedia objects into a single multimedia presentation, with coherent temporal and spatial attributes. To achieve this goal, SMIL provides techniques for

- temporal composition (synchronization),
- spatial composition (layout), and
- association of hyperlinks.

All the descriptors in the SMIL specification follow an HTML-like format. These syntax elements provide an extension on top of HTML and Extended Markup Language (XML) that tells the browser how to position in time and space, and how to link to other content.

SMIL was defined as an extension of XML to specifically address the issue of presenting content comprising several independent media objects. Some major differences distinguish SMIL and MPEG-4. SMIL

- assumes media objects are in separate URLs,
- specifies synchronization at the stream level (beginning and end of a stream, leaving the fine-grain synchronization to the browser implementation), and
- provides no explicit support for 3D objects.

#### BHTML

BHTML is the extension to HTML being specified by the ATSC, the committee for digital broadcasting of the United States Federal Communication Commission. The comparison with MPEG-4 is interesting, since while BHTML is specifically designed for applications like “interactive digital TV with browsing functionality,” the MPEG-4 specification is an alternative candidate for the same application.

The approach taken by BHTML somewhat resembles that of SMIL. Typical HTML content is

complemented by (discrete or continuous) media objects embedded in the same presentation. Since the specification is intended for Web navigation on digital TV platforms, a better definition exists of how to synchronize the continuous media streams (like audio and video) to the other media objects in the scene. For example, the synchronization attributes specify whether several objects shall have a sequential or parallel presentation and, further, what primitive must be used to perform synchronization (clock, frame, or “soft” synchronization by the implementation).

Also different from SMIL, BHTML is being designed specifically to downscale HTML and XML for other aspects considered less compelling for digital TV.

#### MPEG-4 tools and projects

MPEG-4 tools and projects developed or under development in the multimedia industry include the following.

#### Emphasis

The Emphasis project is part of the European cooperative research program Advanced Communications Technologies and Services (ACTS). The objective of the Emphasis project is to firmly establish a European lead in software and silicon technology suitable for meeting the needs of MPEG-4. The project will deliver five key technologies:

1. Systems architecture
2. Optimized software implementation of MPEG-4 audiovisual tools and algorithms
3. Audio and video composition and rendering technology (2.5D and 3D)
4. Specifications for processor and coprocessor architectures that meet the processing demands of MPEG-4 applications
5. An integrated demonstration of a typical MPEG-4 application

#### MoMuSys

The ACTS project MoMuSys (Mobile Multimedia Systems) targets the development and validation of the technical elements necessary to provide new audiovisual functionalities for mobile multimedia. The main objectives are

- New multimedia communications system with

full conversational possibilities and totally new functionalities described by a syntactic descriptive language

- Global field trial with 20 to 60 real terminals and real users for symmetric and asymmetric multimedia over wireless networks
- European MPEG-4 platform addressing all MPEG-4 functionalities

#### Mirador

MPEG-4 Intellectual Property Rights by Adducing and Ordering, or Mirador (also an ACTS project), will define, develop, and promote an extensive framework for copyright protection compliant with the MPEG-4 Intellectual Property Rights group specifications, covering still pictures, video, and audio. The project’s main objectives consist of

- delivering an MPEG-4 compliant watermarking for audio, still picture, and video objects;
- integrating these technologies with MPEG-4 codecs;
- pushing forward the technology so that rights holders obtain a better understanding of watermarking technology; and
- creating the first watermarked MPEG-4 production.

#### COMIQS

Commerce through MPEG4 on the Internet with Quality of Services, or COMIQS (within the ACTS program), targets the technical and service validation of a set of new paradigms. Within the convergence process of interactive multimedia applications—ranging from entertainment to engineering—the COMIQS achievements will be validated in the field of electronic commerce.

The project’s added value and specific contribution reside in the following service and technical features:

- All-media integration in a 3D virtual catalog integrating digital high-quality multimedia with video, photorealistic images, and high-quality sound
- Increased interactivity with content such as objects contained in motion pictures, audio tracks, and 3D objects

## For More Information

IPA: <http://www.arts.gla.ac.uk/IPA/ipa.html>  
 X3D site: <http://www.web3d.org/>  
 SMIL site: <http://www.w3.org/TR/REC-smil/>  
 BHTML site: <http://toocan.philabs.research.philips.com/misc/atsc/bhtml/>  
 Emphasis site: <http://www.fzi.de/esm/projects/emphasis/emphasis.html>  
 MoMuSys site: <http://www.at.infowin.org/ACTS/RUS/PROJECTS/ac098.htm>  
 Mirador site: <http://www.de.infowin.org/ACTS/RUS/PROJECTS/ac302.htm>  
 COMIQS site: <http://www.de.infowin.org/ACTS/RUS/PROJECTS/ac322.htm>  
 Venus site: <http://www.eurescom.de/Public/Projects/p900-series/P922/p922.htm>

- Real-time streaming of all media into the user-sided scene, facilitating navigation and better adaptation of the content presentation to available network resources
- Placement of MPEG-4 into the Internet context, in the spirit of ITU-H.225 and of IETF Audiovisual Transport Group
- Quality-of-service (QoS) management, from best-effort to predicted QoS, encompassing scalable services on a variety of access networks, from narrowband to broadband Internet

### Venus

Virtual Environment with Next Generation Multimedia Systems, or Venus, is a new Eurescom project aimed at developing shared virtual environments where people can meet, talk, and share applications, multimedia contents, documents, and so on. The project also focuses on the development of a technology that allows videoconferencing among many participants.

Video objects, static face models with texture-mapped video, or animated avatars, can represent people in Venus. The choice mainly depends on the target application. The representation of all multimedia information will most likely be based on MPEG-4 format in order to exploit its

- data compression algorithms,
- stream multiplexing capabilities,
- scene description and updating, and
- face and body animation and definition parameters.

### SONG

The SONG (Portals of New Generation) project, developed in the context of the European program Information Society Technologies (IST), will investigate the creation of virtual environments populated by synthetic humanoids (avatars). These avatars, representing real users, will be smart enough to understand and react to users' vocal orders. They will be very photorealistic in appearance and animation (for instance, they will show facial expressions, make gestures, assume body postures, and so forth).

The project will investigate a real-time phoneme recognition engine so that the speakers' voices can be analyzed to extract phonemes that, in turn, can be used to animate the avatars' faces. Some avatars could also embody intelligent agents and be smart enough to act autonomously.

MPEG-4 technology will be heavily used to represent the 3D worlds, the avatars and their real-time animation, the multimedia contents included in the scene, and so on.

### Conclusions

In this two-part article, we introduced and provided a reference framework to help you understand a standard much more complex than its predecessors. (MPEG-1 and MPEG-2 are limited essentially to audiovisual content representation.)

The definition of profiles in the specification, and the structure of the standard in several parts, makes it applicable to a wide variety of applications. To name a few, think of an interactive TV broadcast that looks more like the Web pages users see in a PC environment. Or think of an optimized error-resilient audio and video codec applied in wireless digital communication.

The hype that the computer market and the consumer market are getting closer and closer to each other, with Internet communication being a major driving force, has been around for a few years now. This article on MPEG-4 should help in understanding how a technology specification that guarantees interoperability between systems and applications from different manufacturers or service providers can foster this evolution. MM

### References

1. *Advances in Multimedia: Standards, Systems, and Networks*, Atul Puri and Tsuhan Chen, eds., Marcel Dekker, New York, 1999.

*Readers may contact Battista at bSoft, 156 via Velini, 62100 Macerata, Italy, e-mail bautz@tin.it.*