Context, Objectives, and Process[1]

by Fernando Pereira

Keywords: audiovisual object-based representation, audiovisual objects, MPEG-4, standardization process

> "What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is." [Marr82]

The first video coding standards and their underlying representation models mainly address the vision process by providing video representation in the form of a sequence of rectangular 2D frames, giving users a window to the real world: the television paradigm. However, the process of vision is often just the initial part of the task at hand, because typically humans need and want to see, to take actions after, to interact with the objects identified. A similar reasoning can be made regarding the process of hearing and the corresponding audio representation models [Pere99].

1. Some sections of this chapter are adapted from *Signal Processing: Image Communication*, 15 (4–5) (2000), Fernando Pereira, pp. 271–279, "MPEG-4: Why, What, How and When?", Copyright (2000), with permission from Elsevier Science.

Although the television paradigm dominated audiovisual communications for many years, the situation has been evolving quickly in terms of the ways audiovisual content is produced, delivered, and consumed [KoPC97]. Moreover, hardware and software are becoming increasingly powerful, with microelectronic technology providing new programmable processors, opening new frontiers to the representation technologies used and to the functionalities provided.

Producing content today is easier than ever before. Digital still cameras directly storing in JPEG format have hit the mass market. Together with the first digital video cameras directly recording in MPEG-1 format, this represents a major step for consumer acceptance of digital audiovisual acquisition and compression technology. This step transforms every individual into a potential content producer, capable of creating content that can be easily distributed and published on the Internet. In addition, more content is being synthetically produced—that is, computer generated—and integrated with natural material in truly hybrid audiovisual content. The various pieces of content, digitally encoded, can be successively reused without the quality losses typical of the previous analog processes.

Whereas audiovisual information, notably the visual portion, until recently was carried only over very few networks, the trend is now toward the generalization of visual information in every single network. Moreover, the increasing mobility in telecommunications is a major trend. Mobile connections will not be limited to voice; other types of data, including real-time media, are already emerging. Because mobile telephones are replaced every two to three years, new mobile devices can finally make the decade-long promise of audiovisual communications a reality. Notice that the need for visual communication is much stronger when people are not at home, and so have something to show besides the usual living room—for example, the nice beach where they are vacationing. This reinforces the relevance of audiovisual mobile communications.

The explosion of the World Wide Web and the acceptance of its interactive mode of operation have clearly shown that the traditional television paradigm will no longer suffice for audiovisual services. Users want to have access to audio and video as they now have access to text and graphics. This requires moving pictures and audio of acceptable quality at low bit rates on the Web, and Web-type interactivity with live content. It will be possible to activate relationships between users (in a potentially virtual world) through hyperlinking—the Web paradigm—and to experience interactive immersion in natural and virtual environments—the games paradigm.

As many of the emerging audiovisual applications demanded interworking, the need to develop an open and timely international standard addressing the needs mentioned above became evident. In 1993, MPEG (Moving Picture Experts Group) [MPEG] launched the MPEG-4 project, later formally called *Coding of Audio-Visual Objects*, to address (among other things) the requirements associated with the new applications resulting from these trends [N1177, N4319, N4505].

The need for any standard comes from an essential requirement relevant for all applications involving communication between two or more parts: *interoperability.* Interoperability is thus the requirement expressing the user's dream of exchanging any type of information without any technical barriers, in the simplest way. Without a standard way to perform some of the operations involved in the communication process and to structure the data exchanged, easy interoperability between terminals would be impossible. Having said that, it is clear that a standard shall specify the minimum number of tools needed to guarantee interoperability (because it is important that as many as possible non-normative technical zones exist), to allow the incorporation of technical advances, and thus to increase the lifetime of the standard, as well as to stimulate industrial technical competition. The existence of a standard also has important economic implications, because it allows the sharing of costs and investments and the acceleration of applications deployment.

MPEG has been responsible for the successful MPEG-1 (ISO/IEC 11172) and MPEG-2 (ISO/IEC 13818) standards, which have given rise to widely adopted commercial products and services, such as Video-CD, DVD, digital television, digital audio broadcasting (DAB), and MP3 (MPEG-1 Audio layer 3) players and recorders. The MPEG-4 standard (ISO/IEC 14496) is aimed at defining an audiovisual coding standard to address the emerging needs of the communication, interactive, and broadcasting service models as well as the needs of the mixed service models resulting from their convergence. The apparent convergence of the three traditionally separate application areas—communications, computing, and TV/film/entertainment—was evident in their cross-fertilization, with functionalities characteristic of each area increasingly emerging in the others (e.g., personal communications including video information or entertainment including interactive capabilities).

Following the previous successes—in fact, as a natural consequence of the vision underpinning MPEG-4—MPEG initiated in 1996 another standardization project addressing the problem of describing audiovisual content to allow the quick and efficient searching, processing, and filtering of various types of multimedia material: MPEG-7 (ISO/IEC 15938), officially called *Multimedia Content Description Interface* [N4509]. In fact, digital audiovisual information is more and more accessible to everyone, not only in terms of consumption but also in terms of production. But if it is much easier today to acquire, process, and distribute audiovisual content, it must be equally easy to access the available information, because huge amounts of audiovisual information are being generated all over the world every day. The need for a powerful way to quickly and efficiently identify, search, and filter various types of audiovisual content, by humans or machines (also using non-text-based technologies), directly follows from the urge to efficiently use the available audiovisual content and the difficulty of doing so. MPEG-7 will specify a standard way of describing various types of audiovisual information, including still pictures, video, speech, audio, graphics, 3D models, and synthetic audio, regardless of their representation format (e.g., analog or digital) and storage support (e.g., paper, film, or tape). In comparison with other available or emerging solutions for audiovisual content description, MPEG-7 can be mainly distinguished by (a) being general purpose, meaning its ability to describe content from many application environments; (b) its object-based representation model, meaning the capability of independently describing individual objects within a scene (be it MPEG-4 or any other format); (c) the integration of low-level and high-level features/descriptors into a single description framework, allowing it to combine the power of both types of descriptors; and (d) its extensibility, provided by the Description Definition Language, which allows MPEG-7 to keep growing, to be extended to new application areas, to answer newly emerging needs, and to integrate novel description tools [PeKo99]. The MPEG-7 standard was finalized in the summer of 2001.

Following the development of standards addressing more focused targets, MPEG acknowledged the lack of a big picture that described how the various elements building the infrastructure for the deployment of applications using multimedia content relate to each other, or even if there are missing standard specifications for some of these elements [N4333]. To address this problem, MPEG started the MPEG-21 project (first ISO/IEC 18034, now ISO/IEC 21000), formally called *Multimedia Framework*, with the aim of understanding if and how these various elements fit together, and to discuss which new standards might be required if gaps in the infrastructure exist. Once this work has been carried out, new standards will be developed for the missing elements with the involvement of other bodies, where appropriate; finally, the existing and novel standards will be integrated in the MPEG-21 multimedia framework. The MPEG-21 vision is thus to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities [N4333]. The MPEG-21 multimedia framework will identify and define the key elements needed to support the multimedia value and delivery chain, as well as the relationships between and the operations supported by them [N4511].

After briefly covering the context that motivated the birth of the MPEG-4 project, this chapter presents its major objectives in terms of functionalities, requirements, tools, and applications, as well as its organization and the sequence followed to achieve the goals defined. This chapter also addresses the MPEG modus operandi, notably its mission, principles, and specific approach to the development of standards. Finally, the objectives and working approach of the MPEG-4 Industry Forum will be presented.

1.1 MPEG-4 Objectives

Although MPEG discussions about projects beyond MPEG-2 began as early as May 1991, at the Paris MPEG meeting, it was not until September 1993 that the MPEG Applications and Operational Environments (AOE) subgroup was set up and met for the first time. The main task of this subgroup was to identify the applications and requirements relevant to the far-term, very low bit-rate coding solution to be developed by International Organization for Standardization (ISO)/MPEG as stated in the initial MPEG-4 project description [N271]. At the same time, the near-term hybrid coding solution being developed within the International Telecommunications Union-Telecommunications Standardization Sector (ITU-T) Low Bit-rate Coding (LBC) group started producing the first results (later, the ITU-T H.263 standard [H263]). It was then generally felt that those results were close to the best performance that could be obtained by block-based, hybrid, DCT/motion-compensation video coding schemes.

In July 1994, the Grimstadt MPEG meeting marked a major change in the direction of MPEG-4. Until that meeting, the main goal of MPEG-4 was to obtain a significantly better compression ratio than could be achieved by conventional coding techniques. Few people, however, believed it was possible, in the next five years, to make enough improvements over the LBC standard (H.263 and H.263+) to justify a new standard.[2] So the AOE subgroup was faced with the need to broaden the objectives of MPEG-4, believing that pure compression gains would not be enough to start a new MPEG standardization project. The subgroup then began an in-depth analysis of the audiovisual world trends, based on the convergence of the TV/film/entertainment, computing, and telecommunications worlds. The conclusion was that the emerging MPEG-4 coding standard should support new ways (notably content-based) of communicating, accessing, and manipulating digital audiovisual data.

1.1.1 Functionalities

Following this change of direction and the analysis made, the vision driving the MPEG-4 standard was explained through the eight new or improved functionalities described in the MPEG-4 Proposal Package Description (PPD), prepared by the time of the first MPEG-4 call for proposals in July 1995 [N998]. These eight functionalities came from an assessment of the functionalities that would be useful in future audiovisual applications, but which were not supported (or at least not well supported) by the available coding standards. The eight new or improved functionalities were clustered in three classes related to the three worlds—TV/film/entertainment, computing, and telecommunications—the convergence of which MPEG-4 wanted to address [N998, PeKo96]:

1. Content-based interactivity

- ☞ **Content-based multimedia data access tools:** MPEG-4 shall provide efficient data access and organization based on the audiovisual content. Access tools may be indexing, hyperlinking, querying, browsing, uploading, downloading, and deleting. Sample uses include content-based

2. At least under the acceptable complexity boundaries.

retrieval of information from online libraries and travel information databases.[3]

- ☞ **Content-based manipulation and bitstream editing:** MPEG-4 shall provide syntax and coding schemes to support content-based manipulation and bitstream editing without the need for transcoding. This means the user should be able to select one specific object in the scene/bitstream and change some of its characteristics. Sample uses include home movie production and editing, interactive home shopping, and the insertion of a sign language interpreter or subtitles.
- ☞ **Hybrid natural and synthetic data coding:** MPEG-4 shall support efficient methods for combining synthetic scenes with natural scenes (e.g., text and graphics overlays), the ability to code and manipulate natural and synthetic audio and visual data, and decoder-controllable methods of mixing synthetic data with ordinary video and audio (allowing for interactivity). For example, in virtual reality applications, animations and synthetic audio (e.g., MIDI) can be mixed with ordinary audio and video in games, and graphics can be rendered from different viewpoints.
- ☞ **Improved temporal random access:** MPEG-4 shall provide efficient methods to randomly access, within a limited time and with fine resolution, parts (e.g., frames or objects) from an audiovisual sequence. Example usage: audiovisual data can be randomly accessed from a remote terminal over limited-capacity media, a fast-forward can be performed on a single audiovisual object in the sequence.

2. Compression efficiency

- ☞ **Improved coding efficiency:** The growth of mobile networks creates an ongoing demand for improved coding efficiency; therefore, MPEG-4 set as its target providing subjectively better audiovisual quality compared to existing or other emerging standards (such as H.263), at comparable bit rates. Sample uses include efficient transmission of audiovisual data on low-bandwidth channels and efficient storage of audiovisual data on limited-capacity media, such as chip cards.
- ☞ **Coding of multiple concurrent data streams:** MPEG-4 shall provide the ability to efficiently code multiple views/soundtracks of a scene as well as sufficient synchronization between the resulting elementary streams.[4] For stereoscopic and multiview video applications, MPEG-4 shall include the ability to exploit redundancy in multiple views of the

3. Although this is one of the eight initial MPEG-4 functionalities, it relates more to the tools developed later in the context of the MPEG-7 standard [N4509].
4. This seems to be the only initial MPEG-4 functionality that has not been addressed at the end of 2001, as the companies participating in MPEG have shown no significant interest in it.

same scene, also permitting solutions that allow compatibility with normal video. Sample uses include multimedia entertainment (e.g., virtual reality games and 3D movies), training and flight simulations, multimedia presentations, and education.

3. Universal access

- ☞ **Robustness in error-prone environments:** Because universal accessibility implies access to applications over many wireless and wired networks and storage media, MPEG-4 shall provide an error robustness capability. Particularly, sufficient error robustness shall be provided for low bit-rate applications under severe error conditions. Sample uses include transmission from a database over a wireless network, communicating with a mobile terminal, and gathering audiovisual data from a remote location.
- ☞ **Content-based scalability:** MPEG-4 shall provide the ability to achieve scalability with a fine granularity in content, spatial resolution, temporal resolution, quality, and complexity. Content-scalability may imply the existence of a prioritization of the objects in the scene. Sample uses include user selection of decoded quality of individual objects in the scene and database browsing at different scales, resolutions, and qualities.

These functionalities were essential to shaping the MPEG-4 vision, balancing completely new functionalities with more traditional ones, and thus allowing a bridge from the past to the future not only in terms of functionalities but also in terms of tools and experts.

1.1.2 Requirements

Following the identification of the fundamental MPEG-4 functionalities, MPEG started a requirements development process, which has been continuously evolving since then. The requirements serve to drive the tools development process and assure that the right technology is being specified: Tools will be developed to fulfill the identified requirements, and no tools that do not address any requirement will be defined.

By the middle of 2001, the MPEG-4 requirements were structured as shown in Table 1.1. The requirements are organized in terms of major technical areas, which do not directly correspond either to Parts of the MPEG-4 standard or to MPEG working subgroups. Table 1.1 gives only a general flavor of the requirements (for the details, see [N4319]).

Within each category, the requirements are to be fulfilled by the set of tools selected for the standard, and not all requirements must be addressed by each individual tool. It is the right combination of tools that allows building the algorithms that can address the specific needs of a certain class of applications.

Table 1.1 MPEG-4 requirements [N4319]

Requirements for systems	
Flexibility	Multipoint operation
Multiplexing of audio, visual, and other information	Object content information (OCI)
Composition of audio and visual objects	Video-related metadata
Application texture	Delay
Downloading	Configuration modes
User interaction	Priority of audiovisual objects
Media interworking	Dynamic resource management
Compatibility	Reference to associated MPEG-7 data
Robustness to information errors and losses	File format
Object-based bitstream manipulation and editing	Textual format
Requirements for natural video objects	
Object-based representation	Object-based coding flexibility
Video content	Object-based scalability
Object-based bitstream manipulation and editing	Delay modes
Object-based random access	Formats
Object quality and fidelity	Bit-rate modes
Coding of multiple concurrent data streams	Complexity modes
Robustness to information errors and losses	Still images
	Tandem coding*
Requirements for synthetic video objects	
Types of synthetic video objects	Text overlay
2D/3D mesh compression	Image and graphics overlay
Definition and animation parameter compression	View-dependent texture scalability
Texture mapping	Geometrical transformations
	Video object tracking
Requirements for natural audio objects	
Object-based representation	Robustness to information errors and losses
Audio content	Delay modes
Object-based bitstream editing and manipulation	Complexity modes
Object-based scalability	Bit-rate modes
Object-based random access and user controls	Downmix†
Time scale change	Transcoding
Pitch change	Tandem coding
	Audio formats
	Improved coding efficiency

Table 1.1 MPEG-4 requirements [N4319] (Continued)

Requirements for synthetic audio objects	
Low bit-rate speech	Text to speech
Synthetic speech data	Sound synthesis
Requirements for delivery multimedia integration format (DMIF)	
Connectivity	End-to-end Quality of Service (QoS) management
Transparency	Network-based stream processing and management
Application service enablement	
Requirements for MPEG-J	
Functional requirements	Byte code execution
Byte code delivery	Event mechanism
Authentication	
Requirements for multiuser environments	
Scene graph representation	IPMP and sharing tools
Audiovisual objects and avatars representation	Application programming interfaces
Delivery and stream management	
Requirements for animation framework	
Enhanced texture mapping	Reusability of scene graph nodes and animation streams
Animation support	Persistence
High-level shape representation	Compression of animated objects
Requirements for intellectual property management and protection (IPMP)	
Identification of intellectual property	Intellectual property management and protection interfaces
Intellectual property management and protection hooks	

*Relates to the repeated encoding and decoding (cascading) of audio or video material, either with the same or different codecs.

†Relates to the ability to reduce the number of channels to a configuration with a lower number of channels for presentation purposes (e.g., listening to multichannel audio using stereophonic reproduction).

Although efficient compression was not the only first-priority requirement in MPEG-4 (as it had been for MPEG-1 and MPEG-2), it is clear that it has been a central requirement in MPEG-4 in the sense that, whatever the type of data that had to be represented by binary encoding (e.g., shape for video objects, facial animation parameters for 3D facial models, or even scene composition data), the target was always to reach the smallest number of bits for a certain level of quality. Although some people claim that efficient compression of data is not a must today because of the growing availability of bandwidth, MPEG always acknowledged that bandwidth resources (either for transmission or for storage) are still limited, and thus efficient compression is required. This is even truer for some relevant recent transmission cases, such as the Internet and mobile networks, where bandwidth limitations and efficient compression are major issues. However, in the context of MPEG-4, efficient compression must be balanced against other major requirements, such as those related to interactivity capabilities (which have a price in terms of compression efficiency compared with noninteractive representation schemes), if new functionalities not supported by frame-based coding schemes are to be provided.

As noted, the MPEG-4 requirements' development process has been evolving since the beginning of the standardization process; this evolution has been targeting the inclusion of additional requirements related to functionalities, which fit and complement well the MPEG-4 vision. Because the MPEG-4 standardization process is not yet finished, it cannot be said that all the requirements have been addressed by means of an MPEG-4 tool. However, it is possible to say that all of the not-yet-addressed requirements are either being worked on or should be removed in a short time if the industries do not show sufficient support to move to the technical development phase.

1.1.3 Tools

The major trends mentioned—notably the mounting presence of audiovisual media on all networks, increasing mobility, and growing interactivity—have driven, and continue to drive, the development of the MPEG-4 standard.

To address the identified functionalities and requirements [N4319], a set of tools was developed to perform the following functions [Pere99]:

- ☞ Efficiently represent a number of data types through media codecs
  - ✘ Video from very low bit rates to very high-quality conditions
  - ✘ Music and speech data for a very wide bit-rate range, from transparent music to very low bit-rate speech
  - ✘ Generic dynamic 3D objects as well as specific objects such as human faces and bodies
  - ✘ Synthetic speech and music, including support for 3D audio spaces
  - ✘ Text and graphics

- Provide fine granularity scalability in the quality, temporal, and spatial dimensions
- Provide, in the encoding layer, resilience to residual errors for the various data types, especially under difficult channel conditions such as mobile ones
- Independently represent the various objects in the scene, allowing independent access for their manipulation and reuse
- Compose audio and visual (natural and synthetic) objects into one audiovisual scene in a synchronized way
- Describe the objects and the events in the scene
- Provide interaction and hyperlinking capabilities
- Manage and protect intellectual property on audiovisual content and algorithms, so that only authorized users have access to the content
- Provide a delivery-media-independent representation format, to transparently cross the borders of different delivery environments

The major difference with previous audiovisual coding standards, at the basis of the new functionalities, is the object-based audiovisual representation model that underpins MPEG-4 (see Figure 1.1). An object-based scene is built using individual independent objects that have relationships in space and time. This representation approach offers a number of advantages: First, different object types may have different suitable coded representations—for example, a synthetic moving head is best represented using animation parameters, whereas video benefits from a smart representation of pixel values. Second, it allows the harmonious integration of different types of data into one scene, notably with natural and synthetic origins—for example, an animated cartoon character in a real world environment, or a real person in a virtual studio set. Third, interacting with the objects and hyperlinking from them is

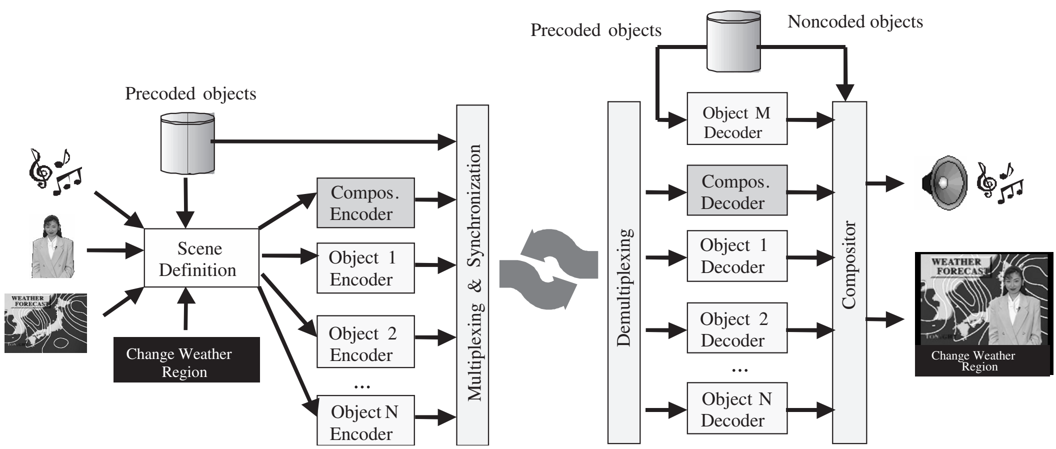

Fig. 1.1 MPEG-4 object-based representation architecture.

now feasible. There are many more advantages—such as the selective spending of bits, the easy reuse of content without transcoding, and the provision of sophisticated coding solutions for scalable content on the Internet—all of them resulting from the adoption of the object-based representation model.

The applications that benefit from the new concepts and functionalities are found in many (and very different) environments [N2724]. Therefore, MPEG-4 is constructed as a toolbox rather than a monolithic standard, using profiles that provide solutions for these different application settings (see Chapter 13). This means that, although MPEG-4 is a standard comprising a vast array of technologies, it is structured so that solutions are available at the measure of the needs. It is the task of each implementer to extract from the MPEG-4 standard the technical solutions adequate to his or her needs (likely a small subset of the standardized tools) by choosing the adequate profiling combination.

Because a standard is always a constraint on freedom, it is important to make it as minimally constraining as possible [Chia97]. To MPEG, this means a standard must offer maximum advantages by specifying the minimum necessary, allowing for competition and evolution of the technology in the *non-normative areas;* these non-normative areas correspond to the tools for which normative specification is not essential for interoperability. On the contrary, normative tools are those defined by the standard whose specification is essential for interoperability. For example, whereas video segmentation and rate control are non-normative tools, the decoding process needs to be normative. The strategy of specifying the minimum for maximum usability ensures that good use can be made of the continuous technical improvements in the relevant technical areas. The consequence is that better non-normative tools can always be used, even after the standard is finalized, and it is possible to rely on competition for obtaining ever-better results. In fact, it is through the non-normative tools that products will distinguish themselves, which only reinforces the importance of this type of tools.

1.1.4 Applications

MPEG-4 wants to address a wide range of applications, many of them completely new, as there are very new functionalities to take benefit from, and many others improved regarding those already available [N2724]. Unlike MPEG-2 where the *killer application* was digital television, in a first approach just understood as the digital translation of the rather old analog version, MPEG-4 does not target a major and exclusive killer application but opens many new frontiers. Playing with audiovisual scenes and creating, reusing, accessing, and consuming audiovisual content will become easier and more powerful. New and richer applications can be developed, for example, in enhanced broadcasting, remote surveillance, personal communications, games, mobile multimedia, and virtual environments. MPEG-4 allows services combining the traditionally different service models: broadcast, (online) interaction, and communication. As such, MPEG-4 addresses *convergence,* understood as the proliferation of audiovisual information in all kinds of services and on all types of (access) networks.

The MPEG-4 Applications document [N2724] describes application examples benefiting from the MPEG-4 technology that will serve as inspiration to the industry to create many more exciting applications. In this sense, MPEG-4 is a technical playground where many application constructions may be built by the manufacturers and service providers. The MPEG-4 Applications document suggests the following applications, using both audio and visual information or just one of them: broadcast, collaborative scene visualization, content-based storage and retrieval, digital amplitude modulation (AM) broadcasting, digital television set-top box, DVD, infotainment, mobile multimedia, real-time communications, streaming video on the Internet/Intranet, studio and television postproduction, surveillance, and virtual meetings.

1.2 Formal Standardization Process

MPEG (Moving Picture Experts Group) formally called *Coding of Moving Picture and Audio* is Working Group 11 (WG11) of Subcommittee 29 (Coding of Audio, Picture, Multimedia and Hypermedia Information) of the ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) Joint Technical Committee 1 (JTC1) [MPEG]. JTC1's scope is standardization in the field of information technology: According to JTC1 [JTC1], information technology includes the specification, design, and development of systems and tools dealing with the capture, representation, processing, security, transfer, interchange, presentation, management, organization, storage, and retrieval of information.

Since the major objective of MPEG is to produce technical specifications, it is important to have an idea, even if brief, about the types of documents that can be produced by a group like MPEG. Through JTC1, MPEG can produce the types of documents presented in the following:[5]

- ☞ **International standards:** These are documents with the technical specification of the standard. After the first edition of a standard, new editions may be published, notably to incorporate in one document the previous edition and all the amendments and corrigenda meanwhile issued. Whenever a new edition for a certain Part of the standard includes new technology (and thus is not just a compilation of previously approved technology), it must be approved following a process similar to amendments.

5. For each type of document, the various successive stages, as well as the minimum time the voting period by national standardization bodies may take for each stage (this means the minimum time to take the promotion of a certain document from one stage to the next), are indicated.

The evolution process for an International Standard is as follows:

- ✘ New work item Proposal (NP): 3 months ballot (with comments)
- ✘ Working Draft (WD): no ballot
- ✘ Committee Draft (CD): 3 months ballot (with comments)
- ✘ Final Committee Draft (FCD): 4 months ballot (with comments)
- ✘ Final Draft International Standard (FDIS): 2 months binary (only yes/no) ballot (failing this ballot [no vote] implies going back to WD stage)
- ✘ International Standard (IS)

FDIS and IS documents are copyrighted by ISO.

☞ **Amendments:** Documents with technical additions or technical changes (but not corrections) to an International Standard; they are edited as delta documents to the standard they amend. Each amendment lists the status of all amendments and technical corrigenda to the current edition of the standard. Amendments are published as separate documents, which means the edition of the IS affected remains in print. When additions to a standard must be produced, it is decided if it is better to publish an amendment or a new edition of the IS that incorporates the additions. When a new edition is published, it includes a full revision of the previous edition, incorporating all the changes corresponding to the amendments and corrigenda meanwhile issued.

The evolution process for an Amendment is as follows:

- ✘ New work item Proposal (NP): 3 months ballot (with comments)
- ✘ Working Draft (WD): no ballot
- ✘ Proposed Draft Amendment (PDAM): 3 months ballot (with comments)
- ✘ Final Proposed Draft Amendment (FPDAM): 4 months ballot (with comments)
- ✘ Final Draft Amendment (FDAM): 2 months binary (only yes/no) ballot (failing this ballot [no vote] implies going back to WD stage)
- ✘ Amendment (AMD)

FDAM and AMD documents are copyrighted by ISO.

☞ **Corrigenda:** These are documents issued to correct technical defects in an IS (or an Amendment).[6] Technical corrigenda usually are not issued for the correction of a few editorial defects. In such cases, corrections can be incorporated in future technical corrigenda. Technical corrigenda are not issued for technical additions, which follow the Amendment procedure.

The evolution process for a corrigendum is as follows:

6. Notice that, formally speaking, an amendment is an IS, as it specifies new technology (and never corrections to previous technology, which is the task of corrigenda).

- ✘ Defect Report (DR): no ballot
- ✘ Draft Technical Corrigendum (DCOR): 3 months ballot (with comments)
- ✘ Technical Corrigendum (COR)

  COR documents are copyrighted by ISO.

☞ **Technical reports:** These documents contain information of a different kind from that normally published as an IS, such as a model/framework, technical requirements and planning information, a testing criteria methodology, information obtained from surveys carried out among national bodies, information on work in other international bodies, or information on the state of the art regarding national body standards on a particular subject [JTC1].

  The evolution process for a Technical Report is as follows:

- ✘ New Work Item Proposal (NP): 3 months ballot (with comments)
- ✘ Working Draft (WD): no ballot
- ✘ Proposed Draft Technical Report (PDTR): 3 months ballot (with comments)
- ✘ Draft Technical Report (DTR): 3 months ballot (with comments)
- ✘ Technical Report (TR)

  DTR and TR documents are copyrighted by ISO.

For all ballots involving comments, MPEG produces a *Disposition of Comments* (DoC) document in which all comments made by national bodies are answered. During the MPEG-4 standard development process, MPEG used all the types of documents listed. Section 1.5 presents an overview of the many documents published in the context of the standardization process for the MPEG-4 standard.

1.3 MPEG MODUS OPERANDI

MPEG is open to experts duly accredited by an appropriate National Standards Body. On average, a meeting is attended by more than 300 experts representing more than 200 companies, universities, and research centers spanning all industry domains with a stake in digital audio, video, and multimedia. On average, more than 20 countries are represented. The group meets three to five times a year. Participants in MPEG-4 include broadcasters, equipment and software manufacturers, digital content creators and managers, telecommunication service providers, and publishers and intellectual property rights managers, as well as university researchers.

MPEG held its first meeting in May 1988 in Ottawa (Canada) and reached its 58th meeting in December 2001 in Pattaya (Thailand). A Convener chairs the MPEG activities: Dr. Leonardo Chiariglione has been the MPEG Convener since the start of MPEG.

Because of its size, an MPEG plenary meeting can be a difficult experience for a newcomer if he or she has no guidance. MPEG meetings are well known for their openness, lively discussions, and detailed dissection of issues. During the meetings, MPEG is organized in several subgroups, each one with a chairman, notably these [N4500]:

- **Specification development subgroups:** Systems, Video, Audio, Multimedia Description Schemes (MDS), and Synthetic Natural Hybrid Coding (SNHC)
- **Auxiliary subgroups:** Requirements, Test, Implementation Studies, and Liaison

To better coordinate the work between the MPEG plenary meetings, MPEG may decide to establish ad hoc groups (AHG). AHGs are established with mandate, membership, chairman, duration, and meeting schedule at the end of an MPEG meeting, for the sole purpose of continuing work between MPEG meetings. They are established by MPEG and report to it; the task of an AHG may cover just preparation of recommendations to be submitted to MPEG, with no decisions made. The duration of an AHG usually is limited to the period between two successive MPEG meetings; AHGs always cease to exist at the start of an MPEG meeting. Participation in AHGs is not restricted to the delegates present at the meeting.

1.3.1 Mission

MPEG's area of work is the "development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio, and their combination, in order to satisfy a wide variety of applications." [N4500]. According to the MPEG *Terms of Reference* [N4500], the MPEG *Programme of Work* is as follows:

- Serve as responsible body within ISO/IEC for recommending a set of standards consistent with the area of work.
- Cooperate with other standardization bodies dealing with similar applications. For this, MPEG creates liaisons with other standardization bodies as well as with other relevant duly constituted organizations (e.g., industrial consortia), exchanging requirements and technical specifications with the aim of reaching the largest possible use of standards.
- Consider requirements for interworking with other applications, such as telecommunications and broadcasting, with other image coding algorithms defined by other SC29 Working Groups and with other picture and audio coding algorithms defined by other standardization bodies.
- Define methods for the subjective quality assessment of audio, moving pictures, and their combination for the purpose of the area of work.

- ☞ Assess characteristics of implementation technologies realizing coding algorithms of audio, moving pictures, and their combination.
- ☞ Assess characteristics of digital storage and other delivery media, targets of the standards developed by WG11.
- ☞ Develop standards for coding of moving pictures, audio, and their combination, taking into account the quality of coded media, effective implementation, and constraints from delivery media.
- ☞ Propose standards for the coded representation of moving picture information.
- ☞ Propose standards for the coded representation of audio information.
- ☞ Propose standards for the coded representation of information consisting of moving pictures and audio in combination.
- ☞ Propose standards for protocols associated with the coded representation of moving pictures, audio, and their combination.

1.3.2 Principles

Because the technological landscape changed from analog to digital, with all the associated implications, it was essential that standard makers acknowledged this change by modifying the way they create standards. Standards must offer interoperability across countries, services, and applications, and not just a system-driven approach by which the value of a standard is limited to a specific, vertically integrated system. This brings us to the toolbox approach by which a standard must provide a minimum set of relevant tools which (after they are assembled according to industry needs) provide the maximum interoperability at a minimum complexity—and very likely cost [Chia97]. The success of MPEG standards is based on this toolbox approach, bounded by the *one functionality, one tool* principle. In summary, MPEG wants to offer users interoperability and flexibility at the lowest complexity and cost.

To develop its standards, MPEG has been following a few principles. The most important of these principles, many of them very well known, are these [Chia97]:

- ☞ **Stick to the deadline:** Because of the importance of allowing industries to make serious planning and investments based on timely standards, MPEG rigorously follows the workplan set at the beginning of each standardization project. Never has an MPEG standard been delayed in reaching IS status compared to the planned dates.
- ☞ ***A priori* standardization:** To avoid becoming a standardization body endorsing industry-developed standards, MPEG identifies the maturity of technologies for standardization before industries make commitments. This approach allows the standard development process to be essentially technical and not biased by specific company interests.

- ☞ **Specify the minimum:** In order that a standard might be useful to several industries (and not especially tuned for any of them), will technically evolve (increasing its lifetime), and will keep space for industrial competition, it is essential that only the minimum set of tools essential for interoperability is specified.
- ☞ **Not systems but tools:** Because MPEG wants to specify standards that are useful for the various industries using the same technology, it is essential that tools rather than systems are specified, leaving the various industries the task of shaping those tools for their systems. To guide these industries, however, some major combinations of tools (profiling concept) are normatively specified so that functionality-driven and not application-driven combinations are adopted; this helps to increase interoperability and share costs.
- ☞ **One functionality, one tool:** To reach interoperability with an acceptable level of complexity justifying industry investments, it is essential that clear choices be made regarding the best tool to provide a certain capability. Past experience has shown that the *options* epidemic (often engaged to satisfy companies' wishes) prevents interoperability, many times preventing the standards from flying and thus destroying everybody's work.
- ☞ **Relocation of tools/algorithms:** So that the tools/algorithms are as useful as possible, not only must they be generic in the sense that they are not shaped to any specific application environment, it must also be possible to locate them at different positions in the final system, as different industries will likely position the same tool/algorithm differently (e.g., depending on the business model they follow).
- ☞ **Verification of the standard:** To check that a standard delivers what it should deliver, it must be evaluated and checked against the identified requirements in the same way a product is checked against the product specification. This verification process typically involves the performance of subjective or task-based tests.

These principles have guided the development of the MPEG-1, MPEG-2, MPEG-4, and MPEG-7 standards with the same success, and they will be applied to the MPEG-21 standard development process.

1.3.3 Standards Development Approach

In MPEG, any standardization project is the result of an exploration attitude by which MPEG proactively looks for the relevant problems to be addressed—for example, emerging applications, functionalities, or even technologies. This proactive exploration attitude allows the standards to be produced in a timely way (and not too late) when applications and technology are fully mature as, by that time, proprietary solutions will be conquering the markets. For this purpose, MPEG holds open seminars to discuss relevant topics with people from industry and academia outside MPEG and creates ad hoc groups with the mandate to study the relevance in terms of standardization of specific topics.

After identifying a relevant area of interest, and in order to fulfill its objectives guided by the preceding principles, MPEG follows a standards development process with the following major steps [Chia97]:

1. **Applications:** Identify relevant applications using input from MPEG members.
2. **Functionalities:** Identify the functionalities needed by the applications.
3. **Requirements**: Describe the requirements following from the functionalities so that common requirements can be identified for different applications.
4. **Common requirements:** Identify which requirements are common across the areas of interest, and which are not common but still relevant.
5. **Specification:** Specify the tools supporting the requirements in three phases.
    i) **Call for proposals:** A public call for proposals is issued, asking all interested parties to submit technology that is relevant to the identified requirements and functionalities.
    ii) **Proposal evaluation:** The proposals are evaluated in a well-defined, adequate, and fair evaluation process, which is published with the call for proposals. The process can entail subjective testing, objective comparison, and evaluation by experts.
    iii) **Technical specification:** As a result of the evaluation, the technology best addressing the requirements is selected; this technology typically does not correspond to a single proposal but to the set of the best tools extracted from all proposals. This is the start of a collaborative process to draft and improve the standard. The collaboration includes the definition and improvement of a *working model*, which embodies early versions of the standard and can include non-normative tools to better and more completely test the normative tools. The working model evolves through comparing different alternative tools with those already in the working model, through the so-called *core experiments* (CE).
6. **Verification:** Verify that the tools developed can be used to assemble the target systems and to provide the desired functionalities with an adequate level of performance. This is done by means of *verification tests.* Until MPEG-4, the verification tests consisted of formal subjective tests aimed at evaluating the quality of either audio or video signals processed using specific MPEG algorithms; in MPEG-4, new types of tests have been performed [N999]. In order to obtain reliable and representative results, the tests are performed using optimized assessment methods and suitable panels of subjects.

This process is not rigid; some steps may be taken more than once and iterations are sometimes needed (as happened in MPEG-4). The time schedule, however, is always closely observed by MPEG. Although all decisions are made by consensus, the process keeps a fast pace, allowing MPEG to provide good technical solutions in a timely manner.

While the period until the proposals are evaluated is called the *competitive phase,* the period after the evaluation is the *collaborative phase*. During the collaborative phase, all the MPEG members collectively improve and complete the most promising tools identified at the proposals' evaluation. The collaborative phase is the major strength of the MPEG process, as hundreds of the top experts in the world, from many companies and universities, work together for a common goal. In this context, it does not come as a surprise that this superteam traditionally achieves excellent technical results, justifying the need for most companies to at least follow the process, if direct involvement is not possible.

Two working tools play a major role in the collaborative development phase that follows the initial competitive phase: the working model and core experiments [N1375]. In MPEG-1, the (video) working model was called Simulation Model (SM); in MPEG-2, the (video) working model was called Test Model (TM); and in MPEG-4, the various working models were called Verification Models (VMs).[7] In MPEG-4, there were independent VMs for the video, audio, synthetic and natural hybrid coding (SNHC), and systems developments. Regarding the MPEG-4 VMs and CEs, it is important to highlight a few points [Pere99].

1.3.3.1 Verification Models A VM is a complete framework defined in text and with a corresponding software implementation, such that an experiment performed by multiple independent parties will produce essentially identical results. VMs are enabled to check the relative performance of different tools, as well as to improve the performance of selected tools. The MPEG-4 VMs were built after screening the proposals. The first VM (for each technical area, e.g., video, audio, and SNHC) was not the best proposal but a combination of the best tools, independent of the proposal to which they belonged. Each VM included normative and non-normative tools to create the *common framework* that allowed performing adequate evaluation and comparison of tools targeting the continuous improvement of the technology included in the VM. After the first VMs were established, new tools were brought to MPEG-4 and evaluated within the VMs following a core experiment procedure. The VMs evolved through versions as CEs verified the inclusion of new techniques or proved that included techniques should be substituted. At each VM version, only the best performing tools were part of the VM. If any part of a proposal was

7. In MPEG-7, the working model is called eXperimentation Model (XM), whereas in MPEG-21 the working model is called sYstems Model (YM).

selected for inclusion in the VM, the proposer had to provide the corresponding source code for integration into the VM software in the conditions specified by MPEG.

1.3.3.2 Core Experiments The improvement of the VMs started with a first set of CEs defined at the conclusion of proposal evaluation [N998]. The CE process allowed for multiple, independent, directly comparable experiments to determine whether a proposed tool had merit. Proposed tools targeted the substitution of a tool in one of the VMs or the direct inclusion in the VM to provide a new relevant functionality. Improvements and additions to the VMs were decided based on the results of CEs.

A CE must be completely and uniquely defined, so that the results are unambiguous. In addition to the specification of the tool to be evaluated, a CE also specifies the conditions to be used, again so the results can be compared. A CE is proposed by one or more MPEG experts and is accepted by consensus, providing that two or more independent experts agreed to perform the experiment.

It is important to realize that neither the text of the VMs nor any of the CEs ended (or will end) up in the standard itself, as they were just working tools to ease the development process. However the VMs' software is the basis for MPEG-4 Part 5: Reference Software [MPEG4-5] presented in the next section. Although it is not easy at this stage to tell how many CEs have been performed in MPEG-4, it is possible to state that they reached their goal by continuously improving and completing the technology to be included in the standard.

1.4 MPEG-4 Standard Organization

The MPEG-4 requirements [N4319] have been addressed by a standard organized in several Parts, each one with multiple editions, amendments, and corrigenda. Amendments to each Part of the standard were also called *Versions* (the first edition itself is Version 1).[8] The list of documents issued as of July 2001 for the various Parts of the MPEG-4 standard is as follows:

☞ **Part 1: Systems.** This Part specifies scene description, multiplexing, synchronization, buffer management, and management and protection of intellectual property [MPEG4-1].

Version 1: First edition (14496-1:1999)

Version 2 (first amendment to first edition): Systems extensions (MPEG-4 file format, BIFS nodes)

8. Amendments to the second editions were sometimes called *extensions*.

Corrigendum 1 to first edition

Second edition of Part 1, including the 1999 edition, Amendment 1, and Corrigendum 1 (14496-1:2001)

Amendment 1 to second edition: Extended BIFS (Flextime)

Amendment 2 to second edition: Textual format (XMT)

Amendment 3 to second edition: Intellectual Property Management and Protection (IPMP) extension

Amendment 4 to second edition: Multiuser worlds and animation framework (AFX)

Amendment 5 to second edition: MP4 extensions (common with WG1/JPEG)

Amendment 6 to second edition: MP4 extensions

Corrigendum 1 to second edition

Corrigendum 2 to second edition

☞ **Part 2: Visual.** This Part specifies the coded representation of natural and synthetic visual objects [MPEG4-2].

Version 1: First edition (14496-2:1999)

Version 2 (first amendment to first edition): Visual extensions

Version 3 (second amendment to first edition): 3D mesh profiles (withdrawn)

Corrigendum 1 to first edition

Corrigendum 2 to first edition

Second edition of Part 2, including the first edition, Amendment 1, and Corrigenda 1 and 2 (14496-2:2001)

Amendment 1 to second edition: Studio profiles

Amendment 2 to second edition: Streaming video profiles

Amendment 3 to second edition: New levels and tools

☞ **Part 3: Audio.** This Part specifies the coded representation of natural and synthetic audio objects [MPEG4-3].

Version 1: First edition (14496-3:1999)

Version 2 (first amendment to first edition): Audio extensions

Corrigendum 1 to first edition

Second edition of Part 3, including the first edition, Amendment 1, and Corrigendum 1 (14496-3:2001)

Corrigendum 1 to second edition

- **Part 4: Conformance Testing.** This Part defines conformance conditions for bitstreams and devices; it is used to test MPEG-4 implementations [MPEG4-4].

  Version 1: First edition (14496-4:2000)

  Version 2 (first amendment to first edition): Extensions to conformance testing

  Corrigendum 1 to first edition

  Second edition of Part 4, including the first edition, Amendment 1, and Corrigendum 1 (14496-4:2001)

  Amendment 1 to second edition: Conformance extensions for studio and streaming video profiles and for Flextime

  Corrigendum 1 to second edition

- **Part 5: Reference Software.** This Part includes software corresponding to most Parts of MPEG-4 (normative and non-normative tools); this means the VMs mentioned in Section 1.3.3.1. This software can be used for implementing compliant products as ISO waives the copyright of the code[9] [MPEG4-5].

  Version 1: First edition (14496-5:1999)

  Version 2 (first amendment to first edition): Reference software extensions

  Corrigendum 1

  Second edition of Part 5, including the first edition, Amendment 1, and Corrigendum 1 (14496-5:2001)

  Amendment 1 to second edition: Reference software extensions for studio and streaming video profiles and for Flextime

- **Part 6: Delivery Multimedia Integration Framework (DMIF).** Part 6 defines a session protocol for the management of multimedia streaming over generic delivery technologies [MPEG4-6].

  Version 1: First edition (14496-6:1999)

  Version 2 (Amendment 1 to first edition): DMIF extensions

  Corrigendum 1

  Second edition of Part 6, including the first edition, Amendment 1, and Corrigendum 1 (14496-6:2000)

9. This does not mean that product developers do not have to license the patents involved.

- **Part 7: Optimized Visual Reference Software.** This Part includes optimized software for visual tools such as fast motion estimation, fast global motion estimation, and fast and robust sprite generation [N4554]; the optimized software is called *Optimization Model* (OM). Unlike the other MPEG-4 standard Parts, this Part is a TR and not a standard specification.

  Version 1: First edition (14496-7:2001)

- **Part 8: Carriage of MPEG-4 Contents over IP Networks.** This Part specifies the mapping of MPEG-4 content into several IP-based protocols [N4427]; it is well known as *4on IP*.

  Version 1: First edition (14496-8:2002)

- **Part 9: Reference Hardware Description.** Part 9 will include portable synthesizable/simulatable very high-speed integrated circuit hardware description language (VHDL) descriptions of MPEG-4 tools [N4218].

  Version 1: First edition (14496-9:2003)

- **Part 10: Advanced Video Coding (AVC).** This Part will specify video syntax and coding tools[10] in the context of a joint project with ITU-T SG16 [N4400], known as Joint Video Team (JVT); this activity used as a starting point the available version of the H.26L video coding specification to address the identified requirements [N4466, N4508].

  Version 1: First edition (14496-10:2003)

Although most parts are IS, Parts 7 and 9 are TRs with informative value. Parts 1 to 3 as well as Parts 6, 8, and 10 specify the core MPEG-4 technology, whereas Parts 4, 5, 7, and 9 are supporting Parts. Parts 1, 2, 3, and 10 are delivery-independent, leaving to Parts 6 and 8 the task of dealing with the idiosyncrasies of the delivery layer.

The major reason to develop the MPEG-4 standard in several rather independent Parts (besides avoiding a single document with several thousand pages) is to allow the various pieces of technology to be useful as stand-alones and thus as much used as possible, even if in conjunction with proprietary technologies. This has been the case, for example, for MPEG-2 Video, which today is being used together with MPEG-2 Systems but not with MPEG-2 Audio in the context of the U.S. digital TV system. This means that within the context of a certain standardization effort (e.g., MPEG-4), whenever a new technological area is addressed that is different from the areas already addressed (e.g., MPEG-4 on IP), an additional Part of the standard is created to allow stand-alone use and a clearer organization of the tools specified by the standard as a whole. However, although the various Parts may be used

10. The working model for this activity is called Joint Model (JM).

independently, they were developed to give optimal results when they are used together.

During the MPEG-4 development process, it was decided to issue successive versions of the several MPEG-4 Parts whenever new tools needed to be added to that Part of the standard. In this context, versions would serve to specify new tools, either offering new functionalities or bringing a significant improvement in terms of functionalities already supported. Formally speaking, the various versions (except the first) of a certain MPEG-4 Part correspond to amendments to that Part of the standard. So, whereas the IS for Systems (issued in 1999) is called Version 1, MPEG-4 Systems Version 2 corresponds to the first MPEG-4 Systems amendment, Version 3 to the Amendment 2 to Systems, and so on. It is important to note that new versions of a Part do not substitute or redefine tools specified in previous versions but simply add more tools. At each stage of specification, a certain Part of the MPEG-4 standard is the set of all the tools specified in all versions for that Part. In that sense, it is common to say that versions are backward-compatible, meaning that Version N may only add new tools and profiles to Version N-1 and not remove or redefine any tool or profile. This implies that existing terminals will always remain compliant, as profiles will not be changed in retrospect. The same reasoning applies to the amendments to the second and further editions of the various Parts of the standard.

1.5 MPEG-4 SCHEDULE

After the initial study phase, in which the major MPEG-4 objectives were identified, notably through the functionalities presented earlier, the first MPEG-4 call for proposals was issued in July 1995 [N998] and answers were received by September/October 1995. The call for proposals asked for relevant video and audio technology addressing the eight MPEG-4 functionalities as described in the MPEG-4 Proposal Package Description [N998]. The technology received was evaluated using subjective tests for complete algorithms and expert panels for single tools [N999]. In the case of algorithms addressing the eight MPEG-4 functionalities, three functionalities (one for each of the three classes) were selected as representative—content-based scalability, improved compression efficiency, and robustness in error-prone environments—and formal subjective tests were conducted for those. For the other five functionalities, proposals were evaluated by expert panels (in the same manner as tools), and the corresponding selected tools were thoroughly examined using the CE procedure.

The video subjective tests were performed in November 1995 at Hughes Aircraft Company, in Los Angeles; the audio subjective tests were performed in December 1995 at CCETT, Mitsubishi, NTT, and Sony. The video expert panels evaluation was performed in October 1995 and January 1996.

At this initial phase, MPEG decided to develop the Systems specification by means of a pure collaborative approach and thus no calls were issued for Systems tools.

After the evaluation of the technology received [M532], choices were made and the collaborative phase started with the most promising tools. In the course of developing the standard, additional calls were issued when insufficient technology was available within MPEG to meet the requirements—for example, for synthetic and hybrid coding tools in July 1996 [N1315], for identification and protection of content in April 1997 [N1714], and for an intermedia file format in October 1997 [N1919]. This is a typical solution when MPEG is missing some technology and there are good indications that the technology exists outside MPEG. Moreover, in order to check that the standardized technology is still among the best available, MPEG may issue calls for proposals to compare the already standardized tools with the most recent developments outside MPEG. This was the case with the calls for proposals on audio [N3992] and video [N4065] coding tools in March 2001. The results of the video tests [N4454] following the video call for proposals led to the creation of the joint project with ITU-T SG 16 to develop MPEG-4 Part 10 [N4400].

At the MPEG January 1996 meeting in Munich, the first MPEG-4 Video Verification Model (VM) was defined [N1172]. In this VM, and for the first time in a standardization process, a video scene was represented as a composition of arbitrarily shaped objects; each object was represented by a sequence of *Video Object Planes* (VOPs), which are the temporal instantiations of a video object at a certain moment. For this, the first MPEG-4 Video VM used ITU-T H.263 coding tools [H263] together with shape coding tools, following the results of the November 1995 MPEG-4 video subjective tests [M532].

A process similar to the one used for video was followed for audio, although with some initial delay due to the involvement of many audio experts in the advanced audio coding (AAC) MPEG-2 work [N1214].

Following this initial phase, the several MPEG-4 VMs evolved using the CE process. A new version of each of the MPEG-4 VMs has been issued at each MPEG meeting—for example, the Video VM was in its 18th version at the Pisa meeting in January 2001 [3908].

As highlighted in the previous section, the last step of the MPEG process is the verification of the technology in the standard aiming at testing the performance of the tools and demonstrating their potentialities. For MPEG-4, the verification step was performed through a set of verification tests addressing various parts of the standard. Many verification tests have already been performed for video and audio tools and profiles (see Chapters 15 and 16 on testing for validation).

For MPEG-4, the process highlighted in the previous sections translated to the time schedule presented in Table 1.2.

Table 1.2 MPEG-4 time schedule*

Date	Event
July 1995 (Tokyo)	Call for proposals on audio and video tools and algorithms [N998] Final version of the MPEG-4 evaluation document [N999]
November 1995	Subjective evaluation of video proposals
December 1995	Subjective evaluation of audio proposals
January 1996 (Munich)	Experts evaluation of video proposals First version of the MPEG-4 Video VM [N1172]
March 1996 (Florence)	First version of the MPEG-4 Audio VM [N1214]
July 1996 (Tampere)	Call for proposals on SNHC tools [N1315]
September 1996 (Chicago)	First version of the MPEG-4 SNHC VM [N1364] Call for proposals on synthetic audio [N1397]
November 1996 (Maceió)	WD, Parts 1, 2, 3, 5, and 6 Call for proposals on audio and video tools and algorithms [N1499]
April 1997 (Bristol)	Call for proposals on identification and protection of content in MPEG-4 [N1714]
October 1997 (Fribourg)	Call for proposals for an MPEG-4 intermedia format [N1919] CD, Parts 1, 2, 3, 5, and 6 WD, Part 4
March 1998 (Tokyo)	FCD, Parts 1, 2, 3, 5, and 6
October 1998 (Atlantic City)	FDIS, Parts 1, 2, 3, and 6
December 1998 (Rome)	CD, Part 4
March 1999 (Seoul)	Version 2, PDAM status, Parts 1, 2, 3, and 6
July 1999 (Vancouver)	Version 2, FPDAM status, Parts 1, 2, 3, and 6 FCD, Part 4 FDIS, Part 5 Version 2, PDAM status, Part 5
October 1999 (Melbourne)	COR 1, DCOR status, Parts 1, 2, and 6
December 1999 (Maui)	Call for proposals for an MPEG-4 textual format [N3157] Version 2, FDAM status, Parts 1, 2, 3, and 6 Amendment 1 to second edition, PDAM status, Part 1

Table 1.2 MPEG-4 time schedule* (Continued)

Date	Event
December 1999 (Maui) (cont)	FDIS, Part 4 Version 2, PDAM status, Part 4
March 2000 (Noordwijker-hout)	Call for proposals for a generic animation framework of synthetic objects [N3341] COR 1, COR status, Parts 1, 2, and 6 Amendment 1 to second edition, PDAM status, Part 2 Amendment 2 to second edition, PDAM status, Part 2 Version 2, FPDAM status, Part 5
May 2000 (Geneva)	Amendment 1 to second edition, FPDAM status, Part 1
July 2000 (Beijing)	Call for proposals for IPMP solutions [N3543] Call for proposals on multiusers worlds technology [N3574] Amendment 1 to second edition, FPDAM status, Part 2 Amendment 2 to second edition, FPDAM status, Part 2 COR 2, DCOR status, Part 2 COR 1, DCOR status, Part 3 Version 2, FPDAM status, Part 4 Version 2, FDAM status, Part 5 COR 1, DCOR status, Part 5
October 2000 (La Baule)	Call for proposals for new tools to further improve video coding efficiency [N3671] WG11 approval of second edition, Part 1 Amendment 1 to second edition, FDAM status, Part 1 Amendment 2 to second edition, PDAM status, Part 1 COR 1, DCOR status, Part 1 (second edition) COR 2, COR status, Part 2 WG11 approval of second edition, Part 6
January 2001 (Pisa)	Amendment 1 to second edition, FDAM status, Part 2 Amendment 2 to second edition, FDAM status, Part 2 COR 1, COR status, Part 3 Version 2, FDAM status, Part 4 COR 1, COR status, Part 5
March 2001 (Singapore)	Call for proposals for new tools for audio coding [N3992] Call for proposals for new tools for video compression technology [N4065] Call for proposals for interpolator compression [N4098] WG11 approval of second edition, Part 3 WG11 approval of second edition, Part 4 Amendment 1 to second edition, PDAM, Part 4 WG11 approval of second edition, Part 5

Table 1.2 MPEG-4 time schedule* (Continued)

Date	Event
March 2001 (Singapore) (cont)	Amendment 1 to second edition, PDAM, Part 5 CD, Part 8
July 2001 (Sydney)	Call for proposals for hardware reference code [N4218] Amendment 2 to second edition, FPDAM status, Part 1 Amendment 3 to second edition, PDAM status, Part 1 COR 1, COR status, Part 1 (second edition) WG11 Approval of Edition, Part 2 Amendment 1 to second edition, FPDAM, Part 5 Amendment 2 to second edition, PDAM, Part 5 PDTR, Part 7 FCD, Part 8
December 2001 (Pattaya)	Amendment 4 to second edition, PDAM status, Part 1 Amendment 5 to second edition, PDAM status, Part 1 COR 2, DCOR status, Part 1 (second edition) Amendment 3 to second edition, PDAM, Part 2 COR 1, DCOR status, Part 3 (second edition) Amendment 1 to second edition, FPDAM, Part 4 COR 1, DCOR status, Part 4 (second edition) Amendment 2 to second edition, FPDAM, Part 5 DTR, Part 7 FDIS, Part 8

*Note: For the first edition of each Part of the standard, Version N corresponds to Amendment N-1.

Looking at Table 1.2, it is possible to conclude that the development of Version 1 of the MPEG-4 standard took about four and a half years between issuing the first call for proposals and the publication of the IS (Version 1). Although MPEG is generally considered as a body adopting challenging workplans, notably by its members, it is worthwhile to reflect about the time it takes to develop a standard, and thus on its chances of success against proprietary solutions in such a quickly moving technical landscape, if at least four years is the time a fast standardization body needs to make a standard available to the industry. This is not to speak about the undefined additional time the companies owning the essential patents will take to set the licensing procedure so that industry can start selling products based on that standard.

MPEG-4 was developed by hundreds of experts from many companies and universities around the world who believe that MPEG-4 technology can power the next generation of multimedia products and services. MPEG-4 Version 1 reached FDIS status at the end of 1998 and thus was technically finished by that time. The following amendments increased the capabilities of the standard in a backward-compatible way and are being developed as the industry needs emerge.

1.6 MPEG-4 Industry Forum

Because MPEG itself is not allowed to deal with any issues besides the development of technical specifications (notably patent identification and licensing), the industry players interested in the deployment of products and applications based on the MPEG-4 standard decided in the spring of 2000 to create the MPEG-4 Industry Forum (M4IF) [M4IF]. The M4IF is a not-for-profit organization for industrial players who want to manufacture, deploy, or use MPEG-4 technology. The philosophy is that all involved parties, even though competitors in some respect, profit from the standard being accepted and taking off. The major goal of the forum is "to further the adoption of the MPEG-4 standard, by establishing MPEG-4 as an accepted standard among users, application developers, service providers, content creators and end users" [M4IF].

The forum is open to all parties agreeing with the forum's objectives and includes a broad, worldwide representation from consumer electronics, computers, and telecommunications companies as well as research institutions. Also, some of the members are business users of MPEG-4. Among the participants are many small companies that develop or deploy MPEG-4 technology. Membership goes beyond the MPEG constituency, partly because some of the smaller companies cannot comply with the requirements for participation in MPEG (which vary from country to country) and also because some companies do not need to be involved in the development phase.

The activities of M4IF generally start where MPEG stops. This includes issues that MPEG cannot deal with because of ISO rules, such as clearance of patents. According to the M4IF statutes, the M4IF purposes shall be pursued by the following means:

- ☞ Promoting the standard and serving as a single point of information on MPEG-4 technology, products, and services.
- ☞ Initiating discussions leading to the potential establishment of patent pools outside M4IF, which should grant a license to an unlimited number of applicants throughout the world under reasonable terms and conditions that are demonstrably free of any unfair competition; includes studying licensing models for downloadable software decoders, such as Internet players.
- ☞ Organizing MPEG-4 exhibitions and tutorials.
- ☞ Creating industrial focus around the standard, for example, by identifying which MPEG-4 profiles are needed for which market.

M4IF sees the creation of patent pools[11] as one of the most important issues to enable the wide-scale adoption of the MPEG-4 standard. That is why the forum has facilitated a number of meetings of patent holders and intends to continue doing so until the pools are well under way toward being established. M4IF will not directly deal with patents, patent pools, and patent licensing. This is strictly a matter for patent owners to resolve and falls outside M4IF's scope. This means M4IF will not sell any MPEG-4 licenses or even determine the licensing policies. It merely acts as a catalyst in getting holders of essential patents to sit together and establish a portfolio of essential worldwide patents that are necessary for the implementation of the MPEG-4 standard in order to provide all MPEG-4 users with fair, reasonable, nondiscriminatory access to the technology under a single license. Also, M4IF discusses possible licensing principles, applicable to hardware and software products, to better understand the needs for licensing in emerging MPEG-4 application domains.

The goals are realized through the open international collaboration of all interested parties, on reasonable terms applied uniformly and openly. M4IF will contribute the results of its activities to appropriate formal standards bodies if applicable. The business of M4IF is not conducted for the financial profit of its members but for their mutual benefit. Any corporation and individual firm, partnership, governmental body, or international organization supporting the purpose of M4IF may apply for membership. Members are not bound to implement or use specific technology standards or recommendations by virtue of participation in M4IF.

M4IF anticipates holding three physical meetings per year, with a slightly higher frequency in the start-up phase. About 100 people have attended these meetings from all over the world. The instructions to join M4IF are available at the M4IF Web site at *www.m4if.org/join.html*.

1.7 SUMMARY

MPEG-1 and MPEG-2 are successful standards that have given rise to widely adopted commercial products, such as CD-interactive, digital audio broadcasting, and digital television. However, these standards are limited in terms of the functionalities provided by the data representation models used.

The MPEG-4 standard opens new frontiers in the way users will play with, create, reuse, access, and consume audiovisual content. The MPEG-4

11. A patent pool is typically a set of patents, belonging to a group of companies, for which a single licensing point exists, facilitating the licensing task of the companies willing to develop applications and products using those patents. The patents in a pool are all relevant to build integrated solutions to address the needs of certain application domains. For example, there is a patent pool for MPEG-2 Systems and MPEG-2 Video main profile@main level tools.

object-based representation approach—in which a scene is modeled as a composition of objects, both natural and synthetic, with which the user may interact—is at the heart of the MPEG-4 technology. Moreover MPEG-4 behaves rather well also in terms of compression. For example, for frame-based video coding, MPEG-4 brings a competing solution from very low bit rates to very high bit rates; in fact there are already MPEG-4 visual levels from 64 kbit/s for the `Simple` profile at Level 0 (adopted by the Third Generation Project Partnership [3GPP] consortium for the third-generation mobile networks applications) up to 1,800 Mbit/s for the `Simple Studio` profile at Level 4.

The MPEG-4 vision and its associated technology provide the means to launch a great diversity of applications, with varying degrees of interactivity, notably for the emerging third-generation mobile networks, for the Internet, and even for digital radio and cable broadcasting networks. It is now up to the application developers and content authors to transform this great technology in content, products, and applications, making the MPEG-4 standard the audiovisual playground of the future.

Whatever will be the success of the MPEG-4 standard in terms of products and applications (likely determined by industrial, economic, legal, and marketing interests), the new concepts underpinning the MPEG-4 standard point in the right direction in terms of representation technology, as they rely on some basic characteristics of the human–world relationship, brought for the first time to the audiovisual representation arena.

1.8 REFERENCES

[Chia97] Chiariglione, Leonardo. "MPEG and Multimedia Communications." *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1): 5–18, February 1997.

[H263] ITU-T Recommendation H.263++. "Video Coding for Low Bit Rate Communication." International Telecommunications Union—Telecommunications Standardization Sector (Geneva), 3, 2000.

[JTC1] JTC1 Home Page. *www.jtc1.org/*

[KoPC97] Koenen, Rob, Fernando Pereira, and Leonardo Chiariglione. "MPEG-4: Context and Objectives." *Signal Processing: Image Communication*, 9(4): 295–304, May 1997.

[M4IF] MPEG-4 Industry Forum Home Page. *www.m4if.org/index.html*

[M532] MPEG Test. *Report of the Ad Hoc Group on MPEG-4 Video Testing Logistics*. Doc. ISO/MPEG M532, Dallas MPEG Meeting, November 1995.

[Marr82] Marr, David. *Vision*. New York: W.H. Freeman, 1982.

[MPEG] MPEG Home Page. *http://mpeg.telecomitalialab.com/*

[MPEG4-1] ISO/IEC 14496-1:2001. *Coding of Audio-Visual Objects—Part 1: Systems*, 2d Edition, 2001.

[MPEG4-2] ISO/IEC 14496-2:2001. *Coding of Audio-Visual Objects—Part 2: Visual*, 2d Edition, 2001.

[MPEG4-3] ISO/IEC 14496-3:2001. *Coding of Audio-Visual Objects—Part 3: Audio*, 2d Edition, 2001.

[MPEG4-4] ISO/IEC 14496-4:2001. *Coding of Audio-Visual Objects—Part 4: Conformance Testing*, 2d Edition, 2001.

[MPEG4-5] ISO/IEC 14496-5:2001. *Coding of Audio-Visual Objects—Part 5: Reference Software*, 2d Edition, 2001.

[MPEG4-6] ISO/IEC 14496-6:2000. *Coding of Audio-Visual Objects—Part 6: Delivery Multimedia Integration Framework (DMIF),* 2d Edition, 2000.

[N1172] MPEG Video. *MPEG-4 Video Verification Model 1.0*. Doc. ISO/MPEG N1172, Munich MPEG Meeting, January 1996.

[N1177] MPEG Convener. *MPEG-4 Project Description*. Doc. ISO/MPEG N1177, Munich MPEG Meeting, January 1996.

[N1214] MPEG Audio. *MPEG-4 Audio Verification Model 1.0*. Doc. ISO/MPEG N1214, Florence MPEG Meeting, March 1996.

[N1315] MPEG Convener. *Call for Proposals and PPD on SNHC Tool*s. Doc. ISO/MPEG N1315, Tampere MPEG Meeting, July 1996.

[N1364] MPEG SNHC. *MPEG-4 SNHC Verification Model 1.0*. Doc. ISO/MPEG N1364, Chicago MPEG Meeting, September 1996.

[N1375] MPEG Convener. *Verification Model (VM) Development and Core Experiments*. Doc. ISO/MPEG N1375, Chicago MPEG Meeting, September 1996.

[N1397] MPEG Convener. *Call for Proposals on Synthetic Audio*. Doc. ISO/MPEG N1397, Chicago MPEG Meeting, September 1996.

[N1499] MPEG Convener. *MPEG-4 Call for Proposals*. Doc. ISO/MPEG N1499, Maceió MPEG Meeting, November 1996.

[N1714] MPEG Convener. *Call for Proposals on Identification and Protection of Content in MPEG-4*, Doc. ISO/MPEG N1714. Bristol MPEG Meeting, April 1997.

[N1919] MPEG Convener. *Call for Proposals for an MPEG-4 Intermedia Format*. Doc. ISO/MPEG N1919, Fribourg MPEG Meeting, October 1997.

[N271] MPEG Convener. *New Work Item Proposal (NP) for Very-Low Bitrates Audiovisual Coding*. Doc. ISO/MPEG N271, London MPEG Meeting, November 1992.

[N2724] MPEG Requirements. *MPEG-4 Applications*. Doc. ISO/MPEG N2724, Seoul MPEG Meeting, March 1999.

[N3157] MPEG Convener. *Call for Proposals for an MPEG-4 Textual Format*. Doc. ISO/MPEG N3157, Maui MPEG Meeting, December 1999.

[N3341] MPEG Convener. *Call for Proposals for a Generic Animation Framework of Synthetic Objects*, Doc. ISO/MPEG N3341. Noordwijkerhout MPEG Meeting, March 2000.

[N3543] MPEG Convener. *Call for Proposals for IPMP Solutions*. Doc. ISO/MPEG N3543, Beijing MPEG Meeting, July 2000.

[N3574] MPEG Convener. *Call for Proposals on Multi-Users Worlds Technology.* Doc. ISO/MPEG N3574, Beijing MPEG Meeting, July 2000.

[N3671] MPEG Convener. *Call for Proposals for New Tools to Further Improve Video Coding Efficiency.* Doc. ISO/MPEG N3671, La Baule MPEG Meeting, October 2000.

[N3908] MPEG Video. *MPEG-4 Video Verification Model,* Version 18. Doc. ISO/MPEG N3908, Pisa MPEG Meeting, January 2001.

[N3992] MPEG Convener. *Call for Proposals for New Tools for Audio Coding.* Doc. ISO/MPEG N3992, Singapore MPEG Meeting, March 2001.

[N4065] MPEG Convener. *Call for Proposals for New Tools for Video Compression Technology.* Doc. ISO/MPEG N4065, Singapore MPEG Meeting, March 2001.

[N4098] MPEG Convener. *Call for Proposals for Interpolator Compression.* Doc. ISO/MPEG N4098, Singapore MPEG Meeting, March 2001.

[N4218] MPEG Convener. *Call for the Submission of Hardware Reference Code for MPEG-4—Part 9*: *Reference Hardware Description.* Doc. ISO/MPEG N4218, Sydney MPEG Meeting, July 2001.

[N4319] MPEG Requirements. *MPEG-4 Requirements*. Doc. ISO/MPEG N4319, Sydney MPEG Meeting, July 2001.

[N4333] MPEG. *Multimedia Framework (MPEG-21)—Part 1: Vision, Technologies and Strategy.* Draft Technical Report, Doc. ISO/MPEG N4333, Sydney MPEG Meeting, July 2001.

[N4400] MPEG. *JVT Terms of Reference*. Doc. ISO/MPEG N4400, Pattaya MPEG Meeting, December 2001.

[N4427] MPEG. *Coding of Audio-Visual Objects—Part 8: Carriage of MPEG-4 Contents Over IP Networks*. Final Draft International Standard, Doc. ISO/MPEG N4427, Pattaya MPEG Meeting, December 2001.

[N4454] MPEG Test. *Results of Subjective Assessment of Responses to Video Call for New Tools to Further Improve Coding Efficiency.* Doc. ISO/MPEG N4454, Pattaya MPEG Meeting, December 2001.

[N4466] MPEG Video. *JVT Coding Joint Working Draft*. Doc. ISO/MPEG N4466, Pattaya MPEG Meeting, December 2001.

[N4500] MPEG Convener. *Terms of Reference*. Doc. ISO/MPEG N4500, Pattaya MPEG Meeting, December 2001.

[N4505] MPEG. *MPEG-4 Overview.* Doc. ISO/MPEG N4505, Pattaya MPEG Meeting, December 2001.

[N4508] MPEG Requirements. *Requirements for JVT.* Doc. ISO/MPEG N4508, Pattaya MPEG Meeting, December 2001.

[N4509] MPEG Requirements. *Overview of the MPEG-7 Standard*. Doc. ISO/MPEG N4509, Pattaya MPEG Meeting, December 2001.

[N4511] MPEG. *MPEG-21 Overview.* Doc. ISO/MPEG N4511, Pattaya MPEG Meeting, December 2001.

[N4554] MPEG. *Coding of Audio-Visual Objects—Part 7: Optimized Visual Reference Software*. Draft Technical Report, Doc. ISO/MPEG N4554, Pattaya MPEG Meeting, December 2001.

[N998] MPEG AOE. *Proposal Package Description (PPD): Revision 3*. Doc. ISO/MPEG N998, Tokyo MPEG Meeting, July 1995.

[N999] MPEG Test. *MPEG-4 Testing and Evaluation Procedures Document*. Doc. ISO/MPEG N999, Tokyo MPEG Meeting, July 1995.

[PeKo96] Pereira, Fernando, and Rob Koenen. "Very Low Bitrate Audiovisual Applications." *Signal Processing: Image Communication*, 9(1): 55–77. November 1996.

[PeKo99] Pereira, Fernando, and Rob Koenen. “MPEG-7: Status and Directions,” (A. Puri and T. Chen, Eds.). *Multimedia Systems, Standards and Networks*. New York: Marcel Dekker, 1999.

[Pere99] Pereira, Fernando. “MPEG-4: Why, What, How and When?” *Signal Processing: Image Communication, Tutorial Issue on the MPEG-4 Standard*, 15(4–5): 271–279. December 1999.