MPEG-4
– opening new frontiers to broadcast services

F. Pereira
Instituto Superior Técnico/Instituto de Telecomunicações

R. Koenen
KPN Research

The recently finalized MPEG-4 standard follows an object-based audio-visual representation model, with integrated hyperlinking and interaction capabilities. It offers support for both natural and computer-generated content.

This article discusses the new possibilities that MPEG-4 technology could offer to broadcast services, and addresses the consequences in terms of new and enhanced applications. MPEG-4 is not designed to replace MPEG-2. On the contrary, MPEG is actively working on a solid integration of MPEG-2 and MPEG-4 technology, which will allow new and attractive services within existing digital broadcasting environments.

1. Context and objectives

Since its introduction in the thirties, audio-visual broadcasting has had an enormous impact on our society. During all that time, the audio-visual representation model has remained the same – the audio-visual signal is broadcast as a periodic sequence of a number of image lines, along with the associated audio: the *television paradigm*.

Neither the introduction of colour, nor even the (MPEG-2 based) digitalization of our television systems has substantially changed this. Of course, the MPEG-2 standard *has* opened up a range of new possibilities, spawning an explosion of new broadcast services, including "free-to-air", subscription, (near-)video-on-demand and pay-per-view. Bandwidth can be deployed far more efficiently and various types of services can be multiplexed together. Limited interactivity has become possible, either using the same delivery channel or by invoking a separate link.

In the last five years, the explosion of the Web, and the acceptance of its interactive mode of operation, have clearly shown that the traditional television paradigm would no longer suffice for audio-visual services. Users will want to have access to audio and video over the Web, just like they now have access to text and graphics. This requires moving pictures and audio of good quality at low bit-rates on the Web, and Web-type interactivity with live content. It should be possible to activate relationships between entities (in a potentially virtual world)

Original language: English
Manuscript received: 11/2/99.

through hyperlinking – the *Web paradigm* – and to experience interactive immersion in natural and virtual environments – the *games paradigm*.

1.1. Paradigm shifts in television viewing

The "Web" and "games" paradigms will inevitably have their impact on television services. What has started with simply displaying an HTML page on a TV screen will evolve to an environment in which content comes from both broadcast and interactive networks, seamlessly integrated on the same screen and in the same sound space. Users will click on a car in an advertisement, and an already downloaded interactive animation will start, allowing them to see the car from all sides. Further clicking will (transparently) establish an on-line connection to the local dealer and will supply price information and an order form, with a (virtual?) sales assistant helping to answer any questions the user may have. Users will also choose broadcast television programmes from on-line TV guides, and ask for on-line information through links sent with programmes.

Enhancements of TV can already be witnessed in the boxes that digitally record programmes based on information from on-line connections. Examples are Replay TV [1], WebTV [2] and Tivo [3].

The increasing mobility in telecommunications is another major trend that is important to MPEG-4. Five years ago, MPEG foresaw that mobile connections would not be limited to voice, but that data, including real-time media, would be next. Because mobile telephones are replaced every two to three years, new mobile devices can finally make the decade-long promise of audio-visual communications turn into reality.

Many audio-visual applications demand interworking, which is best served by an open and timely international standard. In 1993, MPEG [4] launched the MPEG-4 work item – officially called "Coding of audio-visual objects" – to address, among others, the requirements mentioned above [5].

1.2. Capabilities of the MPEG-4 standard

The three main developments (i.e. the mounting importance of audio-visual media on all networks, growing interactivity and increasing mobility) have driven – and still drive – the development of the MPEG-4 standard. To address the identified requirements, a standard was needed that could:

- efficiently represent a number of data types:
  - **video** with a wide range of bit-rates, according to the quality requirements;
  - **audio** with a very wide bit-rate range, according to the quality requirements (e.g. from transparent music down to low-quality speech);
  - **generic dynamic 2-D and 3-D visual objects**, as well as **specific objects** such as human faces and bodies;
  - **speech and music synthesized in the decoder**, including support for 3-D audio spaces;
  - **text** and **graphics**;

- provide – in the encoding layer – resilience to residual errors, for various data types, especially under difficult channel conditions such as in mobile environments;
- independently represent the various objects in the scene, allowing independent access for their manipulation and re-use;
- compose audio and visual objects, both natural and synthetic, into one audio-visual scene;
- describe the objects and the events in the scene;
- provide interaction and hyperlinking capabilities;
- manage and protect the intellectual property rights associated with audio-visual content and algorithms, so that only authorized users could have access.
- provide a representation format that is independent of the media delivery method, one that can transparently cross the borders of different delivery environments.

These requirements are addressed by the six parts of the recently finalized MPEG-4 standard:

Part 1: Systems – which specifies the scene description, multiplexing, synchronization, buffer management, and management and protection of intellectual property [6];

Part 2: Visual – which specifies the coded representation of natural and synthetic visual objects [7];

Part 3: Audio – which specifies the coded representation of natural and synthetic audio objects [8];

Part 4: Conformance Testing – which defines the conformance conditions for bitstreams and devices; this part is used to test MPEG-4 implementations [9];

Part 5: Reference Software – which includes software corresponding to most parts of MPEG-4 (normative and non-normative pieces); it can be used for implementing compliant products [10];

Part 6: Delivery Multimedia Integration Framework (DMIF) – which defines a session protocol for the management of multimedia streaming over generic delivery technologies [11].

Parts 1 to 3 and 6 specify the core MPEG-4 technology, and Parts 4 and 5 are ”supporting parts”. Parts 1, 2 and 3 are delivery-independent, leaving to Part 6 (DMIF) the task of interfacing with the idiosyncrasies of the delivery layer.

The major difference with previous audio-visual standards is the object-based representation model that underpins MPEG-4. An object-based scene can be built using individual objects that have relationships in space and time, offering a number of advantages.

Firstly, different types of objects may have different suitable coded representations: a synthetic moving head is clearly best represented using animation parameters, while camera video benefits from a smart representation of the pixel values. Secondly, it allows harmonious integration of different types of data into one scene: an animated cartoon character in a real world, or a real person in a virtual studio set. Thirdly, interacting with the objects and hyperlinking from them is now feasible.

There are more advantages, such as selective spending of bits, easy re-use of content without transcoding, providing sophisticated schemas for scalable content on the Internet, etc. *Fig. 1.* gives an example of an MPEG-4 scene with objects of various types. *Fig. 2.* gives the “scene graph" for this scene, which is discussed in *Section 2.2*. *Fig. 3.* (on the next page) gives a high-level view of an MPEG-4 system. Note the individual streams for each of the objects.

The applications that benefit from MPEG-4 are found in many – and very different – environments. Therefore, MPEG-4 is constructed as a tool-box rather than a monolithic standard, using profiles that provide solutions in these different settings (*Section 3* further discusses profiling). MPEG-4 can be used to deploy complete new applications or to improve existing ones. Unlike MPEG-2 (digital television), MPEG-4 does not target a major "killer application"; rather, it opens many new frontiers. Playing with audio-visual scenes, creating, re-using, accessing and consuming audio-visual content will become easier. New and richer applications can be developed in, for example, broadcasting, remote surveillance, personal communications, games, mobile multimedia, virtual environments, etc. It allows services with combinations of the three, traditionally different, service models: "broadcast", "(on-line) interaction" and "communication". As such, MPEG-4 addresses "convergence", when defined as the proliferation of multimedia in all kinds of services and on all types of (access) networks.

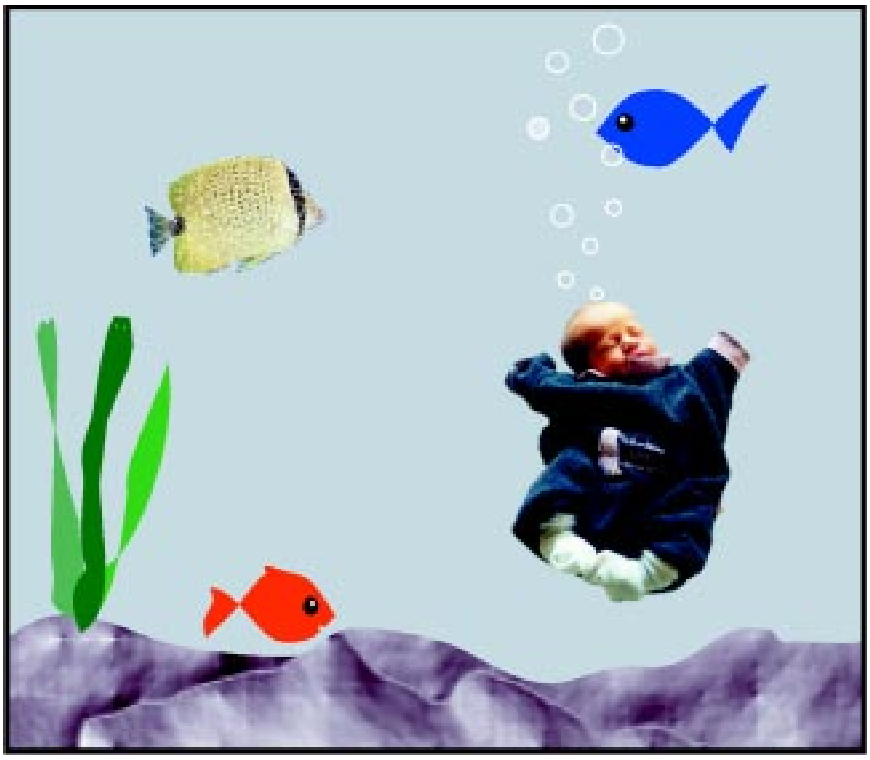

Figure 1
An MPEG-4 scene with objects of various types.

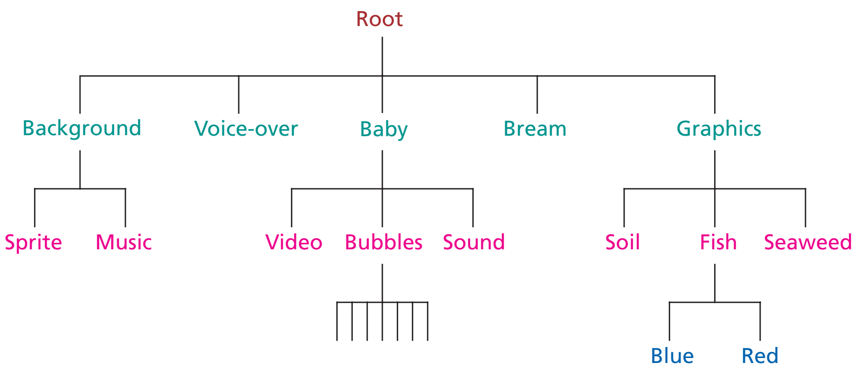

Figure 2
The "scene graph" for the scene shown in *Fig.1*.

Some people seem to think that MPEG intended MPEG-4 as a replacement for MPEG-2. Quite the contrary is true. MPEG-2 will continue to work well for services that use the television paradigm. In fact, MPEG is currently working on a smooth integration of the two standards, to enable usage of MPEG-4 functionalities in MPEG-2 environments. This includes integration of MPEG-2 encoded content (MPEG-2 "objects") in MPEG-4 scenes, and the transport of MPEG-4 content over MPEG-2 systems – possibly to enhance the existing MPEG-2 broadcast services.

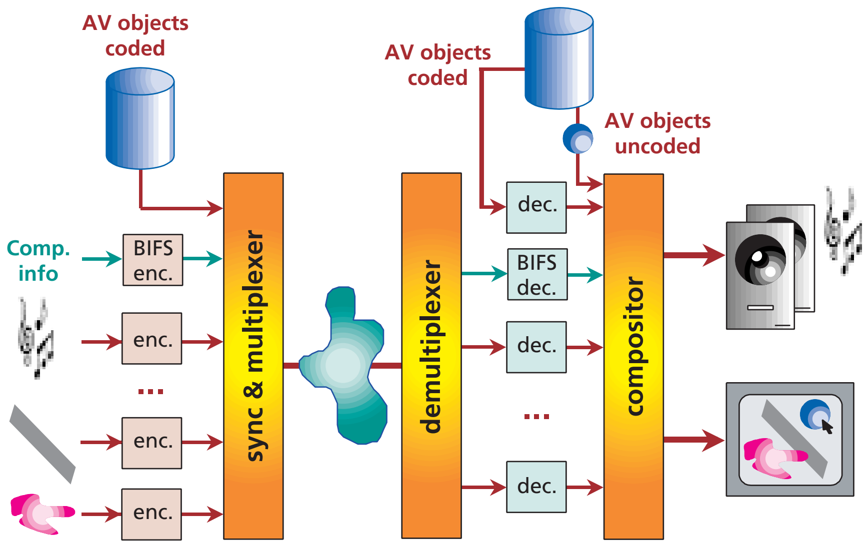

Figure 3
The object-based architecture of an MPEG-4 system.

2. The technological milestones of MPEG-4

MPEG-4 was developed over the past five years by hundreds of experts from many companies spread globally, who believed that the MPEG-4 technology can power their next generation products and services. MPEG-4 Version 1 is now available. MPEG-4 Version 2 will extend the capabilities of the standard in a backwards compatible way, to be ready by the end of 1999. Participants in MPEG-4 represent broadcasters, equipment and software manufacturers, digital content creators and managers, telecommunications service providers, publishers and intellectual property rights managers, as well as university researchers.

We will now highlight a number of MPEG-4 tools, explaining what they can do and how they are new [5].

2.1. Coded representation of media objects

A very important goal of MPEG-4 is the integration of different types of media objects. An efficient coded representation of these objects is a pre-requisite, so each type of object in the scene has its own optimal representation. This is in contrast with, for example, MPEG-2 Video where all objects are merged together, transformed into pixels, and coded using hybrid DCT coding. MPEG-4 defines efficient coded representations for the following types of objects:

- natural audio (including special speech codecs);
- synthetic sound;
- a text-to-speech interface;
- arbitrarily-shaped video and stills;
- facial and body animation;
- generic 2-D and 3-D “computer-generated objects”.

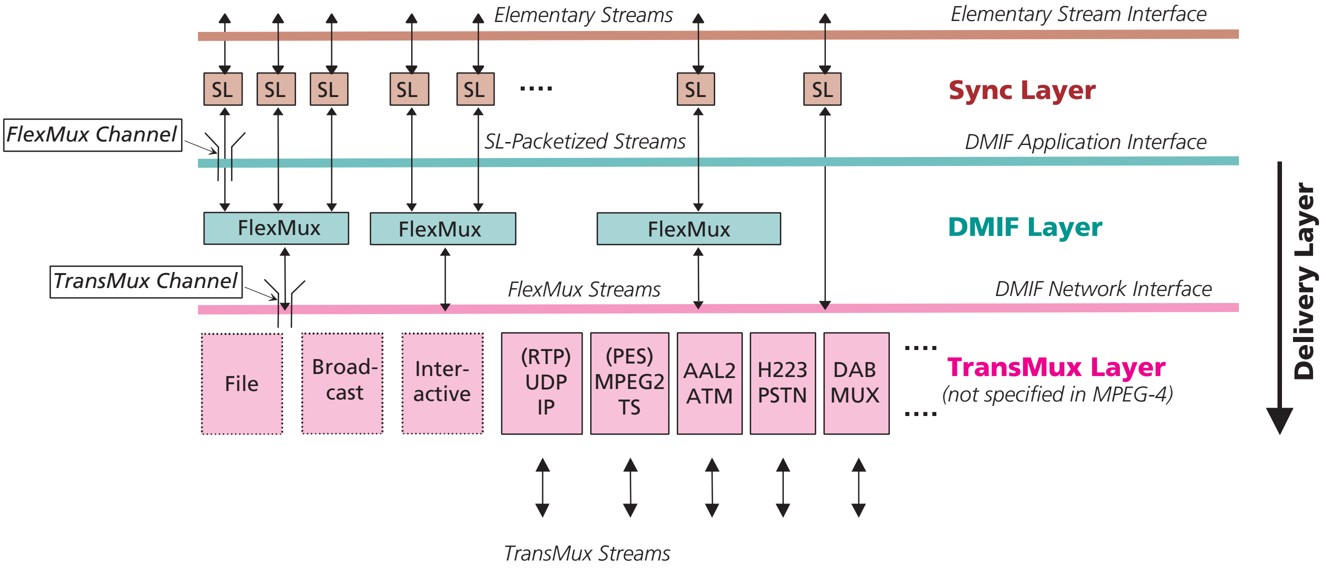

Figure 4
The elementary streams associated with MPEG-4 objects.

Objects have one or more associated *elementary streams* (ESs), in which the coded data are transferred (see *Fig. 4.*). More than one such ES are needed for, e.g., a scalable object with a base quality and one or more enhancement layers, each with its own ES. The coded representation of media objects is optimally efficient, given also other requirements than compression efficiency (e.g. error resilience, random access, easy extraction and editing of an object, or scalability). Because arbitrarily-shaped visual objects need to be supported, MPEG-4 is the first standard with a shape-coding tool. Using it, for example, material that was shot against a blue screen can be represented with explicit shape information. The "shape" can be a contour, or an alpha map defining the transparency for all pixels, allowing good quality blending of several objects together – the background of course also being an object.

Video coding is supported from 5 kbit/s to 10 Mbits/s, and work is currently ongoing to extend this to studio quality. Speech tools work from 2 kbit/s upwards and audio tools are available from 4 kbit/s up to transparent quality at 64 kbit/s per channel.

2.2. Composition, interaction and hyperlinking of objects

Another important innovation is BIFS *(BInary Format for Scenes)*, a binary language for defining the layout of a scene (audio and visual) and for changing and updating the scene with, for example, new objects. BIFS has its roots in the text-based "Virtual Reality Modeling Language" (VRML) [12], adding an efficient binary representation and real-time updates as well as 2-D primitives. BIFS commands are available for changing the properties of objects and for defining their conditional behaviour. The BIFS description of the scene follows a hierarchical structure, which can be represented by a graph: the *scene graph*. *Fig. 2* gives the scene graph for the scene in *Fig. 1*. Each node of the graph is a scene object. As the scene is dynamic, so is the graph structure: relationships can dynamically change and nodes can be added or deleted. The objects can be located in a 2-D or 3-D scene, and 2-D objects can be included in 3-D scenes if required.

BIFS is designed as a collection of so-called *nodes*, which describe the scene and its layout. Like media objects, the BIFS commands have their own elementary stream. An object in the scene is described by one or more BIFS nodes, which may be grouped together using a grouping node. A video object is called a *media object* because it has an associated elementary stream (a media stream) with its coded texture. Text and graphics objects are not media objects, as all their characteristics and behaviour are defined through BIFS nodes, in the BIFS elementary stream.

There are also BIFS nodes for defining interaction and hyperlinking capabilities, to be used for any type of object. They allow the specifying of behaviour, both unconditionally (setting an object in motion, changing its colour) or conditionally in response to user commands (click on an object to make a menu appear, or to view it from a different angle).

2.3. Managing and pro-tecting intellectual property

In order for MPEG-4 to successfully create the audio-visual playground of the future, it is essential that intellectual property in MPEG-4 content can be identified and protected. The required intellectual property management and protection methods are probably as diverse as the applications themselves. Also, the required type and level of protection depends on the content's value, complexity, the delivery model (broadcast being the most difficult) and the associated business models. Therefore, and because protection schemes tend to be broken after some period, MPEG-4 does not standardize any specific protection algorithm. Rather, it includes the MPEG-4 Intellectual Property Management & Protection (IPMP) framework, which is a set of hooks, integrated deeply into the system, which give application builders the ability to construct the most appropriate domain-specific IPMP solution.

2.4. Description and synchronization of media objects

Media objects need streaming data, conveyed in one or more elementary streams. An *object descriptor* identifies which streams belong to one media object, and which decoders should be invoked to decode the data. This is used to handle scalable data. Also, descriptive information about the content – Object Content Information (OCI) – and intellectual property identification can be carried in the object descriptors.

Like in MPEG-2, time stamping of individual access units within elementary streams is used to achieve synchronization. The synchronization layer is designed to identify such access units and to perform the time stamping. This layer identifies the access units in elementary streams, recovers the media object's or scene description's timebase, and performs the actual synchronization. The syntax of this layer is configurable in a number of ways, allowing its use in a broad spectrum of systems.

2.5. Delivery of streaming data

To allow efficient transmission of the possibly many objects in a scene, MPEG-4 defines a two-layer multiplexer: the *FlexMux* and the *TransMux* (Transport Multiplexing).

With a low overhead, the (optional) FlexMux tool groups together elementary streams, because often many streams are needed in a scene and it is highly inefficient to set up separate channels for all of them. Elementary streams requiring similar QoS channels may be combined, reducing the network connections and lowering the end-to-end delay.

The TransMux layer offers the actual transport service, matching the requested QoS. MPEG-4 does not specify any of these TransMuxes, because many already exist; only the interface to this layer is specified. In some cases, MPEG works with other bodies to define the mapping of the data packets and control-signalling to the network. A Real-Time Protocol (**RTP**) mapping is being defined with the Internet Engineering Task Force, for example. Suitable, existing, transport protocol stacks which play the TransMux role are (RTP) / UDP / IP, (AAL5) / ATM, or MPEG-2's Transport Stream. Leaving the choice to the end-user / service provider allows MPEG-4 to be used in a wide variety of operational environments.

3. MPEG-4 profiling

Some people find MPEG-4 too big and difficult to implement. While MPEG-4 is indeed a large standard, profiling permits the specification of less complex (and, hence, less expensive) MPEG-4 devices, targeted to certain classes of applications. Profiling is also used to check whether MPEG-4 devices conform to the standard.

There are a number of different types of profiles:

- **visual profiles** are defined as the *visual object types* that can be used in the scene;
- **audio profiles** do the same for audio;
- **graphics profiles** are defined as the BIFS nodes that specify the graphical elements that can be composed in the scene;
- **scene graph profiles** define the supported scene composition capabilities, such as 2 or 3 dimensionality, interaction capabilities, and support for, e.g., translation, rotation, scaling;
- finally, **object descriptor profiles** define the capabilities of the synchronization layer and object descriptor tools.

Again, like in MPEG-2, *levels* of complexity are also defined: specifications of the constraints (e.g. bit-rate, sampling rate, object memory size, etc.) that go with a profile.

In MPEG-4 Version 1, a number of profiles and levels have been specified, addressing several classes of applications. They range from simple ones targeted at low bit-rates to more advanced profiles which support high-quality and sophisticated interaction. MPEG does not prescribe or advise combinations of the various types of profiles, or provide any guidelines about which profile / levels combinations to use for which application. However, care has been taken that good matches exist between the different dimensions. Choices are left to implementers with their specific needs and constraints. MPEG is of course not unaware of applications: applications guide the profiling process, but care is taken that the usage of the resulting profiles is as broad as possible.

4. Application of MPEG-4 in broadcast services

The digitalization of TV broadcasts is not the end of a revolution, but much more the beginning of a long evolution that will transform the way we watch television. We do not believe the predictions that television, as a real-time event, will largely disappear and that all TV programmes will be downloaded at night to some local storage. Firstly, television as a service *will* change under the influence of digitalization because, once a receiving terminal becomes digital, many options become available. Secondly, set-top boxes and television sets are increasingly being connected to telephone networks, partly as a result of falling telephone charges. Thirdly, the Internet (more specifically the Web) is providing a fast-growing and rich source of interactive possibilities and (increasingly audio-visual) digital content. In this section, we will discuss the potential of MPEG-4 in such an environment.

During the 1998 International Broadcasting Convention (IBC) in Amsterdam, September 1998, a number of experts involved in digital audio-visual services launched the Advanced Interactive Content Initiative (**AIC-I**) [13]. AIC-I intends to bridge the television and Web paradigms, and develop a specification for highly sophisticated interactive 2-D and 3-D content, to be delivered at affordable cost to digital set-top boxes. Version 1.0 of the specification is available now, using available standard technologies: MPEG-4 (as a key technology in offering attractive and advanced consumer applications), VRML and a form of HTML [13].

Before starting its specification work, AIC-I listed in its requirements document, a number of example applications used for deriving the requirements, giving excellent examples of how MPEG-4 technology can enhance the broadcast services. While these are good examples, real life will certainly bring even more interesting ones. Some of the examples below are inspired by the AIC-I documents.

4.1. Programme enhancement

One of the most powerful forms of programme enhancement is the addition of collateral information in the form of text overlays, still images, spoken text (in several languages), 2-D and 3-D graphics or even audio and video. Already today, some programmes use enhancement information with the difference that this information is a mandatory part of the content. With MPEG-4, adding optional information is easy, as different programme elements can be sent as different objects, giving the user the choice to display them. Textual and graphical information can be sent more efficiently and with higher quality when represented as text and graphics, rather than as video pixels. Viewers may even be given the possibility of customiz-

Abbreviations

AAL	ATM adaptation layer	**IEC**	International Electrotechnical Commission
ATM	Asynchronous transfer mode		
DCT	Discrete cosine transform	**ISO**	International Organization for Standardization
DVB	Digital Video Broadcasting		
DVB-SI	DVB - Service Information	**MPEG**	(ISO/IEC) Moving Picture Experts Group
EPG	Electronic programme guide		
HTML	Hyper-text markup language	**QoS**	Quality of service

ing the layout; a user may, for example, choose whether or not to see the answers to a quiz show. Maps, virtual reality fly-throughs, visualization models, etc. may be added to historical, scientific or arts programmes.

The additional information could either be sent together with the broadcast, or retrieved using a different link; but even then it could still be integrated and synchronized with the broadcast itself. Note that overlays, buttons, still pictures, etc. are regular MPEG-4 scene elements, just like audio and video. Also, hyperlinking capabilities can be provided through all objects.

More specific examples of possibilities that are useful for enhancing the programmes and which are offered by MPEG-4 are:

Interactive viewpoint selection

A user can select from multiple camera views of an event, e.g., a sports game or a music show, by "clicking" on a (MPEG-4) button. In multi-channel audio systems, the audio source will change correspondingly. There may also be multiple alternate audio feeds to choose from, e.g., the sports reporter or the crowd. The button could be an iconic video.

Additional information about elements of the scene

Text overlays or still images may provide synchronized background information about scene elements. These "elements" could be actors, players in a sports event, or monuments in a city guided tour. A player's detailed statistics or the history of a building may be displayed by simple clicking or pointing.

Concurrent programme information

This is information that complements the programme. On-demand text overlays could provide a synopsis of the script or could re-cap the story up to the current point in the programme, allowing those users tuning in late to catch up. Graphical overlays could complement the visual information, e.g. by superimposing over the image of an orchestra, different colours according to the various types of instruments. Closed captions, sub-titles and opera "super-titles" could be displayed (in several optional languages), and multiple language audio tracks could be selected. Optional audio objects could give whispered hints in interactive games.

Adding information for special groups

MPEG-4 facial animation could be used to allow lip reading at mere hundreds of bits per second. Body animation could be used to transmit signing information, equally efficiently; the commands are fine-grain enough to animate the hands and arms. At such low cost, even multiple signing languages could be sent.

Of course, these examples can be combined. *Fig. 5.* shows one such example where text and graphics overlays, combined with multi-channel selection and hyperlinking, provide a very innovative version of a sports programme.

Figure 5
Enhanced version of a bicycle race programme *(courtesy TDF/ENST 98)*.

4.2. *Interactive entertainment*

Interaction is without doubt one of the most promising functionalities supported by MPEG-4. Object-based coding, BIFS-defined behaviour and hyperlinking provide new ways to let users interact with other people and with content itself. Examples include broadcast video games with 3-D content, interactive quiz shows that let users compete, multi-view movies, as well as documentaries with educational purposes.

Some applications benefit from an on-line connection, but MPEG-4's BIFS also gives the option to unidirectionally send objects *and* the associated behaviour, thus letting the content respond to user input. In fact, MPEG is working to extend MPEG-4 with Java interfaces to the scene and input devices, among other elements. For example, a user may participate in a quiz that is related to a sports show (giving feedback) or may navigate in a broadcast 3-D virtual environment.

Of course, multiple viewpoints – physical (e.g. different positions) as well as in terms of script (e.g. the point of view of the killer, or of the detective) – along with multiple scenarios and collateral information are features that surely will enrich interactive entertainment. A more incremental approach might consider embedding interactive games in a conventional programme or commercial. A commercial mini-quiz might reward players with discount coupons.

4.3. *Interactive commercials and shopping*

Commercials form an important element of broadcast content, and they can benefit greatly from MPEG-4. New commercials, when demanded by the user, may provide more detailed or additional information about products. Products could be viewed interactively, perhaps even taken apart, and purchased using an on-line connection. Commercials may be targeted and personalised to the user's needs, taste and profile.

Commercials, like other advanced content, can be authored in such a way that basic receivers would display the basic audio-visual information, while advanced receivers would make full use of the interactive possibilities. The advanced version may include a 3-D model of a product, e.g. a car, which is initially downloaded and then used as a component of a more sophisticated commercial. To return to the car example given in *Section 1.1*, the user could interact with the model, changing the colour and viewpoint during a driving demonstration. The model might also be stored for later interaction, or sent through e-mail to a friend with a recommendation. Commercials could contain links to on-line stores, with the opportunity to talk to a sales consultant – in the future, maybe even a virtual consultant.

Finally, a retailer may want to broadcast an interactive version of the products catalogue, to be saved in local storage. Consumers would browse the catalogue, inspect the products by manipulating videos and 3-D models, and then buy the chosen product(s) on-line. The content of commercials could evolve to become increasingly sophisticated, e.g. by enabling the downloading of furniture models which the users could place in a 3-D model of their own living room and then view with different fabrics and colours, under different types of light.

4.4. Advanced electronic programme guides

Since digitalization is bringing an explosion in the amount of available content, it is essential to help users to find the programme of their choice. MPEG-4 technology may help to create user-friendly and powerful, interactive EPGs.

It is already common practice to have EPGs in digital television services. MPEG-4 could be used to create very attractive ones, including small video objects and audio clips for preview. In addition, the programme guide could contain links to relevant information outside the specific channel multiplex, e.g. on the Web. Of course the EPG could also be made available as a web page, with hyperlinks to television programmes.

Both the layout and selection of the content for the programme guide could be user-customised. Advanced features may allow automatic selection of services according to user preferences, using the description provided in the MPEG-4 Object Content Information field or the DVB-SI. In addition, a kind of “interactive magazine” could provide weekly or monthly pay-per-view programmes, and programme listings, across all broadcast services. EPGs could be built using 3-D graphics including the interactive part of a 3-D world.

(Note that the description of audio-visual content is the objective of the next MPEG standard, known as MPEG-7. Advanced electronic programme guides are among MPEG-7’s target applications.)

5. Concluding remarks

The MPEG-4 object-based, interactive, audio-visual representation model opens new frontiers for broadcasting, supporting important features from the Web and computer-games paradigms. MPEG-4 will not replace MPEG-2, but it can provide new and attractive possibilities in existing environments. What the future will bring is as uncertain as ever, and much depends on the creativity of those who make the content.

Fernando Pereira *was born in Vermelha, Portugal, in October 1962. He graduated in Electrical and Computer Engineering from the* Instituto Superior Técnico (IST), Universidade Técnica de Lisboa *in 1985. He received an M.Sc. and a Ph.D. in Electrical and Computer Engineering from IST in 1988 and 1991, respectively. He is currently Professor of the Electrical and Computer Engineering Department of IST, responsible for the participation of IST in many national and international research projects.*

Dr Pereira is a member of the Editorial Board of the Signal Processing: Image Communication Journal, and an Associate Editor of IEEE Transactions of Circuits and Systems for Video Technology. He is a member of the Scientific committees of several international conferences. He has contributed more than eighty papers. He has participated in the work of ISO/MPEG for many years, notably as the head of the Portuguese delegation, and has chaired many Ad Hoc Groups relating to the MPEG-4 and MPEG-7 standards. His current areas of interest are video analyzis, processing, description and representation, and multimedia interactive services.

Rob Koenen *received his "Ingenieur" (MSEE) degree from Delft University of Technology, the Netherlands, in 1989. He studied Electrical Engineering, specializing in information theory. He joined KPN Research in 1990, where he has researched various aspects of audiovisual communication, working as a project manager and later as co-ordinator of the Video Group. His projects have addressed, or still address: image-coding research, audiovisual communication for people with special needs, interactive broadband multimedia for residential users, mobile multimedia, the strategic deployment of new multimedia services, and audiovisual quality assessment. As an MPEG delegate, he has played a key role in the development of the MPEG-4 standard since 1993, and in defining the upcoming MPEG-7 standard since 1995.*

Ir. Koenen now works as a senior adviser/program manager with the Multimedia Technology group of KPN Research. He is the chairman of the MPEG Requirements subgroup and heads the Dutch MPEG delegation. He is also an associate editor of the IEEE Transactions on Circuits and Systems for Video Technology.

MPEG-4 technology can keep television an interesting medium for the interactive generation. Rather than competing with the Internet for viewing time, television services should make use of the new possibilities of on-line media, and MPEG-4 is one of the elements that can make the cross-fertilization happen.

Acknowledgements

The authors would like to thank all MPEG members for the interesting and fruitful discussions in meetings and by e-mail, which has substantially enriched their technical knowledge in the past years. Fernando Pereira acknowledges the support of the Portuguese framework PRAXIS XXI through his participation in MPEG meetings, under the project "*Normalização de métodos avançados de representação de vídeo*".

Bibliography

[1] Replay Network Service
http://www.replaytv.com

[2] WebTV
http://www.webtv.com

[3] http://www.tivo.com

[4] MPEG Home Page
http://www.cselt.it/mpeg

[5] Doc. ISO/MPEG N2725: **MPEG-4 Overview**
Seoul MPEG Meeting, March 1999.

[6] Doc. ISO/MPEG N2501: **Text of ISO/IEC FDIS 14496-1: Systems**
Atlantic City MPEG Meeting, October 1998.

[7] Doc. ISO/MPEG N2502 **Text of ISO/IEC FDIS 14496-2: Visual**
Atlantic City MPEG Meeting, October 1998.

[8] Doc. ISO/MPEG N2503: **Text of ISO/IEC FDIS 14496-3: Audio**
Atlantic City MPEG Meeting, October 1998.

[9] Doc. ISO/MPEG N2550: **Text for CD 14496-4: Conformance testing**
Rome MPEG Meeting, December 1998.

[10] Doc. ISO/MPEG N2205: **Final text for FCD 14496-5: Reference software**
Tokyo MPEG Meeting, March 1998.

[11] Doc. ISO/MPEG N2506: **Text of ISO/IEC FDIS 14496-6: DMIF**
Atlantic City MPEG Meeting, October 1998.

[12] "Virtual Reality Modeling Language"
http://www.web3D.org

[13] Advanced Interactive Content Initiative
http://toocan.philabs.research. philips.com/misc/aici/