MPEG and Multimedia Communications

Leonardo Chiariglione

***Abstract*—Digital television is a reality today, but multimedia communications, after years of hype, is still a catchword. Lack of suitable multi-industry standards supporting it is one reason for the unfulfilled promise.**

The MPEG committee which originated the MPEG-1 and MPEG-2 standards that made digital television possible is currently developing MPEG-4 with wide industry participation. This paper describes how the MPEG-4 standard, with its network-independent nature and application-level features, is poised to become the enabling technology for multimedia communications and will therefore contribute to solve the problems that are hindering multimedia communications.

***Index Terms*—Audio coding, MPEG, multimedia, video coding, VRML.**

I. Introduction

AFTER ten years since the word "multimedia" has entered the techno-vocabulary, five years of "convergence" hype and 2-1/2 years of digital video, we are still struggling to make multimedia communications happen. The reasons of this stalemate are manifold. Here are some of them.

- The terms of the *convergence* issue are not well posed. It is not the *businesses* of telecommunications, entertainment, and computers that are going to converge, but the traditional barriers inherited by the three businesses in the content production and packing, information transport and processing, and user equipment domains that are going to disappear. It makes technical and business sense if the content, transport, and equipment industries converge, less if the mentioned businesses do. The sooner these traditional barriers are removed, the sooner multimedia communications will happen.
- *Digital television* is a technology for better exploitation of bandwidth when used to transmit television signals. The current usage of MPEG-1 and MPEG-2 is restricted to digital television. The few openings toward multimedia communications that were embedded in the standard have been emasculated by the all-out hijacking of the technology by a particular industry.
- There is an antithesis created by the traditional, slow, bottom-up, generic, broadband, network *service*-driven model proper of telecommunication operators and the swift, pragmatic, specific, narrowband, *application*-driven approach proper of the Internet. The bit-transportation industry has invested in the former but the latter seems to provide many of the applications that were intended to be supported by the former.

Communications mean standards, but the production of standards for multimedia communications is beset by the problem that the many industries having a stake in it have radically different approaches to standardization.

A solution to this problem is offered by MPEG with its successful development of the multi-industry MPEG-1 and MPEG-2 standards, even though it is recognized that the new task is vastly more complex than that of the previous two standards. Such MPEG-established standardization principles as "not systems but tools," "one functionality—one tool," "relocatability of tools," "specify the minimum," "*a priori* standardization," "stick to the deadline," etc., if not adopted in practice by other standards bodies, are at least becoming widely known, discussed, and their positive implications are gradually being appreciated in the standards world. The profile/level approach that complements the principles above combines the need of generic technology specifications with the application-specific needs of different industries.

MPEG-4, the current standardization project of MPEG, combines some of the typical features of other MPEG standards with new ones coming from existing or anticipated manifestations of multimedia:

- independence of applications from lower-layer details, as in the Internet paradigm;
- technology awareness of lower layer characteristics (scalability, error robustness, etc.);
- application software downloadability, as in Java and the network computer paradigm;
- reusability of encoding tools and data;
- interactivity not just with an integral audio-visual bitstream but with the individual pieces of information within it called "audio-visual (AV) objects" as in the Web paradigm;
- possibility to hyperlink and interact with multiple sources of information simultaneously as in the Web paradigm but at the AV object level;
- possibility of hyperlinking as in the Web paradigm but at the AV object level;
- capability to handle natural/synthetic and real-time/nonreal-time information in an integrated fashion;
- capability to composite and present information according to user's needs as in virtual reality modeling language (VRML) and computer graphics paradigm in general.

Backward compatibility with MPEG-1 and MPEG-2 will be ensured by the toolkit nature of the standard.

MPEG-4 will become International Standard in November 1998. It can be expected that MPEG-4 will become the

Manuscript received July 10, 1996; revised October 1, 1996. This paper was recommended by Guest Editors Y.-Q. Zhang, F. Pereira, T. Sikora, and C. Reader.

The author is with Multimedia and Video Services, CSELT, I-10148 Torino, Italy.

Publisher Item Identifier S 1051-8215(97)00939-7.

enabling technology for multimedia communications as much as MPEG-2 has become the enabling technology for digital television.

Section II will try to clear the ground from the convergence hype and will identify which parts of the industries are candidate for convergence under the condition of existence of common multi-industry communication standards. Section III will assess the difficulty of the task of producing such common standards in view of the different approaches to standardization of the different industries. Sections IV and V will claim that MPEG, with its successful development of the multi-industry MPEG-1 and MPEG-2 standards, is uniquely qualified to the task and will briefly outline the technical content of the two standards. Sections VI and VII will clarify the scope of applicability of MPEG-1 and MPEG-2, i.e., digital television and will identify the need of a new standard—MPEG-4—to satisfy the requirements coming from new information interaction/consumption paradigms. Section VIII goes into some of the technical details of the MPEG-4 standard, and Section IX lists the additional features that MPEG-4 will need to avoid and some of the problems encountered with the customization of MPEG-2 made by specific industries.

II. About Multimedia

After years of multimedia hype, there is no sign that multimedia communications will happen in the way media gurus had anticipated, i.e., by convergence of telecommunications, entertainment, and computers all adopting digital technology. This is not happening as much as the professions of barber, butcher, and cobbler have not moved a single inch to a convergence point through the millennia in spite of all sharing the common "knife" technology. What is happening is movie makers buying broadcasting companies, telcos buying CATV companies, consumer electronics companies buying movie makers, etc. To do this digital technology convergence is redundant; you need fat wallets, complacent boards of directors, and patient shareholders.

Digital technologies have many benefits, but the real plus is the possibility to replicate in a more economic and compact way the different system components that technologies specific of a field have made possible so far. Some examples follow.

- The vinyl disc, including its predecessors the hard disc and the phonograph cylinder, has existed for over 100 years, but now the compact disc is used by hundreds of million people;
- Analog speech has existed for almost 100 years but now A-law/μ-law pulse code modulation (PCM) is used in the network by billions of people;
- Analog satellite television has been in operation for two decades, but now digital satellite television is being watched by millions of people.

Now, try to ask the layman to tell you the difference between the analog and the digital version! People buy benefits, even if only perceived, and not features.

I am not an unbeliever in convergence, which is part of life as much as divergence, but certainly not of the mentioned *businesses*. The first thing we must do, if we want to have any chance of understanding, forecasting, and, if possible, shaping what is going to happen, is to acknowledge that the three industries of entertainment, telecommunications, and computers do not provide the right dimensions to study the phenomenon. "Entertainment" usually represents a vertical business such as terrestrial broadcasting that produces content and takes care of its delivery up to the customers' homes. "Telecommunications" is another vertical business spanning all communication layers. "Computers" are an inextricable mixture of hardware and software, an underlying technology that is used everywhere, in the telecommunication system as well as in the user devices.

TABLE I
Examples of Integration of Industries

	Content	Transport	Equipment
Terrestrial TV	X	X	
CATV		X	X
Satellite TV	X	X	
Telecommunications		X	X
Movies	X	X	
Consumer Electronics		X	X
Video games	X		X

Better axes are provided by "Content," "Transport," and "Equipment." "Content"—the message—is what matters to the user who foots the entire bill and therefore justifies the existence of the entire system, "Transport" is what is needed to deliver "Content" to people who want it, and "Equipment"—the user device—is what is needed to enable the human user to interact with the system and to convert "Content" into human-consumable form. There are different types of "Content:" movies, TV programs, news, telephone calls, and many ways to pack content in a way that makes users more prone to consume; different types of "Transport:" radio channel, cable, twisted pair at the physical level and emerging ones as middleware; and an almost infinite variety of "Equipment."

The many businesses ("industries") that play a role in making the system work are present in one or more of the three domains, e.g., the broadcasting industry typically integrates content and transport, the community antenna television (CATV) industry transport and equipment, and the video games industry content and equipment.

Table I gives some examples, possibly different from an environment to another, of how the different industries (first column) integrate within themselves the content, transport, and equipment components.

The convergence case can be made, even though I do not personally think the businesses will converge, nor that there is a cogent need for it. But this will not happen because the industries will decide to abandon the technologies proper to their businesses and convert to digital technologies, something they have been doing for a long time (see the examples above) while in search of rationalizing their way of doing business, but because they will decide to do the conversion in such a way that the communication standards of one industry will be compatible with those of the other industries.

And this is a monumental task, seeing the diverging attitudes of the different industries toward standardization that are described in the next section.

III. Multimedia and Standardization

Communications require standards that define the meaning and the syntax given to information at the source when it reaches the destination. Starting from the Morse alphabet, communication standards have become increasingly more sophisticated and the different industries that have been created in the process have developed considerably diverging attitudes.

- The *telecommunications industry* bases its standards on the original consideration that impedance mismatches in passing from the wires of one telephone company to another is not the right way to promote communications, i.e., the telephone companies' business. Even the famous A-law/μ-law (digital speech) dichotomy can be justified, if not praised, for its farsightedness, considering that at the time (the 1960's) digital speech was just a means to optimize transmission in the network, not something to be offered as an end-to-end service to customers.
- The *movie industry* has settled on a small number of film formats each characterized by different audio-visual performance levels. The hardware and software motion picture industry agreed that having the possibility to project a movie everywhere in the world was good for everybody's business.
- The *radio industry* took the commendable approach of defining standards of worldwide reach, but its daughter, the *television industry*, has defined its standards in such a way that users can only watch programs coming from a certain source. Notwithstanding a good 405 lines at 50 Hz television system having been deployed in the United Kingdom in the late 1930's, in the early 1940's the United States established their 525/60 system that improved on the U.K. system by some 20%, and a few years later, Europe established its own 625/50 television system that did not improve bandwidth over National Television Standards Committe (NTSC) at all ($625 \times 25 \simeq 525 \times 30$). With the addition of color to the monochrome signal, the number of "national paths" to television increased dramatically with NTSC, phase alternate line (PAL), Séquentiel Mémoire (SECAM), and their almost countless variants.
- The *CATV industry*, sitting in between television and telecommunications, and by definition a local business, has a schizophrenic attitude toward standards depending on the country in which it operates.
- The *consumer electronics industry* (mostly recording) has taken the most straightforward application of the definition of standards: a freely entered agreement between a manufacturer and a user to sell/buy a certain piece of equipment with which users can play back audio or video from media that are specific to the type of equipment purchased ("format") from a third party that agrees to produce content in that format.
- The *computer industry* takes an attitude very similar to consumer electronics, but considerably more articulated. The purchase of a computer is a freely entered agreement between a manufacturer and a user to provide hardware and some layers of software on top of it so that high-level applications can be developed or purchased from the manufacturer or a third party.
- In the *electronic games industry* the purchase of an electronic game is a freely entered agreement between a manufacturer and a user to sell/buy hardware and software (the latter possibly from a third party) that runs exclusively on the hardware.

So far the different industries have been diverging, but multimedia communications necessarily need some convergence zone that can only be achieved by standardization in key areas. Putting every stake holder together and producing communication standards accepted by all is a big task. Still MPEG succeeded in doing that for its first-generation standards MPEG-1 and MPEG-2, particularly the latter which from now on will be used to indicate both.

IV. The MPEG Approach to Standardization

With MPEG-2, MPEG has produced common audio-visual coding standards that can be used by all industries mentioned in Section III. This has enabled sharing of cost, acceleration of digital audio-visual technology development, and, more fundamentally for users, flow of content unrestricted by built-in technical barriers. If convergence will happen, it will be because all industries will be willing to adopt this kind of single representation of information, shared by all industries.

It is worth looking at the way in which MPEG has operated in its eight years of activity and try to rationalize what has been a successful approach to standardization serving the needs of multiple industries for the general multimedia communication case.

A. Stick to the Deadline

No business can survive if work is done living day by day. This is, unfortunately, the practice of some standards committees. They are in charge of producing something (the something being often itself loosely defined) without a date attached for delivering an output (the standard) or with a date that is just a reference. It would be as if a company promised its customers to deliver something sometime.

Standards are the goods that standards committees sell their customers. As for a company, the goods, of course, have to be of high quality, have to be according to the specification issued by the customers, but, foremost, they have to be delivered by the agreed date.

Standards are not novels, standards are the technology that enables companies to make products (those sold to end users). If a company makes a plan to go to the market by a certain date with a certain product that requires a certain technology, and makes the necessary investments for it, the company—the buyer vis-à-vis the standards committee—is not going to be happy if the standards committee—the supplier vis-à-vis the company—at the due date reports that they are "behind schedule."

MPEG has a strict work plan that specifies, for all parts of a standard, the times when the levels of Working Draft, Committee Draft, Draft International Standards, and International Standards are reached. So far there have been occasional minor slips at *intermediate* steps but no delay in reaching International Standard status compared to the planned dates.

B. A-Priori Standardization

Everybody agrees that standards should be issued by standards bodies for which making standards is the *raison d'être*, however, the inability of many standards committees to deliver on time has forced companies to take shortcuts, so-called "industry standards." These private specifications, possibly endorsed by a few other companies, are often submitted to a standards committee for ratification.

The main problem with such an approach is that the standards committee then becomes a place where discussions cease to be technical, i.e., definition of a technology, but instead become commercial. The issues discussed are no longer those aimed at making a good technical standard but definition of terms of exploitation, fitness of the technology to the current plans of companies, etc. Of course, there is nothing wrong with technology deals between companies, but this is wrong if it is done in a standards committee. MPEG instead takes a very clear attitude.

1) Before industries make commitments, the maturity of a technology for standardization is identified.
2) A Call for Proposals, to which interested companies are free to respond, is usually produced.
3) In all cases the technology is standardized by MPEG experts.

So far, MPEG has successfully applied this principle. Standardization items have been identified well in advance so that it can be claimed that no MPEG standard has endorsed an "industry standard." It must be borne in mind that MPEG standards do not specify complete systems. It is therefore possible that "industry standards" are needed alongside with MPEG standards to make complete products.

C. Not Systems But Tools

The principles described above, applicable to standardization in general, require further ingenuity when they are to be applied to the production of standards that serve multiple industries.

Industries, by definition, need to make vertically integrated specifications in order to make products that satisfy some needs. Audio-visual decoding may well be a piece of technology that can be shared with other communities, but in the event industries need to sell a satellite receiver or a video CD player, these require an integrated standard. But if different industries need the same standard, quite likely they will have different end systems in mind. Therefore, only the components of a standard, the "tools," as they are called in MPEG, can be specified in a joint effort.

The implementation of this principle requires the change of the nature of standards from "system" standards to "component" standards. Industries will assemble the tool specification from the standards body and build their own product specification.

If "tools" are the object of standardization, a new process must be devised to produce meaningful standards. The following sequence of steps has been found to be practically implementable and to produce the desired result.

1) Select a number of target applications for which the generic technology is intended to be specified.
2) List the functionalities needed by each application.
3) Break down the functionalities into components of sufficiently reduced complexity so that they can be identified in the different applications.
4) Identify the functionality components that are common across the systems of interest.
5) Specify the tools that support the identified functionality components, particularly those common to different application.
6) Verify that the tools specified can actually be used to assemble the target systems and provide the desired functionalities.

Still, industry needs some guidance. It is therefore advisable that certain major combinations of tools be specified as normative, making sure that these are not application-specific, but functionality-specific. These standardized sets of tools have been called "profiles" in MPEG-2 Video.

D. Specify the Minimum

In some environments it is very convenient to add to a standard those nice little things that bring a standard nearer to a product specification. This is, for instance, the case of industry standards or when standards are used to enforce the concept of "guaranteed quality" so dear to broadcasters and telecommunication operators because of their "public service" nature.

This practice must be abandoned when a standard is to be used by multiple industries. Only the minimum that is necessary for interoperability can be specified. Going beyond this border line requires a separate agreement involving all participating industries.

E. One Functionality—One Tool

A standard is an agreement to do certain things in a definite way and in abstract terms everybody agrees that tools should be unique. Unfortunately, when people working for a company are in a standards committee, the determination dwindles if they see competing technologies to their company's prevail in the favors of the committee. The usual outcome of a dialectic battle lasting anywhere from an hour to three years is compromising the intellectually accepted principle of one functionality—one tool and, *voilà*, "options" come in. Because of too many signaling options, it took ten years for European integrated services digital network (ISDN) to achieve a decent level of interoperability between different telecommunications operators and, within the same operator, between equipment of different manufacturers. Because of too many options, many standards were stillborn because the critical mass that would have justified the necessary investments by the industry could not be reached.

What constitutes a tool, however, is not always obvious. Single channel and multichannel audio or conventional television and high definition television (HDTV) are components needed in many systems. Defining a single "tool" that does the job of coding both single channel and multichannel audio or conventional television and HDTV may be impractical because the technology has to be designed and manufactured to do things to an extent that in some cases is not needed. The "profile/level" philosophy successfully implemented by MPEG provides a solution: within a single tool one may define different "grades," called "levels" in MPEG.

F. Relocation of Tools

When a standard is defined by a single type of industry there is generally agreement on where a certain functionality resides in the system. In a multi-industry environment this is not possible. Take the case of encryption. Depending on which is your role in the audio-visual distribution chain, you will like to have the encryption function located where it serves your place in the chain best, because encryption is an important value-added function. If the standard endorses your business model, you will adopt the standard, if it does not, you will antagonize it.

Not only must the technology be defined in a generic way, but also in such a way that the technology can be located at different points in the system.

G. Verification of Standard

Once the work is nearing completion it is important to make sure that the work done does indeed satisfy the requirements ("product specification") originally set. MPEG does that through a process called "Verification Tests" with the scope of ascertaining how well the standard produced meets the specification.

V. A Guided Tour to MPEG-1 and MPEG-2

The Moving Picture Coding Experts Group (MPEG) was established in January 1988 with the mandate to develop standards for coded representation of moving pictures, audio, and their combination. It operates in the framework of the Joint ISO/IEC Technical Committee (JTC 1) on Information Technology and is formally WG11 of SC29.

A. MPEG-1

The first standard developed by the group, nicknamed MPEG-1, was the coding of the combined audio-visual signal at a bit rate around 1.5 Mb/s. This was motivated by the prospect that was becoming apparent in 1988 to store video signals on a compact disc with a quality comparable to VHS cassettes.

In 1988, coding of video at such low bit rates had become possible thanks to decades of research in video coding algorithms. These algorithms, however, had to be applied to subsampled pictures—a single field from a frame and only half of the samples in a line—to show their effectiveness. Also, coding of audio, as separate from speech, could rely on R&D work that allowed reduction by 1/6 of the PCM bit rate, typically 256 kb/s for a stereo source, with virtual transparency. Encoded audio and video streams, with the constraint of having a common time base, were combined into a single stream by the MPEG system layer.

TABLE II
MPEG-2 Profile/Level Table

	Simple Profile	Main Profile	SNR Scalable Profile	Spatially Scalable Profile	High Profile	4:2:2 Profile
High Level		X			X	
High-1440 Level		X		X	X	
Main Level	X	X	X		X	X
Low Level		X	X			

MPEG-1, formally known as ISO/IEC 11172, is a standard in five parts. The first three parts are Systems, Video, and Audio, in that order. Two more parts complete the suite of MPEG-1 standards: Conformance Testing, which specifies the methodology for verifying claims of conformance to the standard by manufacturers of equipment and producers of bitstreams, and Software Simulation, a full C-language implementation of the MPEG-1 standard (encoder and decoder).

Manifold have been the implementations of the MPEG-1 standard: from software implementations running on a PC in real time, to single boards for PC's, to the so-called Video CD, etc. The last product has become a market success in some countries: in China alone, 2 million Video CD decoders were sold in 1996 and the number is expected to double the following year.

B. MPEG-2

The second standard developed by MPEG, nicknamed MPEG-2, has the title "Generic Coding of Moving Pictures and Associated Audio." Work on this standard could commence as early as July 1990 because:

- at that time the technical foundations of MPEG-1 had already been laid down;
- extrapolations from the results of MPEG-1 promised a quality comparable to composite TV at about four times the typical MPEG-1 bit rate;
- there were expectations that VLSI technology would be ready to implement a video decoder handling full-size pictures at bit rates up to 10 Mb/s.

Unlike MPEG-1, basically a standard to store moving pictures on a disk at low bit rates, the much larger number of applications of the MPEG-2 standard forced MPEG to develop and implement the "toolkit approach" described above for MPEG-2 Video. Different coding "tools" serving different purposes were developed and standardized. Different assemblies of tools—called "profiles"—were also standardized and could be used to serve different needs. Each profiles had in general different "levels" for some parameters (e.g., picture size). Table II gives the current situation of MPEG-2 Video Profiles and Levels.

MPEG-2 Audio is an extension of MPEG-1 Audio to the multichannel case. This means that an MPEG-1 Audio decoder can decode two channels of the MPEG-2 stream and an MPEG-2 Audio decoder can decode an MPEG-1 Audio stream as if it were an MPEG-1 Audio decoder.

As for MPEG-1, the systems part of the MPEG-2 standard addresses the combination of one or more elementary streams of video and audio as well as other data into single or multiple streams which are suitable for storage or transmission. Two such combinations are specified: Program Stream and Transport Stream.

The Program Stream is analogous to MPEG-1 Systems Multiplex. It results from combining one or more Packetized Elementary Streams (PES), which have a common time base, into a single stream. The Program Stream is designed for use in relatively error-free environments and is suitable for applications which may involve software processing.

The Transport Stream combines one or more PES's with one or more independent time bases into a single stream. Elementary streams sharing a common time base form a program. The Transport Stream is designed for use in environments where errors are likely, such as storage or transmission in lossy or noisy media.

MPEG-2, formally known as ISO/IEC 13818, is also a multipart standard. The first five parts have the same function as the corresponding parts of MPEG-1.

MPEG-2 has been a very successful standard, pieces of equipment that claim conformance to it have been produced by the millions, receivers for digital satellite broadcasting being the most popular. More application domains are anticipated, such as digital receivers for CATV or digital versatile disc (DVD), a new generation of compact disc capable of playing back MPEG-2 bitstreams at a higher and variable bit rate and for a longer time than standard CD's.

ITU-T has collaborated with MPEG in the development of MPEG-2 Systems and Video which have become ITU-T Recommendations for the purpose of broadband visual communications. This means that the same physical document has the value of ISO standard and ITU-T Recommendation.

C. Other MPEG-2 Functionalities

MPEG-2 provides support to a number of technical features, the most important of which are support to content addressing, encryption, and copyright identification.

- The MPEG-2 Systems Transport Stream has been designed so that it can be used to carry a large number of television programs. For this reason it provides support to signal the content of the programs by means of tables that describe which program can be found where. This specification has been extended by regional initiatives to identify more features, such as the nature of the program, the scheduled time, the interval between starting times, etc.
- Copyright protection and management are important features that a system designed to carry audio-visual information must support. MPEG-2 Systems defines two special streams called encryption control message (ECM) and encryption management message (EMM) that carry information that can be used to decrypt information carried by the MPEG-2 Transport Stream, if this has been encrypted. The encryption system itself is not specified by MPEG.
- MPEG-2 Systems provides support for the management of audio-visual works copyrighting. This is done by means of a copyright descriptor that identifies the society that manages the rights of that particular audio-visual work followed by a field that gives the identification number of the work, as assigned by the society. This information enables, for instance, the monitoring of the flow of copyrighted work through a network.

D. Other Parts of MPEG-2

MPEG-2, as described above, provides the enabling technology for a variety of television-based applications, such as satellite broadcasting and CATV, which can now send an average of five times more programs on the same delivery medium if MPEG-2 encoded programs are carried on top of medium-specific modulation schemes.

Other applications, however, require a standardized terminal-to-server protocol to provide a complete working system. This is, for instance, the case when the user needs to interact with the source to select the content he or she wishes to see and hear, such as in video on demand and home shopping.

Part 6 of MPEG-2, titled Digital Storage Media Command and Control (DSM-CC), an International Standard since July 1996, is the specification of a set of protocols which provide the control functions and operations specific to managing MPEG bitstreams. These protocols may be used to support applications in both stand-alone and heterogeneous network environments. In the DSM-CC model, a stream is sourced by a server and delivered to a client, both considered to be users of the DSM-CC network. DSM-CC defines a logical entity called the Session and Resource Manager (SRM) which provides a logically centralized management of the DSM-CC Sessions and Resources.

Part 7 of MPEG-2 is the so-called MPEG-2 Advanced Audio Coding (AAC). The main feature for this standard is its non-backwards compatibility (NBC) with MPEG-1 Audio which provides nearly the same performance as MPEG-1 Audio at one-half the bit rate. The need for such a standard arose from te consideration that the backwards compatability built in MPEG-2 Audio is an important service feature for many applications, such as television broadcasting. This compatibility, however, entails a degree of quality penalty that other applications not requiring backwards compatability need not pay. In November 1996, part 7 has reached the status of Draft International Standard. International Standard status will be reached in April 1997.

Part 8 of MPEG-2 was originally planned to be coding of video when input samples are 10 b in order to provide room for postprocessing. Work on this part was discontinued when the professional video industry that had requested the standard eventually shifted its interests to other domains.

Part 9 of MPEG-2, titled Real-Time Interface (RTI), an International Standard since July 1996, provides a specification for a real-time interface to Transport Stream decoders which may be utilized for adaptation to all appropriate networks carrying Transport Streams. RTI can be used to achieve equipment-level interoperability in consumer electronic, computer, and other domains because it enables the building of network adaptation layers which are guaranteed to provide the required performance, and simultaneously, it enables the building of decoders which are guaranteed to have appropriate behavior of buffers and timing recovery mechanisms.

Part 10 of MPEG-2 is the Conformance Testing of DSM-CC and is still under development.

Other current MPEG activities concern the definition of other MPEG-2 Video Profiles. The 4 : 2 : 2 Profile, completed in January 1996, provides a response to users of professional video equipment and services who were keen to exploit the existing consumer-electronics grade MPEG-2 Video technology for professional applications. The Multiview Profile, completed in November 1996, uses existing video coding tools for the purpose of providing an efficient way to encode two slightly different pictures such as those obtained from two slightly separated cameras shooting the same scene.

VI. MPEG-2 OR DIGITAL TELEVISION

A broadcast TV program is sometimes a simple piece of linear audio and video, basically the output of a microphone and a video camera shooting an outdoor scene, but sometimes it is much more than that. Imagine your favorite evening TV news: you see a computer graphics animation with, say, a rotating globe, an animated "Evening News" text moving along a curve on the screen, the station's logo blinking at the bottom, etc. Then you see a snapshot of the studio, a zoom to the announcer, and then a window containing the bulleted textual summary of the main news items opens and lasts on the screen for a few seconds. Then the announcer presents the first news item with, say, an on-site report and an interview in the original language with subtitles of the interpreted interview at the bottom of the screen and an inserted small moving picture of the announcer. This is in turn followed by a photograph of the individual the news is about with some biographical details, lasting for a few seconds and so on ....

What is the difference between this evening news program and an interactive multimedia application from your PC or a Web page? In terms of richness of multimedia presentation, the TV program is by and large superior. However, you cannot do the most natural thing you do when you are shown the table of content of a Web page: choose the item you want. In a TV program you have to watch and listen to the first item and, if you find something of interest, you have no chance of being shown a button with the indication "click here for more detail," nor can you click on a part of the screen that is sensitive to your mouse.

Does this matter? Depending on whom you are talking to you will be told that couch potatoes do not like interactivity but, oh well, something of it would be a nice addition or that every day millions of people surf the net in search of content that, today, contains a high degree of interactivity but no video, the addition of which would be a great complement to the existing media offered by the Web.

MPEG-2 by itself cannot provide interactivity to a broadcast environment, but it is not a big deal to exploit some "hooks" and enhance the "multimedia look" that will provide some elements that can be perceived as interactivity:

- multiplexing more than one audio channel in a program can be used to offer users the possibility to, say, switch original/interpreted language;
- multiplexing more that one video stream can be used to, say, switch on/off the face of the announcer from a video reportage;
- sending character streams as "other data" in the MPEG multiplex gives the possibility to the user to choose subtitles in the preferred language;
- sending graphics files and animating them to enhance the multimedia look;
- sending the same program at staggered starting times decreasing the time a user has to wait to see the preferred program.

There are many technical ways to do the things listed above. Therefore, a standard is needed if all TV receivers are expected to render the different media correctly.

Now we are at a bifurcation point which is exactly at the same point we heard the two differing views on interactivity. Do we want to define this "multimedia embellishment" specification for a broadcast-only environment or do we want to define it in such a way that the broadcast-only environment is the zero-return channel case of a more general interactivity? In other words, do we want convergence of broadcasting, telecommunication, and computer *technologies* or not?

If we do not want convergence, then the multimedia look can be implemented by exploiting some simple hooks in MPEG-2 Systems that let you multiplex "other data" such as characters, graphic files, etc., with their time and space information along with audio and video. The approach is to treat all additional information sources as supplementary to the audio-visual information.

- A full graphic page will be transmitted as an MPEG-2 I-frame (a mode of MPEG-2 Video to code pictures without dependence on past or future information).
- The temporal and spatial information of rectangles containing graphic information superimposed on top of a natural TV picture will be transmitted by hooks present in MPEG-2 Systems.
- The temporal and spatial information of moving graphics will be transmitted extending the hooks present in MPEG-2 Systems.
- The attributes of the text displayed on the screen will be encoded using some ad-hoc method.

It is clear that no representative of the computer world is ever going to take this solution and extend it to the nonzero return channel case. They have been working for years starting from the other end of the spectrum, building multimedia inch by inch as text + graphics + still pictures + ... with the intention of *including* eventually audio and video as the last step to full multimedia.

In the mind of the author who launched and implemented the idea of the Digital Audio-Visual Council (DAVIC) at the beginning of 1994, DAVIC should have provided the neutral solution acceptable to the broadcasting, telecommunications, and computer worlds. Regrettably, he failed. It was not possible to convince DAVIC people to address the problem in a rational way. Those of broadcast background were unable or unwilling to think of it without making reference to "subtitling." They further aggravated the problem with their inability to agree on a compression scheme for two-dimensional (2-D) arrays of pixels representing graphic information, because they kept saying that none of the one hundred or so graphic formats developed by the computer world suited their needs. They invented a new text coding scheme when a very small subset of HTML, the number of pages of which throughout the world numbers tens of millions, would have amply sufficed. Those of computer background stuck to their "application downloading" paradigm, an impossible marriage with broadcasting. DAVIC people of telecommunication background gathered around the MHEG standard because it fitted their idea of multimedia information representation, that would obviously require a standardized coded representation. In the event DAVIC settled with a double solution, one that contrasts the one-functionality—one tool principle: use of MPEG-2 hooks *and* MHEG. This is convergence of entertainment, telecommunications, and computers at work!

Having so created a dividing line between the technology needed to make interactive and noninteractive multimedia communication, DAVIC has automatically raised the entry threshold for the former. Interactive multimedia communication will happen, but it will not be via an extension of the current information consumption paradigm that is supported by broadcasting. Interactive multimedia communication will have to wait for a new approach that overcomes the current broadcasting/interactive antithesis.

VII. MPEG-4 OR MULTIMEDIA COMMUNICATIONS

With their convergence frenzy, multimedia gurus have forgotten to respond to a simpler, but basic question: what is multimedia communications? Let me try to give my definition.

Multimedia communication is the possibility to communicate audio-visual information that:

1) is natural, synthetic, or both;
2) is real time and nonreal time;
3) supports different functionalities responding to user's needs;
4) flows to and from different sources simultaneously;
5) does not require the user to bother with the specifics of the communication channel, but uses a technology that is aware of it;
6) gives users the possibility to interact with the different information elements;
7) lets the user present the result of his interaction with content in the way suiting his needs.

MPEG-4 is the current MPEG project that is being developed to provide enabling technology for the seven items above. Started in July 1993, it has reached Working Draft level in November 1996, will reach Committee Draft level in November 1997, and International Standard level in November 1998.

Even though the MPEG-4 project predates the Internet frenzy, the motivations at the basis of the project bear a high degree of similarity with some of the topics that make headlines today.

- *Physical Network Independence*: In spite of the word "net," Internet has nothing to do with "network," at least not in the traditional sense of physical-layer telecommunications infrastructure. As soon as a communication link is digitized, you can start using the Internet Protocol (IP), and on top of it transmission control protocol (TCP) or user datagram protocol (UDP), and on top of these protocols the suite of Internet Protocols, such as simple mail transport protocol (SMTP) for mail, hypertext transport protocol (HTTP) for the Web, file transfer protocol (FTP) for file transfer, etc. Internet has smashed vertically-integrated communications: an end user needs not be concerned with the physical nature of the bit pipe, if it is twisted pair, cable, optical fiber, or microwave (of course the Internet access provider and the core network operator do care). TCP/IP can be considered as a part of the communication socket. Already in MPEG-1 and MPEG-2, physical-layer independence has always been assumed, and in MPEG-4 this is just confirmed. Of course independence does not mean unawareness of network peculiarities, the standard must be capable of coping with them.
- *Interactivity*: The Web phenomenon has shown that the capability to search ("surf") the net and interacting with content found in it is indeed a feature users are keen to have. What the Web has not been able to provide is real-time moving pictures and audio. MPEG is the expert body in moving pictures and audio and MPEG-4 must be capable of providing audio and video with the kind of interactivity users have grown accustomed to with the Web. Interactivity, however, is at the level of visual objects that are of arbitrary shapes rather than imposing the "video window" paradigm. MPEG-4 will deliver not only multiple arbitrary shaped video objects, but also individual audio channels associated with objects. Once there is explicit segmentation of objects, there is far greater scope for content developers to produce applications with hitherto unattainable levels of interactivity.
- *Decoding Downloadability*: The success of the Internet has prompted the obvious question: if the Internet is ubiquitous and the bandwidth available to the user is constantly expanding, why should I have my PC loaded with Mbytes of software, most of which I seldom use? Is it not just more effective to download the software I need on demand? Much before the network computer hype, MPEG has come to realize that in many applications, a programmable decoder where decoding tools are downloaded is a preferable solution. MPEG-4 must therefore support downloadability.

VIII. THE MPEG-4 STANDARD

Let us see in more detail some of the features of the MPEG-4 standard.

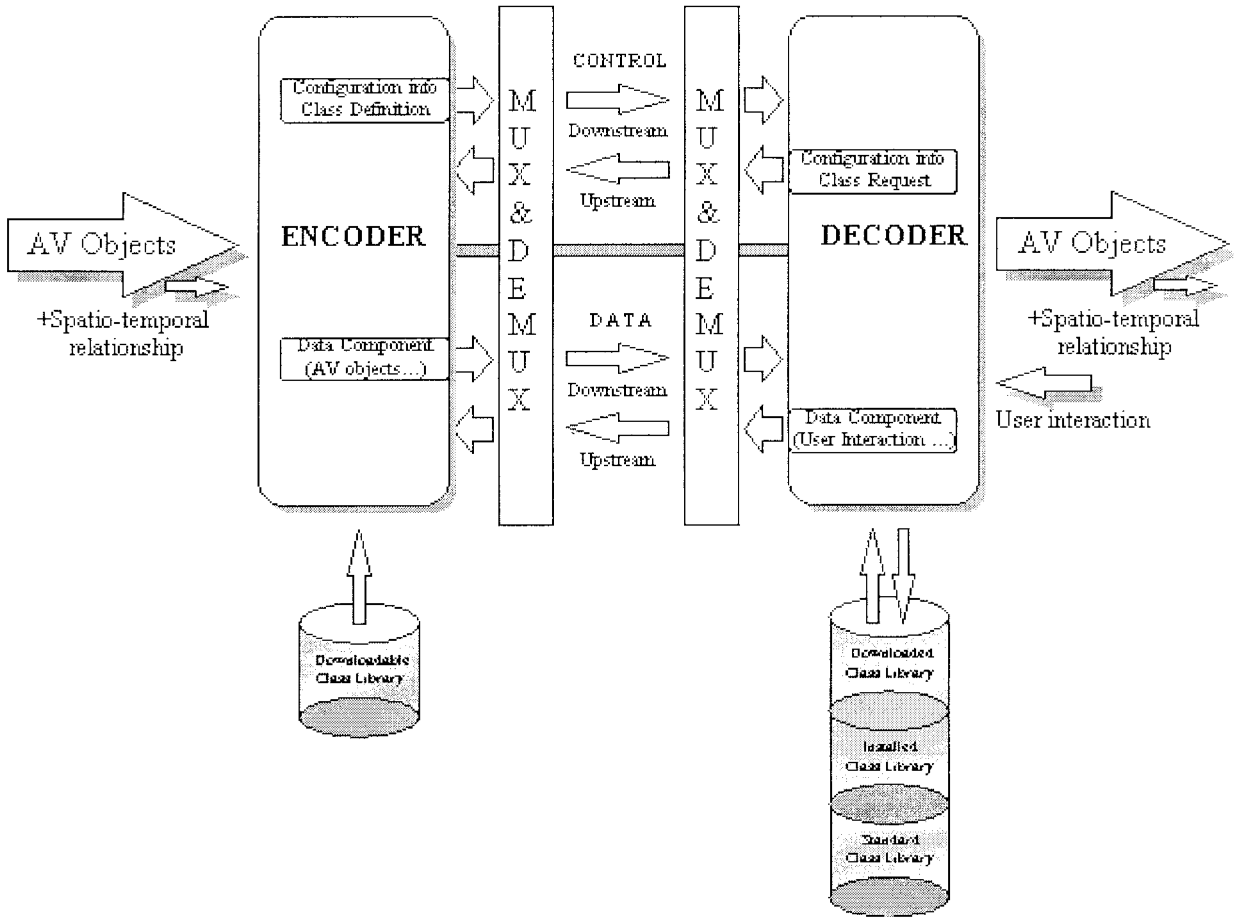

Fig. 1. General MPEG-4 architecture.

A. MPEG-4 Architecture

In the MPEG-4 architecture, one or more AV objects, including their spatio-temporal relationships, if any, are transmitted from a source to an MPEG-4 decoder, as illustrated in Fig. 1. At the source, the AV objects are error protected if necessary, multiplexed, and transmitted downstream. The transmission may occur across multiple channels offering various qualities of service. At the decoder, the AV objects are demultiplexed, error corrected if necessary, decompressed, composited, and presented to the end user. The end user may like to interact with the presentation. Interaction information can be processed locally, or can be transmitted upstream to the encoder for action.

Before the AV objects are transmitted, the source and the decoder must exchange configuration information. The source determines which classes of algorithms, tools, and other objects are needed by the decoder to process the AV objects. Each class of objects can be defined by a data structure plus executable code. The definitions of any missing classes are downloaded to the decoder, where they supplement or override existing class definitions installed or predefined at the decoder. As the decoder executes, new class definitions may be needed in response to user interaction. The decoder can request that the source download specific additional class definitions, possibly in parallel with the transmitted data.

B. Multiplexer

The multiplexer performs the function of combining all the elementary data streams into one output data stream. The demultiplexer defines functionalities needed to recover a system time base, synchronize multiple compressed data streams on decoding, interleave multiple compressed streams into a single stream, and initialize and continuously manage the decoder buffers (Fig. 2).

Some MPEG-4 applications do not involve the serialization function but the stream-synchronization-related functions still apply.

C. Video

In MPEG-1 and MPEG-2, the video information is assumed to be rectangular, of fixed size, displayed at fixed interval. In MPEG-4, the concept of video object (VO), video object layer (VOL), and video object plane (VOP) have been introduced (Fig. 3). VOP represents instances of a given VO. VO and VOP correspond to entities in the bitstream that a user can access and manipulate (with e.g., cut and paste operations). The VOP can have arbitrary shape. At the encoder side, together with the VOP, composition information is sent to indicate where and when each VOP is to be displayed. At the decoder side, the user may be allowed to change the

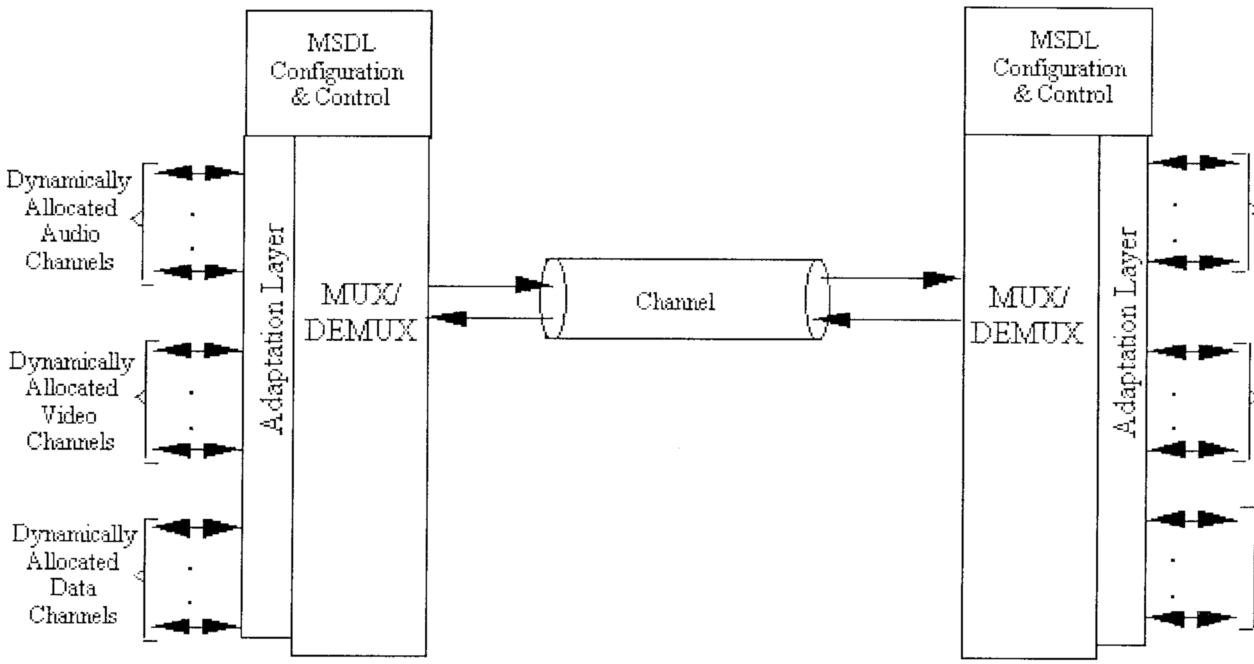

Fig. 2. MPEG-4 multiplexer.

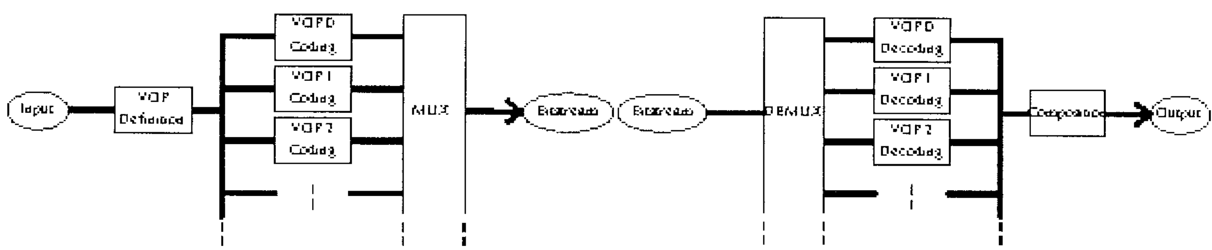

Fig. 3. MPEG-4 video encoder and decoder structure.

composition of the scene displayed by interacting on the composition information.

Video objects overcome the limitation of such standards as JPEG and MPEG where images, still and moving, have been represented as rectangular matrices and compressed with this format, e.g., in MP@ML of MPEG2 video is represented as moving frames each of size e.g., 720×576 pixels. This way of representation destroys the capability to distinguish different elements or objects that make up the picture. For the applications that do not require content/object level interactivity, this type of representation, and coding based on that, has proven to be very efficient. However, it does not provide a convenient way of adding object level interactivity. For example, if one wants to point at a particular person in a scene and get more information about him/her (this is sometimes also called creation of "hot regions" in a scene), it becomes necessary to represent a scene as separate parts called objects. An object does not have to be a person, it can be as broad as foreground or background or a local region containing an advertisement inside a picture.

This type of representation requires that the compression scheme be able to handle arbitrary shapes and also that the receivers not only have the capability to decompress the compressed information but also be able to put them together (composition). Therefore, additional information needs to be sent to the receiver telling it how to compose, or put together, the scene. This additional information is called Alpha Channel.

In a simple case, the Alpha Channel information can be just an on or off (1 or 0) information for a given pixel. In this case, different objects completely hide the information behind them. However, in general, the Alpha Channel can easily consist of 8 b/pixel/frame to allow for various levels of the transparency. As the transmission of such information in uncompressed form can require several Mb/s of bandwidth, it is not practical in many situations. Thus, for this representation and object level interactivity to be successful, a successful scheme is required to compress the Alpha Channel. Driven by these applications, the MPEG-4 Video part will provide efficient compression of objects/parts of a scene and Alpha Channel describing how to put those parts together.

One of the tasks of the System Layer [MPEG-4 system description language (MSDL)], therefore, is to allow the capability to send (or multiplex) the additional information for different objects and bind (or synchronize) it with the

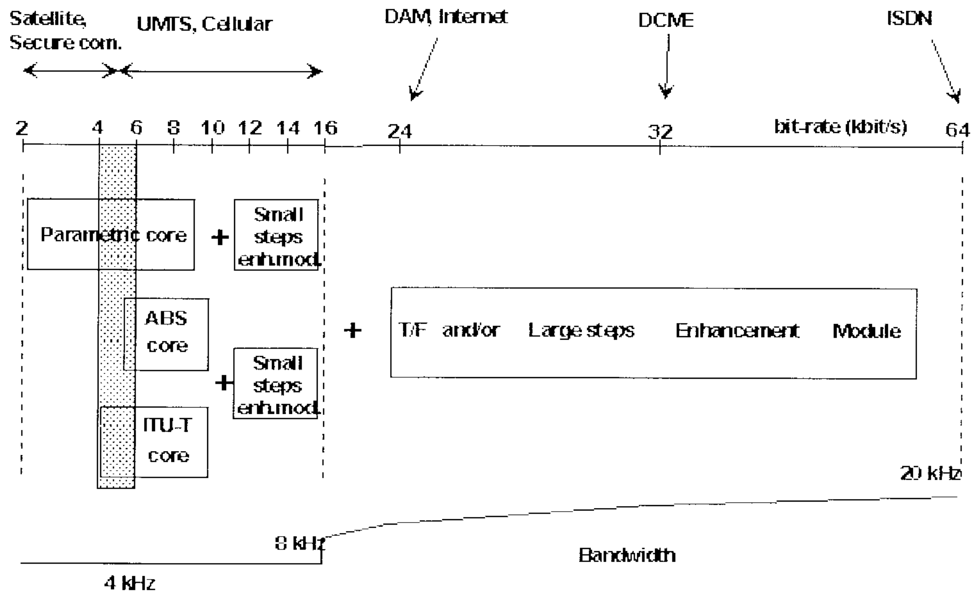

Fig. 4. Bit rates, applications, and techniques considered.

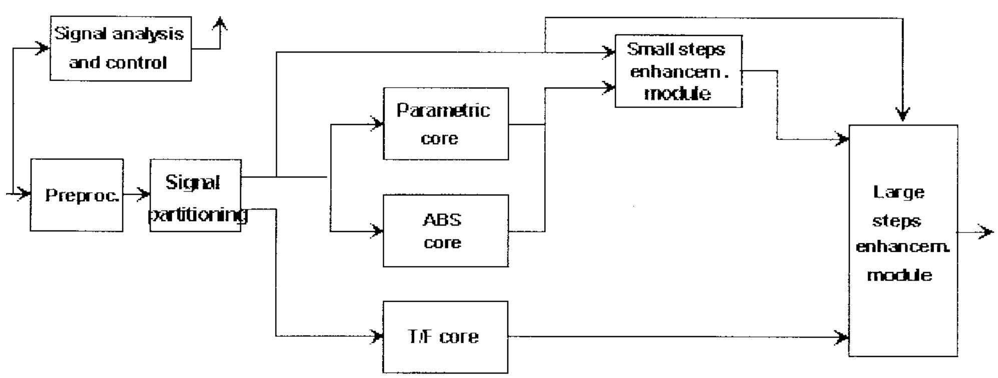

Fig. 5. General block diagram of MPEG-4 Audio.

objects and/or to be able to retrieve additional information about the selected objects within a scene. It should be noted that MPEG-4's field of view is not limited to allow this type of capability for only the natural scenes but also for computer generated information.

Among the main functionalities supported are spatial and temporal scalability and error robustness at VOL and VOP level. Scalability is an important feature when the same audio-visual objects are to be made available through channels of different bandwidth or receivers of different processing capability, or to respond to different user requests. Robustness to errors is also an important feature, as audio-visual communications on radio channels is foreseen to be an important application of MPEG-4.

D. Audio

The AAC standard (part 7 of MPEG-2) is bringing down to 64 kb/s virtual transparency of single channel music which MPEG-1 Audio had set at 128 kb/s. It is expected that interesting performance will be obtained even at bit rates lower than 64 kb/s. NBC is therefore already providing part of the MPEG-4 Audio standard. More work, however, needs to be done in the bit-rate range much lower than 64 kb/s. This is an area where there is a need for a generic technology serving such different applications as satellite and cellular communications, Internet, universal mobile telecommunications system (UMTS), etc. Fig. 4 gives a synthesis of the different bit rates, audio bandwidths, application, and coding techniques currently considered.

Fig. 5 gives a block diagram that is believed to be capable of handling the requirements of the application scenario of Fig. 4.

E. SNHC

So far the communications world has treated synthetically generated contents as a subset of natural contents, e.g., graphics have been communicated by sending as a video. Therefore, there has been no standard for representing and compressing that information separately.

Synthetic–natural hybrid coding's (SNHC's) aim is to treat the synthetically generated contents as another NEW data type from the communication point of view and to standardize how to represent and compress it. As this is the first attempt by a standards committee to do so, it is expected that MPEG-4 will get some initial key work done in this area and more needs will arise that can possibly be taken care of in future phases of MPEG-4 or other standards.

An initial area of focus has been to extend the models available in VRML. In VRML it is relatively easy to create models of things that are not live, like tables, chairs, etc. However, it is virtually impossible to create a good model of the human face and body. For next generation of multimedia communications, that is a very important piece missing in VRML.

MPEG-4 is first working on developing the capability to create representations and models for human faces and bodies. It is working on developing the standardized set of parameters required to model a human face and to also synchronize the facial expression and lip movements with audio. This in addition to VRML or VRML-like language can allow the creation of realistic scenes. After successful completion of this effort, the focus will be on developing standardized techniques and/or parameter sets for texture mapping.

On the audio side, the initial focus is to standardize the parameter set to allow interoperable text-to-speech conversion and synthetic sound.

F. Flexibility

A decoder with all the features described may be unnecessarily expensive. Some decoders may have been designed to support only a subset of all coding tools (e.g., a mobile video phone) or be flexible enough to acquire the particular subset of the tools when it decodes content from one source (e.g., a movie) and acquire a different subset to decode content from another source (e.g., a video game).

This is becoming possible thanks to the progress of VLSI technology that is producing powerful programmable processors at an accelerated rate. As an example, today I can perform real-time decoding of MPEG-1 Audio-Video-System bitstreams at 1.4 Mb/s on my 133 MHz Pentium. The possibility will even materialize one day when coding tools, not belonging to the standardized set, can be downloaded on fully programmable processors.

MPEG-4 defines three capabilities of decoder programmability that support flexibility and extensibility.

- **Flex_0** (nonprogrammable) is a finite set of standardized algorithms of Audio, Video, System decoders made up of standardized Audio, Video, System tools.
- **Flex_1** (flexible) consists of a finite set of standardized Audio, Video, and System tools and their standardized interfaces, which may be flexibly configured into arbitrary algorithms.
- **Flex_2** (extensible) is a standardized mechanism to describe arbitrary algorithms made of arbitrary tools. It should be clear, however, that currently no Flex_2 specification is attempted and may be included only at a later date.

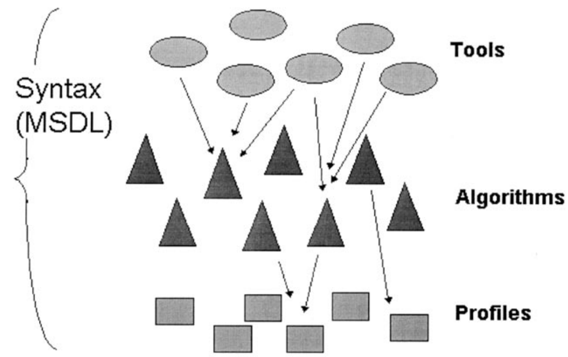

Fig. 6. MPEG-4 tools, algorithms, and profiles.

The Flex_0 case is conceptually similar to the MPEG-2 Profile/Level arrangement. To implement Flex_1, MPEG has started the definition of tool Application Programming Interface (API). The language selected for this purpose is the Java language, which is also used for the purpose of linking the tools together and providing a complete decoder. No commitment is made at this time, however, that Java will be the language eventually retained for these purposes.

Fig. 6 describes how the standardized tool set can be used to assemble algorithms and profiles.

G. Software Implementation of the Standard

In MPEG-1 and MPEG-2 extensive use was already made of simulation programs written in C language to implement the Simulation Models of MPEG-1 and the Test Models of MPEG-2. Parts 5 of both standards give a software implementation of both encoder and decoder. In MPEG-4, a substantial innovation has been made with the definition of a *Reference* Implementation of the MPEG-4 Audio, Video, and System Verification Model (VM), written in C or C++, recognizing the benefits of reference programs implementing the VM's as a means to improve collaboration, speed up development, reduce unnecessary duplication of efforts, and promote eventual acceptance of the standard in the marketplace.

The copyright of decoder source doce modules representing normative elements is released to ISO/IED. Any patent needed to implement the VM in either hardware or software still applies. ISO/IEC gives users of the standard free license to use the normative elements of the possibly modified VM decoder software for use in hardware or software products claiming conformance to the MPEG-4 standard. Copyright is not released for non MPEG-4 conforming products but the original developer retains full right to use the code for his own purposes and inhibit a third party from using the code for non MPEG-4 conforming products.

IX. Is this All that We Need to Make Multimedia Communications?

MPEG-4 will provide a generic technology for multimedia communications applications and services. In the list of issues considered above, some more elements, not currently included in the MPEG-4 considerations, are needed to complete the picture. These are: intellectual property right (IPR) management, security, search of content, and transport protocol.

A. IPR Management

No matter what the different consumption paradigm brought about by the Internet and the Web in particular is, the role of content as the engine that drives authors to produce it and users to consume it will remain intact, but the nature of IPR will not necessarily be the same as today.

In MPEG-2, IPR management is a transposition to the digital world of what is being done today, because of the very superficial role played by the coding algorithm. The fact that content is digital and can be "stamped" with the copyright descriptor gives the advantage that more effective IPR management by automatic processing in the delivery chain becomes possible.

Three important new elements can be discerned in an MPEG-4 scenario.

- In MPEG-4, the way content is represented goes deeper into the semantics of the information itself. There is an obvious IPR on the elementary components (say, a VOP) and on the way they are composited at the source, but then the user can himself make a different compositing and presentation, using information of natural and synthetic origin, real-time and nonreal time, from one or more different sources.
- A second element of difference lies in the nature of the coding algorithms. In MPEG-2 the algorithm is fixed, while in MPEG-4 it may be assembled by the author. Even though every single tool may have one or more pieces of IPR associated with it, what is the nature of the assembly of tools used by a particular author?
- In MPEG-2, a decoder is a fixed piece of silicon capable of performing only certain operations. This case will still exist in MPEG-4 but alongside with it there will be a range of alternative cases culminating with the extreme case of a generic programmable processor whose decoding software will downloaded with the content itself. In this case, "algorithm IPR" can no longer be associated with the decoder hardware but with the content itself.
- Watermarking is a widely used technology today in the analog domain. Watermarking is also possible in the digital domain and may depend on the technical features of the coding algorithm. Appropriate hooks may have to be put in place to allow an effective watermarking of content.

There is a need to provide a solid IPR management mechanism in MPEG-4. This has already started and, needless to say, it sees the involvement of representatives of the contents world.

B. Security

Support for security in the form of encryption (ECM and EMM messages) was already present in MPEG-2. Unfortunately the MPEG-2 specification fell too short of the goal of enabling transparency of the encryption technology to the user. This is becoming a major hindrance to the wide deployment of MPEG-2 services. Moreover MPEG-2 security was designed to provide mostly scrambling functions for a service provider in a broadcasting environment.

These limitations should be avoided in MPEG-4. Security is an essential feature of a communication standard that has several dimensions, scrambling being just one of them.

C. Content Search

MPEG-1 and MPEG-2 have been designed and are widely used to encode content that has a clear identity such as a movie, a documentary, etc. In the current usage of MPEG-2, the so-called "Service Information" describes each piece of content according to well-identified categories, so as to enable search by a user.

This solution serves well the purpose for which it has been designed: to find information of interest in a large but still manageable number of programs. It would be awkward, as an example, to extend the solution for use in content search in the Web. This is, however, the paradigm, if not exactly the environment, in which MPEG-4 will mostly be used.

The lack of suitable search technologies is one of the reason why, in spite of the explosive growth of the Web, many are questioning its business value. The problem is exacerbated by the fact that HTML was just designed as a language to encode text and links without any consideration for the information searching function.

That this limitation should be avoided in MPEG-4 has been clearly identified and a new project is about to start with the nickname MPEG-7. This will address, among others:

- the requirements from different application domains when a need for a certain piece of MPEG-4 encoded information must be searched;
- the definition of appropriate support within the MPEG-4 syntax for search functions;
- the specification of generic tools for search engines for MPEG-4 encoded information.

D. Transport Protocol

Every age has its religion war, and the one that is raging now has two camps called ATM and Internet. In summary, the technical terms of the debate are the following.

- ATM has been designed as digital broadband network capable of carrying all sorts of information. Applications should use the network services provided by ATM via an appropriate ATM adaptation layer (AAL).
- Internet has already developed and deployed a suite of protocols that provide all services that are needed by a range of (narrow band) applications. They sit on the IP which can sit or not on ATM.

MPEG-4 needs not take side in this debate. As much as MPEG-2 can be carried directly on a digital stream at the physical layer, or over ATM, or over TCP/IP, so MPEG-4 shall be usable in all three cases.

X. CONCLUSIONS

This paper has addressed the decade-old problem of multimedia communications, recognizing the unfulfilled promises of this new communication domain and clearly separating the technical issues from the "convergence" hype of the early nineties.

The multi-industry nature of multimedia communications calls for cross-industry standards. The difficulty to deal with industries having so different approaches to standardization has then been recognized, but the successful recipe adopted by MPEG in its MPEG-1 and MPEG-2 standards can be applied again to the new standardization project MPEG-4, which promises to become the enabling technology for multimedia communications.

Acknowledgment

The author would like to thank O. Avaro, FT/CNET, P. Chou, Xerox, T. Ebrahimi, EPFL, P. Fellows, ST, A. Luthra, GI, G. Morrison, BTLabs, P. Haikonen, Nokia, R. Koenen, KPN, K. O'Connel, Motorola, S. Okubo, GCL, P. Schirling, IBM, and A. Tabatabai, Tektronix, for their careful reading of the manuscript and their valuable comments.

References

[1] S. Okubo, K. McCann, and A. Lippman, "MPEG-2 requirements, profiles and performance verification—Framework for developing a generic video coding standard," *Signal Processing: Image Commun.*, vol. 7, pp. 201–209, 1995.

Leonardo Chiariglione was born in Almese, Italy. He received the M.S. degree in electronic engineering from the Polytechnic of Turin, Italy, and the Ph.D. degree from the University of Tokyo, Japan, in 1973.

In 1971 he joined CSELT, Torino, Italy, the corporate research center of the STET Telecommunications Group where he is Head of the Multimedia and Video Services Research Division. In 1986 he originated the HDTV Workshop, an international event targeted at promoting the technical aspects of HDTV overcoming the traditional barriers of specific industry interests. In 1988 he originated the ISO standardization activity known as MPEG (Moving Pictures Experts Group). This group has produced the MPEG-1 and MPEG-2 standards that have triggered the digital audio-visual revolution and DSM-CC, a server to set-top protocol. In 1989 he founded *Image Communications*, a EURASIP bimonthly journal for the development of the theory and practice of image communication, of which he is the Editor-in-Chief. In January 1994 he launched the idea of the Digital Audio-Visual Council (DAVIC). The purpose of DAVIC is the promotion of emerging digital audio-visual applications and services by the timely availability of internationally agreed specifications of open interfaces and protocols that maximize interoperability across countries and applications/services. DAVIC has produced its first specifications (DAVIC 1.0) in December 1995 and is currently working on two new versions: DAVIC 1.1 and DAVIC 1.2. He was President and Chairman of the Board until December 1995. In January 1996 he launched the idea of the Foundation for Intelligent Physical Agents (FIPA), an international nonprofit association of companies and organizations which agree to share efforts to produce in a timely fashion internationally agreed specifications of generic agent technologies that are usable across a large number of applications, providing a high level of interoperability across applications. He is currently the President and Chairman of the Board.