MPEG4: coding for content, interactivity, and universal accessibility

Cliff Reader
Samsung Electronics
3655 North First Street
San Jose, California 95134-1713
E-mail: creader@samsung.com

Abstract. MPEG4 is a natural extension of audiovisual coding, and yet from many perspectives breaks new ground as a standard. New coding techniques are being introduced, of course, but they will work on new data structures. The standard itself has a new architecture, and will use a new operational model when implemented on equipment that is likely to have innovative system architecture. The author introduces the background developments in technology and applications that are driving or enabling the standard, introduces the focus of MPEG4, and enumerates the new functionalities to be supported. Key applications in interactive TV and heterogeneous environments are discussed. The architecture of MPEG4 is described, followed by a discussion of the multiphase MPEG4 communication scenarios, and issues of practical implementation of MPEG4 terminals. The paper concludes with a description of the MPEG4 workplan. In summary, MPEG4 has two fundamental attributes. First, it is the coding of audiovisual objects, which may be natural or synthetic data in two or three dimensions. Second, the heart of MPEG4 is its syntax: the MPEG4 Syntactic Descriptive Language—MSDL.

Subject terms: visual communication and image processing; MPEG4; multimedia; digital video; compression.

Optical Engineering 35(1), 104–108 (January 1996).

1 Introduction

MPEG4 introduces a new generation of coding standards for audiovisual data. Fifty years ago, video standardization meant specifying voltage levels, and end-to-end synchronous timing was required. The motivation was not only interoperability, but the need to define practical, economical implementations. Recently video standardization has meant standardizing a coding algorithm, again motivated in part by the economics of mass production that result from "hard-wired" implementation. Coding schemes such as H.320 and MPEG each have defined a bit-precise syntax for efficiency and ease of decoding, but the coded representation is abstracted from the specifics of the implementation; in other words, it is at a higher layer in the protocol stack than the physical layer. These schemes also have been able to include a limited degree of flexibility, most notably in the form of "profiles" in MPEG2. However, they still code "video," i.e., a regular time sequence of two-dimensional video frames and audio samples. See Fig. 1. In contrast, in the future, we will standardize a communication language for audiovisual objects, and MPEG4 is taking the key step in that direction.

This evolution is made possible to a large extent by advances in technology. Digital TV settop boxes may be built upon the architecture of a PC, and PCs may be enhanced by adding TV tuner cards. Clearly the two are fusing, and what matters is the availability of very high-performance processors and high-capacity memories at consumer-level prices. Today we have standardized specific coding algorithms recognizing that dedicated implementation in hardware is essential to low cost. But the audiovisual system of tomorrow will be built around an embedded PC, and coding will be done in software. That means a big advance in flexibility, and in particular an openness for future development.

A specific consequence is flexibility in presentation, with a decoupling of the update rate and resolution for the audiovisual data themselves from the refresh rate and resolution of the display or sample rate of the audio output. The opportunity exists to present audiovisual objects each at its native resolution or to scale the resolution as appropriate.

Computer technology has already had a major influence through creating tools for postproduction. Today, the leading technology is nonlinear editing systems. With workstation processing power and hard-disk capacities unthinkable only a few years ago, today's all-digital suite generates, manipulates, and composes a wide range of audio and video objects: 3-D titles, video clips, audio clips, computer animation, synthetic audio. The authoring environment is thus already object-oriented with a mix of natural and synthetic data that is 2-D or 3-D.

On the presentation side, computer technology has also led to great progress in video games. Memory technology permits storage of databases large enough to support photorealistic scenes in high-end environments, and remarkably complex scenes even on consumer-level products. The interactive nature of the games means the presentation is also "nonlinear" and consists of manipulating and composing the audio and video objects in the database.

Paper VIS-09 received May 10, 1995; revised manuscript received July 12, 1995; accepted for publication July 31, 1995.
© 1996 Society of Photo-Optical Instrumentation Engineers. 0091-3286/96/$6.00.

SAMSUNG-1038

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 24 Mar 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

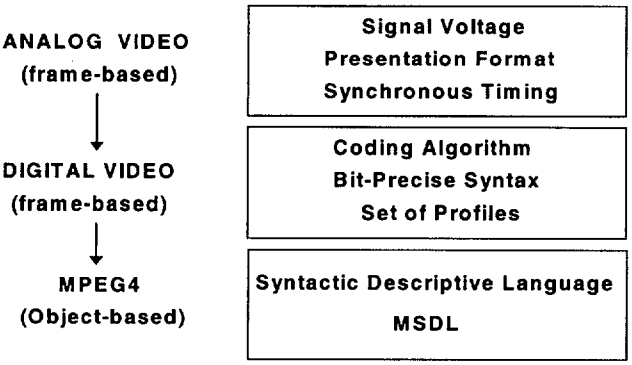


Fig. 1 Evolution of audiovisual standards.

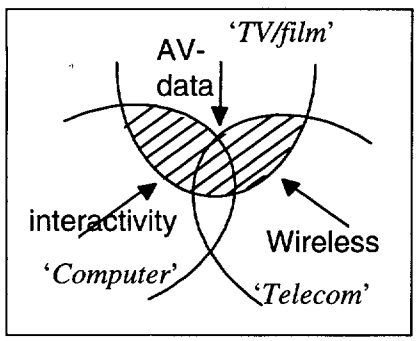


Fig. 2 MPEG4 addresses applications in the shaded area.

Both of these developments, and the projected new capabilities of "interactive TV," suggest a trend from purely linear TV toward "nonlinear TV" from both a content and a technology point of view. MPEG4 is the storage and transmission standard for nonlinear TV.

2 MPEG4 Functionalities and Applications

In the near future, convergence of the worlds of computers and consumer audiovisual products will be accompanied by advances and diversification of communications. High-bandwidth wide-area networks, wireless networks, and huge audiovisual databases offer the potential for an explosion in applications on one hand, and a crisis in data management on the other hand. MPEG4 will tackle the data representation part of this problem and will be an audiovisual coding standard allowing for interactivity, high compression, and universal accessibility. This is illustrated in Fig. 2.

To accomplish this, MPEG4 will address a set of eight new and improved functionalities (in addition to the standard functionalities). These are shown in Table 1, together with examples of applications. This table has been extracted from the MPEG4 Proposal Package Description (PPD) document.[1]

Within this suite of functionalities, several trends are clear. First, there is a prevalence of content-based capabilities. We want to be able to access audiovisual objects, manipulate them, and present them in a highly flexible way. We want to do this at the coded data level. Second, interactivity is important. Third, we want to have these capabilities for both natural and synthetic audiovisual objects, and fourth, we want these capabilities no matter where we are. A comparison between the conventional production, distribution and presentation environment and the new MPEG4-enabled environment is shown in Fig. 3.

In the conventional environment, the scene is fully composed before being fixed into a standard video format. Three-dimensional objects are reduced to two-dimensional projections, the script is fixed, and the presentation is essentially constrained to the spatial and temporal format of the input. In contrast, if audio and video objects are coded at the source, together with (optionally) the script, then flexibility is maintained. If desired, the three-dimensional nature of the objects can be preserved, the script may consist only of a behavior model, and the format of the presentation can be independently chosen. In summary, MPEG4 will not code video as 25/30 frames/s of PAL/NTSC resolution, or audio as 48 ksamples/s. MPEG4 will code audio and video objects at their native resolution; then, by selecting objects and scaling their resolution, a wide variety of presentations can be made.

The other important factor in the applications addressed by MPEG4 is the heterogeneous networking environment we can already foresee, and in particular wireless communication. Mobile applications are compelling, whether for videophone communications or remote access to audiovisual databases. This introduces requirements for tolerance of noisy environments, varying bandwidths, and varying degrees of decoder resources, including display resolution, computation resources, and battery power. MPEG4 will address the problem of channels in which the residual noise is not zero, and provide fine-grained scalability for constrained bit rate and decoder resources.

3 Structure of MPEG4

MPEG4 will not standardize a single algorithm. No such algorithm can exist in view of the range of functionalities and applications to be addressed. Also, there is no need to standardize a single algorithm, when cost-effective systems can be built to switch between algorithms, or even learn new ones. The latter capability also permits future advances in coding techniques to be included in the standard. So MPEG4 will establish an extendible set of coding *tools,* which can be combined in various ways to make *algorithms,* and the algorithms can be customized for specific applications to make *profiles.* This is illustrated in Fig. 4.

Tools, algorithms and profiles are coding *objects* and consist of an independent core and a standard interface. The standard interface guarantees the coding objects can interwork, and the independent core permits proprietary techniques to be invented and made available within the standard. This is analogous to the situation in computer software applications, where independent software vendors (ISVs) can develop and market products that are guaranteed to run, provided they are compatible with the application program interface (API). In this sense, MPEG4 will be the API for the coded representation of audiovisual data.

The "glue" that binds the coding objects together is the MPEG4 Syntactic Descriptive Language (MSDL) which comprises several key components: first, definition of the coding object interface noted above; second, a mechanism to combine coding objects to construct coding algorithms and profiles; and third, a mechanism for downloading new coding objects. The current thinking is for coded data objects

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 24 Mar 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Table 1 Description of functionalities.

Functionality	Description	Example Uses
Content Based Interactivity		
Content-Based Multimedia Data Access Tools	MPEG-4 shall provide data access based on the audio-visual content by using various accessing tools such as indexing, hyperlinking, querying, browsing, uploading, downloading, and deleting.	• Content-based retrieval of information from on-line libraries and travel information databases.
Content-Based Manipulation and Bitstream Editing	MPEG-4 shall provide an 'MPEG-4 Syntactic Description Language' and coding schemes to support content-based manipulation and bitstream editing without the need for transcoding. The MSDL shall be flexible enough to provide extension for future uses.	• Interactive home shopping • Home movie production and editing • Insertion of sign language interpreter or subtitles; • Digital effects (e.g. fade-ins);
Hybrid Natural and Synthetic Data Coding	MPEG-4 shall support efficient methods for combining synthetic scenes or objects with natural scenes or objects (e.g. text and graphics overlays), the ability to code and manipulate natural and synthetic audio and video data, and decoder-controllable methods of compositing synthetic data with ordinary video and audio, allowing for interactivity.	• Animations and synthetic sound can be composited with natural audio and video in a game. • A viewer can translate or remove a graphic overlay to view the video beneath it; • Graphics and sound can be 'rendered' from different points of observation;
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 audio-visual sequence. This includes 'conventional random access at very low bit rates.	• Audio-visual data can be randomly accessed from a remote terminal over limited capacity media. • A 'fast forward' can be performed on a single AV object in the sequence.
Compression		
Improved Coding Efficiency	For specific applications targeted by MPEG-4, MPEG-4 shall provide subjectively better audio-visual quality at comparable bit-rates compared to existing or other emerging standards.	• Efficient transmission of audio-visual data on low-bandwidth channels. • Efficient storage of audio-visual data on limited capacity media, e.g. magnetic disks.
Coding of Multiple Concurrent Data Streams	MPEG-4 shall provide the ability to code multiple views/ soundtracks of a scene efficiently and provide sufficient synchronization between the resulting elementary streams. For stereoscopic video applications, MPEG-4 shall include the ability to exploit redundancy in multiple viewing or hearing points of the same scene, permitting joint coding solutions that allow compatibility with normal audio and video as well as the ones without the compatibility constraint.	• Multimedia entertainment, e.g. virtual reality games, 3D movies; • Training and flight simulations; • Multimedia presentations and education;
Universal Access		
Robustness in Error-Prone Environments	MPEG-4 shall provide an error robustness capability to allow access to applications over a variety of wireless and wired networks and storage media. Sufficient error robustness shall be provided for low bit-rate applications under severe error conditions (e.g. long error bursts).	• Transmitting from a database over a wireless network; • Communicating with a mobile terminal. • Gathering audio-visual data from a remote location.
Content-Based Scalability	MPEG-4 shall provide ability to achieve scalability with a fine granularity in content, quality, (e.g. spatial resolution, temporal resolution), and complexity. In MPEG-4, these scalabilities are especially intended to result in content-based scaling of audio-visual information.	• User or automated selection of decoded quality of objects in the scene; • Database browsing at different content levels, scales, resolutions, and qualities;

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 24 Mar 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

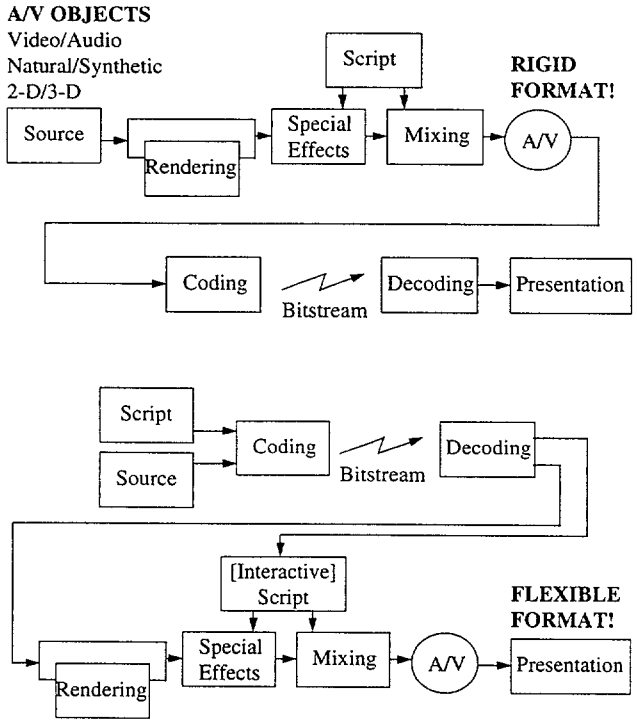


Fig. 3 Coding in the production, distribution and presentation process: (a) conventional video environment, (b) MPEG4 environment.

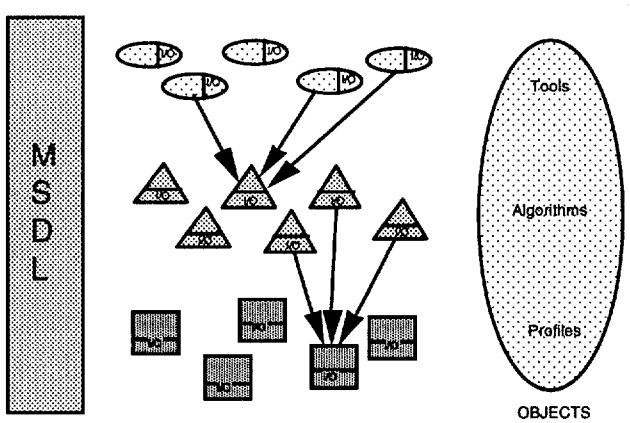


Fig. 4 Structure of MPEG4.

themselves to be described by the coding objects. Collectively these components define a syntax for MPEG4, and the fourth component of MSDL is a set of rules for parsing this syntax. The MSDL is described more fully in the document "Requirements for the MPEG4 SDL."[2]

This structure implies a multiphase transmission of MPEG4 data. At the beginning of an exchange between a user and a database or between two users, there is a configuration phase, during which the coder and decoder determine the coding objects to be used, their configuration, and whether or not both of them have all the required objects. If not, there may be a learning phase, during which coding objects are downloaded. Finally, there is the transmission phase for the communication of the data, which must of course be bit-efficient. The process is illustrated in Fig. 5.

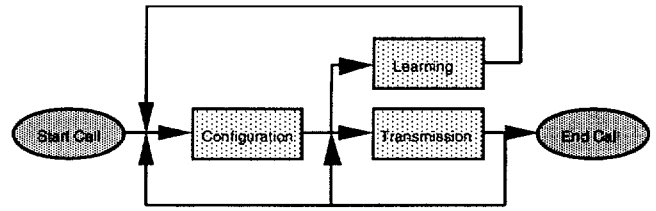


Fig. 5 MPEG4 communication phases.

4 Practical MPEG4 Systems

The general structure of MPEG4 is very flexible, and allows extension of capabilities in the future, but a generic implementation of the standard would produce neither a cost-effective solution to any specific application, nor an efficient solution during the setup phases. One would expect that any specific system would be implemented to address a very small number of very well-defined applications, and that the setup phase would consist only of specifying a default configuration. The learning phase and the full configuration phase should occur only rarely—for example, in the way that software upgrades are done today. However, some help is needed to establish the default configurations; otherwise there may still be too frequent and irritating setup scenarios. For example, if different geographic regions established different default tool sets, then establishing an MPEG4 dialog across the region boundaries would necessitate a setup phase. In the worst case, a failure to agree on which of two competitive tools to use could occur. As a result, the current thinking is for MPEG to standardize an initial set of tools, toward which the implementers and users will gravitate. Most notably, since MPEG2 is definitively the highest-performance coding algorithm for linear TV, the MPEG2 tools will be in the MPEG4 toolset.

At least in the short term, the least expensive implementation for a dedicated application will continue to be dedicated, fixed-function hardware. This is not inconsistent with MPEG4, with the proviso that there be a piece of logic capable of conducting the setup-phase dialog. As a practical matter, virtually all MPEG decoders today are architected with an embedded controller that parses the syntax and implements the higher coding layers in software. It is thus conceivable that an existing MPEG1 or MPEG2 VLSI decoder might become an MPEG4 decoder merely by adding MPEG4 "front-end" microcode.

5 MPEG4 Program of Work

The schedule for MPEG4 is shown in Fig. 6. The principal components of the workplan are development of the syntax, development of the coding tools, and generation of the standard document. The syntax is being developed over the next two years, and currently there is a call for proposals for elements of the MSDL.[3] It is planned for a first draft to be available in January 1996, in time to be used with the coding tools selected from a first round of tests.[3,4] Experience gained from that use will help define the second version of the MSDL, which will occur in November 1996. The syntax will be frozen in July 1997.

Selection and evaluation of the coding tools will be done using the successful competitive-phase—collaborative-phase

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 24 Mar 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

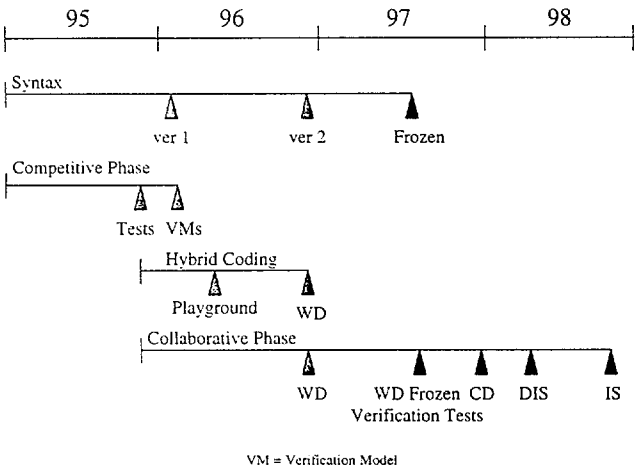


Fig. 6 MPEG4 workplan.

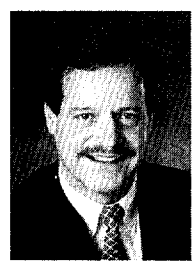
approach used in MPEG1 and MPEG2. Tests are the conclusions of the competitive phases, and past experience has been that no outright winner appears; rather, there are several approaches that can contribute to development of the best result. Convergence to this result is performed by the mechanism of *verification models:* a precisely defined software simulation is used to perform experiments on promising coding techniques. So-called *core experiments* are conducted, in which multiple parties perform the same experiment under identical conditions to evaluate a particular technique. Candidate tools for conventional video and audio have been developed in 1995. Candidate tools for hybrid natural/synthetic audiovisual data will be developed in 1996, after which a Working Draft will be established. The hybrid work will be conducted by constructing a ''playground'' comprising a database of 2-D and 3-D natural and synthetic audiovisual objects. Proponents will be invited to run scripted experiments in this playground, with the results being evaluated for coding efficiency. The emerging standard will be frozen in July 1997, and final approval is expected for November 1998.

6 Conclusions

MPEG has established the world standard for coded audiovisual data, offering the highest available quality and practical implementation. In the emerging MPEG4 standard, this work will be extended to exciting new applications such as interactive TV, and new operational environments such as wireless networks.

References

1. ''MPEG4 proposal package description (PPD),'' revision 3, ISO/IEC JTC1/SC29/WG11 N0998, Jul. 1995.
2. ''Requirements for the MPEG4 SDL'' (draft revision 2.0), ISO/IEC JTC1/SC29/WG11 N1022, Jul. 1995.
3. ''MPEG4 call for proposals,'' ISO/IEC JTC1/SC29/WG11 N0997, Jul. 1995.
4. ''MPEG4 test evaluation procedures document,'' ISO/IEC JTC1/SC29/WG11 N0999, Jul. 1995.



Cliff Reader has more than 20 years experience in digital video coding, image processing, and real-time digital video systems design. His doctoral thesis was written while in residence at the University of Southern California Image Processing Institute, 1971–1973. His doctorate was awarded by the University of Sussex, England in 1974. He has worked in the fields of reconnaissance imaging, medical imaging, and earth resources imaging, and most recently has concentrated in the emerging field of consumer digital video. He has published numerous papers. Since 1990 he has been an active member of the ISO/IEC MPEG community, providing technical contributions to MPEG1 and MPEG2. He was the head of the U.S. delegation for two years, was chief editor of the MPEG1 standard, and is now subcommittee chairman for the Applications and Operational Environments group, better known as MPEG4. He is also the technical expert on MPEG intellectual property, assisting in the establishing of the MPEG Patent Pool. Dr. Reader has been with Samsung Electronics for the past two years as Associate Director for Strategic Planning.

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 24 Mar 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use