A vision made real
Past, present and future of MPEG

By Leonardo Chiariglione

Edition of 2019/05/20

Table of Contents

Why this book?

In a generation, life of the large majority of human beings is incredibly different than the life of the generation before. The ability to communicate made possible by ubiquitous internet and to convey media content to others made possible by MPEG standards can probably be mentioned among the most important factors of change. However, unlike internet about which a lot has been written, little is known about the MPEG group besides its name.

This book wants to make up for this lack of information.

It will talk about the radical transformation that MPEG standards wrought to the media distribution business by replacing a multitude of technologies owned by different businesses with a single technology shared by all; the environment in which it operates; the radically new philosophy that underpins this transformation; the means devised to put the philosophy into practice; the industrial and economic impact of MPEG standards; what new standards are being developed; and what is the future that the author conceives for MPEG as an organisation that plays such an important industrial and social role.

Bottom line, MPEG is about technology. Therefore, the book offers an overview of all MPEG standards and in particular video, audio, media quality, systems and data. This is for those more (but not a lot more) technology-minded.

Important – there are short Conclusions worth reading.

1 Introduction

Media is a word typically used to indicate the means that enable communication, e.g. radio, television, storage media, internet, but this book calls those media *delivery media* and uses the word "media" or "media content" to indicate the *information* conveyed by delivery media.

This distinction is not an abstract distinction, but the core of the MPEG revolution: *shared standards for compressed digital media* that offer individual industries the means to uniquely represent the media content in digital form and the freedom to choose delivery media that best suit their needs.

Chapter 2 The discontinuity of digital technologies analyses four aspects of the media distribution business and their enabling technologies:

1. Analogue media distribution describes the vertical businesses of analogue media distribution;
2. Digitised media describes media digitisation and why it was largely irrelevant to distribution;
3. Compressed digital media describes how industry tried to use compression for distribution;
4. Digital technologies for media distribution describes the potential structural impact of compressed digital media for distribution.

Chapter 3 MPEG and digital media tackles the issue of standardisation of compressed digital media

1. Standards describes the 3 international standards organisations;
2. ISO and IEC standards describes the ISO structure and the ISO/IEC standardisation process;

3. A home for MPEG describes how an independent home for MPEG was found.

Chapter 4 The MPEG principles lays down the principles of the philosophy that MPEG has developed for generic multi-industry standards:

1. Standards need a business model argues that MPEG standardisation is a (non-commercial) business and needs a business model to be successful and long-lasting;
2. Common standards for all countries and industries argues that the transition from vertical businesses requires a process change;
3. Designing common standards for different users argues that industry-agnostic standardisation needs new rules;
4. Technology standards as toolkits highlights the problem that common standards cannot be more burdensome that industry-specific standards;
5. Standards for the market, not the other way around argues that industry-agnostic standardisation is incompatible with the imposition of market-crowned standards;
6. Standards that anticipate the future argues that industry-agnostic standardisation can only serve its purpose if it looks to standards that satisfy future needs;
7. Compete and collaborate argues that industry-agnostic standardisation requires competition to get the best technologies and collaboration to refine and improve them;
8. Industry-friendly standards sets some basic rules to make sure that standards do not restrict but enhance the role of individual companies;
9. Audio and video come together sets the obvious rule that audio and video compression should be part of the same standardisation effort;
10. A glue to keep audio and video together argues that audio-visual systems are made of pieces that need glue to work properly together;
11. Integrated standards as toolkits identifies the need for users to have integrated solutions but also to cherry pick individual components of an integrated solution;
12. One step at a time sets the strategy to respond to challenges in small steps;
13. Separate the wheat from the chaff identifies the need for industry components to test implementation for conformance to an industry-agnostic standard;
14. Technology is always on the move identifies the need for a standardisation process where new technology may make some standards obsolete;
15. Research for MPEG standards identifies the need for an organic relationship between research and MPEG standardisation;
16. Standards as enablers, not disablers argues that industry members have the right to have their legitimate requests satisfied by industry-agnostic standardisation;
17. Never stop working together claims that to achieve standards developed collaboratively in a timely fashion, collaborative work should not be limited to official meetings;
18. The nature and borders of compression claims that compression is the enabling technology but its nature and borders change with time.

Chapter 5 The MPEG operation studies how the principles of Chapter 4 have been put into an operational practice within the rules of the ISO/IEC Directives:

1. The MPEG organisation describes MPEG's internal organisation;
2. Organisation of work describes how the expertise of MPEG participants is organised to develop standards;
3. How MPEG develops standards describes the specific steps MPEG undertakes to develop standards'
4. The ecosystem drives MPEG standards explains how MPEG is part of an ecosystem where all participants play a role;
5. Standardisation and product making compares the task of developing a standard with the task of developing a product;

6. Standards are living beings explains why publishing a standard is a new beginning, not the end of standardisation work;
7. Standards and uncertainty concludes that is no recipe to design a guaranteed successful MPEG standard;
8. MPEG communicates identifies communication as the most important bond keeping the ecosystem together.

Chapter 6 The MPEG success looks at three sides of success

1. What has been done looks at those MPEG standards that have the largest impact in terms of use;
2. The MPEG success in numbers (of units and dollars) looks at the numbers – of devices and dollars – triggered by MPEG standards
3. MPEG is working for more success describes the current MPEG work plan.

Chapter 7 Planning for the future of MPEG analyses the current context in which MPEG operates and how it can morph to face new challenges while staying true to its tradition.

The remaining chapters are dedicated to the (even not too) technically-inclined MPEG fans who want to have views of the inside of MPEG standards.

Chapter 8 Video compression in MPEG analyses video compression standards from four viewpoints:

1. Forty years of video compression is a thorough scan on 40 years of ITU and MPEG video compression standards;
2. More video features looks at the process that added new functionalities at each MPEG standard;
3. Immersive visual experiences looks at how " immersive experience" has been enhanced at each MPEG standard;
4. Video can be green describes how video compression can become more energy thrifty without sacrificing the user experience.

Chapter 9 Audio compression in MPEG is a thorough scan on 30 years of MPEG audio compression standards.

Chapter 10 Quality assessment describes the ever more important process of assessing quality after applying media compression.

Chapter 11 Systems standards keep the pieces together describes the role of the Systems components across most MPEG standards.

Chapter 12 Data compression takes the view of applying compression to a variety of data types:

1. Meaningful data can be compressed analyses of data types that MPEG has handled in all its standards
2. Moving intelligence around identifies "artificial neural networks" as a data type of growing importance and size that benefits from compression;
3. MPEG standards for genomics describes how a variety of MPEG compression tools have been successfully applied to the compression of DNA reads from high-speed sequencing machines
4. Compression of other data raises again the issue of "data compression" being a field of endeavour

Chaper 13 The MPEG standards gives an often concise overview of most of the ~200 standards developed by MPEG.

Chapter 14 gives the MPEG work plan.

Chapter 15 gives the official 5-digit ISO numbers of the MPEG standard

Chapter 16 Conclusions draws short but meaningful conclusions on the MPEG story.

2 The discontinuity of digital technologies

This chapter analyses four aspects of the media distribution business and their enabling technologies:

1. Analogue media distribution describes the vertical businesses of analogue media distribution;
2. Digitised media describes media digitisation and why it was largely irrelevant to distribution;
3. Compressed digital media describes how industry tried to use compression for distribution;
4. Digital technologies for media distribution describes the potential structural impact of compressed digital media for distribution.

Analogue media distribution

In the 1980's media were analogue, the sole exception being music on compact disc (CD). Different industries were engaged in the business of distributing media: telecommunication companies distributed music, cable operators distributed television via cable, terrestrial and satellite broadcasters did the same via terrestrial and satellite networks and different types of businesses distributed all sort of recorded media on physical support (film, laser discs, compact cassette, VHS/Betamax cassette, etc.).

Even if the media content was exactly the same, say a movie, the baseband signals that represented the media content were all different and specific of the delivery media: film for the theatrical vision, television for the terrestrial or satellite network or for the cable, a different format for video cassette. Added to these technological differences caused by the physical nature of the delivery media, there were often substantial differences that depended on countries or manufacturers.

Figure 1 depicts the vertical businesses of the analogue world when media distribution was a collection of industry-dependent distribution systems each using their own technologies for the baseband signal. The figure is simplified because it does not take into account the country- or region-based differences within each industry.

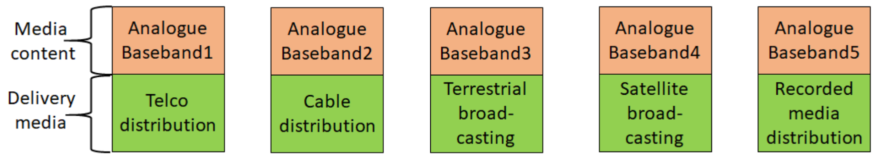


Figure 1 – Analogue media distribution

Digitised media

Since the 1930's the telecom industry had investigated digitisation of signals (speech at that time). In the 1960's technology could support digitisation and ITU developed G.711, the standard for digital speech, i.e. analogue speech sampled at 8 kHz with a nonlinear 8 bits quantisation. For several decades digital speech only existed in the (fixed) network, but few were aware of it because the speech did not leave the network as bits.

It was necessary to wait until 1982 for Philips and Sony to develop the Compact Disc (CD) which carried *digital stereo audio*, specified in the "Red Book": analogue stereo audio sampled at 44.1 kHz with 16 bits linear. It was a revolution because consumers could have an audio quality that did nor deteriorate with time.

In 1980 a digital video standard was issued by ITU-R. The luminance and the two colour-difference signals were sampled at 13.5 and 6.75 MHz, respectively, at 8 bits per sample yielding an exceedingly high bitrate of 216 Mbit/s. It was a major achievement, but digital television never left the studio if not as bulky magnetic tapes.

The network could carry 64 kbit/s of digital speech, but no consumer-level delivery media of that time could carry the 1.41 Mbit/s of digital audio and much less the 216 Mbit/s of digital video. Therefore, in the 1960s studies on compression of digitised media begun in earnest.

Compressed digital media

In the 1980's compression research yielded its first fruits:

1. In 1980 ITU approved Recommendation T.4: Standardization of Group 3 facsimile terminals for document transmission. In the following decades hundreds of million Group 3 facsimile devices were installed worldwide because, thanks to compression, transmission time of an A4 sheet was cut from 6 min (Group 1 facsimile), or 3 min (Group 2 facsimile) to about 1 min.
2. In 1982 ITU approved H.100 (11/88) Visual telephone systems for transmission of videoconference at 1.5/2 Mbit/s. Analogue videoconferencing was not unknown at that time because several companies had trials, but many hoped that H.100 would enable diffused business communication.
3. In 1984 ITU started the standardisation activity that would give rise to Recommendations H.261: Video codec for audio-visual services at p x 64 kbit/s approved in 1988.
4. In the mid-1980s several CE laboratories were studying *digital video recording* for magnetic tapes. One example was the European Digital Video Recorder (DVS) project that people expected would provide a higher-quality alternative to the analogue VHS or Betamax videocassette recorder, as much as CDs were supposed to be a higher-quality alternative to LP records.
5. Still in the area of recording, but for a radically new type of application – *interactive video on compact disc* – Philips and RCA were independently studying methods to encode video signals at bitrates of 1.41 Mbit/s (the output bitrate of CD).
6. In the same years CMTT, a special Group of the ITU dealing with transmission of radio and television programs on telecommunication networks, had started working on a standard for transmission of *digital television* for "primary contribution" (i.e. transmission between studios).
7. In 1987 the Advisory Committee on Advanced Television Service was formed to devise a plan to introduce HDTV in the USA and Europe was doing the same with their HD-MAC project.
8. At the end of the 1980's RAI and Telettra had developed an HDTV codec for satellite broadcasting that was used for demonstrations during the Soccer World cup in 1990 and General Instrument had showed its Digicipher II system for terrestrial HDTV broadcasting in the bandwidth of 6 MHz used by American terrestrial television.

Digital technologies for media distribution

The above shows how companies, industries and standards committees were jockeying for a position in the upcoming digital world. These disparate and often uncorrelated initiatives betrayed the mindset that guided them: future distribution of digital media would have an arrangement similar to the one sketched in *Figure 1* for analogue media: the "baseband signal" of each delivery medium would be digital, thus using new technology, but different for each industry and possibly for each country/region.

In the analogue world these scattered roles and responsibilities were not particularly harmful because the delivery media and the baseband signals were so different that unification had never been attempted. But in the digital world unification made a lot of sense.

MPEG was conceived as the organisation that would achieve unification and provide generic, i.e. domain-independent digital media compression. In other words, MPEG envisaged the completely different set up depicted in *Figure 2*.

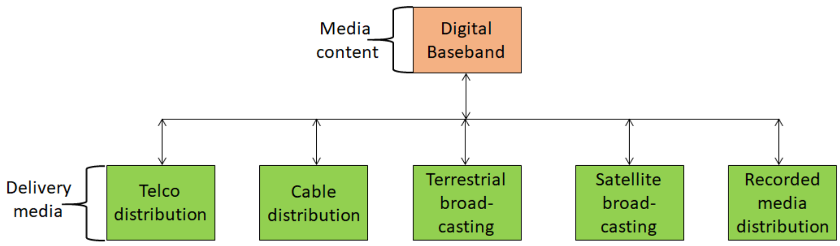


Figure 2 – Digital media distribution (à la MPEG)

In retrospect that was a daunting task. If its magnitude had been realised, it would probably never had been started.

3 MPEG and digital media

The first problem to be tackled was that the MPEG "digital baseband" had to be based on standards, and international ones because they had to have global validity. The second problem to be addressed was that the standards should not be industry specific.

The question then was: where should those standards be developed? The answer is provided by the following:

1. Standards describes the 3 international standards organisations;
2. ISO and IEC standards describes the ISO structure and the ISO/IEC standardisation process;
3. A home for MPEG describes how an independent home for MPEG was found.

Standards

Standards have a special place in industry because they represent convergence points where the parties involved, who typically are in competition, find it convenient to agree on a single solution.

Standards bodies exists at the international level:

1. *International Telecommunication Union* (ITU) for matters related to telecommunication and broadcasting
2. *International Electrotechnical Commission* (IEC) for electrotechnical matters
3. *International Organisation for Standardisation* (ISO) for everything else.

Many other standards organisations or industry fora exist at national/regional or industry level.

ITU

The International Telecommunication Union (ITU) is the result of the 1934 merge between the International Telegraph Convention of 1865 and the International Radiotelegraph Convention of 1906, and today is an agency of the United Nations. This is reflected in the two main branches of the ITU: ITU-T and ITU-R. The former deals with standards for global telecommunications excluding radio communication because this is the purview of ITU-R.

IEC

The International Electrotechnical Commission (IEC) is a not-for-profit organisation founded in 1906. It develops International Standards in the fields of electrotechnology, e.g. power generation, transmission and distribution to home appliances and office equipment, semiconductors, fibre optics, batteries, solar energy, nanotechnology and marine energy.

ISO

The International Organization for Standardization (ISO) is an international non-governmental standard-setting organisation founded in 1947 and composed of representatives from various national standards organizations.

ISO is well known for its family of quality management systems standards (ISO 9000), environmental management standards (ISO 14000) and Information Security Management Systems standards (ISO 27000). There are more than 20,000 ISO published standards.

ISO is a huge organisation whose technical branch is structured, as is the IEC's, in Technical Committees (TC). The first 3 active TCs are: TC 1 Screw threads, TC 2 Fasteners and TC 4 Rolling bearings. The last 3 TCs in order of establishment are TC 322 Sustainable finance, TC 323 Circular economy and TC 324 Sharing economy.

Between these two extremes there is a large number of TCs, e.g., TC 35 Paints and varnishes, TC 186 Cutlery and table and decorative metal hollow-ware, TC 249 Traditional Chinese medicine, TC 282 Water reuse, TC 297 Waste collection and transportation management, etc.
Most TCs are organised in working groups (WG). They are tasked to develop standards while TCs retain key functions such as strategy and management. In quite a few cases the area of responsibility is so broad that a horizontal organisation would not be functional. In this case a TC may decide to establish Subcommittees (SC) which include WGs tasked develop standards.
Figure 3 is an organigram of ISO.

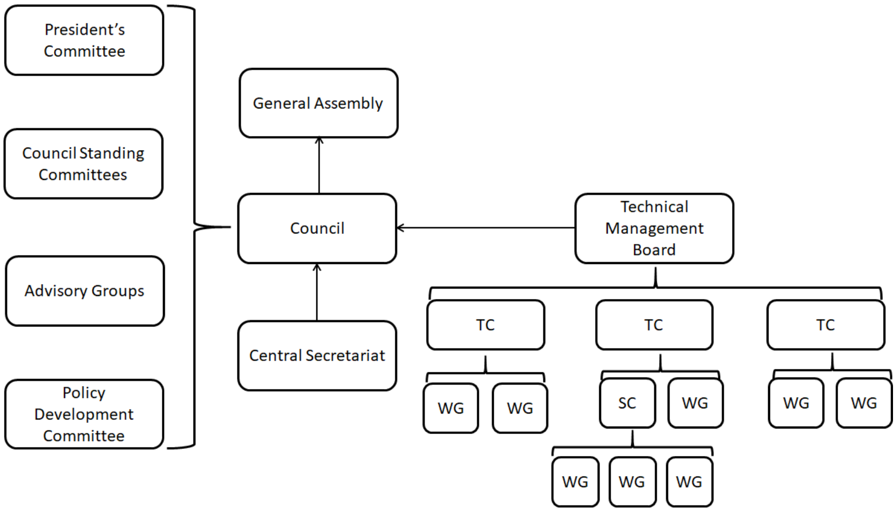


Figure 3 – ISO governance structure

ISO and IEC standards

The development process

ISO and IEC share the standard development process which can be summarised as follows:

1. Submission and balloting of a New Work Item Proposal (NWIP) of a new project meant to lead to an International Standard (IS) or Technical Report (TR). The former contains normative clauses, the latter is informative
2. Development of a Working Draft (WD, possibly several versions of it
3. Balloting of the Committee Draft (CD, when the WD has achieved sufficient maturity)
4. Balloting of the Draft International Standard (DIS, after resolving comments made by National Bodies)
5. Balloting of the Final Draft International Standard (FDIS, after resolving comments made by National Bodies)

The last ballot is yes/no. No comments allowed.
Amendments (AMD) extend a standard. The same steps as above are carried out with the names Proposed Draft Amendment (PDAM), Draft Amendment (DAM) and Final Draft Amendment (FDAM).
If an error is discovered, a Corrigendum (COR) is produced. This only goes through two stages: Draft Corrigendum (DCOR) and Corrigendum (COR).
A Technical Report, a document without normative clauses, goes through two stages of approval: Proposed Draft Technical Report (PDTR) and Technical Report (TR).

Consensus

ISO/IEC mandates that working groups operate based on consensus when they develop standards. Consensus is defined as

General agreement characterised by the absence of sustained opposition to substantial issues by any important part of the concerned interests and by a process that involves seeking to take into account the views of all parties concerned and to reconcile any conflicting arguments.
NOTE — Consensus need not imply unanimity.

Patent policy
ISO, IEC and ITU share a common policy vis-à-vis patents in their standards. Using few imprecise but hopefully clear words (as opposed to many precise but unclear words), the policy is:
1. It is good if a standard has no patents or if the patent holders allow use of their patents for free (with an "Option 1" declaration);
2. It is accepted if a standard has patents, but the patents holders only allow use of their patents on fair and reasonable terms and non-discriminatory conditions (with an "Option 2" declaration);
3. It is not permitted to have a standard with patents whose holders do not allow use of their patents (with an "Option 3" declaration).

A home for MPEG
When the MPEG idea took shape in July 1987, the selection of a home to implement the idea was the primary concern. The idea was spoilt for choices as shown by the list of international committees in *Table 1* that were created for various reasons – regulation or simply need for an independent technical reference – to cater to the needs of standards by the different industries.

Table 1 – Media-related standards committees (1980's)

ITU-T	Speech	SG XV WP 1
	Video	SG XV WP 2
ITU-R	Audio	SG 10
	Video	SG 11
IEC	Recording of audio	SC 60 A
	Recording of video	SC 60 B
	Audio-visual equipment	TC 84
	Receivers	SC 12A and G
ISO	Photography	TC 42
	Cinematography	TC 36

Since MPEG was conceived to be industry-neutral, committees already developing standards in the "media" area were considered unsuitable because the represented "vested interests". The choice fell on ISO TC 97 Data Processing who had SC 2 Character sets and Information Coding who included WG 8 Coding of Audio and Picture Information.
In 1987 ISO/TC 97 Data Processing merged with IEC/TC 83 Information technology equipment. The resulting (joint) technical committee was called ISO/IEC JTC 1 Information Technology. SC 2 with its WGs, including WG 8, became part of JTC 1. MPEG was established as an Experts Group on Moving Pictures of ISO/IEC JTC 1/SC 2/WG 8 in 1988.
Note that Experts Group is an organisational entity not officially recognised in the ISO organigram.
In 1991 SC 2/WG 8 seceded from SC 2 and became SC 29. WG 8's Moving Picture Experts Group (MPEG) became WG 11 Coding of audio, picture, multimedia and hypermedia information (but everybody in the industry, and even in the general public, calls it MPEG).

4 The MPEG principles

The challenging goals that MPEG set to itself 3 decades ago have been successfully achieved across many generations of digital media standards implemented in countless products, applications and services by many different industries in interoperable ways. How could this happen?

A significant part of the answer lies in the fact that MPEG has developed a philosophy based on the notion of generic multi-industry standards that is has thoroughly applied.

Purpose of this chapter is to examine the different aspects of that philosophy:

1. Standards need a business model argues that MPEG standardisation is a (non-commercial) business and needs a business model to be successful and long-lasting;
2. Common standards for all countries and industries argues that the transition from vertical businesses requires a process change;
3. Designing common standards for different users argues that industry-agnostic standardisation needs new rules;
4. Technology standards as toolkits highlights the problem that common standards cannot be more burdensome that industry-specific standards;
5. Standards for the market, not the other way around argues that industry-agnostic standardisation is incompatible with the imposition of market-crowned standards;
6. Standards that anticipate the future argues that industry-agnostic standardisation can only serve its purpose if it looks to standards that satisfy future needs;
7. Compete and collaborate argues that industry-agnostic standardisation requires competition to get the best technologies and collaboration to refine and improve them;
8. Industry-friendly standards sets some basic rules to make sure that standards do not restrict but enhance the role of individual companies;
9. Audio and video come together sets the obvious rule that audio and video compression should be part of the same standardisation effort;
10. A glue to keep audio and video together argues that audio-visual systems are made of pieces that need glue to work properly together;
11. Integrated standards as toolkits identifies the need for users to have integrated solutions but also to cherry pick individual components of an integrated solution;
12. One step at a time sets the strategy to respond to challenges in small steps;
13. Separate the wheat from the chaff identifies the need for industry components to test implementation for conformance to an industry-agnostic standard;
14. Technology is always on the move identifies the need for a standardisation process where new technology may make some standards obsolete;
15. Research for MPEG standards identifies the need for an organic relationship between research and MPEG standardisation;
16. Standards as enablers, not disablers argues that industry members have the right to have their legitimate requests satisfied by industry-agnostic standardisation;
17. Never stop working together claims that to achieve standards developed collaboratively in a timely fashion, collaborative work should not be limited to official meetings;
18. The nature and borders of compression claims that compression is the enabling technology but its nature and borders change with time.

Standards need a business model

MPEG in not engaged in a commercial business, but it does have a business model that has guided its existence as a group and driven its standard-development work.

When MPEG started, there were committees who developed standards that avoided the use of essential patents. A successful example of this approach is the well-known JPEG standard (ISO/IEC 10918).

The field of video coding, however, was (and more so is today) very different because industry and academia had worked on video compression technologies for some 3 decades and had filed many patents (at that time they could already be counted by the thousands) covering a wide range of basic video coding aspects. A video coding standard for which only Option 1 declarations were made (the we call here "Option 1 standard") was certainly possible but would probably have been unattractive because of its low performance compared with state of the art codecs.
Therefore, MPEG decided that it would develop standards with the best performance, without consideration of the IPR involved. Such standards would be widely adopted by the market and patent holders would get royalties from their use. If a patent holder did not want to allow that to happen, they could make an Option 3 declaration and MPEG would remove the technology.
MPEG's success in the last 30 years proves that its business model has worked as designed. More than that, most patent holders have been and keep on re-investing the royalties they get from existing standards in more technologies for future standards.
The MPEG "business model" has created a standard-producing machine (MPEG) that feeds itself with new technologies.

Common standards for all countries and industries

In the world of analogue technologies, a combination of scarce availability of broadband communication, deliberate policies and natural separation between industries that had had until then little in common, favoured the definition of country-based or industry-based standards.
In the late 1980's many industries, regions and countries had realised that the state of digital technologies had made enough progress to enable a switch from analogue to digital.
The first steps toward digital video by countries and industries showed that the logic of the analogue world was still ruling: different countries and industries tried their own ways independently. Several companies had developed prototypes, regional initiatives were attempting to develop formats for specific countries and industries, some companies were planning products and some standards organisations were actually developing standards for their industries.
MPEG jumped in the scene at a time the different trials had not had the time to solidify, and the epochal analogue-to-digital transition gave MPEG a unique opportunity to execute its plan. The MPEG proposal of a *generic* standard, i.e. a common technology for all industries, caught the attention because it offered global interoperability, created global markets – in geography and across industries – and placed the burden of developing the costly VLSI technology on an industry accustomed to do that. Today the landscape has changed beyond recognition. Still the revolutionary idea of that time is taken as a matter of fact.

Designing common standards for different users

MPEG knew that it was *technically* possible to develop generic standards that could be used in all countries of the world and in all industries that needed compressed digital media. MPEG saw that all actors affected – manufacturers, service providers and end users – would gain if such generic standards existed. However, before treading its adventurous path, MPEG did not know whether it was *procedurally* possible to achieve that goal. But it gambled and succeeded.
Clarifying to oneself the purpose of an undertaking is a good practice that should apply to any human endeavour. This good practice is even more necessary when a group of like-minded people work on a common project – a standard. When the standard is not designed by and for a single industry but by and for many, keeping this rule is vital for the success of the effort. When the standard involves disparate technologies, whose practitioners are not even accustomed to talk to one another, complying with this rule is a prerequisite.
Starting from its early days MPEG has developed a process designed to achieve the common understanding that underpins the technical work: 1) describe the environment (context and objec-

tives), 2) single out a set of exemplary uses of the target standard (use cases), and 3) identify requirements.
MPEG does have an industry of its own. Therefore, MPEG uses its Requirements subgroup to develop generic requirements and to interact with trade associations and standards organisations of the main industries affected by its standards. Communicating its plans, the progress of its work and the results achieved more actively than other groups is vital to MPEG who uses a large number of tools to achieve that goal (see MPEG communicates).
The network of liaisons and, sometimes, joint activities (e.g., as in Genome Compression) is the asset that has allowed MPEG to achieve many of its goals, certainly the most challenging ones.

Technology standards as toolkits
The basic media compression technology should be shared by all industries, but individual industries do not necessarily need the same functions or the same performance. Therefore, industry-agnostic standardisation need to be aware of all requirements of all the industries the standard will serve but must have the means to flexibly allocate technology and performance to one industry without encumbering other unconcerned industries. MPEG has made its standards toolkit-based and has successfully applied Profiles and Levels, a notion developed in the 1980's in the context of Open System Interconnection (OSI), to create industry-specific standard configurations with a high level of interoperability.

Standards for the market, not the other way around
Standards are ethereal entities, their impact may be unpredictable but, when it is there, it can be very concrete. This was true and already well understood in the world of analogue media. At that time a company that had developed a successful product would try to get a "standard" stamp on it, share the technology with its competitors and enjoy the economic benefits of its "standard" technology.
MPEG reshuffled the existing order of steps. Instead of waiting for the *market* to decide which technology would win – an outcome that very often had little to do with the value of the technology or the product – MPEG offered its standard development process where the collaboratively defined "best standard" is developed and assessed by *MPEG experts* who decide which individual technology wins on the basis of criteria defined a priori. Then the technology package – the standard – developed by MPEG is taken over by the industry.
At a given time MPEG standards are consistently the best standards. Those who have technologies selected to be part of MPEG standards reap the benefits and most likely will continue investing in new technologies for future standards.

Standards that anticipate the future
When MPEG was specifying MPEG-1, the technology to implement the standard was not available. MPEG made a bet that industry would be able to to design circuits on silicon that could execute the complex operations and build products of which there was no concrete evidence but only guesses: interactive video on CD (not so educated) and digital audio broadcasting (much touted). Ironically, neither products really took off, at least in that time frame, but other products that relied on MPEG-1 technologies – Video CD and MP3 – were (the former) and still are (the latter) extremely successful.
When technology moves fast, or actually accelerates, waiting is a luxury no one can afford. MPEG-1 and MPEG-2 were standards whose enabling technologies were already considered by some industries and MPEG-4 (started in 1993) was a bold and successful attempt to bring media into the IT world (or the other way around). That it is no longer possible to wait is shown by MPEG-I, the current undertaking where MPEG is addressing standards for interfaces that are still shaky or just hypothetical. Having standards that lead – as opposed to trail – technology is a tough trial-

and-error game. However, it is the only one possible for digital media. MPEG has practiced this game for the last 30 years.
The alternative is to stop making standards for digital media because if MPEG waits until market needs are clear, the market is already full of incompatible solutions and there is no room left for standards.
Anticipating market needs is in the DNA of MPEG standards. With each of its standards MPEG is betting that a certain standard technology will be adopted. In Standards and uncertainty we show how some MPEG standards are extremely successful and other less so.

Compete and collaborate
MPEG favours competition to the maximum extent possible. This is achieved by calling for solutions that must be comprehensively described, i.e. without black boxes, to qualify for consideration. MPEG's line-up of aggressive "judges" (meeting participants, especially other proponents) assess the merit of proposed technologies.
Extending competition beyond a certain point, however, is counterproductive and prevents the group from reaching the goal with the best results.
MPEG develops and uses a software platform assembling the candidate components selected by the "judges" – called Test Model – as the platform where participants can work on improving the different areas of the Test Model.
Core Experiments is the tool that allows experts to improve the Test Model by adding step by step the software that implements the accepted technologies. Core Experiments were first defined in March 1992 as "a technical experiment where the alternatives considered are fully documented as part of the test model, ensuring that the results of independent experimenters are consistent". This definition applies unchanged to the work being done today.
If the MPEG mission is to provide the best standards to industry via competition, why should MPEG standards be shielded from it? Probably the earliest example of application of this principle is provided by MPEG-2 part 3 (Audio). When backward compatibility requirements did not allow the standard to yield the performance of algorithms not constrained by compatibility, MPEG issued a Call for Proposals and developed MPEG-2 part 7 (Advanced Audio Codec). Later the algorithms evolved and became the now ubiquitous MPEG-4 AAC. Had MPEG not made this decision, probably we would still have MP3 everywhere, but no other MPEG Audio standards. The latest example is Essential Video Coding (EVC), a standard not designed to offer the best performance, but a good performance with good licensability prospects.
Working on generic standards means that reasonable requests – the best unconstrained multichannel audio quality on this case – cannot be dismissed. MPEG tried to achieve that with the technology it was working on – backward-compatible multichannel audio coding – and failed. The only way to respond to the request was to work on a new technology.

Industry-friendly standards
Making MPEG standard friendly to a set of disparate industries is difficult task but a necessity that MPEG has managed in 3 ways:

1. *Display formats*: Since the appearance of television cameras and displays in the 1920's, industry and governments have created tens of television formats, mostly around the basic NTSC, PAL and SECAM families. Even in the late 1960's, when the Picturephone service was deployed, AT&T invented a new 267-line format, with no obvious connection with any of the existing video formats. As MPEG wanted to serve all markets, it decided that it would just *support any display format*, leaving display formats outside of MPEG standards.
2. *Serving one without encumbering others*. Industry may like the idea of sharing the cost of an enabling technology but not at the cost of compromising individual needs. MPEG standards share the basic technology but provide the necessary flexibility to the many different users of

MPEG standards with the notion of *Profiles and Levels*. With Profiles MPEG defines subsets of general interoperability, with Levels it defines different grades of performance within a Profile.

3. *Standards apply only to decoders*; encoders are implicitly defined and have ample margins of implementation freedom. By restricting standardisation to the decoding functionality, MPEG extends the life of its standards and, at the same time, allows industry players to compete on the basis of their constantly improved encoders.

Audio and video come together

Because of the way audio and video industries had developed – audio for a century and video for half a century – people working on the corresponding technologies tended to operate in almost "watertight" compartments, be they in academia, research or companies. That attitude had some justification in the analogue world because the relevant technologies were indeed different and there was not so much added value in keeping the *technologies* together, considering the big effort that would be needed to keep the *experts* together.

However, the digital world no longer justified keeping the two domains separate because of so many commonalities of technologies. That is why MPEG, just 6 months after its first meeting, kicked off the Audio subgroup after successfully assembling the best experts in a few months of work.

This injection of new technologies with the experts carrying them was not effortless. When transformed into digital, audio and video signals are bits and bits and bits, but the sources are still different and influence how they are compressed. Audio experts shared some high-level compression technologies – Subband and Discrete Cosine Transform – but video is (was) a time-changing *2D signal*, often with "objects" in it, while audio is (was) a time-changing *1D signal*. More importantly, audio experts were driven by other concerns than video such as the way the human hearing process handles the data coming out of the frequency analysis performed by the human cochlea and other niceties of the human hearing process.

The audio work has never been "dependent" on the video work. MPEG audio standards can have a stand-alone use (i.e. they do not assume that there is a video associated with it), but there are very few uses of MPEG video standard that do not need an MPEG Audio standard. So it was necessary to keep the two together and it is even more important to do so now when both video and audio are both time-dependent *3D signals* and users are going to actually create their own experiences.

A glue to keep audio and video together

Having audio and video together does not necessarily mean that audio and video will play together in the right way if they are stored or transmitted over a channel.

The fact that MPEG established a Digital Storage Media subgroup and a Systems subgroups 18 months after its foundation signals that MPEG has always been keenly aware of the issue that a bitstream composed by MPEG audio and video bitstreams need to be transported to be played back as intended by the bitstream creator. In MPEG-1 it was a bitstream in a controlled environment, in MPEG-2 it was a bitstream in a noisy environment, from MPEG-4 times on it was on IP, in MPEG-DASH it had to deal with unpredictability of the Internet Protocol in the real world.

During its existence the issue of multiplexing and transport formats have shaped MPEG standards. Without a Systems subgroup, efficiently compressed audio and video bitstreams would have remained floating in the space without a standard means to plug them into real systems.

Integrated standards as toolkits

Six months after its inception, MPEG had non only realised that digital media is not just video (although this is the first component that catches attention), but it is also audio (no less challenging

and with special quality requirements). In 12 months, it had realised that bits do not float in the air but that a stream of bits needs some means to adapt the stream to the mechanism that carries it (in MPEG-1 the CD-ROM). If the transport mechanism is analogue (as it was 25 years ago and, to some extent, still is today), the adaptation is even more challenging. Later MPEG also realised that a user interacts with the audio-visual bits it receives (even though it is so difficult to understand what exactly is the interaction that the user wants). With its MPEG-2 standard MPEG was able to provide the industry with a complete Audio-Video-Systems (and DSM-CC) solution whose pieces could also be used independently.
That was possible because MPEG could attract, organise and retain the necessary expertise to address such a broad problem area and provide not just a solution that worked, but the best that technology could offer at the time.
By the time it developed its earliest standards such as MPEG-1 and MPEG-2, MPEG had to assemble disparate technology competences that had probably never worked together in a project. With its example, MPEG has promoted the organisational aggregation of audio and video research in many institutions where the two were separate.
When it developed MPEG-4 (a standard with 34 parts), MPEG has assembled its largest ever number of competences ranging from audio and video to scene description, Hardware Description Languages, fonts, timed text and more. MPEG keeps competences organisationally separate in different in MPEG subgroups. However, it retains the flexibility to combine and deploy the needed resources to respond to specific needs.
Most MPEG standards are composed of the 3 key elements – audio, video and systems – that make an audio-visual system and some, such as MPEG-4 and MPEG-I, even include 3D Graphic information and the way to combine all the media. However, the standards allow maximum usage flexibility:

1. A standard can be directly used as complete solutions, e.g. like in VCD where Systems, Video and Audio are used
2. The components of the standard can be used individually, e.g. like in ATSC A/53 where Systems and Video are from MPEG, and Audio is from and external source
3. The standard does not specify a technology but only an interface to different implementations of the technology, e.g. like in the case of MPEG-I, for which MPEG will likely not standardise a Scene Description but just indicate how externally defined technologies can be plugged into the system
4. A standard does not specify the solution but only the components of a solution, e.g. like in the case of Reconfigurable Video Coding (RVC) where a non-standard video codec can be assembled using an MPEG standard.

MPEG wants to satisfy the needs of all customers, even those who do not want to use its standards but other specifications. MPEG standards can signal how an external technology can be plugged into a set of other native MPEG technologies. With one caveat: customer has to take care of the integration of the external technology. That MPEG will not do.

One step at a time

Before MPEG came to the fore many players were trying to be “first” and “impose” their early solutions to other countries, industries or companies. If the newly born MPEG had proposed itself as the developer of an ambitious generic digital media technology standard applicable to all industries, the proposal would have been seen as far-fetched and most likely the initiative would have gone nowhere.
Instead, MPEG started with a moderately ambitious project: a video coding standard for interactive applications on digital storage media (CD-ROM) at a rather low bitrate (1.5 Mbit/s) targeting the market covered by the video cassette (VHS/Beta) with the addition of interactivity.
Moving one step at a time has been MPEG policy for MPEG-1 and all its subsequent standards.

Separate wheat from chaff

In human societies parliaments make laws and tribunals decide if a specific human action conforms to the law. In certain regulated environments (e.g. terrestrial broadcasting in many countries) there are standards and entities (authorised test laboratories) who decide whether a specific implementation conforms to the standard. MPEG has neither but, in keeping with its "industry-neutral" mission, it provides the technical means – namely, tools for conformance assessment, e.g. bitstreams and reference software – for industries to use in case they want to establish authorised test laboratories for their own purposes.

Technology is always on the move

The Greek philosopher Heraclitus is reported to have said: τὰ πάντα ῥεῖ καὶ οὐδὲν μένει (everything flows and nothing stays). Digital technologies do more than that, they not only do not stay, but move fast and actually accelerate.

MPEG is well aware that the technology landscape is constantly changing, and this awareness informs its standards. Until HEVC – one can even say, including the upcoming Versatile Video Coding (VVC) standard – video meant coding a rectangular area (in MPEG-4, a flat area of any shape, in HEVC it can be a video projected on a sphere). The birth of immersive visual experiences is not without pain, but they are happening and MPEG must be ready with solutions that take this basic assumption into account. This means that, in the technology scenario that is taking shape, the MPEG role of "anticipatory standards" is ever more important and challenging to achieve.

Digital media is one of the most fast-evolving digital technology areas because most of the developers of good technologies incorporated in MPEG standards invest the royalties earned from previous standards to develop new technologies for new standards. As soon as a new technology shows interesting performance, which MPEG assesses by issuing Calls for Evidence (CfE) or the context changes offering new opportunities, MPEG swiftly examines the case, develops requirements and issues Calls for Proposals (CfP).

This has happened for most of its video and audio compression standards. A paradigmatic case of a standard addressing a change of context is MPEG Media Transport (MMT) that MPEG designed having in mind a broadcasting system for which the layer below it is IP, unlike MPEG-2 Transport Stream, originally designed for a digitised analogue channel (but also used for transport over IP as in IPTV).

Research for MPEG standards

MPEG is not in the research business. However, without research there would be no MPEG. The MPEG work plan is a great promoter of corporate/academic research because it pushes companies to improve their technologies to enable them to make successful responses to CfPs.

One of the reasons of MPEG success, but also of the difficulties highlighted in this book, is that MPEG standardisation is a process closer to research than to product design.

Roughly speaking, in the MPEG standardisation process, research happens in two phases:

1. In companies, in preparation for CfEs or CfPs (MPEG calls this competitive phase)
2. In MPEG in what is called collaborative phase, i.e. during the development of Core Experiments (of course this research phase is still done by the companies, but in the framework of an MPEG standard under development).

The MPEG collaborative phase offers another opportunity to do more research. This has apparently a more limited scope, because it is in the context of optimising a subset of the entire scope of the standard, but the sum of many small optimisations can provide big gains in performance. The shortcoming of this process is the possible introduction of a large number of IP items for a gain that some may well consider not to justify the added IP onus to complexity. With its *MPEG-5* project MPEG is trying to see if a suitably placed lower limit to performance improvements can help solve the problems identified in the HEVC standard.

Rethinking what we are
MPEG started as a "club" of Telecommunication and Consumer Electronics companies. With MPEG-2 the "club" was enlarged to Terrestrial and Satellite Broadcasting, and Cable TV concerns. With MPEG-4, IT companies joined forces. Later, a large number of research institutions and academia joined (today they count for ~25% of the total membership). With MPEG-I, MPEG faces new challenges because the demand for standards for immersive services and applications is there, but technology immaturity deprives MPEG of its usual "anchors".
Thirty years ago, MPEG invented itself and, subsequently, morphed to adapt to the changed conditions while keeping its principles intact. If MPEG will be able to continue to do as it did in the last 30 years, it can continue to support the industry it serves in the future, no matter the changes of context.

Standards as enablers, not disablers
MPEG standards are not "owned" by a specific industry. Therefore MPEG, keeping faith to its "generic standards" mission, tries to accommodate all legitimate *functional* requirements when it develops a *new* standard. MPEG assesses each requirement for its merit (value of functionality, cost of implementation, possibility to aggregate the functionality with others etc.). Profiles and Levels are then used to partition the application space in response to specific industry needs.
The same happens if an industry comes with a legitimate request to add a functionality to an *existing* standard. The decision to accept or reject a request is only driven by the value brought by the proposal, as substantiated by use cases, not because an industry gets an advantage or another is penalised.

Never stop working together
Development of basic technology is a private job, but collaboration is mandatory when requirements are defined or once a Test Model is available. This happens easily during meetings, but meetings are short events surrounded by times where official collaboration is not possible. Since its early days, MPEG has made massive use of ad hoc groups to progress collaborative work, with limitation such as that work of an ad hoc group is limited to preparation of recommendations to MPEG.

The nature and borders of compression
What is the meaning of compression? Is it "less bits is always good" or can it also be "as few *meaningful* bits as possible is also good"? The former is certainly desirable but, as the nature of information consumption changes and compression digs deeper in the nature of information, compressed representations that offer easier access to the information embedded in the data becomes more valuable.
What is the scope of application of MPEG compression? When MPEG started the MPEG-1 standards work, the gap between the telecom and the CE industries (the first two industries in attendance at that time) were as wide as between the media industry and, say, the genomic industries today. Both are digital now and the dialogue gets easier.
With patience and determination MPEG has succeeded creating a common language and mindset in the media industries. This is an important foundation of MPEG standards. The same amalgamation can continue and be achieved – not in a day – between MPEG and other industries.

5 The MPEG operation

Over the years, MPEG has developed a large number of standards and specifications following the principles described in the Chapter 4 and implemented them in an efficient way. This operational practice obviously respects the general rules laid down in the ISO/IEC directives.

1. The MPEG organisation describes MPEG's internal organisation;
2. Organisation of work describes how the expertise of MPEG participants is organised to develop standards;
3. How MPEG develops standards describes the specific steps MPEG undertakes to develop standards
4. The ecosystem drives MPEG standards explains how MPEG is part of an ecosystem where all participants play a role;
5. Standardisation and product making compares the task of developing a standard with the task of developing a product;
6. Standards are living beings explains why publishing a standard is a new beginning, not the end of standardisation work;
7. Standards and uncertainty concludes that is no recipe to design a guaranteed successful MPEG standard;
8. MPEG communicates identifies communication as the most important bond keeping the ecosystem together.

5.1 The MPEG organisation

The MPEG membership

MPEG's most precious assets are its members. As MPEG is an ISO working group, to be entitled to attend and actively participate, experts and/or their companies must be members of one of the 162 national standards organisations who are members of ISO (or as an approved liaison officer from another organisation). Respondents to an MPEG Call for Evidence (CfE) or Call for proposals (CfP) are allowed to attend the meeting where their proposal is presented for the first time. To continue attending, however, proponents are required to become formal members, and it is certainly in their interest if they want their proposals to stay part of the standard.

The MPEG workflow

In its 30 years MPEG has created and disbanded several groups created for specific needs and *Figure 4* depicts the current MPEG organisation. More about the history of the MPEG groups and their chairs can be found here.

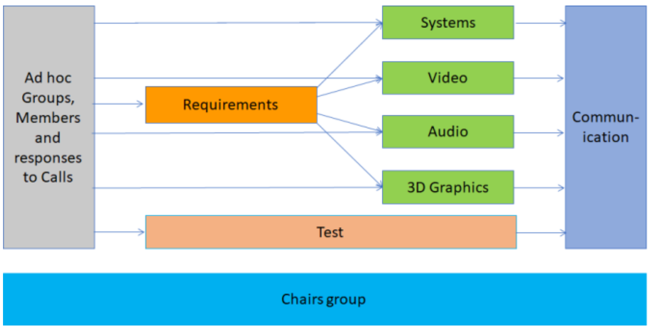


Figure 4 – The MPEG workflow

New ideas are typically submitted to the Requirements group. If they are considered worth of further explorations, they are typically discussed in ad hoc groups (AhG) after the meeting. An AhG will report its findings at the following MPEG meeting. After a few iterations, MPEG will produce a Call for Evidence (CfE) or a Call for Proposals (CfP) with due publicity in the Press Release. At that time, an AhG is charged with the task to disseminate the CfE or CfP, prepare for the logistics of the tests and perform a first assessment of the responses.

This does not mean that the Requirements group is no longer involved in this standard because the group typically continues the development of Use Cases and Requirements while the technical work progresses. When necessary, requirements are reviewed with the appropriate technology groups and may give rise to new CfPs whose outcome is fed into the work of the relevant technology groups after an assessment.
Today the task of the Test group is no longer confined to the assessment of the quality of proposals at CfE and CfP time. Designing and executing appropriate quality tests in support to Core Experiments has become the norm especially in the Video group.
If a Call requests Proposals, the Requirements group, in conjunction with the Test group and possibly one or more technology group, reviews the result of the AHG and makes a final assessment. If the technologies submitted are judged to be sufficient to initiate the development of a standard, the activity is transferred to that/those group(s) for standard development.
The current MPEG Chairs are given in *Table 1*. They come from a variety of countries (CN, DE. FR, IT, KR, US) and organisations (small/medium size and large companies, and Universities).

Table 2 – MPEG subgroup chairs

Group name	**Chair name**	**Affiliation**
Requirements	Jörn Ostermann	Leibniz University Hannover
Systems	Youngkwon Lim	Samsung
Video	Lu Yu	Zhejiang University
	Jens-Rainer Ohm	RWTH Aachen
	Gary Sullivan	Microsoft
Audio	Schuyler Quackenbush	Audio Research Labs
3DGC	Marius Preda	Institut Mines Télécom
Test	Vittorio Baroncini	GBTech
Communication	Kyuheon Kim	Kyunghee University

Meeting documents
Feeding documents into the MPEG process is of vital importance because the number of submissions uploaded can easily cross the level of 1000-1500 documents per meeting. Today it is hard to imagine how MPEG could operate if it were still "paper-based". MPEG shed its skin in October 1995, when Pete Schirling (then with IBM) developed a document management system that allowed members to upload their documents and download those uploaded by all other members. Wo Chang (NIST) in 2000 and Christian Tulvan (Institut Mines Télécom) in 2005 took over and improved the system that ensures the high efficiency of MPEG operation.

How to know what happens where
For MPEG experts a place with so many "hotbeds" discussing requirements, and assessing, integrating and testing media technologies is exciting, but how can MPEG experts know what happens when and where at a meeting? Christian Tulvan has come to help again and designed a system that answers that question. MPEG members can have a full view of which meetings are held where and when to discuss which topics or documents. The design is responsive so MPEG experts can get the information from their smartphones.
Figure 5 shows how the service looked like at MPEG 124 (October 2018). This particular view shows all meetings of all groups, but filtered views are possible. By clicking on a meeting, the user obtains details (room, links to documents to be discussed etc.).

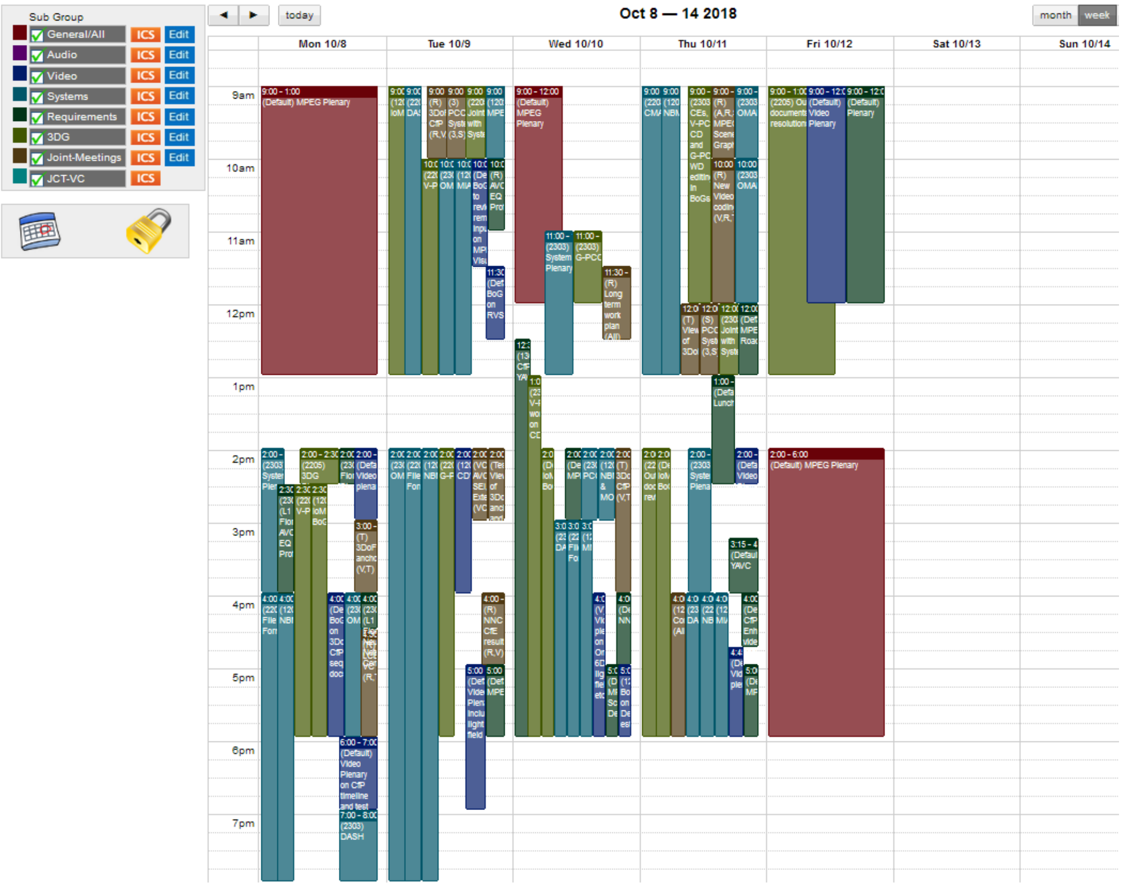
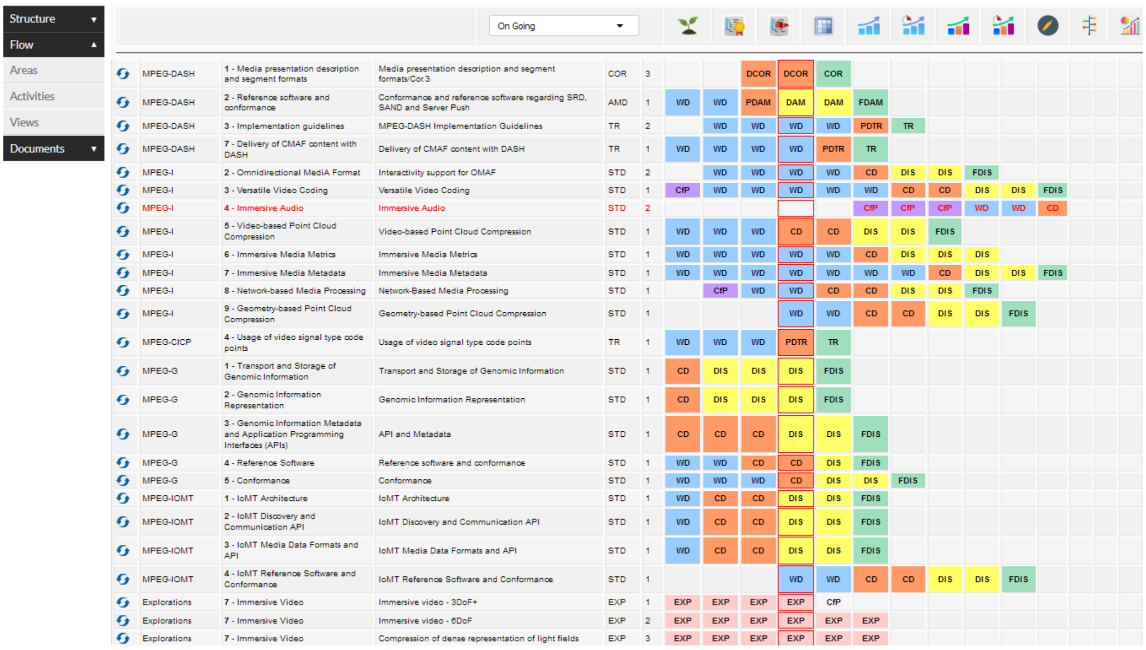


Figure 5 – Coping with many parallel meetings



MPEG-DASH	1 - Media presentation description and segment formats	Media presentation description and segment formats/Cor.3	COR	3			DCOR	DCOR	COR						
MPEG-DASH	2 - Reference software and conformance	Conformance and reference software regarding SRD, SAND and Server Push	AMD	1	WD	WD	PDAM	DAM	DAM	FDAM					
MPEG-DASH	3 - Implementation guidelines	MPEG-DASH Implementation Guidelines	TR	2		WD	WD	WD	WD	PDTR	TR				
MPEG-DASH	7 - Delivery of CMAF content with DASH	Delivery of CMAF content with DASH	TR	1	WD	WD	WD	WD	PDTR	TR					
MPEG-I	2 - Omnidirectional MediA Format	Interactivity support for OMAF	STD	2		WD	WD	WD	WD	CD	DIS	DIS	FDIS		
MPEG-I	3 - Versatile Video Coding	Versatile Video Coding	STD	1	CfP	WD	WD	WD	WD	WD	CD	CD	DIS	DIS	FDIS
MPEG-I	4 - Immersive Audio	Immersive Audio	STD	2						CfP	CfP	CfP	WD	WD	CD
MPEG-I	5 - Video-based Point Cloud Compression	Video-based Point Cloud Compression	STD	1	WD	WD	WD	CD	CD	DIS	DIS	FDIS			
MPEG-I	6 - Immersive Media Metrics	Immersive Media Metrics	STD	1	WD	WD	WD	WD	WD	CD	DIS	DIS	DIS		
MPEG-I	7 - Immersive Media Metadata	Immersive Media Metadata	STD	1	WD	WD	WD	WD	WD	WD	WD	CD	DIS	DIS	FDIS
MPEG-I	8 - Network-based Media Processing	Network-Based Media Processing	STD	1		CfP	WD	WD	CD	CD	DIS	DIS	FDIS		
MPEG-I	9 - Geometry-based Point Cloud Compression	Geometry-based Point Cloud Compression	STD	1				WD	WD	CD	CD	DIS	DIS	FDIS	
MPEG-CICP	4 - Usage of video signal type code points	Usage of video signal type code points	TR	1	WD	WD	WD	PDTR	TR						
MPEG-G	1 - Transport and Storage of Genomic Information	Transport and Storage of Genomic Information	STD	1	CD	DIS	DIS	DIS	FDIS						
MPEG-G	2 - Genomic Information Representation	Genomic Information Representation	STD	1	CD	DIS	DIS	DIS	FDIS						
MPEG-G	3 - Genomic Information Metadata and Application Programming Interfaces (APIs)	API and Metadata	STD	1	CD	CD	CD	DIS	DIS	FDIS					
MPEG-G	4 - Reference Software	Reference software and conformance	STD	1	WD	WD	CD	CD	DIS	FDIS					
MPEG-G	5 - Conformance	Conformance	STD	1	WD	WD	WD	CD	DIS	DIS	FDIS				
MPEG-IOMT	1 - IoMT Architecture	IoMT Architecture	STD	1	WD	CD	CD	DIS	DIS	FDIS					
MPEG-IOMT	2 - IoMT Discovery and Communication API	IoMT Discovery and Communication API	STD	1	WD	CD	CD	DIS	DIS	FDIS					
MPEG-IOMT	3 - IoMT Media Data Formats and API	IoMT Media Data Formats and API	STD	1	WD	CD	CD	DIS	DIS	FDIS					
MPEG-IOMT	4 - IoMT Reference Software and Conformance	IoMT Reference Software and Conformance	STD	1				WD	WD	CD	CD	DIS	DIS	FDIS	
Explorations	7 - Immersive Video	Immersive video - 3DoF+	EXP	1	EXP	EXP	EXP	EXP	CfP						
Explorations	7 - Immersive Video	Immersive video - 6DoF	EXP	2	EXP	EXP	EXP	EXP	EXP	EXP					
Explorations	7 - Immersive Video	Compression of dense representation of light fields	EXP	3	EXP	EXP	EXP	EXP	EXP	EXP					

Figure 6 – An example timeline of MPEG standards

Project management
Another major development made by Christian provides the chairs with a series of functionalities to manage the MPEG workplan and timeline. Some of these, like the one below, are open to MPEG members.
Figure 6 shows the MPEG timeline for the portion related to the MPEG-I, MPEG-CICP and MPEG-G standards. On the top there are icons to create new activities, show different table previews, provide project-wise views and more.
Another remarkable functionality is the possibility to create the document that collects all the results from a meeting. Group chairs enter their own results independently and the system organises them by standards and integrates them in a single document for approval at the Friday plenary. This has made the review of decisions shorter and allowed MPEG to slash the time taken by plenaries.

MPEG assets
MPEG defines as assets the data associated with the development of standards: URIs, Schemas, Content, Software and Conformance testing data. The first two are publicly available.
By Content assets we mean the collection of all test data (content) used for CfEs and CfPs, and in subsequent Core Experiments. MPEG relies on the good will of companies interested in a standard to provide relevant test data. These are typically licensed for exclusive use in MPEG standardisation because they typically have high value, e.g. from the content and technology viewpoints. Content owners license their content directly to individual MPEG experts.
Software assets is the collection of all software developed or under development as reference software of MPEG standards. With the approach taken by MPEG to develop most of its standards from reference software and give it a normative status, its importance can hardly be overstated.
Conformance testing data is the collection of all data that can be used to test an implementation for conformance to an MPEG standard, published or under development. The process of developing conformance testing suites is painstaking and depends on the good will of MPEG members and their companies. On the other hand, conformance testing suites are vital for the creation of ecosystems of interoperable implementations.
Content, Software and Conformance testing data for most data used in the development of MPEG standards in the last 30 years are stored in the Media Repository, Software Repository and Conformance Testing Repository hosted by Institut Mines Télécom. We are talking of a total of 8 TeraBytes of data.

Conclusions
This article can only give an idea of the level of commitment and dedication of so many MPEG members – with the support of their companies. Of course, MPEG standards are produced by MPEG experts, but the MPEG process works because so many other people and organisations contribute, some in a visible and some others in a less visible form, to make MPEG what it is.

5.2 Organisation of work

No one, reading MPEG standards, should deny that the spectrum of MPEG standards is an impressive set of disparate technologies integrated to cover fields connected by the common thread of technologies underpinning Data Compression: Coding of Video, Audio, 3D Graphics, Fonts, Digital Items, Sensors and Actuators Data, Genome, and Neural Networks; Media Description and Composition; Systems support; Intellectual Property Management and Protection (IPMP); Transport; Application Formats; API; and Media Systems.
How are all these technologies specified and integrated in MPEG standards to respond to industry needs? This chapter will try and answer this question by starting, as many novels do, from the end (of an MPEG meeting).

When an MPEG meeting closes, the plenary approves the results of the week. This marks the end of formal collaborative work within the meeting. Back in 1990 MPEG developed a mechanism – called "ad hoc group" (AhG) – that would allow experts to continue a form of collaboration. AhGs allow MPEG experts to continue working together, albeit with limitations:

1. In the *scope*, i.e. an AhG may only work on the areas identified by the mandates (in Latin *ad hoc* means "for the specific purpose"). Of course, experts are free to work individually on anything and in any way that pleases them and submit their independent results at the next meeting;
2. In the *purpose*, i.e. an AhG may only prepare recommendations – in the scope of its mandates – to be submitted to MPEG. This is done at the beginning of the following meeting, after which the AhG is disbanded;
3. In the *method of work*, i.e. an AhG operates under the leadership of one or more Chairs. It is clear that the success of an AhG depends very much on the attitude and activity of its members.

On average some 25 AhGs are established at each meeting. There is no one-to-one correspondence between MPEG activities and AhGs. AhGs are actually great opportunities to explore new and possibly cross-domain ideas.

Examples of AhG titles are

1. Scene Description for MPEG-I
2. System technologies for Point Cloud Coding (PCC)
3. Network Based Media Processing (NBMP)
4. Compression of Neural Networks (NNR).

What happens between MPEG meetings

An AhG uses different means to carry out collaborative work: by using email reflectors, by teleconferencing and by holding physical meetings. The last can only be held if they were scheduled in the AhG establishment form. Unscheduled physical meetings may only be held if those who subscribed to the AhG unanimously agree.

Most AhGs hold scheduled meetings on the weekend that precedes the next MPEG meeting. These are very useful to coordinate the results of the work done and to prepare the report that all AhGs must make to the MPEG plenary on the following Monday.

It should be noted that AhG meetings, including those in the weekend preceding the MPEG meeting, are not formally part of an MPEG meeting.

MPEG chairs meet on the *Sunday* evening to review the progress of AhG work, coordinate activities impacting more than one Subgroup and plan activities to be carried out during the week including identification of joint meetings

An MPEG meeting at a glance

Plenary meetings

During an MPEG week MPEG holds 3 plenaries

1. On *Monday* morning: to make everybody aware of the results of work carried out since the last meeting and to plan work of the week. AhG reports are a main part of it as they are presented and, when necessary, discussed;
2. On *Wednesday* morning to make everybody aware of the work done in all subgroups in the first two days and to plan work for the next two days;
3. On *Friday* afternoon to approve the results of the work of Subgroups, including liaison letters, establishment of new AhGs etc.

Subgroup and Breakout Group

Subgroups start their meetings on Monday afternoon. They review their own activities and kick off work in their areas. Each subgroup assigns activities to breakout groups (BoG) who meet with

their own schedules to achieve the goals assigned. Each Subgroup may hold other brief meetings to keep everybody in the Subgroup in sync with the general progress of the work.
For instance, the activities of the Systems Subgroups are currently (March 2019): File format, DASH, OMAF, OMAF and DASH, CMAF and MIAF, MPEG Media Transport, Network Based Media Processing and PCC Systems.
On Friday morning all Subgroups approve their own results. These are automatically integrated in the general document to be approved by the MPEG Plenary on Friday afternoon.

Joint meetings
A key MPEG feature is the immediate availability of the necessary technical expertise to discuss matters that cross organisational boundaries. The speed of development and quality of MPEG standards would hardly be possible if it were not possible to deploy the necessary expertise to address multi-faceted issues in a timely manner.
Let's take as an example this video: the basketball player is represented as a compressed dynamic point cloud in a 360° video. The issue of integrating the two components may be raised at a meeting and/or discussed by the Chairs where the need for a joint meeting is identified. This is proposed to the MPEG plenary and held with the participation of Requirements, Systems and 3D Graphics Coding experts. The outcome of such a meeting may be anything from the acknowledgement that "we don't know well enough yet", to the identification of technologies that can simply be developed as a collaborative effort of MPEG experts or require a CfP.
For example, the table below lists the joint meetings that the Systems Subgroup held with other Subgroups at the January 2019 meeting.

Table 3 – Joint meetings of Systems Subgroup with other Subgroups

Systems meeting with	Topics
Reqs, Video, VCEG	SEI messages in VVC
Audio, 3DG	Scene Description
3DG	Systems for Point Cloud Compression
3DG	API for multiple decoders
Audio	Uncompressed Audio
Reqs, JVET, VCEG	Immersive Decoding Interface

NB: VCEG is Video Coding Experts Group of ITU-T Study Group 16, not an MPEG Subgroup.

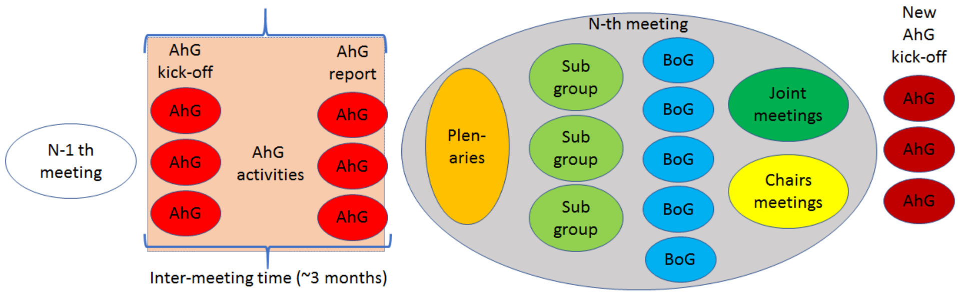


Figure 7: MPEG work from the end of a meeting to the end of the next meeting

Chairs meeting
The Chairs meet on *Tuesday* evening to assess the result of the first two days of work, review the work plan and time lines based on the expected outcomes and identify the need of new joint meet-

ings, and on *Thursday* evening to wrap up the expected results and review the preliminary version of the results of the meeting.

A bird's eye view of an MPEG meeting

Figure 7 depicts the workflow described in the paragraphs above, starting from the end of the N-1 th meeting to the end of the N-th meeting.

What is "done" at an MPEG meeting?

There are around 500 of the best worldwide experts attending an MPEG meeting, an event that mobilises an incredible amount of brain power. The following explains how this brain power is directed.

An example – the video area

Let's take as example the work done in the Video Coding area at the March 2019 meeting by looking at the 3 columns of
Table 4:

1. The **standards** on which work is done (currently Video works on MPEG-H, MPEG-I, MPEG-CICP, MPEG-5 and Explorations)
2. The names of the **activities**
3. The types of **documents** resulting from the activities (see the legend after
4. *Table 4* for an explanation of the acronyms).

Table 4 – Documents produced in the video coding area

Std	Activity	Document type
H	High Efficiency (HEVC)	TM, CE, CTC
I	Versatile Video Coding (VVC)	WD, TM, CE, CTC
	3 Degrees of Freedom + (3DoF+) coding	CfP, WD, TM, CE, CTC
CICP	Usage of video signal type code points (Ed. 1)	TR
	Usage of video signal type code points (Ed. 2)	WD
5	Essential Video Coding	WD, TM, CE, CTC
	Low Complexity Enhancement Video Coding	CfP, WD, TM, CE, CTC
Expl	6 Degrees of Freedom (6DoF) coding	EE, Tools
	Coding of dense representation of light fields	EE, CTC

Legend

TM	Test Model, software implementing the standard (encoder & decoder)
WD	Working Draft
CE	Core Experiment, i.e. definition of and experiment that should improve performance
CTC	Common Test Conditions, to be used by all CE participants
CfP	Call for Proposals (this time no new CfP produced, but reports and analyses of submissions in response to CfPs)
TR	Technical Report (ISO document)
EE	Exploration Experiment, an experiment to explore an issue because it si not mature enough to be a CE
Tools	Other supporting material, e.g. software developed for common use in CEs/EEs

The outcome of an MPEG meeting

Figure 8 gives the number of activities for each type of activity defined in the legend (and others that were not part of the work in the video area).

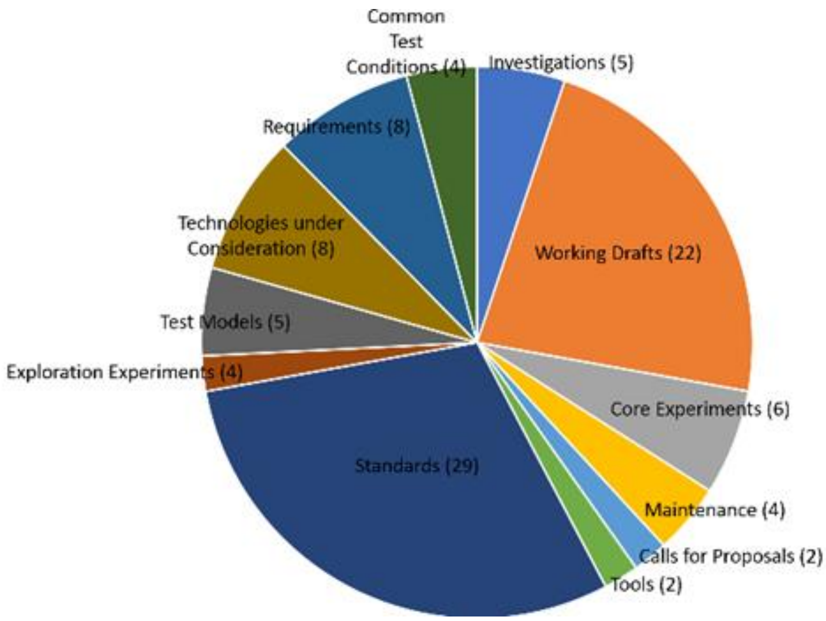


Figure 8: Activities at an MPEG meeting

For instance, out of a total of 97 activities:

1. 29 relate to *processing of standards* through the canonical stages of Committee Draft (CD), Draft International Standard (DIS) and Draft International Standard (FDIS) and the equivalent for Amendments, Technical Reports and Corrigenda. In other words, at every meeting MPEG is working on ~10 "deliverables" (i.e. standards, amendments, technical reports or corrigenda) at different stages of the approval process;
2. 22 relate to *working drafts*, i.e. "new" activities that have not entered the approval stages;
3. 8 relate to *Technologies under Consideration*, i.e. new technologies that are being considered to enhance existing standards;
4. 8 relate to *requirements*, typically for new standards;
5. 6 relate to *Core Experiments;*
6. Etc.

Figure 8 does not provide a quantitative measure of "how many" documents were produced for each activity or "how big" they were. As an example, Point Cloud Compression has 20 Core Experiments and 8 Exploration Experiments under way, while MPEG-5 EVC has only one large CE. An average value of activity at the March 2019 meeting is provided by dividing the number of output documents (212), by the number of activities (97), i.e. 2.2.

Conclusions

MPEG holds quarterly meetings with an attendance of ~500 experts. If we assume that the average salary of an MPEG expert is 500 $/working day and that every expert stays 6 days (to account for attendance at AhG meetings), the industry investment in attending MPEG meetings is 1.5 M$/meeting or 6 M$/year. Of course, the total yearly investment made by the industry is more than that.

With the meeting organisation described above MPEG tries to get the most out of industry investment in MPEG standardisation.

5.3 How MPEG develops standards

The discovery of algorithms that enable better and/or new audio-visual user experiences may trigger the development and deployment of highly rewarding solutions. By joining MPEG, companies owning good technologies have the opportunity to make them part of a standard that will help industry develop top performance and interoperable products, services and applications. *Figure 9* depicts how MPEG has extended the ISO standard development process to suit the need of a field where sophisticated technologies are needed to make viable standards.

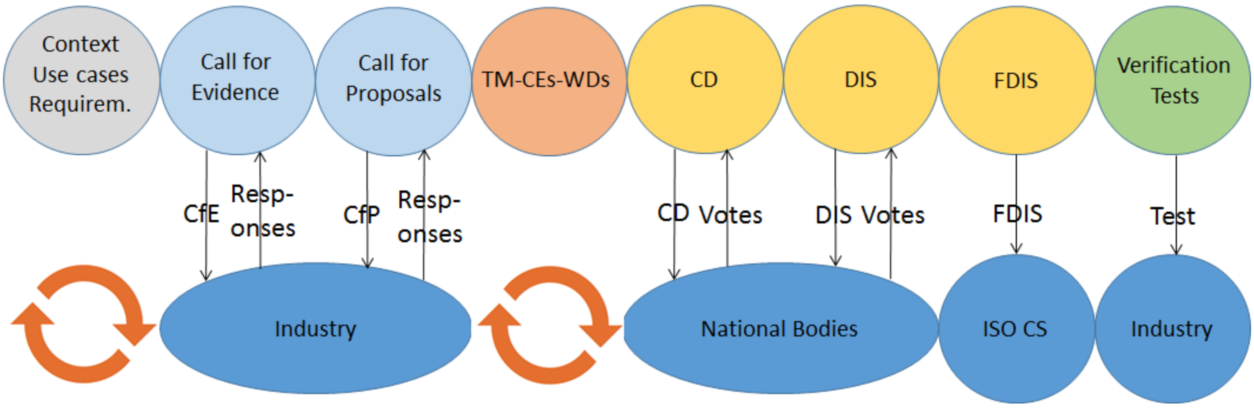


Figure 9: The MPEG process to develop standards

Members bring ideas to MPEG for presentation and discussion at MPEG meetings. If the idea is found interesting and promising, an ad hoc group is typically established to explore further the opportunity until the next meeting. The proposer of the idea is typically appointed as chair of the ad hoc group. In this way MPEG offers proposers the opportunity to become the *MPEG entrepreneurs* who can convert a collective experts' idea into an MPEG standard.

To get to a standard it usually takes some time: the newer the idea, the more time it may take to make it actionable by the MPEG process. After the idea has been clarified, the first step is to understand:

1) the context for which the idea is relevant;
2) the use cases the idea offers advantages for;
3) the requirements that a solution should satisfy to support the use cases.

Even if idea, use context, use cases and requirements have been clarified, it does not mean that technologies necessarily exist out there that can be assembled to provide the needed solution. For this reason, MPEG typically produces and publishes a *Call for Evidence* (CfE) - sometimes more than one - with attached context, use cases and requirements. The CfE requests companies who think they have technologies satisfying the requirements to demonstrate what they can achieve. At this stage companies are *not requested to describe* how the results have been achieved, they are *only requested to show* the results. In many cases respondents are requested to use specific test data to facilitate comparison between different demonstrations. If MPEG does not have those test data, it will request industry (non-necessarily only MPEG members) to provide them.

If the result of the CfE is positive, MPEG will move to the next step and publish a Call for Proposals (CfP), with attached context, use cases, requirements, test data and evaluation method. The CfP requests companies who have technologies satisfying the requirements to submit responses to the CfP where they demonstrate the performance *and* disclose the exact nature of the technologies that achieve the demonstrated results.

Let's see how this process has worked out in the specific case of neural network compression (*Table 5*).

We can see that Use Cases and Requirements are updated at each meeting and made public so as to get feedback even from people not part of the ISO process, Test data are requested to the industry and the Evaluation Framework is developed well in advance of the CfP. In this particular case it has taken 18 months just to move from the idea of compressing neural networks to CfP responses.

Table 5 –Example of idea up to Call for Proposals

Mtg	YY	MM	Actions
120	17	Oct	Presentation of idea of compressing neural networks
			Approval of CNN in CDVA (document)
121	18	Jan	Release of Use cases and requirements
122	18	Apr	New release of Use cases and requirements
			Release of Call for Test Data
			Approval of Draft Evaluation Framework
123	18	Jul	Release of Call for Evidence, Use cases and requirements
			New release of Call for Test Data
			New version of Draft Evaluation Framework
124	18	Oct	Approval of Report on the CfE
			Release of CfP, Use cases and requirements, Evaluation Framework
126	19	Mar	Responses to Call for Proposals
			Evaluation of responses based on Evaluation Framework
			Definition of Core Experiments

A meeting where CfP submissions are due is typically a big event for the community involved. Knowledgeable people say that such a meeting is often more intellectually rewarding than attending a conference. How could it be otherwise if participants not only can understand and discuss the technologies but also see and judge their actual performance? Everybody feels like being part of an engaging process of building a "new thing". Read here an independent view.

If MPEG concludes that the technologies submitted and retained are sufficient to start the development of a standard, the work is moved from the Requirements subgroup, which typically handles the process of moving from idea to proposal submission and assessment, to the appropriate technical group. If not, the idea of creating a standard is – maybe temporarily – dropped or further studies are carried out or a new CfP is issued calling for the missing technologies.

If work proceeds, members involved in the discussions need to decide which technologies are useful to build the standard. Results are there but questions pop up from all sides. Those meetings are good examples of the "survival of the fittest", a principle that applies to living beings as well as to technologies.

Eventually the group identifies useful technologies from the proposals and builds an initial Test Model (TM) of the solution. This is the starting point of a cyclic process where MPEG experts do the following

1. Identify critical points of the TM;
2. Define which experiments – called Core Experiments (CE) – should be carried out to improve TM performance;
3. Review members' submissions;
4. Review CE technologies and adopt those bringing sufficient improvements.

At the right time (which may very well be the meeting where the proposals are reviewed or several meetings after), the group produces a Working Draft (WD). The WD is continually improved following the 4 steps above.

The start of a new standard is typically not without impact on the MPEG ecosystem. Members may wake up to the need to support new requirements or they realise that specific applications may require one or more "vehicles" to embed the technology in those application or they may conclude that the originally conceived solution needs to be split in more than one standard.

These and other events are handled by convening joint meetings between the group developing the technology and technical "stakeholders" in other groups.

Eventually MPEG is ready to "go public" with a document called Committee Draft (CD). However, this only means that the solution is submitted to the national standard organisations – National Bodies (NB) – for consideration. NB experts vote on the CD with comments. If a sufficient number of positive votes are received (this is what has happened for all MPEG CDs so far), MPEG assesses the comments received and decides on accepting or rejecting them one by one. The result is a new version of the specification – called Draft International Standard (DIS) – that is also sent to NBs where it is assessed again by national experts, who vote and comment on it. MPEG reviews NB comments for the second time and produces the Final Draft International Standard. This, after some internal processing by ISO, is eventually published as an International Standard.

MPEG typically deals with complex technologies that companies consider "hot" because they are urgently needed for their products. As much as in companies the development of a product goes through different phases (alpha/beta releases etc., in addition to internal releases), achieving a stable specification requires many reviews. Because CD/DIS ballots may take time, experts may come to a meeting reporting bugs found or proposing improvements to the document out for ballot. To take advantage of this additional information, that the group scrutinises for its merit, MPEG has introduced an unofficial "mezzanine" status of the document called "Study on CD/DIS" where proposals for bug fixes and improvements are added to the document under ballot. These "Studies on CD/DIS" are communicated to the NBs to facilitate their votes on the official documents under ballot.

Let's see in the table below how this part of the process has worked for the Point Cloud Compression (PCC) case. Only the most relevant documents have been retained.

Table 6 – An example from CfP to FDIS

Mtg	YY	MM	Actions – Approval of
120	17	Oct	Report on PCC CfP responses
			7 CEs related documents
121	18	Jan	3 WDs
			PCC requirements
			12 CEs related documents
122	18	Apr	2 WDs
			22 CEs related documents
123	18	Jul	2 WDs
			27 CEs related documents
124	18	Oct	V-PCC CD
			G-PCC WD
			19 CEs related documents
			Storage of V-PCC in ISOBMFF files
128	19	Oct	V-PCC FDIS
129	20	Jan	G-PCC FDIS

I would like to draw your attention to "Approval of 27 CEs related documents" in the July 2018 row. The definition of each of these CE documents requires lengthy discussions by involved experts because they describe the experiments that will be carried out by different parties at different locations and how the results will be compared for a possible decision. It should not be a surprise if some experts work from Thursday until Friday morning to get documents approved by the Friday plenary.

I would also like to draw your attention to "Approval of Storage of V-PCC in ISOBMFF files" in the October 2018 row. This is the start of a Systems standards that will enable the use of PCC in specific applications, as opposed to being just a "data compression" specification.

The PCC work item is currently seeing the involvement of some 80 experts. If you think that there were more than 500 experts at the October 2018 meeting, you should have a pretty good idea of the complexity of the MPEG machine and the amount of energies poured in by MPEG members.
In most cases MPEG performs "Verification Tests" on a standard produced to provide the industry with precise indications of the performance that can be expected from an MPEG compression standard. To do this the following is needed: specification of tests, collection of appropriate test material, execution of reference or proprietary software, execution of subjective tests, test analysis and reporting.
Very often, as a standard takes shape, new requirements for new functionalities are added. They become part of a standard either through the vehicle of "Amendments", i.e. separate documents that specify how the new technologies can be added to the base standard, or "New Editions" where the technical content of the Amendment is directly integrated into the standard in a new document.
As a rule, MPEG develops reference software for its standards. The reference software has the same normative value as the standard expressed in human language. Neither prevails on the other. If an inconsistency is detected, one is aligned to the other.
MPEG also develops conformance tests, supported by test suites, to enable a manufacturer to judge whether its implementation of 1) an *encoder* produces correct data by feeding them into the reference decoder or 2) a *decoder* is capable to correctly decode the test suites.
Finally, it may happen that bugs are discovered in a published standard. This is an event to be managed with great attention because industry may already have released implementations on the market.
MPEG is definitely a complex machine because it needs to assess if an idea is useful in the real world, understand if there are technologies that can be used to make a standard supporting that idea, get the technologies, integrate them, develop the standard and test the goodness of the standard for the intended purpose. Often it also has to provide integrated audio-visual solutions where a line-up of standards nicely fit to provide a system specification.
MPEG needs to work like a company developing products, but it is not a company. Fortunately, one can say, but also unfortunately because it has to operate under strict rules that apply uniformly to thousands of ISO committees, working for very different industries with very different mindsets.

5.4 The ecosystem drives MPEG standards

Standards making changes with time

In days long bygone, standardisation in what today we would call the “media industry” followed a rather simple process. A company wishing to attach a “standard” label to a product that had succeeded in the market made a request to a standards committee whose members, typically from companies in the same industry, had an interest in getting an open specification of what had until then been a closed proprietary system. A good example is offered by the video cassette player for which two products from two different companies, ostensibly for the same functionality – VHS and Betamax – were approved by the same standard organisation – the International Electrotechnical Committee (IEC) and by the same committee – SC 60 B at that time.
Things were a little different in the International Telecommunication Union (ITU) where ITU-T (then called CCITT) had a Study Group where the telecommunication industry – represented by the Post and Telecommunication Administrations of the member countries, at that time the only ones admitted to the committee – requested a standard (called recommendation in the ITU) for digital telephony speech. ITU-T ended up with two different specifications in the same standard: one called A-law and the other called µ-law.
In ITU-R (then called CCIR) National Administrations were operating, or had authorised various entities to operate, television broadcasting services (some had even started services before WW II) and were therefore unable to settle on even a limited number of television systems. The only thing they could do was to produce a document called Report 624 Television Systems that collected the

3 main television systems (NTSC, PAL and SECAM) with tens of pages where country A selected, e.g., a different frequency or a different tolerance of the colour subcarrier than country B or C. Not unaware of past failures of standardisation and taking advantage of the radical technology discontinuity, MPEG took a different approach to standardisation which can be expressed by the synthetic expression "one functionality – one tool". To apply this expression to the example of ITU-T's A-law – µ-law dichotomy, if MPEG had to decide on a standard for digital speech, it would

1. Develop requirements
2. Select speech samples to be used for tests
3. Issue a Call for Proposals (CfP)
4. Run the selected test speech with the proposals
5. Subjectively assess the quality
6. Check the proposals for any issue such as complexity etc.
7. Create a Test Model with the proposals
8. Create Core Experiments (CE)
9. Iterate the Test Model with the results of CEs
10. Produce WD, CD, DIS and FDIS

The process would be long – an overkill in this case because a speech digitiser is just a simple analogue-to-digital (A/D) converter – but not necessarily longer that waiting for a committee to decide on competing proposals with the goal of selecting only one. MPEG's result would be a single standard providing seamless bitstream interoperability without the need to convert speech from one format to another when speech moves from one environment (country, application etc.) to another.

If there were only the 10 points listed above, the MPEG process would not be much more complex than the ITU's. The real difference is that MPEG does not have the mindset of the telecom industry who had decided A-law – µ-law digital speech 50+ years ago because it serves a large number of industries.

Where lies the MPEG difference

MPEG is different because it would address speech digitisation taking into consideration the needs of a *range of other industries* who intend to use and hence want to have a say in how the standard is made: Consumer Electronic (CE), Information Technology (IT), broadcasting, telecommunications and more. Taking into account so many views is a burden for those developing the standard, but the standard eventually produced is abstracted from the small (or big) needs that are specific of individual industries. Profiles and Level allow an industry not to be overburdened by technologies introduced to satisfy requirements from other industries that are irrelevant (and possibly costly) to that industry. Those who need the functionality, not matter what the cost, can do it with different profiles and levels.

Figure 10 depicts how MPEG has succeeded in its role of "abstracting" the needs of client "digital media" industries currently served by MPEG.

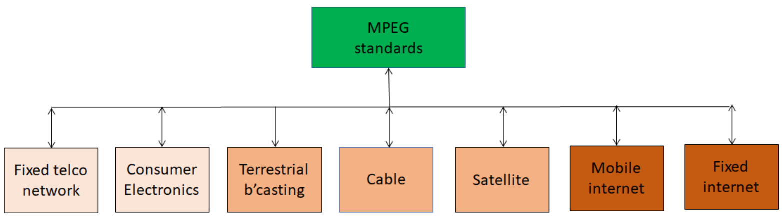


Figure 10: MPEG and its client "digital media" industries

Figure 10, however, does not describe all the ecosystem actors. In MPEG-1 the Consumer Electronics industry was typically able to develop by itself the technology needed to make products that used the MPEG-1 standard. With MPEG-2 this was less the case and independent companies offering encoding and decoding chips sprang up. Today the industry implementing (as opposed to using or selling products based on) MPEG standards has grown to be a very important element of the MPEG ecosystem. This industry typically provides components to companies who actually manufacture a complete product (sometimes this happens inside the same company, but the logic is the same).

MPEG standards can be implemented using various combinations of software, hardware and hybrid software/hardware technologies. The choice for hardware is very wide: from various integrated circuit architectures to analogue technologies. The latter choice is for devices with extremely low power consumption, although with limited compression. Just about to come are devices that use neural networks. Other technologies are likely to find use in the future, such as quantum computing or even genomic technologies.

The MPEG component industries

Figure 11 identifies 3 "layers" in the MPEG ecosystem.

The red arrows show the flow of Requirements and Standards and the violet arrows show the flow of Implementation requests and Implementations.

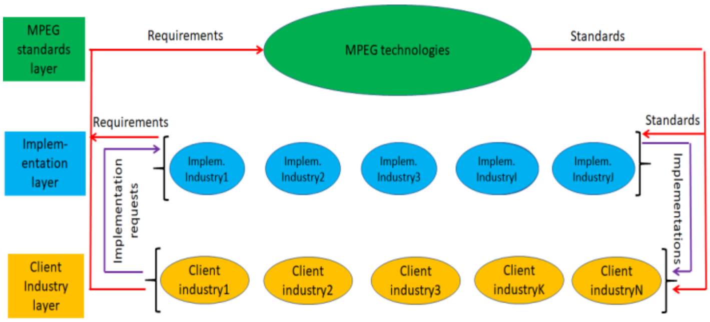


Figure 11: MPEG, its Client and Implementation Industries

Client industries in need of a standard provide requirements. However, the "Implementation layer" industries, examples of which have been provided above, also provide requirements. The MPEG layer eventually develops standards that are fed to the Client Industry layer that requested it, but also to the Implementation layer. Requests to implement a standard are generated by companies in the Client industry layer and directed to companies in the Implementation layer who eventually deliver the implementations to the companies requesting them. Conformance testing typically plays a role in assessing conformance of an implementation to the standard.

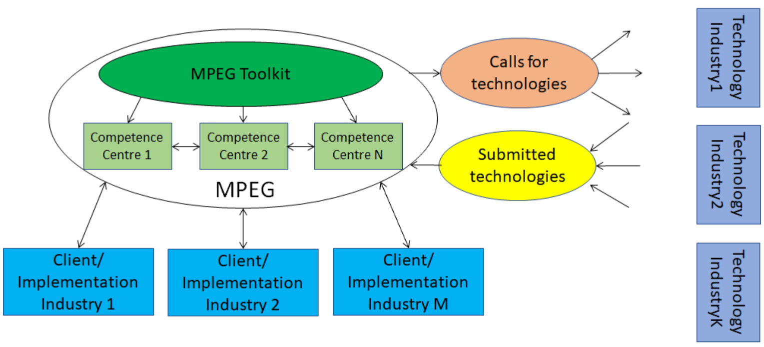


Figure 12: The MPEG process in the MPEG ecosystem

How the MPEG process takes place

Figure 11 does not fully describe the MPEG ecosystem. More elements are provided by *Figure 12* which also describes how the MPEG process actually takes place.

The new elements highlighted by *Figure 12* are

1. The *MPEG Toolkit* assembling all technologies that have been used in MPEG standards
2. The *MPEG Competence Centres* mastering specific technology areas and
3. The *Technology industries* providing new technologies to MPEG by responding to CfPs.

In the early days the Implementation Industries did not have a clear identity and could be considered part of the Client and Implementation Industries. Today, as highlighted above, the providers of basic technologies are well identified and separate industries that may not implement or use the standards.

Using *Figure 12* it is possible to describe how the MPEG process unfolds (the elements of the MPEG ecosystem are in italic).

1. MPEG receives a request for a standard from a *Client Industry* (or more than one)
2. The Requirements *Competence Centre* develops requirements by interacting with *Client Industries* and *Implementation Industries*
3. MPEG issues CfPs (Calls for technologies in the figure)
4. *Technology Industries* respond to CfPs by submitting technologies
5. MPEG mobilises appropriate *Competence Centres*
6. *Competence Centres* develop standards by selecting/adapting submitted technologies and existing technologies (drawn from the toolkit)
7. MPEG updates the toolkit with new technologies.

MPEG's role cannot be described by the simple "Standards Provider – Client Industry" relationship. MPEG is a complex ecosystem that works because all its entities play the role that is proper to them.

MPEG standards are glued together

MPEG started with the idea of a standard for video compression and soon it became clear that an audio compression standard was needed at the same time because in most applications video without audio is not really useful. Then it also became clear that playing audio and video after delivery in the way intended by the creator did not come for free and so came the need for Systems standards. Moving on, a File Format also became a necessity and today, when MPEG develops the MPEG-I standard, a lot more technologies are found necessary to hold the pieces of the system together.

MPEG has devised an organisation of work that allows it to deploy the necessary level of expertise in specific technology areas, e.g. in video and audio coding and file format. At the same time, however, the organisation allows it to identify where interfaces between different media subsystem are needed so that users of the standard do not discover unpleasant surprises when they integrate solutions.

Figure 13 is a good model of how most MPEG standards are developed. Different groups with different competences develop different parts of a standard, say, MPEG-I. Some parts are designed to work together with others in systems identified in the Context-Objectives-Use cases phase. However, many parts are not tightly bound because in general it is possible to use them separately. In other cases, there are parts of different origin that must work tightly together and here is where MPEG provides the "glue" by using ad hoc groups, joint meetings, chairs meeting etc.

This work method was developed and refined over the years and has served well to provide industry with standards that can be used as individual components or as a system.

Here are examples of how standards for different purposes have achieved the goal:

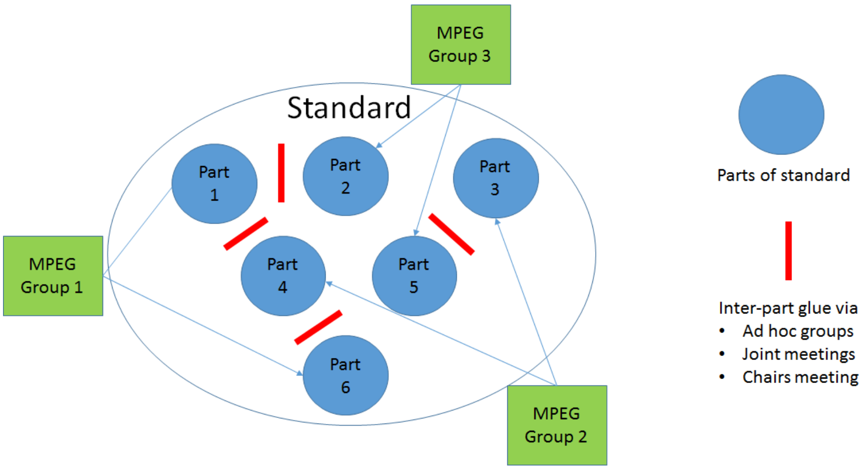


Figure 13: Structure of MPEG standards

1. MPEG-7 – *Multimedia content description interface* – was driven by the idea of a world rich of audio-video-multimedia descriptors that would allow users to navigate the large amount of media content expected at that time and that we have today. Content descriptors were expressed in verbose XML, a tool at odds with the MPEG bit-thrifty purposes. So MPEG developed the first standard for XML compression, a technology adopted in many fields which is consistently used by all MPEG-7 descriptors.
2. Of MPEG-A – *Multimedia Application Formats* – is remarkable the Common Media Application Format (CMAF) standard. Several technologies drawn from different MPEG standards are restricted and integrated to enable efficient delivery of large scale, possibly protected, video applications, e.g. streaming of televised events. CMAF Segments can be delivered once to edge servers in content delivery networks, then accessed from cache by thousands of streaming video players without additional network backbone traffic or transmission delay.
3. MPEG-V – *Media context and control* – is another typical example. The work was initiated in the wake of the success of Second Life, a service with virtual objects that looked like it could take over the world. The purpose of part 4 of MPEG-V Virtual world object characteristics was not to standardise a Second Life like service but the interfaces that would allow a user to move assets from one virtual space to another virtual space. Other parts of MPEG-V concern formats and interfaces to enrich the audio-visual user experience with, say, a breeze when there is a little wind in the movie, a smell when you are in a field of violets etc. So far, this apparently exciting extension of the user experience in a virtual world did not fly, but MPEG-V provides a solid communication framework for sensors and actuator that finds use in other standards.
4. MPEG-H – *High efficiency coding and media delivery in heterogeneous environments* – is another integrated standard of which a user can decide to take only the video part (HEVC) or the audio part (3D Audio). Another part – MPEG Media Transport (MMT) – shows how it is possible to innovate without destabilising existing markets. MPEG-2 Transport Stream (TS) has been in use for 25 years and will continue to be used for the foreseeable future. But MPEG-2 TS shows the signs of time because it has been designed for a one-way channel – an obvious choice 25 years ago – while so much video distribution today happens on two-way channels. Therefore, MMT uses IP transport instead of MPEG-2 TS transport and achieves content delivery unification in both one-way and two-way distribution channels.

Conclusions

MPEG is a complex ecosystem that has successfully operated for decades serving the needs of the growing number of its component industries. Everything human is perfectible, but whoever wants to lay their hands-on MPEG should remember the following

1. Dividing MPEG by Client Industries would mean losing the commonality of technologies;
2. Dividing MPEG by Implementation Industries would make no sense, because in principle any MPEG standard must be implementable with several different technologies;
3. Dividing MPEG by Competence Centres would mean losing the interactions between them.

5.5 Standardisation and product making

In this chapter the work done by a company developing a product is compared with the work of MPEG developing a standard, the product that MPEG delivers to its customers.

Let's see how a company could decide to make a new product:

1. A new product idea is proposed
2. The product idea is supported by market studies
3. Technology is available/accessible to make products
4. Design resources are available
5. "Product board" approves the project
6. Design is developed.

Let us see the corresponding work flow of an MPEG standard (look at How MPEG develops standards to have more details about the process):

1. An idea is proposed/discussed at a meeting
2. Idea is clarified in context and objectives
3. Use cases of the idea are developed
4. Requirements are derived from use cases
5. A Call for Evidence (CfE) is issued to check that technologies meeting the requirements exist
6. A Call for Proposals (CfP) is issued to get the necessary technologies
7. National Bodies (NB) approve the project
8. The standard is developed.

Let us compare and align the two processes because there are significant differences next to similarities:

Table 7 – Comparison of company and MPEG operation

#	Steps in a company product	Steps in an MPEG standard
1	A new product idea is proposed	Idea is aired/proposed at a meeting
2	Market studies support product	Context & objectives of idea drafted
		Use cases developed
3	Product requirements are developed	Requirements derived from use cases
4	Technology is available/accessible	Call for Evidence is issued
		Call for Proposals is issued
5	Design resources are available	MPEG looks for those interested
6	"Product board" approves the product	NBs approve the project
7	Design is developed	Test Model developed
		Core Experiments carried out
		Working Drafts produced
		Standard improved in NB balloting
8	The design is approved	NBs approve the standard

Comparing the two processes one can see that next to a high-level similarity, there are differences:

1. *Product proposal*: processes are hard to compare. Any company has its own processes. In MPEG, proposals can come to the fore spontaneously from any member.
2. *Proposal justification:* processes are hard to compare. Any company has its own specific means to assess the viability of a proposed new product. In MPEG, when enough support exists,

it first documents the context in which the idea would be applied and for what purposes. Then MPEG develops use cases to prove that a standard implementing the idea would support the use cases better than it is possible today or make possible use cases that today are not. As an entity, MPEG does not make "market studies" (because it does not have the means). It relies instead on members bringing relevant information into the committee when "context and objectives" and "use cases" are developed.

3. *Requirements definition*: happens under different names/ processes in companies and in MPEG.
4. *Technology availability* is quite different. A company may owns a technology as a result of some R&D effort or because it has acquired it. If it does not have a technology for a product, it either develops it or acquires it. MPEG "owns" a body of technologies, but typically a new proposal requires new technologies. While MPEG members may know that technologies are actually available, they may not be allowed to talk about it. Therefore, in general MPEG needs two steps: 1) to become aware of technology (via CfE) and 2) to have the technology available (via CfP). In some cases, like in Systems standards, MPEG members may develop the technology collaboratively from a clean sheet of paper.
5. *Design resource availability* is very different in the two environments. If a company sees a product opportunity, it may have the means to deploy the appropriate resources. If MPEG sees an opportunity for a standard, it has no means to "command" members to do something because members report to their companies, not to MPEG. It would be great if some MPEG members who insist on MPEG pursuing certain opportunities without offering resources to achieve them understood this.
6. *Product approval*: is very different in the two environments. Companies have their own internal processes to approve products. In MPEG the project for a new standard is approved by the shareholders, i.e. by the NBs, the simple majority of which must *approve* the project *and* a minimum of five NBs must *commit* resources to execute it.
7. *Design development:* is very different in the two environments. Companies have their own internal processes to design a new product. In MPEG work obviously stops at the design phase but it entails the following steps: 1) Test Model creation, 2) Core Experiments execution, 3) Working Drafts development and 4) Standard improvement though NB balloting.
8. *Design approval:* is very different in the two environments. Companies have their own internal processes to approve the design of a new product. In MPEG, again, the shareholders, i.e. the NBs, approve the standard with a qualified majority.

5.6 Standards are living beings

A standard is developed, published and used. Is that the end of the story for MPEG? It is not the end but the continuation of the same story. Indeed good standard must satisfy industry needs, but their use stimulates more needs that the standard must satisfy.

Here I will describe how this impacts the life of one of MPEG's most prestigious standards: MPEG-2 Systems, which has turned 26 in November 2018 and has played a major role in creating the digital world that we know.

What is MPEG-2 Systems?

When MPEG started, standards for compressed video and later audio were the immediate goal. But it was clear that the industry needed more than that. So, after starting MPEG-1 video compression and audio compression, MPEG soon started to investigate "systems" aspects. Seen with today's eyes, the interactive CD-ROM target of MPEG-1 was an easy problem because all videos on a a CD-ROM are assumed to have the same time base, and bit delivery is error free and on-time because the time interval between a byte leaving the transmitter is the same as the time interval at its arrival at the receiver.

In July 1990, even before delivering the MPEG-1 standard (which happened in November 1992), MPEG started working on the much more challenging "digital television" problem. This can be described as: the deliver of a package of digital TV programs with different time bases and associated metadata over a variety of analogue channels – terrestrial, satellite and cable. Of course operators expected to be able to do the same operations in the network that the television industry had been accustomed to do in the several decades since TV distribution had become common place. A unique group of experts from different – and competing – industries with their different cultural backgrounds and many countries, and some of them with the common experience of designing from scratch the MPEG-1 Systems standard, started the design the MPEG-2 Systems standards, again from a blank sheet of paper.

The impact of MPEG-2 Systems

MPEG-2 Systems is the container and adapter of the digital audio and video information to the physical world. It is used every day by billions of people who receive TV programs from a variety of sources, analogue delivery media and, often, digital as well (e.g. IPTV).

MPEG-2 Systems was approved in November 1994, at a time when some companies who could not wait had already made implementations before the formal release of the standard. That date, however, far from marking the "end" of the standard, signaled the beginning of a story that continues unabated today. Indeed, in the 26 years after its release, MPEG-2 Systems has been constantly evolving, while keeping complete backward compatibility with the original 1994 specification.

MPEG-2 Systems in action

So far MPEG has developed 34 MPEG-2 Systems amendments (i.e. additions of functionality to an existing standard), 3 additional amendments are close to completion and one is planned. After a few amendments are developed, ISO requests that they be integrated in a new edition of the standard. So far 7 MPEG-2 Systems editions have been produced covering the transport of non-MPEG-2 native media and non-media data. This is an incomplete list of the transport functionality added:

1. *Audio*: MPEG-2 AAC, MPEG-4 AAC and MPEG-H 3D
2. *Video*: MPEG-4 Visual, MPEG-4 AVC and its extensions (SVC and MVC), HEVC, HDR/WCG, JPEG2000, JPEG XS etc.
3. *Other data*: streaming text, quality metadata, green metadata etc.
4. *Signaling*: format descriptor, extensions of the transport stream format (e.g. Tables for splice parameters, DASH event signaling, virtual segment etc.), etc.

Producing an MPEG-2 Systems amendment is a serious job. There is a need for experts with the full visibility of a 26 years old standard (i.e. don't break what works) and the collaboration of experts of the carrier (MPEG-2 Systems) and of the data carried (audio, video etc.). MPEG can respond to the needs of the industry because it has the expertise of all components available in house.

MPEG-2 Systems Amendments

The table below reports the full list of MPEG-2 Systems amendments, The 1st column gives the edition, the 2nd column the sequential number of the amendment of that edition, the 3rd column the title of the amendment and the 4th the dates of the approval stages.

Conclusions

MPEG-2 Systems is probably one of MPEG standards least "visible" to its users. Still it is one of the most important enablers of digital television distribution applications impacting the life of billions of people and tens of thousands of professionals. Its continuous support is vital for the well-being of the industry.

Table 8 – The MPEG-2 Systems Amendments

E	A	Title	Date
1	1	Format descriptor registration	95/11
	2	Copyright descriptor registration	95/11
	3	Transport Stream Description	97/04
	4	Tables for splice parameters	97/07
	5	Table entries for AAC	98/02
	6	4:2:2 @HL splice parameters	
	7	Transport of MPEG-4 content	99/12
2	1	Transport of Metadata	02/10
	2	IPMP support	03/03
	3	Transport of AVC	03/07
	4	Metadata Application Format CP	04/10
	5	New Audio P&L Signaling	04/07
3	1	Transport of Streaming Text	06/10
	2	Transport of Auxiliary Video Data	
	3	Transport of SVC	08/07
	4	Transport of MVC	09/06
	5	Transport of JPEG2000	11/01
	6	MVC operation point descriptor	11/01
	7	Signaling of stereoscopic video	12/02
	8	Simplified carriage of MPEG-4	12/10
4	1	Simplified carriage of MPEG-4	12/07
	2	MVC view, MIME type etc.	12/10
	3	Transport of HEVC	13/07
	4	DASH event signaling	13/07
	5	Transport of MVC depth etc.	14/03
5	1	Timeline for External Data	14/10
	2	Transport of layered HEVC	15/06
	3	Transport of Green Metadata	15/06
	4	Transport of MPEG-4 Audio P&L	15/10
	5	Transport of Quality Metadata	16/02
	6	Transport of MPEG-H 3D Audio	16/02
	7	Virtual segment	16/10
	8	Signaling of HDR/WCG	17/01
	9	Ultra-Low-Latency & JPEG 2000	17/07
	10	Media Orchestration & sample variants	
	11	Transport of HEVC tiles	
6	1	Transport of JPEG XS	
	2	*Carriage of associated CMAF boxes*	

5.7 Standards and uncertainty

What happens to a standard when it leaves MPEG? The answer is not so different from the answer to the question "what happens to a product when it leaves the assembly line?". The market response to the company product or to the MPEG standard is anybody's guess. Some products/standards are widely successful, some fare so and so, and some are simply rejected by the market. Companies deploy significant resources to allow them to put in place other strategies to reduce the number of failures, but it is a reality that even companies darling of the market stumble from time to time.

MPEG is no exception. Remember the famous phrase attributed to John Wanamaker: "Half the money I spend on advertising is wasted; the trouble is I don't know which half".

In How MPEG develops standards the MPEG process is described: once an idea is launched, context and objectives of the idea are identified; use cases submitted and analysed; requirements derived from use cases; and technologies proposed, validated for their effectiveness for eventual incorporation into the standard.

Some people complain that MPEG standards contain too many technologies supporting "non-mainstream" use cases. Such complaints are understandable but misplaced. MPEG standards are designed to satisfy the needs of different industries and what is a must for some, may well not be needed by others, a problem addressed by Profiles and Levels.

It is true that there are a few examples where some technologies in an otherwise successful standard get unused. Was adding such technologies a mistake? In hindsight yes, but at the time a standard is being developed the future is anybody's guess and MPEG does not want to find out later that one of its standards misses a functionality that was deemed to be necessary in some use cases and that technology could support at the time the standard was developed.

For sure there is a cost in adding the technology to the standard – and this is borne by the companies proposing the technology – but there is no burden to those who do not need it because they can use another profile.

Let us see how MPEG has managed the uncertainty surrounding its standards by considering some examples.

1. The *MPEG-1* project was driven by the idea of video interactivity on CD and digital audio broadcasting. MPEG-1 did not have commercial success for both targets. However, Video CD, not even in the radar when MPEG-1 was *started*, used MPEG-1 and sold 1 billion units (and tens of billion CDs). MP3, too, was also not in the radar when MPEG-1 was *approved* and some members even argued against the inclusion of such a "complex" technology into the standard. I doubt there is anybody now regretting the decision to make MP3 part of the MPEG-1 standard. If there is, it is for completely different reasons. The reason why the standard was eventually successful is that MPEG-1 was designed as a system (VCD is exactly that), but its parts were designed to be usable as stand-alone components (as in MP3).
2. The second case is *MPEG-2*. The project was driven by the idea of making television digital. When the first 3 MPEG-2 parts (Systems-Video-Audio) were consolidated, the possibility to use MPEG-2 for interactive video services on the telecom and cable networks became real. MPEG-2 Audio did not fare well in broadcasting (the demand for multichannel was also not there), but it did fare well in other domains. In any case many thought that MPEG-1 Audio delivered just enough. MPEG-2 AAC did fare well in broadcasting and laid the ground for the 20-year long MPEG-4 Audio ride. MPEG started the Digital Storage Media Command and Control (DSM-CC) standard (part 6 of MPEG-2) whose carousel is used in broadcasting because it provides the means for a set top box to access various types of information that a broadcaster sends/updates at regular intervals.
3. *MPEG-4* is rich in relevant examples.
    a. The *MPEG-4 model* was a 3D scene populated by "objects" that could be 1) static or dynamic, 2) natural or synthetic, 3) audio or visual in any combination. BIFS (the MPEG name for the 3D scene technology, an extension of VRML) did not fly (but

VRML did not fly either). However, 10 years later the Korea-originated Digital Multimedia Broadcasting technology, which used BIFS scaled down to 2D, had a significant success in adding services to radio broadcasting.

b. Much of the *MPEG-4 visual* work was driven by the idea of video "objects" which, along with BIFS, did not fly (the standard specified video objects but did not say how to make them, because that was an "encoder issue"). For a few years, MPEG-4 video was used in various environments. Unfortunately, the main intended of MPEG-4 Visual use – video streaming – was hampered by the "content fees" clause of the licensing terms that the target industry did not consider acceptable.
c. On the other hand, Part 10 of MPEG-4 *Advanced Video Coding (AVC)* was very successful, especially because patent holders did not repeat some of the mistakes they had made for MPEG-4 Visual.
d. None of the 3 Option 1 MPEG-4 video coding standards did fly, showing that in ISO today it is not practically possible to make a media-related standard that does not require onerous licensing of thirty party technologies.
e. The MPEG-4 Audio Parametric coding for high-quality audio did not fly, but a particular tool in it – Parametric Stereo (PS) – could very efficiently encode stereo music as a mono signal plus a small amount of side-information. MPEG combined the PS tool with HE-AAC and produced HE-AAC v2, an audio decoder that is on board of billions of mobile handsets today as it enables transmission of a stereo signal at 32 kb/s with very good audio quality.

There is no recipe to design a guaranteed successful standard.

5.8 MPEG communicates

MPEG standards are great communication tools and MPEG itself is – or tries to be – a good communicator, using all available media.

MPEG web site

Almost everything that will be mentioned in this post can be found, not necessarily in an easy way, in the MPEG web page: press releases, MPEG column, video tutorials, white papers, investigations and technical notes, ad hoc groups, events, social networks, liaisons and meetings. Of course, MPEG standards will not be found there. They can be purchased from the ISO web site or from an ISO National Body.

Press releases

Since its early days, MPEG took care of informing the world of its work. At the beginning this only was done on major occasions. Now MPEG publishes press releases systematically at every meeting. The news considered the most important gets the headlines and all other achievements get a mention. The rule is to mention in a press release all Calls for Evidence (CfE) and Call for Proposals (CfP), and the standards under development that reach the stage of Committee Draft (CD) or Final Draft International Standard (FDIS). News are prepared by the relevant group chairs, edited and distributed to the press by Prof. Christian Timmerer of University of Klagenfurt.

Let's have as look at the press release from MPEG 126 (Geneva, CH) to have an example.

1. The headline news is "Three Degrees of Freedom Plus (3DoF+) – Evaluation of responses to the Call for Proposal and start of a new project on Metadata for Immersive Video" because we believe this will be a very important standard.
2. The second news is "Neural Network Compression for Multimedia Applications – Evaluation of responses to the Call for Proposal and kick-off of its technical work".

3. The result of a video compression-related CfP "Low Complexity Enhancement Video Coding – Evaluation of responses to the Call for Proposal and selection of a Test Model for further development"
4. News on "Multi-Image Application Format (MIAF) promoted to FDIS"
5. Anticipation of a new CfP "3DoF+ Draft Call for Proposal goes Public".

The press release page containing all press release since January 2006 is here.
If you want to be added to the distribution list, please send an email to Christian in Austria.

MPEG Column

With its "column" MPEG tries to facilitate understanding of its standards. At every meeting, brief notes are published to explain the purpose and, to some extent, the working of MPEG standards.
Want to know about High Dynamic Range (HDR)? the Common Media Application Format (CMAF)? The new standard to view omnidirectional video (OMAF)? By going here you will see all articles published in the column and will have the opportunity to get answers to many questions on these and other MPEG standards.
The MPEG column is managed by the professional journalist Philip Merrill in the USA who is able to understand MPEG standards *and* explain them to the public.
Do you think an article on a particular standard would be of interest? Please send email to the chairman of the Communication Group Prof. Kyuheon Kim of Kyunhee University in Korea. We will do our best to satisfy your request.

Video tutorials

How could MPEG miss the opportunity to have its own series of "MPEG standards tutorials", of course using its audio and video compression standards? By going to Video tutorials on MPEG standards you will be able to understand what MPEG has done to make its standards "green", what is the Multimedia Preservation Application Format that manages multimedia content over the ages, what is the High Efficiency Video Coding (HEVC) standard, what is MPEG-H 3D Audio, what is MPEG Media Transport (MMT) used in ATSC 3.0, what is Dynamic Adaptive Streaming over HTTP (DASH) for audio-visual streaming over the internet and much more.
The content is delivered by the best MPEG experts in the field. The videos that you see are the result of the shooting and post-processing performed by Alexis Tourapis in USA.

White papers, investigations and technical notes

MPEG makes its best efforts to provide the smoothest entry path to its standards and publicly accessible papers are functional to this strategy. White papers are published with the goal to:
1. Communicate that MPEG is investigating some promising ideas as in Investigations about parts of standards
2. Describe the purpose of an entire suite of standards, as in the case of White papers about standards or for a single part of a standard like in White papers about parts of standards
3. Provide specific guidance about use of standards as in Technical notes about parts of standards.

As a rule, the purpose of white papers is not to describe the technology but about what the standard is for, the problems it solves and the benefits that MPEG expects users will get from using it.
Investigations, White papers and Technical notes can be found here.

Standards

ISO is a private association registered in Switzerland. Standards are developed pro bono by participants in the working groups, but the cost of the organisation is covered by the sale of standards and other sources. Therefore, you should not expect to find ISO standards on the public MPEG web page. If you need a standard, you should go to the ISO web site where you can easily buy online all the standards on sale. In some cases, MPEG requests ISO to make standard public

because the standard is particularly relevant or because the standard is already publicly available (as is the case of all standards developed jointly with ITU-T).
MPEG posts all public documents at Standard documents from the last meeting, e.g. Use cases, Requirements, Calls for Evidence, Calls for Proposals, Working Drafts up to Committee Drafts, Verification Test results and more. MPEG does this because it wants to make sure that the industry is aware of, can comment on, and contribute to the development of standards.

Ad hoc groups

Since 1990 MPEG has created ad hoc groups (AhG). According to the rules "AhGs are established for the sole purpose of continuing work between consecutive MPEG meetings", but they are a unique way to have work done outside MPEG meetings in a coordinated way. The scope of an AhG, however, is limited by the rule: "The task of an AHG may only cover preparation of recommendations to be submitted to MPEG".
AhGs are not permanent organisations but established at the end of a meeting and last until the following meeting, To have a feeling of what AhGs are about you can see the AhGs established at 126th MPEG meeting.
AhGs are mentioned as part of MPEG communication because anybody can join the email reflector of an AhG and even attend AhG meetings.

MPEG events

In certain cases, MPEG organises events open to the public. Some events are held during and co-located with an MPEG meeting, but events outside MPEG meetings are also held. These are some of the goals an MPEG event can have:

1. To present a particular standard under development or just released as in Workshop on MPEG-G (Shenzhen, CN, October 2018)
2. To introduce the MPEG workplan such as MPEG Workshop on Immersive Services Roadmap (Gwangju, KR, January 2018)
3. To demonstrate what the industry is doing with an MPEG standards such as OMAF Developers' Day (Gwangju, KR. January 2018)
4. To frame a particular segment of the MPEG activity in the general context of the industry such as Workshop on standard coding technologies for immersive visual experiences (Gothenburg, SE, July 2019) .
5. To report on state of work and plan such as Workshop on standard coding technologies for immersive visual experiences.

Reference software and conformance

MPEG standards can be made public by a decision of the ISO Central Secretariat. MPEG requests ISO to make all reference software and conformance standards publicly available. The rationale of this request is that if developers look at the reference software, they need to buy the standard to make sure that theirs is a conforming implementation. To create a healthy ecosystem of interoperable products, services and applications – the conformance testing suites, too, must make freely available. This entices more users to buy the standard.

Social networks

MPEG has a Twitter account @MPEGgroup (https://twitter.com/mpeggroup). This is used by a group of *MPEG social champions* to spread information on the currently hottest MPEG topics: MPEG-I Requirements, Neural Network Compression, MPEG-G, OMAF, File format, Network Based Media Processing, MPEG-I Visual (3DoF+ and 6DoF), Audio, Point Cloud Compression, Internet of Media Things. Subscribe to receive brief notes on MPEG-related news and events.

Liaisons

MPEG develops standards technologies that are used by many industries all over the world. MPEG requests to liaise with many standards committees and industry fora for several purposes such as to get use cases and requirements, to jointly develop standards, to promote adoption of the standard once it has been developed and to receive further requests for new functionalities.

Here are some of the organisations MPEG has liaisons with: the Third Generation Partnership Project (3GPP), Advanced Television Systems Committee, Inc. (ATSC), Internet Engineering Task Force (IETF), Society of Motion Picture and Television Engineers (SMPTE), Audio Engineering Society (AES), European Broadcast Union (EBU), Hybrid Broadcast Broadband TV (HbbTV), Society of Cable Telecommunications Engineers (SCTE), World Wide Web Consortium (W3C) and many more.

Administrative documents

At every meeting MPEG publishes several documents – that I call "administrative" for lack of a better name – but are very important because they include organisational information. The following documents relate to MPEG 124 (Macau SAR, October 2018):

1. List of ad hoc groups established at the meeting
2. Call for patents related to standards under development
3. List of MPEG standards produced since day 1, those being worked on and those planned
4. Work plan with a summary description of all activities under way including explorations
5. Timeline with planned dates of development of all standard
6. Terms of reference
7. Schemas
8. URIs.

6 The MPEG success

This chapter analyses MPEG success from the points of view:

1. What has been done looks at those MPEG standards that have the largest impact in terms of use;
2. The MPEG success in numbers (of units and dollars) looks at the numbers – of devices and dollars – triggered by MPEG standards
3. MPEG is working for more success describes the current MPEG work plan.

6.1 What has been done

If I exchange words with taxi drivers in a city somewhere in the world, one of the questions I am usually asked is: "where are you from?". As I do not like straight answers, I usually ask back "where do you think I am from?" It usually takes time before the driver gets the information he asked for. Then the next question is: "what is your job?". Again, instead of giving a straight answer, I ask the question: "do you know MPEG?" Well, believe it or not, 9 out of 10 times the answer is "yes", often supplemented by an explanation decently connected with what MPEG is.

Wow! Do we need a more convincing proof that MPEG has conquered the minds of the people of the world?

The interesting side of the story, though, is that, even if the name MPEG is known by billions of people, it is not a trademark. Officially, the word MPEG does not even exist. When talking to ISO you should say "ISO/IEC JTC 1/SC 29/WG 11" (next time, ask your taxi driver if they know this letter soup). The last insult is that the mpeg.org domain is owned by somebody who just keeps it without using it.

Should all this be of concern? Maybe for some, but not for MPEG. What I have just talked about is just one aspect of what MPEG has always been. Do you think that MPEG was the result of high-level committees made of luminaries advising governments to take action on the future of media? You are going to be disappointed. MPEG was born haphazardly (read here, if you want to know how). Its strength is that it has been driven by the idea that the epochal transition from analogue to digital should not become another PAL-SECAM-NTSC or VHS-Betamax trap and that the format of digital media did not have to be different industry-by-industry.

In 30 years MPEG has grown 20-fold, changed the way companies do business with media, made music liquid, multiplied the size of TV screens, brought media where there were stamp-size displays, made internet the primary delivery for media content, created new experiences, shown that its technologies can successfully be applied beyond media…

MPEG-1 & MPEG-2

MPEG was the first standards group that brought digital media to the masses. In the 2nd half of the 1990's the MPEG-1 and MPEG-2 standards were converted to products and services as the list below will show (not that the use of MPEG-1 and MPEG-2 is confined to the 1990's).

- Digital Audio Broadcasting: in 1995, just 3 years after MPEG-1 was approved, DAB services began to appear in Europe with DAB receivers becoming available some time later.
- Portable music: in 1997, 5 years after MPEG-1 was approved, Saehan Information Systems launched MPMan, probably the first portable digital audio player for the mass market that used MP3. This was followed by a long list of competing players until the mobile handset largely took over that function.
- Video CD: in the second half of the 1990's Video CD (VCD) spread especially in South East Asia until the MPEG-2 based DVD, with its superior quality, slowly replaced it. VCD uses all 3 parts of MPEG-1 (layer 2 for audio).
- Digital Satellite broadcasting: in June 1994 DirecTV launched its satellite TV broadcasting service for the US market, even before MPEG released the MPEG-2 standard in November of that year! It used MPEG-2 and its lead was followed by many other regions who gradually converted their analogue broadcast services to digital (and the process has not be complete globally even now!).
- Digital Cable distribution: in 1992 John Malone launched the "500-channel" vision for future cable services and MPEG gave the cable industry the means to make that vision real.
- Digital Terrestrial broadcasting:
    - In 1996 the USA Federal Communications Commission adopted the ATSC A/53 standard. It took some time, however, before achieving wide coverage of the country, and for other countries following the ATSC standard.
    - In 1998 the UK introduced Digital Terrestrial Television (DTT).
    - In 2003 Japan started DTT services using MPEG-2 AAC for audio in addition to MPEG-2 Video and TS.
    - DTT is not deployed in all countries yet, and there are regularly news of a country switching to digital, the MPEG way of course.
- Digital Versatile Disc (DVD): toward the end of the 1990's the first DVD players were put to market. They used MPEG-2 Program Stream (part 1 of MPEG-2) and MPEG-2 Video, and a host of audio formats, some from MPEG.

MPEG-4

In the 1990s the Consumer Electronics industry provided devices to the broadcasting and telecom industries. and devices for package media. The shift to digital services called for the IT industry to join as providers of big servers for broadcasting and interactive services (even though in the 1990's the latter did not take off). The separate case of portable audio players provided by startups did not fit the established categories, neither did the initiation of the consumer device business by Apple which shed "computer" from its name.

MPEG-4 played the fundamental role of bringing the IT industry under the folds of MPEG as a primary player in the media space.

- Internet-based audio services: The great original insight of Steve Jobs and other industry leaders transformed Advanced Audio Coding (AAC) from a promising technology to a standard that dominates mobile devices and internet services
- Internet video: MPEG-4 Visual, with the MP4 nickname, did not become "MP3 for video". Still it was the first example of digital media on the internet as DivX (a company name). Its hopes to become the streaming video format for the internet were dashed by MPEG-4 Visual licensing terms, the first example of ill-influence of technology rights on an MPEG standard
- Video for all: MPEG-4 Advanced Video Coding (AVC) became a standard adopted in all areas and countries. Broadcasting, internet distribution, package media (Blu-ray) and more.
- Media files: the MP4 File Format is the general structure for time-based media files, that has become another ubiquitous standard at the basis of modern digital media.
- Advanced text and graphics: the Open Font Format (OFF), based on the OpenType specification, revised and extended by MPEG, is universally used.

MPEG-7

- Media production and archiving metadata: are satisfied by the MPEG-7 Audio Visual Description Profile (AVDP), developed with participation of the broadcasting industry.

MPEG-A

- Format for encrypted, adaptable multimedia presentation: is provided by the Common Media Application Format (CMAF), a format optimised for large scale delivery of protected media with a variety of adaptive streaming, broadcast, download, and storage delivery methods including DASH and MMT.
- Interoperable image format: the Multi-Image Application Format (MIAF) enables precise interoperability points for creating, reading, parsing, and decoding images embedded in HEIF.

MPEG-B

- Generic binary format for XML: is provided by Binary format for XML (BiM), a standard used by products and services designed to work according to ARIB and DVB specifications.
- Common encryption for files and streams: is provided by Common Encryption (CENC) defined in two MPEG-B standards – Part 7 for MP4 Files and Parts 9 for MPEG-2 Transport Stream. CENC is widely used for the delivery of video to billions of devices capable to access internet-delivered stored files, MPEG-2 Transport Stream and live adaptive streaming.

MPEG-H

- IP-based television: MPEG Media Transport (MMT) is the "transport layer" of IP-based television. MMT assumes that delivery is achieved by an IP network with in-network intelligent caches close to receiving entities that adaptively packetise and push the content to receiving entities. MMT has been adopted by the ATSC 3.0 standard and is currently being deployed in countries adopting ATSC standards and also used in low-delay streaming applications.
- More video compression, siempre!: has been provided by High Efficiency Video Coding (HEVC), the AVC successor yielding an improved compression up to 60% compared to AVC. Natively, HEVC supports High Dynamic Range (HDR) and Wider Colour Gamut (WCG). However, its use is plagued by a confused licensing landscape
- Not the ultimate audio experience, but close: MPEG-H 3D Audio is a comprehensive audio compression standard capable of providing very satisfactory immersive audio experiences in broadcast and interactive applications. It is part of the ATSC 3.0 standard.
- Comprehensive image file format: High Efficiency Image File Format (HEIF) is a file format for individual HEVC-encoded images and sequences of images. It is a container capable of

storing HEVC intra-images and constrained HEVC inter-images, together with other data such as audio in a way that is compatible with the MP4 File Format. HEIF is widely used and supported by major OSs and image editing software.

MPEG-DASH
Streaming on the unreliable internet: Dynamic Adapting Streaming on HTTP (DASH) is the widely used standard that enables a media client connected to a media server via the internet to obtain instant-by-instant the version, among those available on the server, that best suites the momentary network conditions.

Conclusions
Figure 14 intends to attach some concreteness to the ideas illustrated above by showing some of the most successful MPEG standards issued from 31 years of MPEG activity.
An entity at the lowest layer of the ISO hierarchy has masterminded the transition of media from the analogue to the digital world. Its standards underpin the evolution of digital media, foster the creation of new industries and offer unrelenting growth to old and new industries worth in excess of 1 trillion USD per year, as will be shown in the next chapter.

Figure 14: Some successful MPEG standards

6.2 The MPEG success in numbers

Introduction
In this chapter the talk is not about a supposedly “big business” with MPEG standards, but about the business triggered by MPEG standards. When possible, the breadth and size of businesses for which MPEG standards have an enabling role will be identified.

Direct business generated by MPEG standards
In this section the talk is about the “primary business”, i.e. the business that relies directly on MPEG standards.

Sale of standards
Experts attending MPEG meetings draft standards. The copyright of the standards belongs to ISO and IEC who do business with MPEG standards by selling them, as sales of its standards are their major source of revenues. The number of copies of MPEG standards is not known, but the revenue could easily be in the order of M$.

Companies participating in the development of MPEG standards
Hundreds of companies and organisations send experts to MPEG meetings. Some companies contribute technologies because their primary goal is that MPEG produce the standard(s) they need. Other companies do not have products or services that could benefit from MPEG standards but possess significant amount of IP. This is the case of universities who currently contribute ~1/4 of the experts attending MPEG meetings, but also the case of so-called Non-Performing Entities (NPE).
Columbia University raised the attention of many in the mid 1990's when MPEG LA, the MPEG-2 patent pool administrator, put Columbia University in their list of holders of essential MPEG-2 patents. Today the MPEG LA website mentions at least 10 HEVC licensors who are Universities or research centres.
It is not easy to assess the total value of revenues from patents that are essential to practise MPEG standards, but it is probably in excess of 1 B$/year.

Companies developing integrated circuits for MPEG standards
The early days of MPEG (end of 1980's) were beset by the problem of availability of integrated circuits implementing MPEG-1 Video. A VLSI subgroup (later renamed Implementation Studies Group) was set up to study the implementation impact of the adoption of particular technologies. The large market of MPEG-2 chip kicked off an important branch of the silicon industry.
In the context of the MPEG-4 standard, MPEG developed Part 9 Reference Hardware Description where some parts of the reference software are described by alternative blocks described in a Hardware Description Language (HDL) form. The fundamental objective of the reference HW description was the support of such mixed SW/HW standard description with appropriate platforms, e.g. a programmable board that can be plugged into a standard SW environment.
There is no public information on the size of the industry developing VLSI chips that implement MPEG standards.

Companies developing software for MPEG standards
Starting from MPEG-4, MPEG adopted the practice of specifying all MPEG standards in two languages: one natural (English) and the other in a computer language. The two specifications are both normative and if one specification is discovered not to be aligned with the other, MPEG has to decide which version should be the basis and align the other to the former.
Today the licence of the reference software is a BSD licence modified to indicate that use of the software may require the use of essential patents held by third parties. The licence was first adopted for the MPEG eXtensible Middleware (MXM) standard. The purpose of the reference software is to provide an alternative specification, it is not meant to be suitable for use in products, even though some reference software is of very high quality and performance.
There are several open source implementations of MPEG standards, such as Ffmpeg which is used in many commercial implementations.
There is of course a large number of proprietary implementations developed by software companies. However, there is no public information on the size of the industry developing software that implements MPEG standards.

Indirect business generated by MPEG standards
MPEG standards have created the conditions for the first mass-deployable digital video solutions and provided new and standard solutions for digital audio. Over the years, alternative proprietary solution have been offered to the market, but MPEG continues to be the reference source, especially for hardware-based devices.

The following provides publicly available information on the number of devices/users and/or the size of the business in 3 areas: Devices, Video Surveillance and Content. For each entry the web page where the information was found is provided.

Devices

- In 2017 the global **laptop** market size was valued at 102 B$ and is estimated to expand at a CAGR of 0.4% during the 2018-2025 period (Grand View Research)
- In 2018 global **smartphone** sales reached 522 B$, with 1.44 billion units and an average selling price (ASP) of 384 $. Smartphones, mobile phones and wearables account for 44% of the 1.2 T$ technical consumer goods (TCG) market (GfK)
- In 3Q18 36 million **tablets** were sold worldwide (IDC)
- In 2015 58 million **digital cameras** were sold worldwide (Statista)
- By 2022 In-Car Infotainment Market is expected to garner 34 B$, with a CAGR of 13.3% during the forecast period 2016 – 2022 (Allied Market Research)
- By 2024 the global **capture and production equipment** market is expected to generate revenues of ~38 B$, growing at a CAGR of ~4% in 2018-2024 (Arizton)
- In 1H18 the global market for **television (TV)** saw total sales of 45 B€ (~50 B$). A total of 238 million devices are expected to be sold in 2018 (GfK)
- In 2015 the global **Set Top Box** (STB) market size was estimated at 17 B$ and is anticipated to witness a significant growth over the 2018-2024 period (Grand View Research)
- In 2016 the global **video conferencing** market was valued at 5 B$ and is expected to expand at a CAGR of 7.9% from 2018 to 2026 to reach 10.5 B$ (Transparency Market Research)
- In 2017, the global Smart **Commercial Drones** market size was 1.4 B$ and it is expected to reach 18 B$ by the end of 2025, with a CAGR of 83.3% during 2018-2025 (MarketWatch).

Video surveillance

- In 2014 there were 245 million professionally installed **video surveillance cameras** active and operational globally, 20% of which are estimated to be network cameras and ~2% HD CCTV cameras (IHS)
- In 2017 the global **video surveillance** market generated a revenue of 32 B$ and is estimated to achieve 37 B$ in 2018 (BIS Research)
- The **video surveillance** market size is projected to grow at a CAGR of 16.14% during the 2018 to 2023 period and reach 77 B$ by 2023 (BIS Research)
- In 2018 the **Video Surveillance** market size is estimated to be 37 B$ to grow to 68 B$ by 2023, at a CAGR of 13.1% between 2018 and 2023 (MarketsandMarkets)

Content

- In 2018 revenues of **games, films and music** were 138 B$, 42 B$ and 19 B$, respectively (League of Professional esports)
- The **In-flight Entertainment** & Connectivity (IFEC) market is projected to grow from 5 B$ in 2018 to 7.65 B$ by 2023, at a CAGR of 8.72% from 2018 to 2023 (MarketsandMarkets).
- In 2018 global revenues from **traditional pay-tv and OTT TV episodes and movies** will reach 265 B$; up from 254 B$ in 2017 and 234 B$ in 2015 (Broadband TV News)
- In 2022 the global **pay TV market** including satellite, cable, and IPTV services is expected to generate 295 B$ in 2022 (Broadband TV News)
- In 2017 the global **social media** market was valued at 34 B$ (PRNewswire)
- By end-2018 the total **TV subscriptions** will reach 1.5 billion; up by 38% from 1 billion in 2015. In the same period **SVOD subscriptions** will reach 0.47 billion (Broadband TV News)

- In 2019 worldwide revenue from **TV advertising** will reach 177 B$ and would account for 34% of the global ad revenue in 2019, hence the second advertising medium in importance (Statista)
- In 2019 the **Television Production** market size in the US is 40 B$ and is expected to increase 3.7% in 2019 (IBISWorld)
- In 2018 the **Enterprise Video** market size is expected to be 13.5 B$ and grow to 29 B$ in 2023, at a CAGR of 7.9% during the 2018-2023 period (MarketsandMarkets)

Table 9 –Evolution of IP traffic (Cisco)

IP Traffic	2016	2021	CAGR
Global	1.2 ZB	3.3 ZB	24%
% of PC traffic	46%	25%	10%
% of SM traffic	13%	33%	49%
% of video traffic	73%	82%	31%
% of live video	0.9%	13%	23%

Content transport

Television continues to have its own distribution infrastructure, i.e. terrestrial distribution, the cable and satellite networks. Advanced countries have digitised terrestrial distribution and cable distribution. The majority of satellite channels are now digital.

Table 9, from Cisco Visual Networking Index: Forecast and Trends, 2016–2021 White Paper, provides the amount of IP traffic in the period 2016-2021.

Conclusions

The size of the economic value of products, services and applications enabled by MPEG standards is a good measure of their impacts. The numbers reported in Indirect business generated by MPEG standards show that MPEG standards enable a variety of businesses with a turnover worth more than 1 T$ per year. Of course, MPEG is not the only provider of specifications for digital media systems, but in certain segments such as TV sets, mobile devices and PC, support of MPEG standards is a must.

According to Michelle Abraham - S&P Global Market Intelligence, in 2018

- The value of the installed base of tablets, smartphones, STB, and digital TVs is 2.8 T$[1].
- Units shipped: 1.4 billion smartphones, 221 million digital TVs, 296 million STBs, 155 million tablets, 72 million streaming media devices (sticks and players)
- The revenues of global pay-tv are ~228 B$.

Of course there are more businesses in the indirect business triggered by MPEG standards.

6.3 MPEG is working for more success

MPEG has been a success gives a brief description of the MPEG standards that have been (and often continue to be) extremely successful. As it is difficult to single out those that will be successful in the future 😉, the reasonable thing to do is to show the entire MPEG work plan (*Figure 15*). At the risk of making the wrong bet 😉, the following will introduce some of the most high-profile standards under development, subdivided in the three categories: Media Coding, Systems and Tools, and Beyond Media. If you are interested in MPEG standards, however, you have better to become acquainted with all ongoing activities. In MPEG sometimes the last become the first.

[1] Personal communication

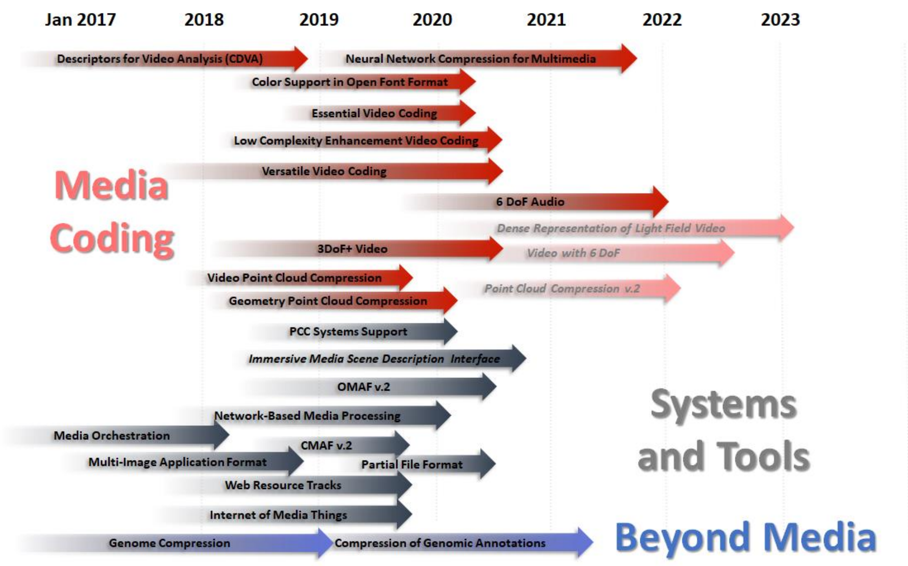


Figure 15: The MPEG work plan (March 2019)

Media Coding

1. Versatile Video Coding (VVC): is the flagship video compression activity that will deliver another round of improved video compression. It is expected to be the platform on which MPEG will build new technologies for immersive visual experiences (see below).
2. Enhanced Video Coding (EVC): is the shorter-term project with less ambitious goals than VVC's. EVC is designed to satisfy urgent needs from those who need a standard with a hopefully simplified IP landscape.
3. Immersive visual technologies: investigates technologies applicable to visual information captured by different camera arrangements, as described in [Immersive visual experiences](#).
4. Point Cloud Compression (PCC): refers to two standards capable of compressing 3D point clouds captured with multiple cameras and depth sensors called V-PCC and G-PCC. The algorithms in both standards are scalable, progressive and support random access to point cloud subsets. V-PCC is lossy and G-PCC is currently losslessSee [Immersive visual experiences](#) for more details.
5. Immersive audio: MPEG-H 3D Audio supports a 3 Degrees of Freedom or 3DoF (yaw, pitch, roll) experience at the movie "sweet spot". More complete user experiences, however, are needed, i.e. 6 DoF (adding x, y, z). These can be achieved with additional metadata and rendering technology.

Systems and Tools

1. Omnidirectional media format (OMAF): is a format supporting the interoperable exchange of omnidirectional (Video 360) content. In v1 for users can only Yaw, Pitch and Roll their head, but v2 will support limited head translation movements. More details in [Immersive visual experiences](#).
2. Storage of PCC data in MP4 FF: MPEG is developing systems support to enable storage and transport of compressed point clouds with DASH, MMT etc.
3. Scene Description Interface: MPEG is investigating the interface to the scene description (not the technology) to enable rich immersive experiences.

4. Service interface for immersive media: Network-based Media Processing will enable a user to access potentially sophisticated processing functionality made available by a network service via standard API.
5. IoT when Things are Media Things: Internet of Media Things (IoMT) will enable the creation of networks of intelligent Media Things (i.e. sensors and actuators)

Beyond Media

1. Standards for biotechnology applications: MPEG is finalising all 5 parts of the MPEG-G standard and establishing new liaisons to investigate new opportunities.
2. Coping with neural networks everywhere: shortly (25 March 2019) MPEG is developing Compression of neural networks for multimedia content description and analysis (part 17 of MPEG-7) after receiving responses to its Call for Proposals for Neural Network Compression (see Moving intelligence around).

7 Planning for the future of MPEG

MPEG has been working without interruption for 31 years. In this time span, new industries have continuously joined, existing technologies have continuously morphed and new ones come to the fore. Today MPEG faces another major change and has to "shed its skin" one more time.

MPEG is about compression

MPEG has produced 5 generations of video compression standards, each generation offering more compression (and more features, see More video features). More of the same is expected from the 6th generation (VVC) that MPEG plans to approve as an FDIS in July 2020, 7.5 years after it approved HEVC as an FDIS.
Will industry keep on demanding more compression? I would like to answer with a resolute yes as I believe that there will always be a need for more compression, but not always and not necessarily of the "old way". If we take the example of light field, we see that more compression is needed if distribution of that kind of information is going to become real within the current (human) generation. MPEG has produced an equivalent number of audio coding standards, but the answer to the question "do we need more audio compression?" is, at least in the current time frame, that the compression we have is enough but we need new standards for other non-compression features (6DoF, in the case of audio). The trend is less noticeable in video, but 3DoF+ moves in the same direction as audio, and the exact nature of 6DoF video is still unclear.
MPEG is developing its first standard for dense point clouds. It is an interesting combination of video coding and 3D Graphics coding technologies and MPEG has started working on a possible unified view of 3D video and point clouds. MPEG is also developing its first standard that compresses sparse point clouds (G-PCC) which is based on a different technology.
There is a need to understand the potential of compression technologies in audio, video and point clouds, interact with industry and develop a vision of compression standards for the years to come.

MPEG is also about systems aspects

This book has re-iterated the role of systems aspects as the enabler of the success of many MPEG standards and the future will not change that role. Actually, the trend toward immersive media will require an even deeper integration between compressed media and the systems that permeate them. This can be seen from the requirements that are being identified in an activity called "Immersive Media Access and Delivery" where four dimensions are identified:

1. Time (as usual)
2. Space (i.e. the ability to retrieve only the relevant parts of the media),
3. Quality (i.e. the ability to access portions of media with the desired quality)

4. Object (i.e. the ability to access specific parts of specific objects of interest).

MPEG is lucky to have an organisation that enables cross-domain interaction, but the efficiency of this organisation must be further improved to enable the needed deeper interaction.

Is MPEG only about media?

My answer is no: it is not and should not only be about media. In its past 30 years MPEG has shown that it has been able to add domains of expertise and to talk the language of those domains. If today all the media domains speak the same (technical) language, that is due in non-small part to the efforts made by MPEG to understand the needs of different industries, convert them to requirements, develop the technologies and quantise the standards into profiles and levels. This workflow has been in operation for 27 years, namely since MPEG invented profiles and levels and consistently applied them to talk to different communities using the same language.

The challenges in talking to a new industry, however, should never be underestimated. MPEG has spent more than 3 years talking to, and identifying and validating requirements with the genomic community before starting the development of the MPEG-G standard. This process was carried out jointly with ISO TC 276, a major representative of that community.

MPEG should be ready to respond to communities that request its unique expertise in data compression.

The business context is changing

MPEG does not have much to say on business matters, but its work is highly influenced by business, e.g. in terms of requirements, experts support, timelines etc.

In the first 10 years of its existence MPEG ruled the field of audio and video coding. In the following 10 years some proprietary solutions popped up, but the field was still largely dominated by MPEG standards. In the last 10 years MPEG has seen proprietary solutions getting more strength and occupying areas that used to be exclusively covered by MPEG standards.

Market spoke and market is right. MPEG should not complain about competition. Operating in a competitive environment is always beneficial because it pushes those affected to do a better job.

This is the easy part of the story. The question, though, is "should MPEG continue undeterred on its way or should it not rethink its role?"

Organisation matters

For decades MPEG has managed a large working group, designed multi-threaded work programs, sought and received support by industry, developed close to 200 specifications and hundreds of amendments, drove the evolution and enabled the relentless expansion of the digital media business.

MPEG did all that as a working group, i.e. as the lowest organisational unit in ISO that the ISO/IEC directives recommend to be "reasonably limited in size". The reality is that already in 1989 MPEG had 100 members, in 1999 it had 300 members and in 2019 it has a membership of 1500 with 500 experts attending its quarterly meetings.

For 30 years MPEG has played a role much above its status, even though I do not think there should be complaints about the results. MPEG could continue as a working group for another 30 years but from now on it should have a status equivalent to the role it plays. ISO have tolerated the "MPEG exception", but now they have decided that there should be no exception.

In this book I have explained why the solution of simply making working groups out of MPEG subgroups is not going to work. This will simply kill MPEG.

Something more creative is needed

Innovation, not revolution

MPEG should become a Subcommittee (SC). The existing subgroups developing standards should become Working Groups (WG) and the subgroups not developing standards should become

Advisory Groups (AG). Requirements would become *Technical requirements AG* and would continue to play the role of holding the integrated view of all MPEG standards. Test would become MPEG should take advantage of the change and establish a *Market needs AG*, supported by National Body participation, to improve adherence of its standards to market needs. In the new organisation, proposals for new work would be assessed from two points of view: technical feasibility (i.e. technical requirements, as done so far) and market needs. The Technical requirements AG would still issue Calls for Evidence, but the decision to issue a Call for Proposals would be made by the Subcommittee (whose members are appointed by National Bodies) on the basis of coordinated reports made by the Market needs and Technical requirements AGs. Technical requirements would remain in charge of issuing Calls for Proposals.
Today individual groups manage their liaisons but in the new organisation the *Liaison and Communication AG* would coordinate outgoing liaisons and communications.

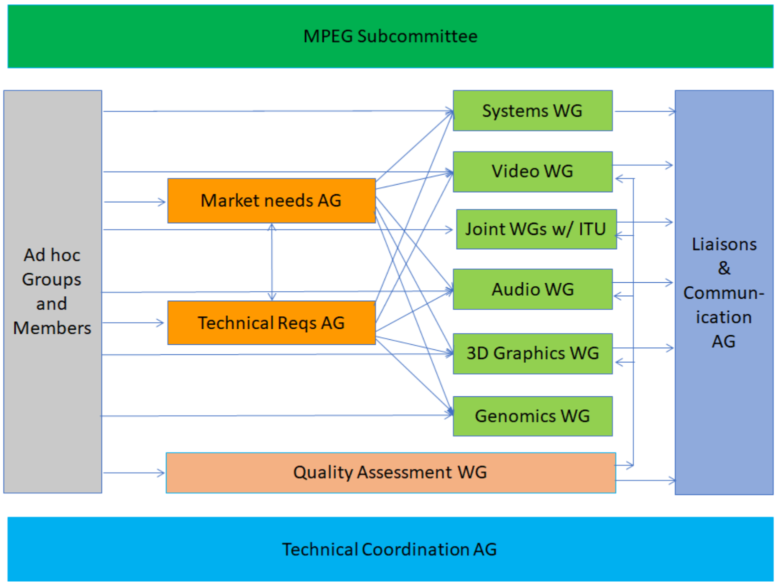


Figure 16 –The proposed new MPEG organisation

The current "Chairs group" would become the *Technical Coordination AG*. Its members would be appointed by the Subcommittee. They would include the Working Group convenors and/or suitably appointed WG members, the Convenor of the Technical Coordination AG and the chair of the new subcommittee.
Ad hoc groups would continue to be interim organisational units established to address specific, possibly cross-domain, issues. They would report about their findings at the meeting that follows the meeting that established them.
The WGs would include the existing technical subgroups, e.g. the *Quality Assessment WG* that would continue offering its services to the media compression WGs, and the Genomics WG.
The Subcommittee would hold quarterly meetings with its WGs/AGs and would hold sessions as required by the need to progress the work of its WGs/AGs.

8 Video compression in MPEG

For about 150 years, the telephone service has provided a socially important communication means to billions of people. For at least a century the telecom industry has strived to offer a more complete user experience (as we would call it today) to its customers by adding a visual component to speech.

Probably the first large scale attempt at offering the new audio-visual service was AT&T's PicturePhone in the mid 1960's. The service was eventually discontinued but the idea of expanding the telephone service with a video service caught the attention of telephone companies. Many expected that digital video-phone or video-conference services on the emerging digital networks would guarantee the success that the PicturePhone service did not have and research in video coding was funded in many research labs of the telephone companies.
In this chapter the efforts to develop compressed digital video standards will be presented along 3 dimensions: the sequences of standards, the features added to compresssion standards and the drive to immersive visual experiences.

1. Forty years of video compression is a thorough scan on 40 years of ITU and MPEG video compression standards;
2. More video features looks at the process that added new functionalities at each MPEG standard;
3. Immersive visual experiences looks at how " immersive experience" has been enhanced at each MPEG standard;
4. Video can be green describes how video compression can become more energy thrifty without sacrificing the user experience.

8.1 Forty years of video compression

Introduction

This chapter will tell the story of how compressed digital video standards gave rise to the ever-improving digital video experience that the world is experiencing in ever greater numbers.

First Video Coding Standard

The first international standard that used video coding techniques – ITU-T Recommendation H.120 – originated from the European research project called COST 211. H.120 was intended for video-conference services, especially on satellite channels. H.120 was approved in 1984 but was implemented in a limited number of specimens.

Second Video Coding Standard

The second international standard that used video coding techniques – ITU-T Recommendation H.261 – was intended for audio-visual services and was approved in 1988. This project signaled the maturity of video coding standardisation that left the old and inefficient algorithms to enter the DCT/motion compensation age.
For several reasons H.261 was implemented by a limited number of manufacturing companies for a limited number of customers.

Third Video Coding Standard

Television broadcasting has always been – and, with challenges, continues to be so today – a socially important communication tool. Unlike audio-visual services that were mostly a strategic target on the part of the telecom industry, television broadcasting in the 1980's was a thriving industry served by the Consumer Electronic (CE) industry providing devices to hundreds of millions of consumers.
The idea that originated ISO MPEG-1, the third international standard that used video coding techniques. was intended for interactive video applications on CD-ROM. The MPEG-1 standard was released by MPEG in November 1992. Besides the declared goal, the intention was to popularise video coding technologies by relying on the manufacturing prowess of the CE industry. MPEG-1 was the first example of a video coding standard developed by two industries that had had until that time very little in common: telecom and CE (terminals for the telecom market were developed by a special industry with few contacts with the CE industry).

Fourth Video Coding Standard

Even though in the late 1990's MPEG-1 Video eventually reached the 1 billion units sold with the nickname "Video CD", especially in the Far East, the big game started with the fourth international standard that used video coding techniques – ISO MPEG-2 – whose target was "digital television". The number of industries interested in it made MPEG a crowded WG: telecom had always sought to have a role in television, CE was obviously interested in having existing analogue TV sets replaced by shining digital TV sets or at least supplemented by a set top box, satellite broadcasters and cable were very keen on the idea of hundreds of TV programs in their bouquets, terrestrial broadcasters had different strategies in different regions but eventually joined, as well as the package media sector of the CE industry, with their tight contacts with the movie industry.

This explains why the official title of MPEG-2 is "Generic coding of moving pictures and associated audio information" to signal the fact that MPEG-2 could be used by all the industries that, at that time, had an interest in digital video, a unique feat in the industry.

Fifth and Sixth Video Coding Standards

Remarkably, MPEG-2 Video and Systems were specifications jointly developed by MPEG and ITU-T. The world, however, follows the dictum of the Romance of Three Kingdoms (三國演義): 話說天下大勢. 分久必合, 合久必分. Adapted to the context this can be translated as "in the world things divided for a long time shall unite, things united for a long time shall divide". So, the MPEG and ITU paths divided in the following phase. ITU-T developed its own H.263 Recommendation "Video coding for low bit rate communication" and MPEG developed its own MPEG-4 Visual standard, part 2 "Coding of audio-visual objects". The conjunction of the two standards is a very tiny code that simply tells the decoder that a bitstream is H.263 or MPEG-4 Visual. A lot of coding tool commonality exists, but not at the bitstream level.

H.263 focused on low bitrate video communication, while MPEG-4 Visual kept on making real the vision of extending video coding to more industries: this time Information Technology and Mobile. MPEG-4 Visual was released in 2 versions in 1999 and 2000, while H.263 went through a series of updates documented in a series of Annexes to the H.263 Recommendation. H.263 enjoyed some success thanks to the common belief that it was "royalty free", while MPEG-4 Visual suffered a devastating blow by a patent pool that decided to impose "content fees" in their licensing term.

Seventh Video Coding Standard

The year 2001 marked the return to the second half of Romance of Three Kingdoms' dictum: 分久必合 (things separated for a long time shall unite), even though it was not too 久 (long time) since they had divided, certainly not on the scale intended by the Romance of Three Kingdoms. MPEG and ITU-T (through its Video Coding Experts Group – VCEG) joined forces again in 2001 and produced the seventh international video coding standard in 2003. The standard is called Advanced Video Coding by both MPEG and ITU, but is labelled as AVC by MPEG and as H.264 by ITU-T. "Reasonable" licensing terms ensured AVC's long-lasting success in the market place that continues to this day.

Eighth Video Coding Standard

The eight international standard dealing with video coding stands by itself because it is not a standard with "new" video coding technologies, but a standard that enables a video decoder to build a decoder matching the bitstream using standardised tools represented in a standard form available at the decoder. The technique, called Reconfigurable Video Coding (RVC) or, more generally, Reconfigurable Media Coding (RMC), because MPEG has applied the same technology to 3D Graphics Coding as well, is enabled by two standards: ISO/IEC 23002-4 Codec configuration representation and ISO/IEC 23003-4 Video tool library (VTL). The former defines the methods and general principles to describe codec configurations. The latter describes the MPEG VTL and specifies the Functional Units that are required to build a complete decoder for the following standards: MPEG-4 Simple Profile, AVC Constrained Baseline Profile and Progressive High Profile, MPEG-4 SC3DMC, and HEVC Main Profile.

Ninth Video Coding Standard

In 2010 MPEG and VCEG extended their collaboration to a new project: High Efficiency Video Coding (HEVC). A few months after the HEVC FDIS had been released, the HEVC Verification Tests showed that the standard had achieved 60% improvement over AVC, 10% more than originally planned. After that, HEVC has been enriched with a number of features that at the time of development were not supported by previous standards such as High Dynamic Range (HDR) and Wide Colour Gamut (WCG), and support to Screen Content and omnidirectional video (video 360). Unfortunately, technical success did not translate into full market success because adoption of HEVC is still hampered – 6 years after its approval by MPEG - by an unclear licensing situation.

Tenth Video Coding Standard

The target of MPEG standards until AVC had always been “best performance no matter what is the IPR involved" (of course, if IPR holders allow) but, as the use of AVC extended to many domains, it was becoming clear that there was so much “old” IP (i.e. more than 20 years old) that it was technically possible to make a standard whose IP components were Option 1.

In 2013 MPEG released the FDIS of WebVC, strictly speaking not a new standard because MPEG had simply extracted what was the Constrained Baseline Profile of AVC and made it a separate standard with the intention of making it Option 1. The attempt failed because some companies confirmed their Option 2 patent declarations already made against the AVC standard.

Eleventh Video Coding Standard

WebVC has not been the only effort made by MPEG to develop an Option 1 video coding standard (i.e. a standard for which only Option 1 patent declarations have been made). A second effort, called Internet Video Coding (IVC), was concluded in 2017 with the release of the IVC FDIS. Verification Tests performed showed that IVC exceeded in performance the best profile of AVC, by then a 14 years old standard. Three companies made Option 2 patent declarations that did not contain any detail about the allegedly infringed technologies. Therefore, MPEG could not remove the technologies in IVC that the companies claimed infringed their patents.

Twelfth Video Coding Standard

MPEG achieved a different result with its third attempt at developing an Option 1 video coding standard. The proposal made by a company in response to an MPEG Call for Proposals was reviewed by MPEG and achieved FDIS stage with the name of Video Coding for Browsers (VCB). However, a company made an Option 3 patent declaration that, like those made against IVC, did not contain any detail that would enable MPEG to remove the allegedly infringing technologies. Eventually ISO did not publish VCB.

Today ISO and IEC have disabled the possibility for companies to make Option 3 patent declarations without details (a policy that ITU, even before, had not allowed). As the VCB approval process has been completed, it is not possible to resume the study of VCB if MPEG does not restart the process. Therefore, VCB is likely to remain unpublished and therefore not an ISO standard.

Thirteenth Video Coding Standard

For the fourth time MPEG and ITU are collaborating in the development of a new video coding standard with the target of a 50% reduction of bitrate compared to HEVC. The development of Versatile Video Coding (VVC), as the new standard is called, is still under way and involves some

250 experts attending VVC sessions. MPEG expects Versatile Video Coding (VVC) to reach the FDIS stage in July 2020 for the key compression engine. Other components, such as high-level syntax or SEI messages will likely be released later.

Fourteenth Video Coding Standard

Thirteen is a large number for video coding standards but this number should be measured against the number of years covered – close to 40. In this long period of time we have gone from 3 initial standards that were mostly application/industry-specific (H.120, MPEG-1 and H.261) to a series of generic (i.e. industry-neutral) standards (MPEG-2, MPEG-4 Visual, MPEG-4 AVC and HEVC) and then to a group of standards that sought to achieve Option 1 status (WebVC, IVC and VCB). Other proprietary video coding formats that have found significant use in the market point to the fact that MPEG cannot stay forever in its ivory tower of "best video coding standards no matter what". MPEG has to face the reality of a market that becomes more and more diversified and where – unlike the golden age of a single coding standard – there is no longer one size that fits all. At its 125th meeting MPEG has reviewed the responses to its Call for Proposals on a new video coding standard that sought proposals with a simplified coding structure and an accelerated development time of 12 months from working draft to FDIS. The new standard will be called MPEG-5 Essential Video Coding (EVC) and is expected to reach FDIS in April 2020.

The new video coding project will have a base layer/profile which is expected to be Option 1 with a performance ~30% more than AVC and a 2nd layer/profile that has already a performance ~25% better than HEVC. Licensing terms are expected to be published by patent holders within 2 years. VCEG has decided not to work with MPEG on this coding standard. Are we back to the 合久必分 (things combined for a long time must split) situation? Partly so because the MPEG-VCEG collaboration in VVC is continuing. In any case VVC will provide a compression performance 50% more than HEVC's.

Fifteenth Video Coding Standard

If there was a need to prove that there is no longer "one size fits all" in video coding, just look at the new project called "Low Complexity Enhancements Video Coding (LCEVC)" MPEG has started working on. It is not about a "new video codec", but a technology capable to extend the capabilities of an existing video codec. A typical usage scenario is the addition of, say, the high definition capability to deployed set top boxes that cannot be recalled. LCEVC is expected to reach FDIS in July 2020.

Sixteenth Video Coding Standard

Point Clouds are not really the traditional "video" content as we know it, namely sequences of "frames" at a frequency that is sufficiently high to fool the eye into believing that the motion is natural. In point clouds, motion is given by dynamic point clouds that represent the surface of objects moving in the scene. For the eye, however, the end-result is the same: moving pictures displayed on a 2D surface, whose objects can be manipulated by the viewer (to do this, however, a system layer is required, and MPEG is already working on it).

MPEG is working on two different technologies: the first one uses HEVC to compress projections of portions of a point cloud (and is therefore well-suited for entertainment applications because it can rely on an existing HEVC decoder) and the second one uses computer graphics technologies (and is currently more suited to automotive and similar applications). The former will become FDIS in January 2020 and the latter in April 2020.

Seventeenth and Eighteenth Video Coding Standards

Unfortunately, the crystal ball gets blurred as we move into the future. MPEG tries to face the reality by investigating several technologies capable to provide solutions for alternative immersive experiences. After providing HEVC and OMAF for 3DoF experiences (where the user can only have roll, pitch, and yaw movement of the head), MPEG is working on OMAF v2 for 3DoF+ experiences (where the user can have a limited translation of the head). The video for OMAF v2 will be provided by MPEG-I Part 7 Immersive Media Metadata. As the title say, this is not about compression, but metadata that enable the viewer not to suffer from parallax errors The FDIS is expected in July 2020. Other investigations regard 6DoF (where the user can have full translation of the head) and light field.

Conclusions

The last 40 years have seen digital video converted from a dream into a reality that involves billions of users every day with an incredible lowering to the threshold to access video creation. This long ride is represented in the figure that ventures into the next steps of the ride.

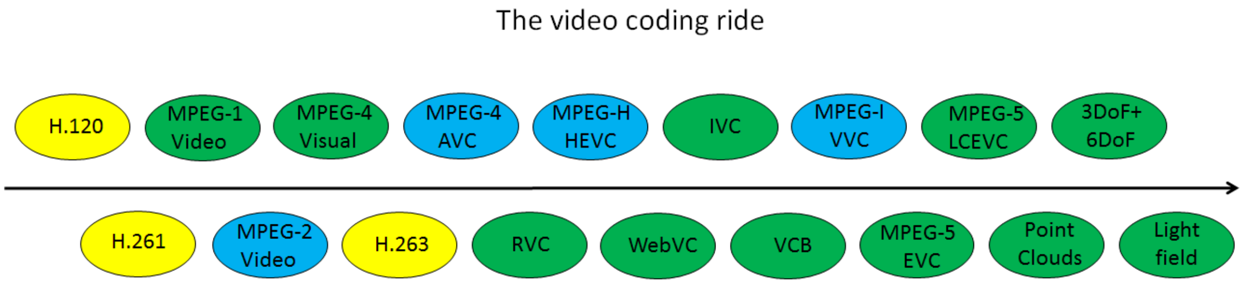


Figure 17 – 30 years of video coding at MPEG
Legend: yellow=ITU-T only, green= MPEG only, turquoise= joint with ITU-T

MPEG keeps working to make sure that manufacturers and content/services providers have access to more and better standard visual technologies for an increasingly diversified market of increasingly demanding users.

8.2 More video features

In the preceding chapter the intense 40-year long history of ITU and MPEG video compression standards was presented. In this chapter the focus is on how over the years, MPEG standards have been offering more functionalities in addition to video compression and how the next generations of standard will add even more.

Table 10 – MPEG video coding standards and functionalities

Standard Title Year	**1 Video 1992**	**2 Video 1994**	**4 Visual 1998**	**4 AVC 2003**	**4 WebVC 2014**	**4 IVC 2016**	**4 VCB 2017**	**H HEVC 2013**	***I VVC 2020***	***5 EVC 2020***	***5 LCEVC 2020***	***I V-PCC 2020***	***I G-PCC 2020***	***I OMAF 2020***
Progressive	•	•	•	•	•	•	•	•	•	•	•			
Interlace		•	•	•	•									
Scalablility		•	•	•				•	•					
Multiview		•	•	•				•	•					
>4:2:0		•	•	•				•	•					
Objects			•											
Error resilience			•	•	•			•	•					
>8 bits				•				•	•	•	•			
Wide Colour Gamut				•				•	•					
HDR				•				•	•					
3DoF				•				•	•					
Option 1					(•)	(•)	[•]			*{•}*				
SCC								•	•					
Two layers											•			
Point clouds												•	•	
3DoF+														•
6DoF														
Light field														

Progressive video
In 1988 MPEG started its first video coding project for interactive video applications on compact disc (MPEG-1). Input video was assumed to be **progressive** (25/29.97 Hz, but it also supported more frame rates) and spatial resolution was Source Image Format (SIF), i.e. 240 or 288 lines of 352 pixels each, depending on whether the original video was 525 of 625 lines. The syntax supported spatial resolutions up to 16 Kpixels. Obviously progressive scanning is a feature that all MPEG video coding standards have supported since MPEG-1. The (obvious) exception is point clouds because there are no "frames".

Table 10 gives an overview of all MPEG video compression standards – past, present and planned. Those in italic have not reached Final Draft International Standard (FDIS) level.

Interlaced video
In 1990 MPEG started its second video coding project – MPEG-2 – targeting digital television. Therefore, the input was assumed to be **interlaced** (frame rate of 50/59.94 Hz, but it also supported other frame rates) and spatial resolution was standard/high definition, and up. The resolution space was quantised by means of levels, the second dimension after profiles. MPEG-4 Visual and Advanced Video Coding (AVC) are the two last standards with specific interlace tools. An attempt was made to introduce interlace tools in High Efficiency Video Coding (HEVC) but the technologies presented did not show appreciable improvements when compared with progressive tools. HEVC does have some indicators (SEI/VUI) to tell the decoder that the video is interlaced.

Scalability, multiview and higher croma resolution
MPEG-2 was the first standard to tackle **scalability** (High Profile), **multiview** (Multiview Profile) and **higher croma resolution** (4:2:2 Profile). Several subsequent video coding standards (MPEG-4 Visual, AVC and HEVC) also support these new features. Versatile Video Coding (VVC) is expected to do the same, probably not in version 1.

Video objects and error resilience
MPEG-4 Visual supports coding of **video objects** and **error resilience**. The first feature has remained specific to MPEG-4 Visual. Most video codecs allow for some error resilience (e.g. with the introduction of slices in MPEG-1). However, MPEG-4 Visual – mobile communication being one relevant use case – was the first to specifically consider error resilience as a tool. MPEG-2 first tried to develop 10-bit support and the empty part 8 of MPEG-2 is what is left of that attempt.

Wide Colour Gamut, High Dynamic Range, 3 Degrees of Freedom
WCG, HDR and 3DoF are all supported by AVC. These functionalities were first introduced in HEVC, later added to AVC and are planned to be supported in VVC as well. WCG allows to display a wider gamut of colours, HDR allows to display pictures with brighter regions and with more visible details in dark areas, SCC allows to achieve better compression of non natural (synthetic) material such as characters and graphics and 3DoF (also called Video 360) allows to represent pictures projected on a sphere.

More than 8 quantisation bits
AVC supports **more than 8 quantisation bits** extended to 14 bits. HEVC even support 16 bits. VVC, EVC and LCEVC are expected to also support more than 8 quantisation bits.

Option 1
WebVC was the first MPEG attempt at defining a video coding standard that would **not require a licence that involves payment of fees** (Option 1 in ISO language, legal language more complex

than this). Strictly speaking, WebVC is not a new standard because MPEG has simply extracted what was the Constrained Baseline Profile in AVC (originally, AVC tried to define an Option 1 profile but did not achieve the goal and did not define the profile) with the hope that WebVC could achieve Option 1 status. The attempt failed because some companies confirmed their Option 2 patent declarations (i.e. a licence is required to use the standard) already made against the AVC standard. The brackets in the figure convey this fact.
Video Coding for Browsers (VCB) is the result of a proposal made by a company in response to an MPEG Call for Proposals for Option 1 video coding technology. Another company made an Option 3 patent declaration (i.e. unavailability to license the technology). As the declaration did not contain any detail that could allow MPEG to remove the allegedly infringing technologies, ISO did not publish VCB as a standard. The square brackets in the figure convey this fact.
Internet Video Coding (IVC) is the third video coding standard intended to be Option 1. Three Option 2 patent declarations were received and MPEG has declared its availability to remove patented technology from the standard if specific technology claims will be made. The brackets convey this fact.

Two-layer video coding
Finally, Essential Video Coding (EVC), part 1 of MPEG-5 (however, the project has not been formally approved by ISO yet), is expected to be a **two-layer** video coding standard. The EVC Call for Proposals requested that the technologies provided in response to the Call for the first (lower) layer of the standard to be Option 1. Technologies for the second (higher) layer are Option 2. The curled brackets in the figure convey this fact.
Low Complexity Enhancement Video Coding (LCEVC) is another two-layer video coding standard. Unlike EVC, however, in LCEVC the lower layer is not tied to any specific technology and can be any video codec. The goal of the 2nd layer is to extend the capability of an existing video codec. A typical usage scenario is to give a large amount of already deployed standard definition set top boxes that cannot be recalled the ability to decode high definition pictures.

Screen Content Coding
SCC allows to achieve better compression of non-natural (synthetic) material such as characters and graphics. It is supported by HEVC and is planned to be supported in VVC and possibly EVC.
3D point clouds
Today technologies are available to capture **3D point clouds,** typically with multiple cameras and depth sensors producing up to billions of points for realistically reconstructed scenes. Point clouds can have attributes such as colors, material properties and/or other attributes and are useful for real-time communication, Geographic Information System (GIS), Computer Aided Design (CAD) and cultural heritage applications. MPEG-I part 5 will specify lossy compression of 3D point clouds employing efficient geometry and attributes compression, scalable/progressive coding, and coding of point clouds sequences captured over time with support of random access to subsets of the point cloud.
Other technologies capture points clouds, potentially with low density of points. These allow users to freely navigate in multi-sensory 3D media spaces. Such representations require a large amount of data, that cannot be transmitted on today’s networks. Therefore, MPEG is developing a second, geometry-based PCC standard, as opposed to the previous one which is video-based, for efficient lossless compression of sparse point clouds.

Three Degrees of Freedom +
3DoF+ is a term used by MPEG to indicate a usage scenario where the user can have translational movements of the head. As a user in a 3DoF scenario sees annoying parallax errors if they move the head too much, MPEG is developing a standard that specifies appropriate metadata that help

present the best image based on the viewer's position if available, or to synthesise a missing one, if not available.

Six Degrees of Freedom and Lightfield

6DoF indicates a use scenario where the user can freely move in a space and enjoy a 3D virtual experience that matches the one in the real world. **Lightfield** refers to new devices that can capture a spatially sampled version of a light field that has both spatial and angular light information in one shot. The size of captured data is not only larger but also different than traditional camera images. MPEG is investigating new and compatible compression methods for potential new services.

Conclusions

In 30 years compressed digital video has made a lot of progress, e.g., bigger and brighter pictures with less bitrate and other features. The end point is nowhere in sight.

Thanks to Gary Sullivan and Jens-Rainer Ohm for useful comments to this chapter.

8.3 Immersive visual experiences

With the exclusion of the first video coding standards, MPEG has always supported more than rectangular monovision. This chapter explores the evolution of this endeavour over the years.

The early days

MPEG-1 did not have big ambitions (but the outcome was not modest at all ;-). MPEG-2 was ambitious because it included scalability – a technology that reached maturity only some 10 years later – and multiview. As depicted in *Figure 18*, multiview was possible because, when two close cameras point to the same scene, it is possible to exploit intraframe, interframe and interview redundancy.

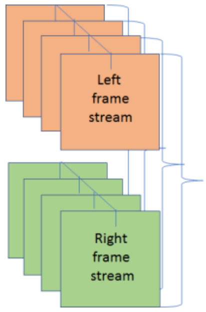


Figure 18: Redundancy in multiview video

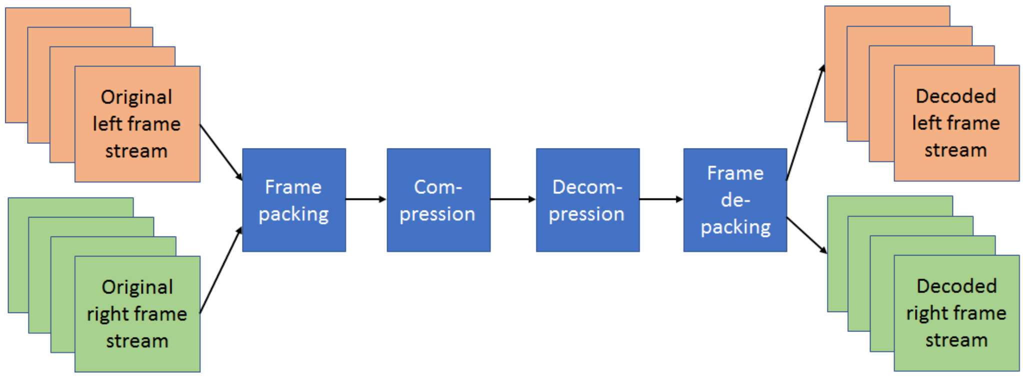


Figure 19 – Frame packing in AVC and HEVC

Both MPEG-2 scalability and multiview saw little take up and both MPEG-4 Visual and AVC had multiview profiles. AVC included Multiview Video Coding (MVC) which was adopted by the Blu-ray Disc Association. The rest of the industry, however, took another turn as depicted in *Figure 19.*

If the left and right frames of two video streams are packed in one frame, regular compression can be applied to the packed frame. At the decoder, the frames are decompressed and then de-packetised to yield the two video streams.
This is a practical but less that optimal solution. Unless the frame size of the codec is doubled, the horizontal or the vertical resolution is compromised depending on the frame-packing method used. Because of this, a host of other more sophisticates, but eventually not successful, frame packing methods have been introduced into the AVC and HEVC standards. The relevant information is carried by Supplemental Enhancement Information (SEI) messages, because the specific frame packing method used is not normative as it applies to the "display" part of the process.
The HEVC standard, too, supports 3D vision with tools that efficiently compress depth maps, and exploit the redundancy between video pictures and associated depth maps. Unfortunately use of HEVC for 3D video has also been limited.

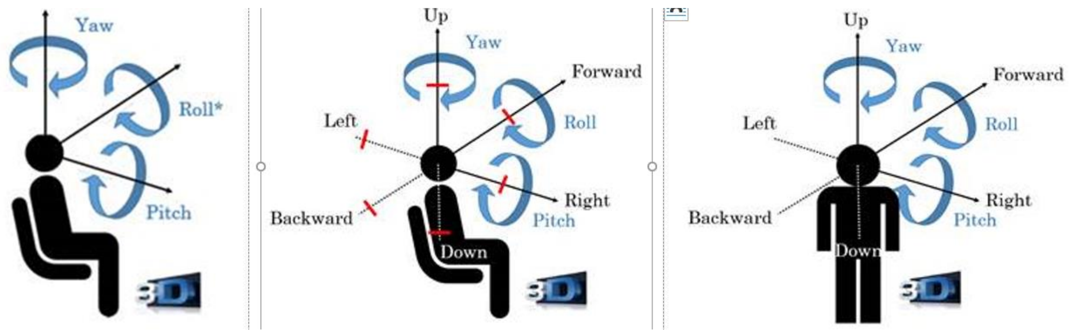


Figure 20 – 3DoF (left), 3DoF+ (centre) and 6DoF (right)

MPEG-I
The MPEG-I project – ISO/IEC 23090 Coded representation of immersive media – was launched at a time when the word "immersive" was prominent in many news headings. *Figure 20* gives three examples of immersivity where technology challenges increase moving from left to right.
In 3 Degrees of Freedom (3DoF) the user is static, but the head can Yaw, Pitch and Roll. In 3DoF+ the user has the added capability of some head movements in the three directions. In 6 DoF the user can freely walk in a 3D space.
Currently there are several activities in MPEG that aim at developing standards that support some form of immersivity. While they had different starting points, they are likely to converge to one or, at least, a cluster of points (hopefully not to a cloud😉).

OMAF
Omnidirectional Media Application Format (OMAF) is not about compressing but about storing and delivering immersive video. Its main features are:
1. Support of several projection formats in addition to the equi-rectangular one
2. Signalling of metadata for rendering of 360° monoscopic and stereoscopic audio-visual data
3. Use of MPEG-H video (HEVC) and audio (3D Audio)
4. Several ways to arrange video pixels to improve compression efficiency
5. Use of the MP4 File Format to store data
6. Delivery of OMAF content with MPEG-DASH and MMT.

MPEG has released OMAF v.1 in 2018 and is now working on v.2. The standard is published as ISO/IEC 23090-2.

3DoF+

If the current version of OMAF is applied to a 3DoF+ scenario, the user experience is affected by parallax errors that are more annoying the larger the movement of the head.
To address this problem, MPEG is working on a specification of appropriate metadata (to be included in the red blocks in *Figure 21*) to help the Post-processor to present the best image based on the viewer's position if available, or to synthesise a missing one, if not available.

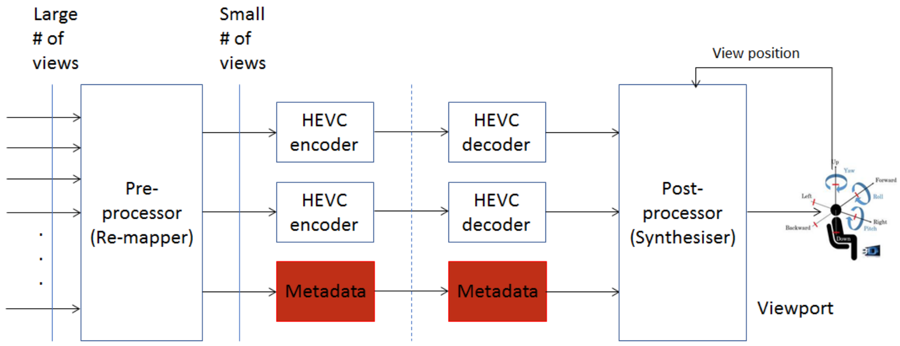


Figure 21: 3DoF+ use scenario

The 3DoF+ standard will be added to OMAF which will be published as 2nd edition. Both standards are planned to be completed in July 2020.

VVC

Versatile Video Coding (VVC) is the latest of MPEG video compression standards supporting 3D vision. Currently VVC does not specifically include full-immersion technologies, as it only supports omnidirectional video as in HEVC. However, VVC could not only replace HEVC in *Figure 21*, but also be the target of other immersive technologies as will be explained later.

Point Cloud Compression

3D point clouds can be captured with multiple cameras and depth sensors. The points can number a few thousands up to a few billions with attributes such as colour, material properties etc. MPEG is developing two different standards whose choice depends on whether the points are dense (Video-based PCC) or sparse (Geometry-based PCC). The algorithms in both standards are scalable, progressive and support random access to subsets of the point cloud. V-PCC is lossy and G-PCC is currently lossless. See here for an example of a Point Cloud test sequence being used by MPEG for developing the V-PCC standard.
MPEG plans to release Video-based Point Cloud Compression as FDIS in January 2020 and Geometry-based Point Cloud Compression as FDIS in April 2020.
Next to PCC compression, MPEG is working on Carriage of Point Cloud Data with the goal to specify how PCC data can be stored in ISOBMFF and transported with DASH, MMT etc.

6DoF

MPEG is carrying out explorations on technologies that enable 6 degrees of freedom (6DoF). The reference diagram for that work is what looks like a minor extension of the 3DoF+ reference model (see *Figure 22*). However, it may have huge technology implications.

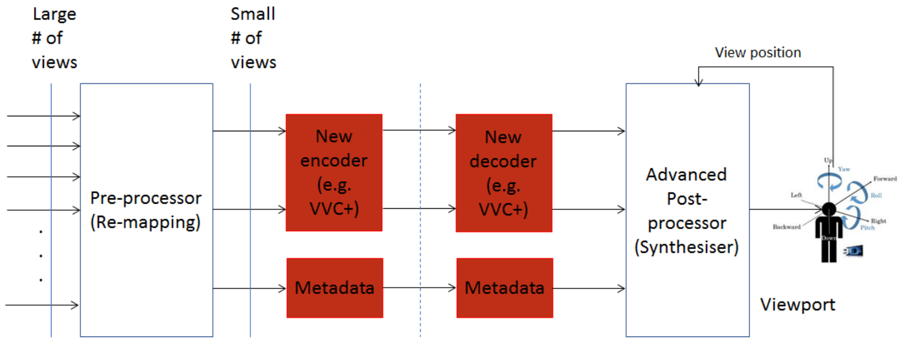


Figure 22: 6DoF use scenario

To enable a viewer to freely move in a space and enjoy a 3D virtual experience that matches the one in the real world, we still need some metadata as in 3DoF+ but likely also additional video compression technologies that could be plugged into the VVC standard.

Light field

The MPEG Video activity is all about standardising efficient technologies that compress digital representations of sampled electromagnetic fields in the visible range captured by digital cameras. Roughly speaking we have 4 types of camera:

1. Conventional cameras with a 2D array of sensors receiving the projection of a 3D scene
2. An array of cameras, possibly supplemented by depth maps
3. Point clouds cameras
4. Plenoptic cameras whose sensors capture the intensity of light from a number of directions that the light rays travel to reach the sensor.

Technologically speaking, #4 is an area that has not been shy in promises and is delivering some of them. However, economic sustainability for companies engaged in developing products for the entertainment market has been a challenge.

MPEG is currently engaged in Exploration Experiments (EE) to check

1. The coding performance of Multiview Video Data (#2) for 3DoF+ and 6DoF, and Lenslet Video Data (#4) for Light Field
2. The relative coding performance of Multiview coding and Lenslet coding, both for Lenslet Video Data (#4).

However, MPEG is not engaged in checking the relative coding performance of #2 data and #4 data because there are no #2 and #4 test data for the same scene.

Conclusions

In good (?) old times MPEG could develop video coding standards – from MPEG-1 to VVC – by relying on established input video formats. This somehow continues to be true for Point Clouds as well. On the other hand, Light Field is a different matter because the capture technologies are still evolving and the actual format in which the data are provided has an impact on the actual processing that MPEG applies to reduce the bitrate.

MPEG has bravely picked up the gauntlet and its machine is grinding data to provide answers that will eventually lead to one or more visual compression standards to enable rewarding immersive user experiences.

MPEG is holding a "Workshop on standard coding technologies for immersive visual experiences" in Gothenburg (Sweden) on 10 July 2019. The workshop, open to the industry, will be an opportunity for MPEG to meet its client industries, report on its results and discuss industries' needs for immersive visual experiences standards.

8.4 Video can be green

Introduction

MPEG has given humans the means to add significant more effectiveness and enjoyment to their lives. If billions of people, however, can stream video to anywhere at any time of the day, global energy consumption can only increase. Enhanced experiences provided by newer features such as High Dynamic Range further adds energy consumption in the display. More sophisticated compression algorithms consume more energy, even though this is mitigated by more advanced circuit geometry.

In 2013 MPEG issued a Call for Proposal on "Green MPEG" requesting technologies that enable reduction of energy consumption in video codecs and 3 years later MPEG released ISO/IEC 23001-11 Green Metadata. This was followed by a number of ancillary activities.

It should be clear that Green Metadata should not be seen as an attempt at solving the global problem of energy consumption. More modestly Green Metadata seeks to reduce power consumption in the encoding, decoding, and display process while preserving the user's quality of experience (QoE). Green Metadata can also be used to reduce the QoE in a controlled way.

The standard does not require changing the operation of a given encoder or decoder (i.e. changing the video coding standard). It just requires to be able to "access" and "influence" appropriate operating points of any or the encoder, decoder or display functionalities.

A systems view

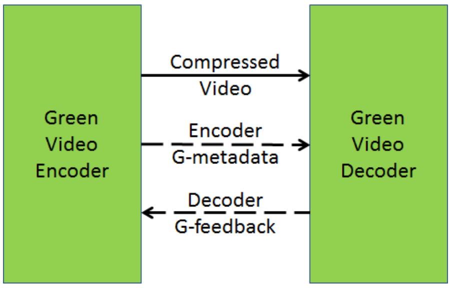


Figure 23 – Conceptual diagram of a green encoder-decoder pair

Green Metadata target metadata suitable to influence the video encoding, decoding and display processes. The framework, however, could be easily generalised by replacing "video" and "display" with "media" and "presentation". However, the numerical results obtained in the video case cannot be directly extrapolated to other media.

Let's start from *Figure 25* representing a conceptual diagram of a green encoder-decoder pair.

The Green Video Encoder (GVE), is a regular video encoder that generates a compressed video bitstream and also a stream of metadata (G-Metadata) for use by a Green Video Decoder (GVD) to reduce power consumption. When a return channel is available (e.g. on the internet), the GVD may generate feedback information (G-Feedback) that the GVE may use to generate a compressed video bitstream that demands less power for the GVD to decode.

To understand the scope of Green Metadata standardisation, it is worth digging a little bit in the following high-level diagram and see what is the new "green component" that is added. *Figure 24* helps to understand such green components.

The GVE generates G-Metadata packaged by the G-Metadata Generator for transmission to a GVD. The GDV G-Metadata Extractor extracts the G-Metadata payload and passes the GVE G-Metadata to the GVD Power Manager along with G-Metadata coming from the GVD. The GVD Power Manager, based on the two G-Metadata streams and possibly other input such as user's input (not shown in figure), may send

1. Power Control data to the Video Decoder to change its operation
2. G-Feedback data to the G-Feedback Generator to package it for transmission to the GVE.

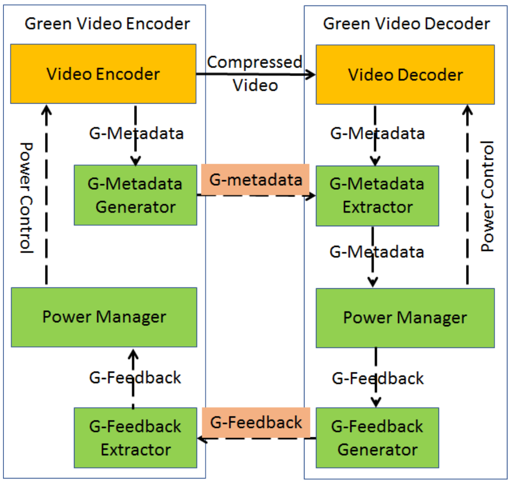


Figure 24 – Inside a green encoder-decoder pair

At the GVE side, G-Feedback Extractor extracts the G-Feedback data and passes them to the GVE Power Manager. This may send Power Control data to the Video Encoder to change its operation. To examine a bit more in detail how G-Metadata can be used, it is helpful to dissect the Video Encoder and Decoder pair as shown in *Figure 25*.

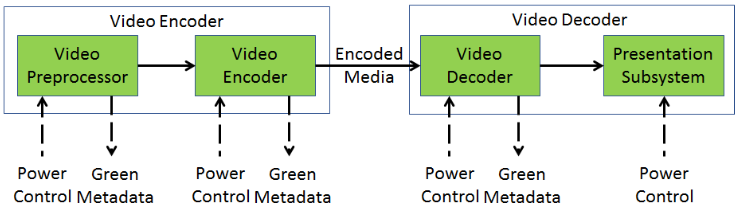


Figure 25 – Inside the encoder and decoder

The Video Encoder is composed of a Media Preprocessor (e.g. a video format converter) and a Media Encoder. The Video Decoder is made of a Media Decoder and a Presentation Subsystem (e.g. to drive the display). All subsystems send G-Metadata and receive Power Control. The Presentation Subsystem only receives Power Control.
What is standardised in Green Metadata? As always in MPEG, the minimum that is required for interoperability. This means the Encoder Green Metadata and the Decoder Green Feedback (in red in *Figure 24*) that are exchanged by systems which are potentially manufactured by different entities. Other data formats inside the GVE and the GVD are a matter for GVE and GVD manufacturers to decide because they do not affect interoperability but may affect performance. In particular, the logic of the Power Manager that generates Power Control can be the differentiating factor between implementations.

Achieving reduced power consumption
In the following the 3 areas positively affected by the use of the Green Metadata standard – encoder, decoder and display – will be illustrated.

1. *Encoder*. By using a segmented delivery mechanism (e.g. DASH), encoder power consumption can be reduced by encoding video segments with alternate high/low quality. Low-quality segments are generated by using lower-complexity encoding (e.g. fewer encoding modes and reference pictures, smaller search ranges etc.). Green Metadata include the quality of the last picture of each segment. The video decoder enhances the low-quality segment by using the metadata and the last high-quality video segment.
2. *Decoder*. Lowering the frequency of a CMOS circuit implementing a video decoder reduces power consumption because this roughly increases linearly with the clock frequency and quadratically with the voltage applied. In a software decoder picture complexity can be used to control the CPU frequency. One type of Green Metadata signals the duration and degree of complexity of upcoming pictures. This can be used to select the most appropriate setting and offer the best QoE for a desired power-consumption level.
3. *Display*. The display adaptation technique known as backlight dimming reduces power consumption by dimming the LCD backlight while RGB values are scaled in proportion to the dimming level (RGB values do not have a strong influence on power consumption).

Green Metadata need to be carried
ISO/IEC 23001-11 only specifies the Green Metadata. The way this information is transported depends on the specific use scenarios (some of them are described in Context, Objectives, Use Cases and Requirements for Green MPEG).
Two transports have been standardised by MPEG. In the first, Green Metadata is transported by a Supplementary Enhancement Information (SEI) message embedded in the video stream. This is a natural solution since Green Metadata are due to be processed in a Green Video Decoder that includes a regular video decoder. In this case, however, transport is limited to decoder metadata, not display metadata. In the second, suitable for a broadcast scenario, all Green Metadata is transported in the MPEG-2 Transport Stream.

Conclusions
In this first attempt at tackling power consumption in the Green Metadata standard, MPEG efforts have been rewarded: with the currently standardised metadata 38% of video decoder power and 12% of video encoder power can be saved without affecting QoE and up to 80% of power can be saved with some degradation. Power saving data were obtained using the Google Nexus 7 platform and the Monsoon power monitor, and a selection of test material.
Interested readers can know more by visiting the MPEG web site and, more so, by purchasing the Green Metadata standard from the ISO website or from a National Body.

9 Audio compression in MPEG

Introduction
Because of obvious technology reasons, the electrical representation of sound information happened before the electrical representation of visual information. The same is true of the services that used that representation to distribute sound information. The digital representation of audio, too, happened before the digital representation of video.
In the early 1980s, the Compact Disc (CD) allowed record companies to distribute digital audio for the *consumer* market, while the D1 digital tape, available in the late 1980's, was meant for the exclusive use of professional applications such as in the studio. Interestingly, the appearance of compression technologies happened in reverse order: *compressed digital video* standards happened before *compressed digital audio* by some 10 years
This statement can be a source of dispute, if a proper definition of *Audio* is not adopted. In this article by Audio we mean sound in the human audible range not generated by a human phonatory

system or for any other sound source for which a sound production model is not available or not used. Indeed, *digital speech* happened in professional applications (the trunk network) some 20 years before the CD. ITU-T G.721 "32 kbit/s adaptive differential pulse code modulation (ADPCM)" dates back to 1984, the same year ITU H.120 was approved as a recommendation.
In this chapter we only talk about audio and not speech. If speech were included this chapter would be a overwhelmed by the large number of speech compression standards. Therefore, this chapter only deals with audio compression standards where audio does not include speech. There is a single exception that will be mentioned later.
Unlike video compression where ITU-T is the non-MPEG body that develops video coding standards, in audio compression standards MPEG dominance is total. Indeed ITU-R, who does need audio compression for its digital audio broadcasting standards, prefers to rely on external sources, including MPEG.

MPEG-1 Audio

Those interested in knowing why and how a group – MPEG – working in video compression ended up also working on audio compression (and a few more other things) can look here. The kick off of the MPEG Audio group took place on 1-2 December 1988, when, in line with a tradition that at that time had not been fully established yet, a most diverse group of audio coding experts met in Hannover and kick-started the work that eventually gave rise to the MPEG-1 Audio standard that MPEG released in November 1992.
Very often, the Audio group in MPEG is the forerunner of things to come. In this instance the first is that while the broadcasting world shunned the low-resolution MPEG-1 Video compression standard, it very much valued the MPEG-1 Audio compression standard. The second is that, unlike video, which essentially relied on the same coding architecture, the Audio Call for Proposals had yielded two classes of algorithms, one that was a well-established, easier to implement but less performing and the other that was more recent, harder to implement but more performing. The work to merge the two technologies was painstaking but eventually the standard included 3 layers (a notion that MPEG first and then the industry later called profiles) where both technologies were used.
Layer 1 was used in Digital Compact Cassette (DCC), a product discontinued a few years after its introduction, Layer 2 was used in audio broadcasting and as the audio component of Video CD (VCD). Layer 3 (MP3) is not in particular need of an introduction 😉. As revised in the subsequent MPEG-2 effort, MP3 provided a user experience with no perceivable difference as compared to the original CD signal for most content at 128 kbit/s from a CD source of 1.44 Mbit/s, i.e with a compression of 11:1.

MPEG-2 Audio

The main goal of this standard, approved in 1994, was multi-channel audio with the key requirement that an MPEG-1 Audio decoder should be able to decode a stereo component of an MPEG-2 Audio bitstream. Backward compatibility is particularly useful in the broadcasting world because an operator can upgrade to a multi-channel services without losing the customers who only have an MPEG-1 Audio decoder.

MPEG-2 AAC

Work on MPEG-2 Advanced Video Coding (AAC) was motivated by the request of those who wished to provide the best possible audio quality without backward compatibility constraints. This meant that layer 2 must decode both layer 1 and 2, and layer 3 must decode all layers. MPEG-2 AAC, released in April 1997, is built upon the MP3 technology and can provide perceptually transparent audio quality at 128 kbit/s for a stereo signal, and 320 kbit/s for a 5.1 channel signal (i.e. as in digital television).

MPEG-4 AAC
In 1998 MPEG-4 Audio was released with the other 2 MPEG-4 components – Systems and Visual. Again MPEG-4 AAC is built on MPEG-2 AAC. The dominating role of MP3 in music distribution was shaken in 2003 when Apple announced that its iTunes and iPod products would use MPEG-4 AAC as primary audio compression algorithm. Most PCs, smart phones and later tablets could play AAC songs. Far from using AAC as a pure *player* technology, Apple started the iTunes service that provides songs in AAC format packaged in the MPEG-4 File Format, with filename extension ".m4a".

AAC-LD
In 1999 MPEG released MPEG-4 amendment 1 with a low delay version of AAC, called Low Delay AAC (AAC-LD). While a typical AAC encoder/decoder has a one-way latency of ~55 ms (transform delay plus look-ahead processing), AAC-LD achieves a one-way latency of only 21 ms by simplifying and replacing some AAC tools (new transform with lower latency and removal of look-ahead processing). AAC-LD can be used as a conversational codec, with a signal bandwidth and perceived quality of a music coder with excellent audio quality at 64 kb/s for a mono signal.

MPEG-4 HE-AAC
In 2003 MPEG released the MPEG-4 High Efficiency Advanced Audio Coding (HE-AAC), as amendment 1 to MPEG-4. HE-AAC helped to consolidate the role of the mobile handset as the tool of choice to access very good audio quality stereo music at 48 kbit/s, more than a factor of 2.5 better than AAC, for a compression ratio of almost 30:1 relative to the CD signal.
HE-AAC adds the spectral bandwidth replication (SBR) tool to the core AAC compression engine. Since AAC was already widely deployed, this permitted extending this base to HE-AAC by only adding the SBR tool to existing AAC implementations.

MPEG HE-AAC v2
In the same year 2003, but 9 months later, MPEG released the MPEG HE-AAC v2 profile. This originated from a tools contained in amendment 2 to MPEG-4 (Parametric coding for high-quality audio). While the core parametric coder did not enjoy wide adoption, the Parametric Stereo (PS) tool in the amendment could very efficiently encode stereo music as a mono signal plus a small amount of side-information. HE-AAC v2, the combination of PS tool with HE-AAC, enabled transmission of a stereo signal at 32 kb/s with very good audio quality.
This profile was also adopted by 3GPP under the name Enhanced aacPlus. Adoption by 3GPP paved the way for HE-AAC v2 technology to be incorporated into mobile phones. Today, more than 10 billion mobile devices support streaming and playout of HE-AAC v2 format songs. Since HE-AAC is built on AAC, these phone also support streaming and playout of AAC format songs.

ALS and SLS
In 2005 MPEG released two algorithms for lossless compression of audio, MPEG Audio LosslesS coding (ALS) and Scalable to LosslesS coding (SLS). Both provide perfect (i.e. lossless) reconstruction of a standard Compact Disc audio signal with a compression ratio approximately 2:1. An important feature of SLS is that it has a variable compression ratio: it can compress a stereo signal to 128 kb/s (11:1 compression ratio) with excellent quality as an AAC codec but it can achieve lossless reconstruction with a compression ratio of 2:1 by increasing the coded bitrate (i.e. by decreasing the compression ratio) in a continuous fashion.

MPEG Surround
ALS/SLS were the last significant standards in MPEG-4 Audio, which is MPEG's most long-lived audio standard. First issued in 1999, 20 years later (in 2019) MPEG has issued its Fifth Edition.

After "closing the MPEG-4 era", MPEG created the MPEG-D suite of audio compression standards. The first of these was MPEG Surround, issued in 2007. This technology is a generalised PS of HE-AAC v2 tool in the sense that, MPEG Surround can operate as a 5-to-2 channel compression tool or as an M-to-N channel compression tool. This "generalised PS" tool is followed by a HE-AAC codec. Therefore, MPEG Surround builds on HE-AAC as much as HE-AAC builds on AAC. MPEG Surround provides an efficient bridge between stereo and multi-channel presentations in low-bitrate applications. It has very good compression while maintaining very good audio quality and also low computational complexity. While HE-AAC can transmit stereo at 48 kbit/s, MPEG Surround can transmit 5.1 channel audio within the same 48 kbit/s transmission budget. The complexity is no greater than stereo HE-AAC's. Hence MPEG Surround is a "drop-in" replacement for stereo services to extend them to 5.1 channel audio!

AAC-ELD

In 2007 MPEG released Enhanced Low Delay AAC (AAC-ELD) technology. This combines tools from other profiles: SBR and PS from HE-AAC v2 profile and AAC-LD. The new codec provides even greater signal compression with only a modest increase in latency: AAC-ELD provides excellent audio quality at 48 kb/s for a mono signal with a one-way latency of only 32 ms.

SAOC

In 2010 MPEG released MPEG-D Spatial Audio Object Coding (SAOC) which allows very efficient coding of a multi-channel signal that is a mix of objects (e.g. individual musical instruments). SAOC down-mixes the multi-channel signal, e.g. stereo to mono, codes and transmits the mono signal along with some side-information, and then up-mixes the received and decoded mono signal back to a stereo signal such that user perceives the instruments to be placed at the correct positions and the resulting stereo signal to be the same as the original. This is done by exploiting the fact that at any instant in time and any frequency region one of the instruments will tend to dominate the others so that in this time/frequency region the other signals will be perceived with much less acuity, if at all. SAOC analyses the input signal, divides each channel into time and frequency "tiles" and then decides to what extent each tile dominates. This is coded as side information.

An example SAOC application is teleconferencing, in which a multi-location conference call can be mixed at the conference bridge down to a single channel and transmitted to each conference participant, along with the SAOC side information. At the user's terminal, the mono channel is up-mixed to stereo (or 3 channels – Left-Center-Right) and presented such that each remote conference participant is at a distinct location in the front sound stage.

USAC

Unified Speech and Audio Coding (USAC), released in 2011, combines the tools for speech coding and audio coding into one algorithm, i.e. the tools from MPEG AAC (exploiting the means of human perception of audio) with the tools from a state-of-the-art speech coder (exploiting the means of human production of speech). Therefore, the encoder has both a perceptual model and a speech excitation/vocal tract model and every 20 ms selects the music/speech coding tools. USAC achieves high performance for any input signal, be it music, speech or a mix of speech and music. In the tradition of MPEG standards, USAC extends the range of "good" performance down to as low as 16 kb/s for a stereo signal and provides higher quality as the bitrate is increased. The quality at 128 kbit/s for a stereo signal is slightly better that MPEG-4 AAC so USAC can replace AAC, because its performance is equal or better than AAC at all bit rates. USACcan similarly code multichannel audio signals, and can also optimally encode speech content.

DRC
MPEG-D Dynamic Range Control (DRC) is a technology that gives listeners the ability to control the audio level. It can be a post-processor for every MPEG audio coding technology and modifies the dynamic range of the decoded signal as it is being played. It can be used to reduce the loudest part of a movie so as not to disturb neighbours, to make the quiet portions of the audio louder in hostile audio environments (car, bus, room with many people), to match the dynamics of the audio to that of a smart phone speaker output, which typically has very limited dynamic range. The DRC standard also plays the very important function of normalizing the loudness of the audio output signal, which may be mandated in some regulatory environments. DRC was released in 2015 and extended in 2017 as Amendment 1 Parametric DRC, gain mapping and equalization tools.

3D Audio
MPEG-H 3D Audio, released in 2015, is part of the typical suite of MPEG tools: Systems, Video and Audio. It provides very efficient coding of immersive audio content, typically from 11 to 22 channels of content. The 3D Audio algorithms can actually process any mix of channels, objects and Higher Order Ambisonics (HOA) content, where objects are single-channel audio whose position can be dynamic in time and HOA can encode an entire sound scene as a multi-channel "HOA coefficient" signal.
Since 3D Audio content is immersive, it is conceived as being consumed as a 360-degree "movie" (i.e. video plus audio). The user sits at the center of a sphere ("sweet spot") and the audio is decoded and presented so that the user perceives it to be coming from somewhere on the surrounding sphere. MPEG-H 3D audio also can be presented via headphones (not every consumer has an 11 or 22 channel listening space). Moreover MPEG-H 3D Audio supports use of a default or personalised Head Related Transfer Function (HRTF) to allow the listener to perceive the audio content as if it is from sources all around the listener, just as it would be when using loudspeakers. An added feature of 3D Audio playout to headphones, is that the audio heard by the listener can remain at the "correct" position when the user turns his or her head. In other words, a sound that is "straight ahead" when the user is looking straight ahead is perceived as coming from the left if the user turns to look right. Hence, MPEG-H 3D Audio is already a nearly complete solution for Video 360 applications.

Immersive Audio
This activity (to be released as an FDIS sometime in 2022) is part of the emerging MPEG-I Immersive Audio standard. MPEG is still defining the requirements and functionality of this standard, which will support audio in Virtual and Augmented Reality applications. It will be based on MPEG-H 3D Audio, which already supports a 360 degree view of a virtual world from one listener position ("3 degrees of freedom" or 3DoF) that the listener can move his or her head left, right, up, down or tilted left or right (so-called "yaw, pitch roll"). The Immersive Audio standard will add three additional degrees of freedom, i.e., permit the user to get up and walk around in the Virtual World. This additional movement is designated "x, y, z," so that MPEG-I Immersive Audio supports 6 degrees of freedom (6 DoF) which are "yaw, pitch roll and x, y, z." It is envisioned that MPEG-I Immersive Audio will use MPEG-H 3D Audio to compress the audio signals and specify additional metadata and technology to render the audio signals in a fully flexible 6 DoF way.

Conclusions
MPEG is proud of the work done by the Audio group. For 30 years the group has injected generations of audio coding standards into the market. In the best MPEG tradition, the standards are generic in the sense that can be used in audio-only or audio+video applications and often scalable, with a new generation of audio coding standards building on previous ones.
This long ride is represented in *Figure 26* that ventures into the next step of the ride.

Today MPEG Audio already provides a realistic 3DoF experience in combination with MPEG Video standards. More will be needed to provide a complete and rewarding 6DoF experience, but MPEG's ability to draw the necessary multi-domain expertise from its membership promises that the goal will be successfully achieved.

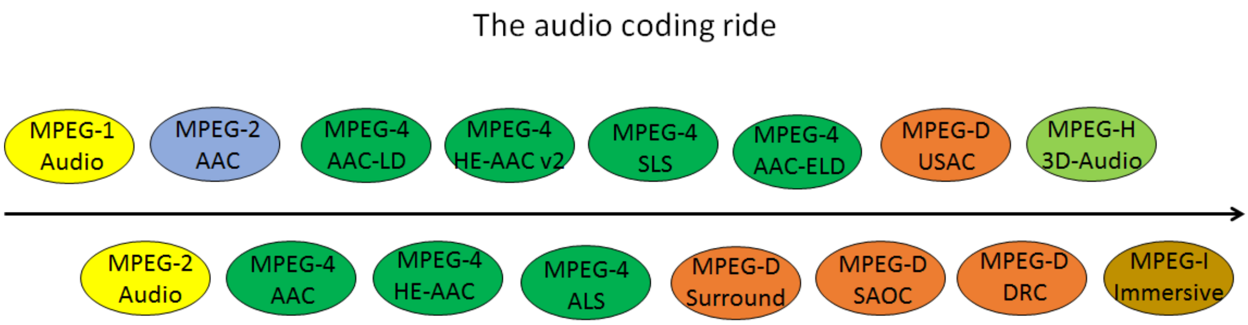


Figure 26: 30 years of MPEG Audio Coding standards
Legend: yellow= 1st generation, dim blue=AAC anticipation, green=AAC, red= MPEG-D, light green=MPEG-H, brown=MPEG-I

Acknowledgements
This chapter would not have been possible without the competent assistance – and memory – of Schuyler Quackenbush, the MPEG Audio Chair.

10 Quality assessment

Introduction

The description of the MPEG workflow highlights the role of quality assessment across the entire MPEG standard life cycle: at the time of issuing a Call for Evidence (CfE) or a Call for Proposals (CfP), carrying out Core Experiments (CE) or executing Verification Tests.
We should consider, however, that in 30 years the coverage of the word “media” has changed substantially.
Originally (1989-90) the media types tested were Standard Definition (SD) 2D rectangular video and stereo audio. Today the video data types include also High Definition (HD), Ultra High Definition (UHD), Stereoscopic Video, High Dynamic Range (HDR) and Wide Colour Gamut (WCG), and Omnidirectional (Video 360).
Today the video information can be 2D, but also multiview and 3D: stereoscopic, 3 degrees of freedom + (3DoF+), 6 degrees of freedom (6DoF) and various forms of light field. Audio has evolved to different forms of Multichannel Audio and 6DoF. Recently Point Clouds have been added to the media types for which MPEG has applied subjective quality measurements to develop compression standards.
This chapter goes inside the work that comes together with *subjectively* assessing the quality of media compression.

Preparing for the tests

Even before MPEG decides to issue a CfE or CfP for compression of some type of media content, viewing or listening to content may take place to appreciate the value of a proposal. When a Call is issued, MPEG has already reached a pretty clear understanding of the use cases and requirements (at the time of a CfE) or the final version of them (at the time of a CfP).

The first step is the availability of appropriate test sequences. Sequences may already be in the MPEG Content Repository or are spontaneously offered by members or are obtained from industry representatives by issuing a Call for Content.
Selection of test sequences is a critical step because MPEG needs sequences that are suitable for the media type and are representative of the use cases. Moreover, test sequences must be in a number that allows MPEG to carry out meaningful and realistic tests.
By the CfE or CfP time, MPEG has also decided what is the standard against which responses to the CfE or CfP should be tested. For example, in the case of HEVC, the comparison was with AVC and, in the case of VVC, the comparison was with HEVC. In the case of Internet Video Coding (IVC) the comparison was with AVC. When such a reference standard does not exist (this was the case for, e.g., all layers of MPEG-1 Audio and Point Cloud Compression), MPEG uses the codec built during the exploratory phase that groups together state of the art tools.
Once the test sequences have been selected, the experts in charge of the reference software are asked to run the reference software encoder and produce the "anchors", i.e. the video sequences encoded using the "old" standard proposals are to be compared with. The anchors are made available in an FTP site so that anybody intending to respond to the CfE or CfP can download them.
The set up used to generate the anchors are documented in "configuration files" for each class of submission. In the case of video, these are (SD/HD/UHD, HDR, 360º) and design conditions (low delay or random access). In order to have comparable data, all proponents must obviously use the same configuration files when they encode the test sequences using their own technologies.
As logistic considerations play a key role in the preparation of quality tests, would-be proponents must submit formal statements of intention to respond to the CfE or CfP to the Requirements group chair (currently Jörn Ostermann) and Test group chair (currently Vittorio Baroncini), and the chair of the relevant technology group, 2 months before the submission deadline.
At the meeting preceding the one in which responses to the CfP/CfE are due, an Ad hoc Group (AhG) is established with the task of promoting awareness of the Call in the industry, to carry out the tests, draft a report and submit conclusions on the quality tests to the following MPEG meeting.

Carrying out the tests

The actual tests are entirely carried out by the AhG under the leadership of AhG chairs (typically the Test chair and a representative of the relevant technology group).
Proponents send the Test Chair their files containing encoded data on hard disk drives or on an FTP site by the deadline specified in the CfE/CfP.
When all drives are received, the Test chair performs the following tasks:

1. Acquire special hardware and displays for the tests (if needed)
2. Verify that the submitted files are all on disk and readable
3. Assign submitted files to independent test labs (sometimes even 10 test labs are concurrently involved is a test run)
4. Make copies and distribute the relevant files to the test labs
5. Specify the tests or provide the scripts for the test labs to carry out.

The test labs carry out a first run of the tests and provide their results for the Test chair to verify. If necessary, the Test chair requests another test run or even visits the test labs to make sure that the tests will run properly. When this "tuning" phase has been successfully executed, all test labs run the entire set of tests assigned to them using test subjects. Tens of "non-expert" subjects may be involved for days.

Test report

Test results undergo a critical revision according to the following steps

1. The Test chair collects all results from all test labs, performs a statistical analysis of the data, prepares and submits a final report to the AhG

2. The report is discussed in the AhG and may be revised depending on the discussions
3. The AhG draws and submits its conclusions to MPEG along with the report
4. Report and conclusions are added to all the material submitted by proponents
5. The Requirements group and the technology group in charge of the media type evaluate the material and rank the proposals. The material may not be made public because of the sensitivity of some of the data.

This signals the end of the competition phase and the beginning of the collaboration phase.

Other media types

The process above has been described having specifically rectangular 2D or 360º video in mind. Most of the process applies to other media types with some specific actions to be made for each of them, e.g.

1. *3DTV*: for the purpose of 3D HEVC tests, polarised glasses as in 3d movies and autostereoscopic displays were used;
2. *3DoF+*: a common synthesiser was used in the 3DoF+ tests to synthesise views that are not available at the decoder;
3. *Audio*: in general subjects need to be carefully trained for the specific tests;
4. *Point Clouds*: videos generated by a common presentation engine consisting of point clouds animated by a script (by rotating the object and seeing it from different viewpoints) were tested for quality as if they were natural videos (NB: there were no established method to assess the quality of point clouds before. It was demonstrated that the subjective tests converged to the same results as the objective measurements).

Verification tests

Verification tests are executed with a similar process. Test sequences are selected and compressed by experts running reference software for the "old" and the "new" standard. Subjective tests are carried out as done in CfE/CfP subjective tests. Test results are made public to provide the industry with guidance on the performance of the new standard. See as examples the Verification Test Report for HDR/WCG Video Coding Using HEVC Main 10 Profile and the MPEG-H 3D Audio Verification Test Report.

Conclusions

Quality tests play an enabling role in all phases of development of a media compression standard. For 30 years MPEG has succeeded in mobilising – on a voluntary basis – the necessary organisational and human resources to perform this critical task.

This chapter has provided a window on an aspect of MPEG life that is little known but instrumental to offer industry the best technology so that users can have the best media experience.

11 Systems standards keep the pieces together

All time that is spent saying that MPEG would not it be what it has become if the Systems aspects had not been part of most of its standards is well spent. This chapter adds more – well spent – time.

MPEG was not the first to deal with the problem of delivering digital audio and video bundled together. In the second half of the 1990's ITU-T used the first digital streams made possible by ITU-T Recommendations to deal with the problem of transmitting audio-visual services using the basic ISDN access at 2B+D (2x64 kbit/s) or at the primary ISDN access.

Figure 27 depicts the solution specified in ITU Recommendation H.221. Let's assume that we have 2 B channels at 64 kbit/s (Basic Access ISDN). H.221 creates on each B channel a Frame Structure of 80 bytes, i.e. 640 bits repeating itself 100 times per second. Each bit position in an octet can be considered as an 8 kbit/s sub-channel. The 8th bit in each octet represents the 8th sub-channel, called the Service Channel.

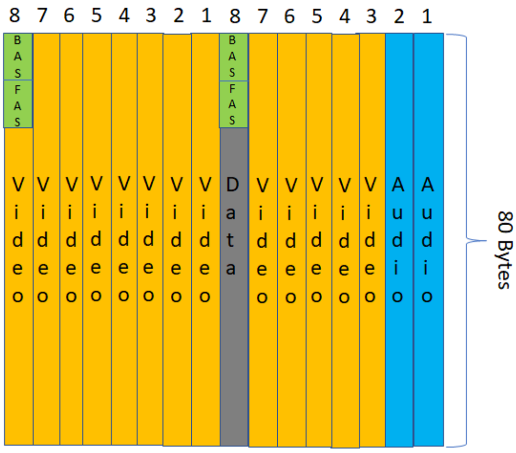


Figure 27: ITU Recommendation H.221

Within the Service Channel, bits 1-8 are used by the Frame Alignment Signal (FAS) and bits 9-16 are used by the Bit Alignment Signal (BAS). Audio is always carried by the first B channel, e.g. by the first 2 subchannels, and Video and Data by the other subchannels (less the bitrate allocated to FAS and BAS).

MPEG-1 Systems

The solution depicted in *Figure 28* bears the mark of the *transmission* part of the telecom industry that had never been much friendly to packet communication. That is why MPEG in the late 1990's had an opportunity to bring some fresh air in this space. Starting from a blank sheet of paper (at that time MPEG still used paper 😉) MPEG designed a flexible packet-based multiplexer to convey in a single stream compressed audio, video and clock information in such a way as to enable audio-video synchronisation.

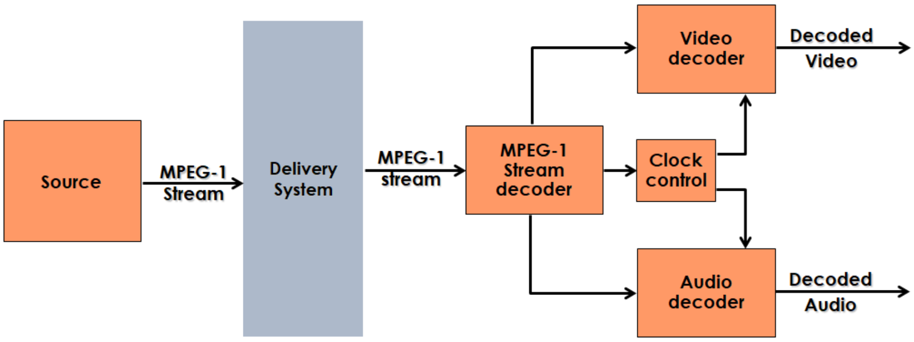


Figure 28: MPEG-1 Systems

The MPEG Systems revolution took time to take effect. Indeed, the European EU 95 project used MPEG-1 Audio layer 2, but designed their frame-based multiplexer for the DAB service.

MPEG-2 Systems

In the early 1990's MPEG started working on another blank sheet of paper. MPEG had designed MPEG-1 Systems, but the MPEG-2 requirements were significantly different. While in MPEG-1 audio and video (possibly many of them in the same stream) had a common time base, the main users of MPEG-2 wanted a system that could deliver a plurality of TV programs, possibly coming from different sources (i.e. with different time bases) and typically with a lot of metadata related to the programs, not to mention some key business enabler such as conditional access information. Moreover, unlike MPEG-1 where it was safe to assume that the bits issuing would travel without errors from a Compact Disc to a demultiplexer, in MPEG-2 it was mandatory to assume that the transmission channel was anything but error-free.

MPEG-2 Transport Stream (TS) provides efficient mechanisms to multiplex multiple audio-visual data streams into one delivery stream. Audio-visual data streams are packetised into small fixed-size packets and interleaved to form a single stream. Information about the multiplexing structure is interleaved with the data packets so that the receiving entity can efficiently identify a specific stream. Sequence numbers help identify missing packets at the receiving end, and timing information is assigned after multiplexing with the assumption that the multiplexed stream will be delivered and played in sequential order.

The structure of the transport bitstream, organised in fixed-length packets of 188 bytes of which 184 bytes are used for the payload, is depicted in *Figure 29*.

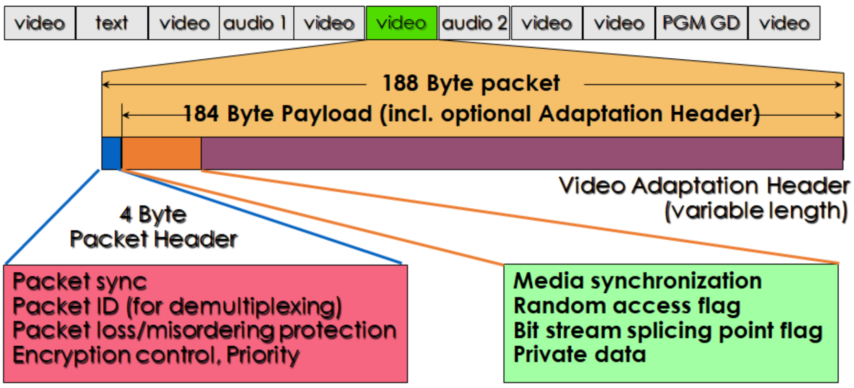


Figure 29 – The MPEG-2 packet format

Figure 30 shows thatMPEG-2 Systems is actually two specifications in one.

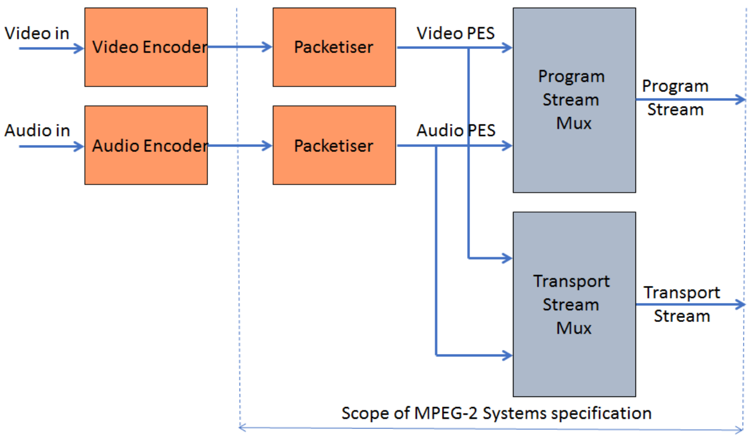


Figure 30: MPEG-2 Systems

The Transport Stream (TS) is a fixed-length packet-based transmission system designed for digital television distribution on error-prone physical channels, while the Program Stream (PS) is a packet-based multiplexer with many points in common with MPEG-1 Systems. `While TS and PS share significant common information, moving content from one to the other may not be immediate.

MPEG-4 Systems

MPEG-4 gave MPEG the opportunity to experience an epochal transition in data delivery. When MPEG-2 Systems was designed, Asynchronous Transfer Mode (ATM) was high on the agenda of the telecom industry and MPEG considered it a possible vehicle to transport MPEG-2 TS streams on telecommunication networks. Indeed, the Digital Audio-Visual Council (DAVIC) designed its specifications on the assumption that MPEG-2 based Video on Demand services would be carried on ATM because at that time IP was still unknown to the telecom (at least to the transmission part), broadcast and consumer electronics worlds.

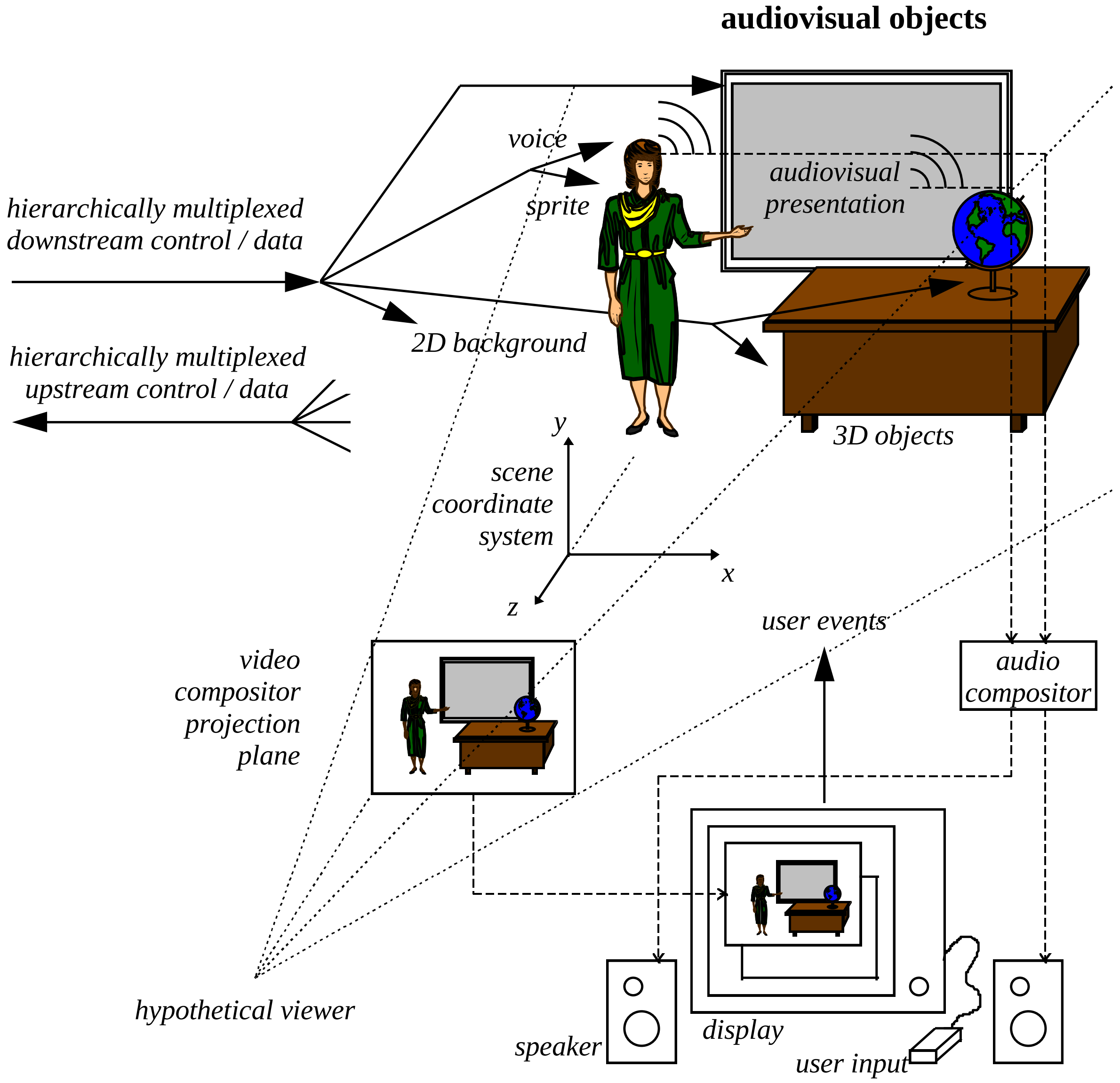


Figure 31 – An example of MPEG-4 scene

The MPEG-4 Systems work was a completely different story than MPEG-2 Systems. An MPEG4 Mux (M4Mux) was developed along the lines of MPEG-1 and MPEG-2 Systems. However, MPEG had to face an unknown world where many transports were surging as possible candidates. Today, 25 years later, the choice is clear, but at that time MPEG was unable to make choices – and it would not even be its task. Therefore MPEG developed the notion of Delivery Multimedia Integration Framework (DMIF), where all communications and data transfers between the data source and the terminal were abstracted through a logical API called the DMIF Application Interface (DAI), independent of the transport type: storage (local interactivity), network (remote interactivity) and broadcast (one way interactivity.

MPEG-4 Systems, however, was about more than interfacing with transport and multiplexing. The MPEG-4 model was a 3D space populated with dynamic audio, video and 3D Graphics objects. *Figure 31* provides an example of MPEG-4 scene with 3D graphic objects and 2D video sprites.

Binary Format for Scenes (BIFS) was the technology designed to provide the scene description functionality, obviously in compressed form.

Figure 32 shows the 4 MPEG-4 layers: Transport, Synchonisation, Compression and Composition.

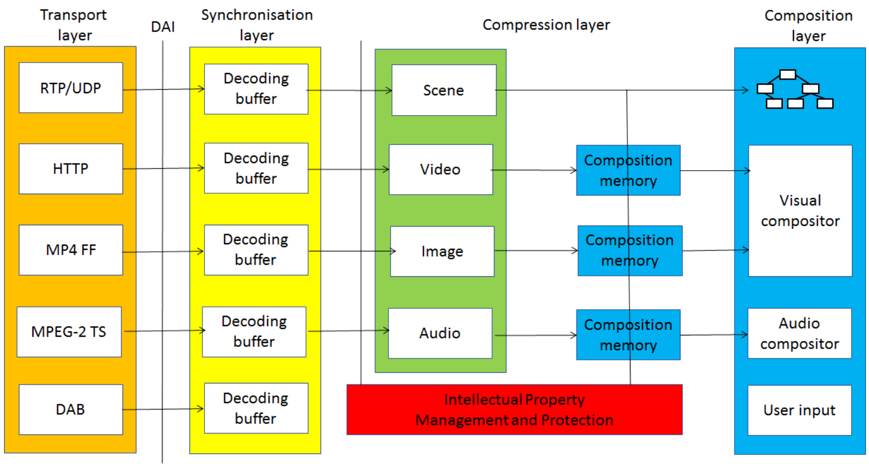


Figure 32 – MPEG-4 Systems

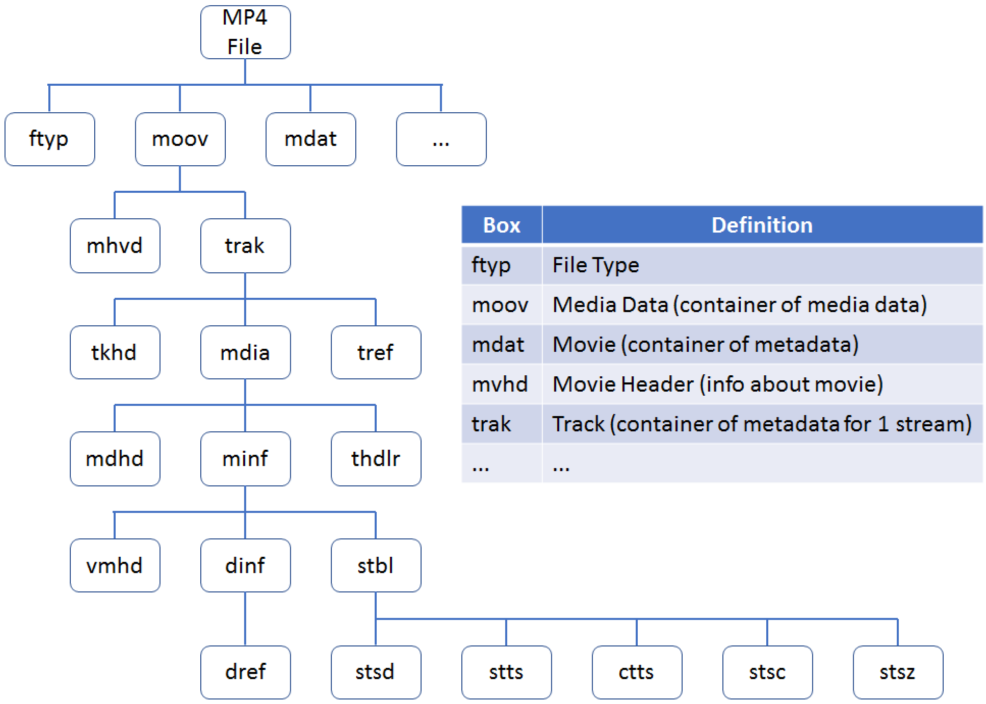
MPEG-4 File Format

For almost 10 years – until 1997 – MPEG was a group making intense use of IT tools (in the form of computer programs that simulated encoding and decoding operation of the standards it was developing) but was not an “IT group”. This can be seen from the fact that, until that time, it had not developed a single file format. Today MPEG can claim to have another such attribute (IT group) along with the many others it has.

In those years MP3 files were already being created and exchanged by the millions, but the files did not provide any structure, if not the one designed by the Audio group. The MP4 File Format, officially called ISO Base Media File Format (ISO BMFF), filled that gap as it can be used for editing, HTTP streaming and broadcasting.

Let’s have a high-level look to understand the sea that separates MP3 files from the MP4 FF. MP4 FF contains tracks for each media type (audio, video etc.), with additional information: a four-character the media type ‘name’ with all parameters needed by the media type decoder. “Track selection data” helps a decoder identify what aspect of a track can be used and to determine which alternatives are available.

Data are stored in a basic structure called box with attributes of length, type (expressed by 4 printable characters), possibly version and flags. No data can be found outside of a box. *Figure 33* shows a possible organisation of an MP4 file



Box	Definition
ftyp	File Type
moov	Media Data (container of media data)
mdat	Movie (container of metadata)
mvhd	Movie Header (info about movie)
trak	Track (container of metadata for 1 stream)
...	...

Figure 33: Boxes in an MP4 File

MP4 FF can store:

1. Structural and media data information for timed presentations of media data (e.g. audio, video, subtitles);
2. Un-timed data (e.g. meta-data);
3. Elementary stream encryption and encryption parameter (CENC);
4. Media for adaptive streaming (e.g. DASH);
5. High Efficiency Image Format (HEIF);
6. Omnidirectional Media Format (OMAF);
7. Files partially received over lossy links for further processing such as playback or repair (Partial File Format);
8. Web resources (e.g. HTML, JavaScript, CSS, …).

The first two features were in the original specification. The last two are still under development. All others were added in the years following 2001 when MP4 FF was approved.

MPEG-7 Systems

With MPEG-7, MPEG made the first big departure from media compression and turned its attention to media description including ways to compress that information. In addition to descriptors for visual audio and multimedia information, MPEG-7 includes a Systems layer used by an application, say, to navigate a multimedia information repository, to access coded information coming from a delivery layer in the form of coded descriptors (in XML or in BiM, MPEG's XML compression technology).

Figure 34 illustrates the operation of an MPEG-7 Systems decoder where it is shown that an MPEG-7 Systems decoder operates in two phases:

1. Initialisation when DecoderInit initialises the decoder by conveying description format information (textual or binary), a list of URIs that identifies schemas, parameters to configure the Fragment Update decoder, and an initial description. The list of URIs is passed to a schema resolver that associates the URIs with schemas to be passed to Fragment Update Decoder.

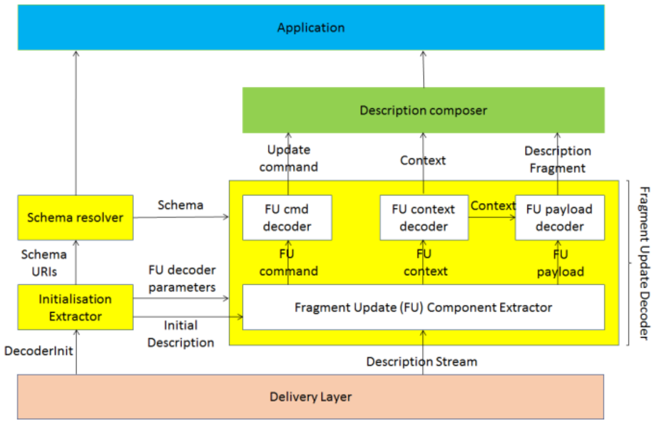


Figure 34: MPEG-7 Systems

2. Main operation, when the Description Stream (composed of Access Units containing fragment updates) is fed to the decoder which processes
    a. Fragment Update Command specifying the update type (i.e., add, replace or delete content or a node, or reset the current description tree);
    b. Fragment Update Context that identifies the data type in a given schema document, and points to the location in the current description tree where the fragment update command applies; and
    c. Fragment Update Payload conveying the coded description fragment to be added to, or replaced in the description.

MPEG-E

MPEG Multimedia Middleware (M3W), also called MPEG-E, is an 8-part standard defining the protocol stack of consumer-oriented multimedia devices, as depicted in *Figure 35*.

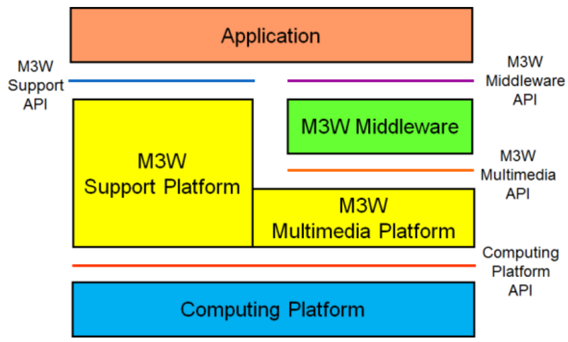


Figure 35: MPEG Multimedia Middleware (M3W)

The M3W model includes 3 layers:

1. *Applications:* non part of the specifications but enabled by the M3W Middleware API;
2. *Middleware:* consisting of
    a. M3W middleware exposing the M3W Middleware API;
    b. Multimedia platform supporting the M3W Middleware by exposing the M3W Multimedia API;
    c. Support platform providing the means to manage the lifetime of, and interaction with, realisation entities by exposing the M3W Support API (it also enables management of support properties, e.g. resource management, fault management and integrity management);
3. *Computing platform*: whose API are outside of M3W scope.

MPEG-M

Multimedia service platform technologies (MPEG-M) is a 5-Part standard that specifies two main components of a multimedia device, called *peer* in MPEG-M.

As shown in *Figure 36*, the first component is API: High-Level API for applications and Low-Level API for network, energy and security.

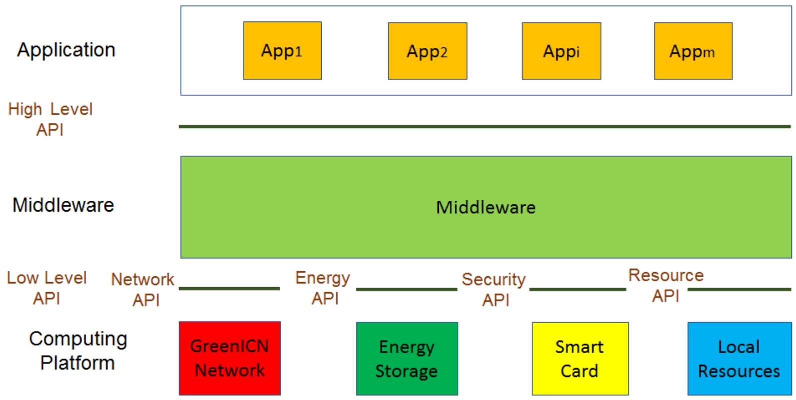


Figure 36: High-Level and Low-Level API

The second components is the middleware called MPEG eXtensible Middleware (MXM) that utilises MPEG multimedia technologies (*Figure 37*).

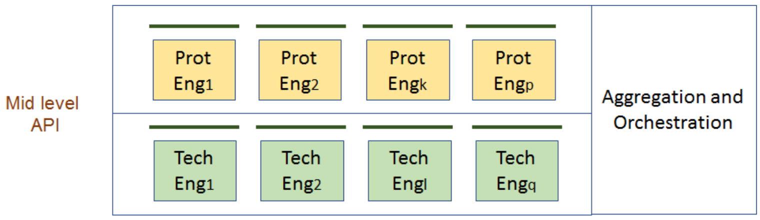


Figure 37: The MXM architecture

The Middleware is composed of two types of engine. *Technology Engines* are used to call locally available functionalities defined by MPEG standards such as creating or interpreting a licence attached to a content item. *Protocol Engines* are used to communicate with other peer, e.g. in case a peer does not have a particular Technology Engine that another peer has. For instance, a peer can use a Protocol Engine to call a licence server to get a licence to attach to a multimedia content item. The MPEG-M middleware has the ability to create chains of Technology Engines (Orchestration) or Protocol Engines (Aggregation).

MMT

MPEG Media Transport (MMT) is part 1 of High efficiency coding and media delivery in heterogeneous environments (MPEG-H). It is the solution for the new world of broadcasting where content can be delivered over different channels each with different characteristics, e.g. one-way (traditional broadcasting) and two-way (the ever more pervasive broadband network). MMT assumes that the Internet Protocol is the transport protocol common to all channels.

Figure 38 depicts the MMT protocol stack

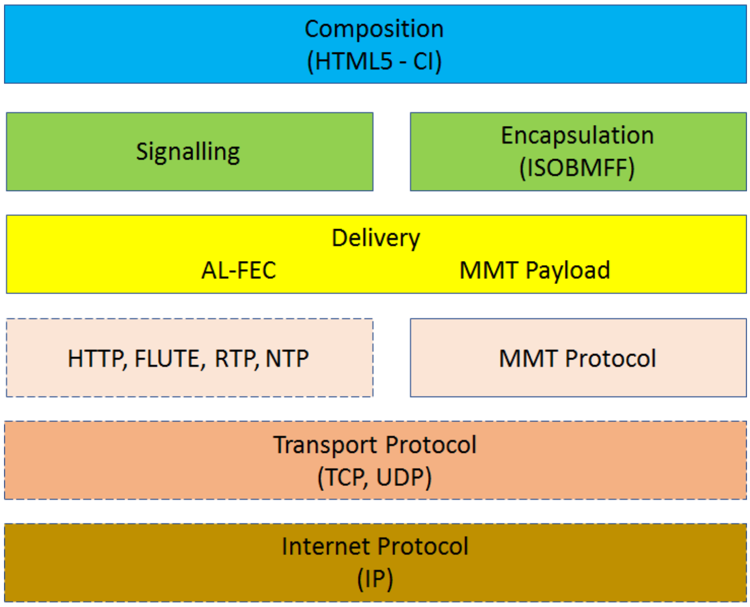


Figure 38: The MMT protocol stack

Figure 39 focuses on the MMT Payload, i.e. on the content structure. The MMT Payload has an onion-like structure. Starting from the inside:

1. Media Fragment Unit (MFU), the atomic unit which can be independently decoded;
2. Media Processing Unit (MPU), the atomic unit for storage and consumption of MMT content (structured according to ISO BMFF), containing one or more MFUs;
3. MMT Asset, the logical unit for elementary streams of multimedia component, e.g. audio, video and data, containing one or more MPU files;
4. MMT Package, a logical unit of multimedia content such as a broadcasting program, containing one or more MMT Assets, also containing
   a. Composition Information (CI), describing the spatio-temporal relationships that exists among MMT Assets;
   b. Delivery Information, describing the network characteristics.

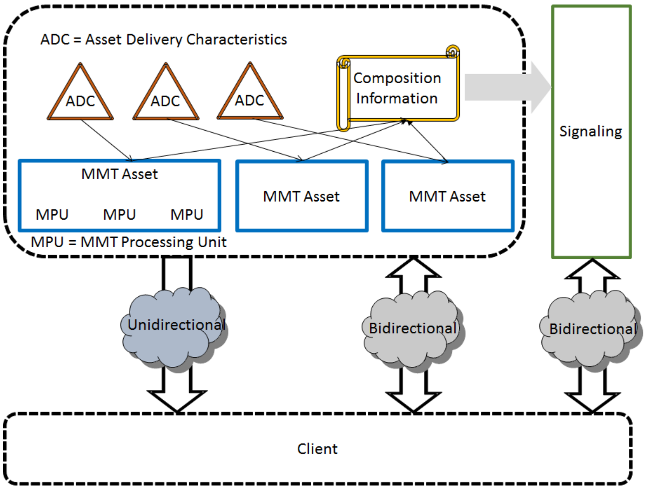


Figure 39: Structure of MMT Payload

MPEG-DASH

Dynamic adaptive streaming over HTTP (DASH) is another MPEG Systems standard that was motivated by the popularity of HTTP streaming and the existence of different protocols used in different streaming platforms, e.g. different manifest and segment formats. By developing the DASH standard for HTTP streaming of multimedia content, MPEG has enabled a standard-based client to stream content from any standard-based server, thereby enabling interoperability between servers and clients of different make.

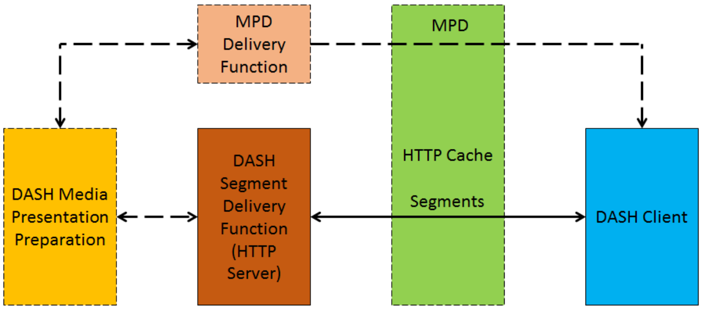


Figure 40 – DASH model

As depicted in *Figure 40*, the multimedia content is stored on an HTTP server in two components: 1) Media Presentation Description (MPD) which describes a manifest of the available content, its various alternatives, their URL addresses and other characteristics, and 2) Segments which contain the actual multimedia bitstreams in form of chunks, in single or multiple files.

A typical operation of the system would follow the steps

1. DASH client obtains the MPD;
2. Parses the MPD;
3. Gets information on several parameters, e.g. program timing, media content availability, media types, resolutions, min/max bandwidths, existence of alternatives of multimedia components, accessibility features and the required protection mechanism, the location of each media component on the network and other content characteristics;
4. Selects the appropriate encoded alternative and starts streaming the content by fetching the segments using HTTP GET requests;
5. Fetches the subsequent segments after appropriate buffering to allow for network throughput variations
6. Monitors the network bandwidth fluctuations;
7. Decides how to adapt to the available bandwidth depending on its measurements by fetching segments of different alternatives (with lower or higher bitrate) to maintain an adequate buffer.

DASH only defines the MPD and the segment formats. MPD delivery and media encoding formats containing the segments as well as client behavior for fetching, adaptation heuristics and content playing are outside of MPEG-DASH's scope.

Conclusions

This is not an exhaustive presentation of MPEG Systems work. Still the description will reveal the amount of work that MPEG has invested in Systems aspects, sometimes per se, and sometimes to provide adequate support to users of its media coding standards. Different standards compete for the palm of the most successful MPEG Systems standards. For sure the MPEG-2 Systems standard is a strong competitor: 9 editions have been produced to keep up with continuous user demands for new functionalities.

Not including the fact that MPEG-2 Systems has received an Emmy Award 😉.

12 Data compression

In its 31 years of activity MPEG has developed standards for compression of audio, video, 3D graphics. What is less known is that MPEG has worked on compression of many other data types. Most of them had an ancillary role to the media data, but DNA reads from high speed sequencing machines are a data type to which compression has been successfully applied.
Some people claim that MPEG is not equipped to deal with compression of other data than media data. This sentence can be true of false depending on which point in time you make this statement. In 1988 MPEG – Moving Picutes – was ill equipped to deal with Audio compression, but then experts came and the epoch-marking MP3 audio compression standard was produced. A priori MPEG did not know how to compress meshes or point clouds, but appropriate experts came in and good standards have been and are produced. MPEG did not know much about how to compress DNA reads, but with appropriate education it could do that.
Compression is not a universal technology, but can be adapted to many different domains.
This chapter takes the view of applying compression to a variety of data types:

1. Meaningful data can be compressed analyses of data types that MPEG has handled in all its standards
2. Moving intelligence around identifies "artificial neural networks" as a data type of growing importance and size that benefits from compression;
3. MPEG standards for genomics describes how a variety of MPEG compression tools have been successfully applied to the compression of DNA reads from high-speed sequencing machines
4. Compression of other data raises again the issue of "data compression" being a field of endeavour

12.1 Meaningful data can be compressed

Introduction

Obviously video is a high-profile topic to MPEG people – MP stands for Moving Pictures (not Motion Pictures, which is another story) and Audio is another high-profile topic. This should not be a surprise given that the official MPEG title is “Coding of Moving Pictures and Audio”.
What is less known, but potentially very important, is the fact that MPEG has already developed a few standards for compression of a wide range of other data types. Point Cloud is the data type that is acquiring a higher profile by the day, but there are many more types, as represented by the table below where standards are represented by the MPEG acronyms and the numbers in the respective columns are the part numbers.

Table 11 – Data types and relevant MPEG standards

	Standards												
Data types	**1**	**2**	**4**	**5**	**7**	**B**	**C**	**D**	**G**	**H**	**I**	**IoMT**	**V**
Video	2	2	2-9-10-29-31-33	1-2	3-13-15		4			2	3-7		
Audio	3	3-7	3		4			1-2-3-4		3	4		
2D/3D-Meshes		2	16				5						
Face/Body Animation		2											
Scene graph			11										
Font			18										
Multimedia					5								
Neural Networks					17								
XML						1							
Genome									2-6				
Point Clouds											5-9		
Sensors/Actuators												3	2-3-4-5-6

Video

The chapters Forty years of video coding, More video features and Immersive visual experiences provide a detailed history of video compression in MPEG from two different perspectives. Here the video-coding related standards produced or being produced are listed.

Table 12 – MPEG Video-related standards

Standards	Part	Description
MPEG-1	Part2	Widely used standard
MPEG-2	Part2	Widely used standard
MPEG-4	Part2	Visual used for video on internet and for movie compression
	Part9	Reference Hardware Description, a standard that supports a reference hardware description of the standard expressed in VHDL (VLSI Hardware Description Language), a hardware description language used in electronic design automation.
	Part10	Advanced Video Coding is the ubiquitous video coding standard
	Part29	Web Video Coding was planned to be Option 1, not achieved
	Part31	Video Coding for Browsers, was planned to be Option 1, discontinued
	Part33	Internet Video coding, was planned to be Option 1, not achieved because there are 3 Option 2 patent declarations. When affected technologies will be disclosed MPEG will remove them from the standard
MPEG-5	Part1	Essential Video Coding will have a base layer/profile which is expected to be Option 1 and a second layer/profile with a performance ~25% better than HEVC. Licensing terms are expected to be published by patent holders within 2 years
	Part2	Low Complexity Enhancement Video Coding (LCEVC) will be a two-layer video coding standard. The lower layer is not tied to any specific technology and can be any video codec; the higher layer is used to extend the capability of an existing video codec
MPEG-7	is about Multimedia Content Description. There are different tools to describe visual information	
	Part3	Visual is a form of compression as it provides tools to describe Color, Texture, Shape, Motion, Localisation, Face Identity, Image signature and Video signature
	Part13	Compact Descriptors for Visual Search can be used to compute compressed visual descriptors of an image. An application is to get further information about an image captured e.g. with a mobile phone
	Part15	Compact Descriptors for Video Analysis allows to manage and organise large scale data bases of video content, e.g. to find content containing a specific object instance or location
MPEG-C	is a collection of video technology standard that do not fit with other standards	
	Part4	Media Tool Library is a collection of video coding tools (called Functional Units) that can be assembled using the technology standardised in MPEG-B Part 4 Codec Configuration Representation
MPEG-H	Part2	High Efficiency Video Coding is the latest MPEG video coding standard with an improved compression of 60% compared to AVC
MPEG-I	is the new standard, mostly under development, for immersive technologies	

	Part 3	Versatile Video Coding is the ongoing project to develop a video compression standard with an expected 50% more compression than HEVC
	Part7	MPEG-I part 7 Immersive Media Metadata is the current project to develop a standard for compressed Omnidirectional Video that allows limited translational movements of the head.
Explorations	Expl1	6 Degrees of Freedom (6DoF)
	Expl2	Light field

Audio

The chapter Audio compression in MPEG provides a detailed history of audio compression in MPEG. Here the audio-coding related standards produced or being produced are listed.

Table 13 – MPEG Audio-related standards

Standards	Part	Description
MPEG-1	Part3	Audio produced, among others, the foundational digital audio standard better known as MP3
MPEG-2	Part3	Audio extended the stereo user experience of MPEG-1 to Multichannel
	Part7	Advanced Audio Coding is the foundational standard on which MPEG-4 AAC is based
MPEG-4	Part3	Advanced Audio Coding (AAC) currently used by some 10 billion devices and software applications growing by half a billion unit every year
MPEG-7	is about Multimedia Content Description.	
	Part4	There are different tools to describe audio information
MPEG-D	is a collection of different audio technologies	
	Part1	MPEG Surround provides an efficient bridge between stereo and multi-channel presentations in low-bitrate applications as it can transmit 5.1 channel audio within the same 48 kbit/s transmission budget
	Part2	Spatial Audio Object Coding (SAOC) allows very efficient coding of a multi-channel signal that is a mix of objects (e.g. individual musical instruments)
	Part3	Unified Speech and Audio Coding (USAC) combines the tools for speech coding and audio coding into one algorithm with a performance that is equal or better than AAC at all bit rates. USAC can code multichannel audio signals and can also optimally encode speech content
	Part4	Dynamic Range Control is a post-processor for any type of MPEG audio coding technology. It can modify the dynamic range of the decoded signal as it is being played

2D/3D Meshes

Polygons meshes can be used to represent the approximate shape of a 2D image or a 3D object. 3D mesh models are used in various multimedia applications such as computer game, animation, and simulation applications. MPEG-4 provides various compression technologies

Table 14 – MPEG 2D/3D-related standards

Standards	Part	Description
MPEG-4	Part2	Visual provides a standard for 2D and 3D Mesh Compression (3DMC) of generic, but static, 3D objects represented by first-order (i.e., polygonal) approximations of their surfaces. 3DMC has the following characteristics: 1. Compression: Near-lossless to lossy compression of 3D models 2. Incremental rendering: No need to wait for the entire file to download to start rendering 3. Error resilience: 3DMC has a built-in error-resilience capability 4. Progressive transmission: Depending on the viewing distance, a reduced accuracy may be sufficient
	Part16	Animation Framework eXtension (AFX) provides a set of compression tools for Shape, Appearance and Animation

Face/Body Animation

Imagine you have a face model that you want to animate from remote. How do you represent the information that animates the model in a bit-thrifty way? MPEG-4 Part 2 Visual has an answer to this question with its Facial Animation Parameters (FAP) defined at two levels:

- High level
  - Viseme (visual equivalent of phoneme)
  - Expression (joy, anger, fear, disgust, sadness, surprise)
- Low level: 66 FAPs associated with the displacement or rotation of the facial feature points.

In *Figure 41* the feature points affected by FAPs are indicated as a black dot. Other feature points are indicated as a small circle.

It is possible to animate a default face model in the receiver with a stream of FAPs or a custom face can be initialised by downloading Face Definition Parameters (FDP) with specific background images, facial textures and head geometry.

MPEG-4 Part 2 uses a similar approach for Body Animation.

Scene Graphs

So far MPEG has never developed a Scene Description technology. In 1996, when the development of the MPEG-4 standard required it, it took the Virtual Reality Modelling Language (VRML) and extended it to support MPEG-specific functionalities. Of course, compression could not be absent from the list. So the Binary Format for Scenes (BiFS), specified in MPEG-4 Part 11 Scene description and application engine was born to allow for efficient representation of dynamic and interactive presentations, comprising 2D & 3D graphics, images, text and audiovisual material.

The representation of such a presentation includes the description of the spatial and temporal organisation of the different scene components as well as user-interaction and animations.

In MPEG-I scene description is playing again an important role. However, MPEG this time does not even intend to pick a scene description technology. It will define instead some interface to a scene description parametres.

Fonts

Many thousands of fonts are available today for use as components of multimedia content. They often utilise custom design fonts that may not be available on a remote terminal. In order to insure faithful appearance and layout of content, the font data have to be embedded with the text objects as part of the multimedia presentation.

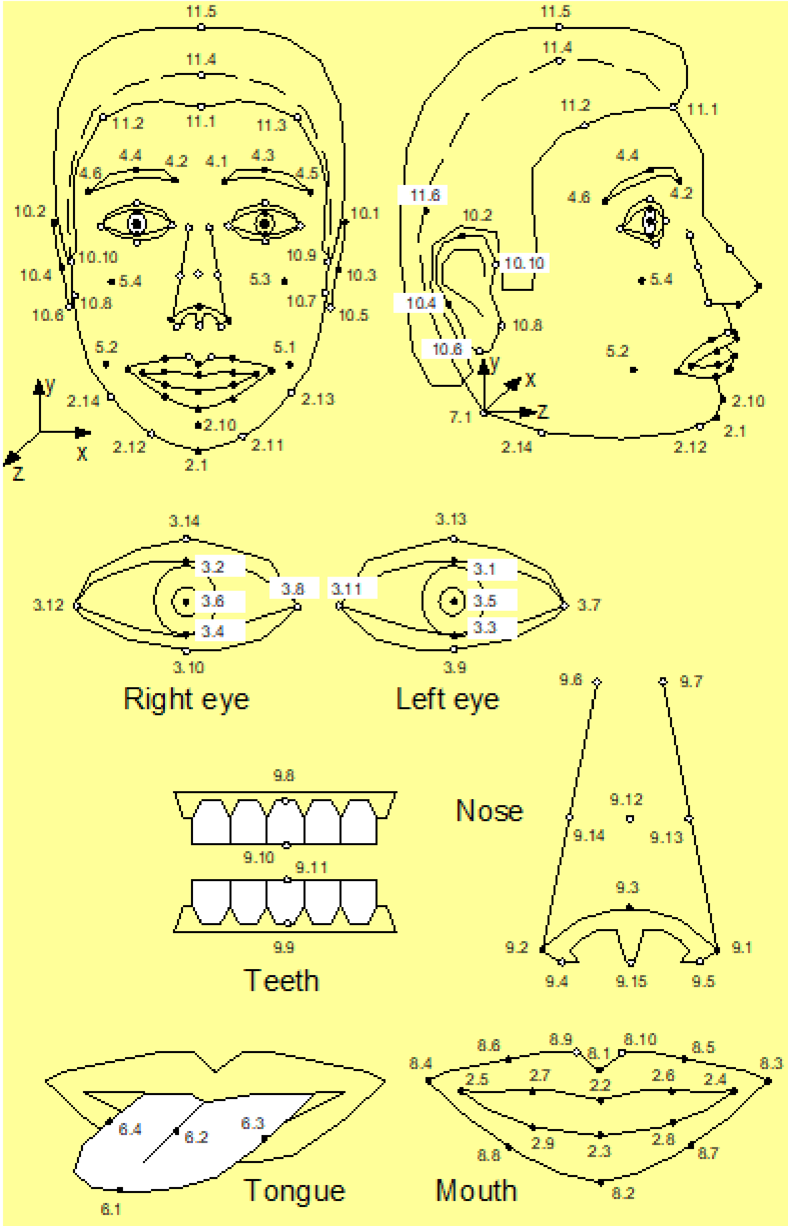


Figure 41: Facial Animation Parameters

MPEG-4 part 18 Font Compression and Streaming defines and provides two main technologies:

- OpenType and TrueType font formats
- Font data transport mechanism – the extensible font stream format, signalling and identification

Multimedia

Multimedia is a combination of multiple media in some form. Probably the closest multimedia "thing" in MPEG is the standard called Multimedia Application Formats. However, MPEG-A is an integrated package of media for specific applications and does not define any specific media format. It only specifies how you can combine MPEG (and sometimes other) formats.

MPEG-7 part 5 Multimedia Description Schemes (MDS) specifies the different description tools that are not visual and audio, i.e. generic and multimedia. By comprising a large number of MPEG-7 description tools from the basic audio and visual structures MDS enables the creation of the structure of the description, the description of collections and user preferences, and the hooks for adding the audio and visual description tools. This is depicted in *Figure 42*.

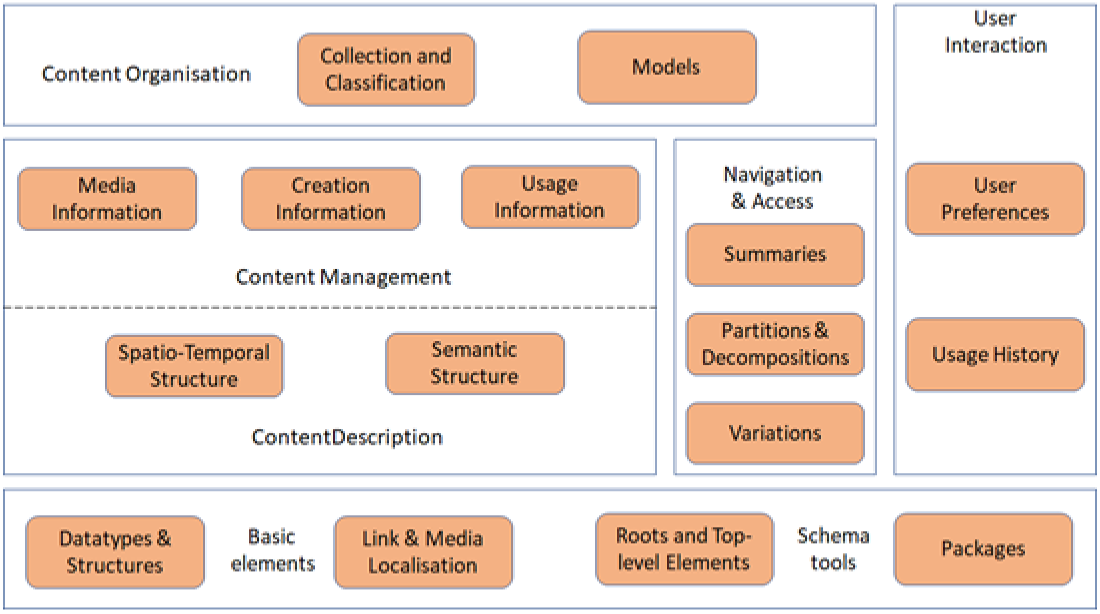


Figure 42: The different functional groups of MDS description tools

Neural Networks
Requirements for neural network compression have been exposed in Moving intelligence around. After 18 months of intense preparation with development of requirements, identification of test material, definition of test methodology and drafting of a Call for Proposals(CfP), at the March 2019 (126th) meeting, MPEG analysed nine technologies submitted by industry leaders. The technologies proposed compress neural network parameters to reduce their size for transmission, while not or only moderately reducing their performance in specific multimedia applications. MPEG-7 Part 17 Neural Network Compression for Multimedia Description and Analysis are the standard name, the part number and the title given to the new standard.

XML
MPEG-B part 1 Binary MPEG Format for XML (BiM) is the current endpoint of an activity that started some 20 years ago when MPEG-7 Descriptors defined by XML schemas were compressed in a standard fashion by MPEG-7 Part 1 Systems. Subsequently MPEG-21 needed XML compression and the technology was extended in MPEG-21 Part 15 Binary Format.
In order to reach high compression efficiency, BiM relies on schema knowledge between encoder and decoder. It also provides fragmentation mechanisms to provide transmission and processing flexibility, and defines means to compile and transmit schema knowledge information to enable decompression of XML documents without a priori schema knowledge at the receiving end.

Point Clouds
3D point clouds can be captured with multiple cameras and depth sensors with points that can number a few thousands up to a few billions, and with attributes such as colour, material properties etc.
MPEG is developing two different standards whose choice depends on whether the point cloud is dense (this is done in MPEG-I Part 5 Video-based Point Cloud Compression) or less so (MPEG-I Part 9 Geometry-based PCC). The algorithms in both standards are lossy, scalable, progressive and support random access to subsets of the point cloud.

MPEG plans to release Video-based PCC as FDIS in January 2020 and Geometry-based PCC Point Cloud Compression as FDIS in April 2020.

Sensors/Actuators

In the middle of the first decade of 2000's, MPEG recognised the need to address compression of data from sensor and data to actuators when it considered the exchange of information taking place between the physical world where the user is located and any sort of virtual world generated by MPEG media.

Therefore, MPEG undertook the task to provide standard interactivity technologies that allow a user to

- Map their real-world sensor and actuator context to a virtual-world sensor and actuator context, and vice-versa, and
- Achieve communication between virtual worlds.

Figure 43 describes the context of the MPEG-V Media context and control standard.

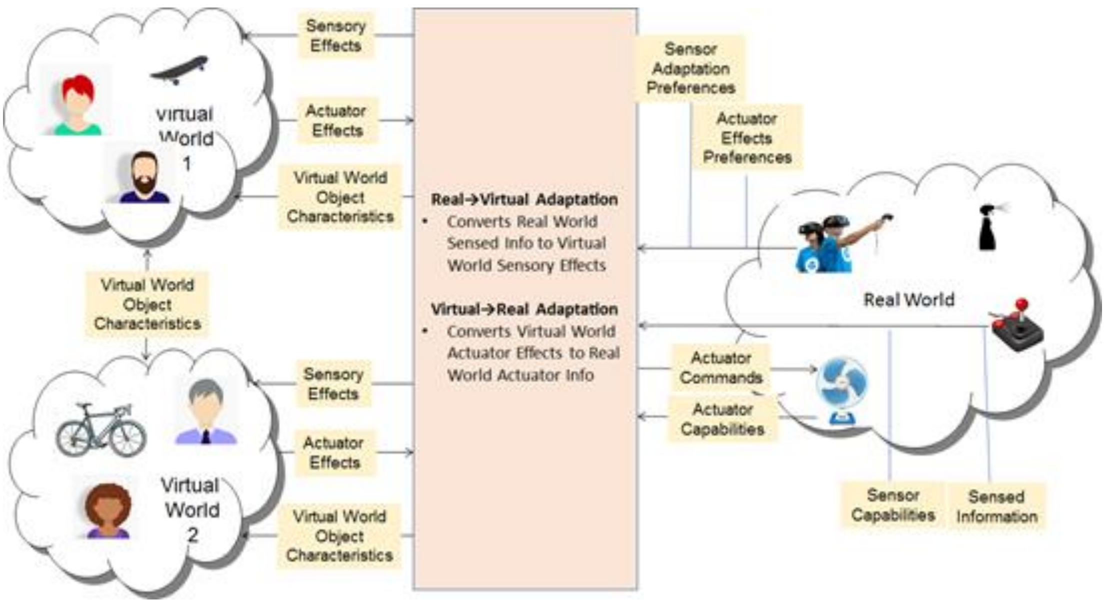


Figure 43: Communication between real and virtual worlds

Standard	Part	Description
MPEG-V	Part2	Control information specifies control devices interoperability (actuators and sensors) in real and virtual worlds
	Part3	Sensory information specifies the XML Schema-based Sensory Effect Description Language to describe actuator commands such as light, wind, fog, vibration, etc. that trigger human senses
	Part4	Virtual world object characteristics defines a base type of attributes and characteristics of the virtual world objects shared by avatars and generic virtual objects
	Part5	Data formats for interaction devices specifies syntax and semantics of data formats for interaction devices – Actuator Commands and Sensed Information – required to achieve interoperability in controlling interaction devices (actuators) and in sensing information from interaction devices (sensors) in real and virtual worlds

	Part6	Common types and tools specifies syntax and semantics of data types and tools used across MPEG-V parts

MPEG-IoMT Internet of Media Things is the mapping of the general IoT context to MPEG media developed by MPEG. MPEG-IoMT Part 3 – IoMT Media Data Formats and API also addresses the issue of media-based sensors and actuators data compression.

Genome

The chapter MPEG standards for genomics presents the technology used in MPEG-G Genomic Information Representation. Many established compression technologies developed for compression of other MPEG media have found good use in genome compression. MPEG is currently busy developing the MPEG-G reference software and is investigating other genomic areas where compression is needed. More concretely MPEG plans to issue a Call for Proposal for Compression of Genome Annotation at its July 2019 (128th) meeting.

12.2 Moving intelligence around

Introduction

Artificial intelligence has reached the attention of mass media and technologies supporting it – Neural Networks (NN) – are being deployed in several contexts affecting a growing number of end users, e.g. in their smart phones.

If a NN is used locally, it is possible to use existing digital representation of NNs (e.g., NNEF, ONNX). However, these formats miss vital features for distributing intelligence, such as compression, scalability and incremental updates.

To appreciate the need for compression let's consider the case of adjusting the automatic mode of a camera based on recognition of scene/object obtained by using a properly trained NN. As this area is being intensely investigated, very soon there will be new better trained versions of the NN or new NNs with additional features. However, as the process to create the necessary "intelligence" usually takes time and labour (skilled and unskilled), in most cases the new created intelligence must be moved from the centre to where the user handset is. With today's NNs reaching a size of several hundred Mbytes and growing, a scenario where millions of users clog the network because they are all downloading the latest NN with great new features looks likely.

This article describes some elements of the MPEG work plan to develop one or more standards that enable compression of neural networks. Those wishing to know more please read Use cases and Requirements, and Call for Proposals.

About Neural Networks

An (artificial) Neural Network is a system composed of connected nodes each of which can

- Receive input signals from other nodes
- Process them
- Transmit an output signal to other nodes.

Nodes are typically aggregated into layers, each performing different functions. Typically, the "first layers" are rather specific of the signals (audio, video, various forms of text information etc.). MPEG is addressing the compression of NNs trained with multimedia data for classification or analysis purposes.

Nodes can send signals to subsequent layers but, depending on the type of network, also to the preceding layers.

Training is the process of "teaching" a network to do a particular job, e.g. recognising a particular object or a particular word. This is done by presenting to the NN data from which it can "learn". Inference is the process of presenting to a trained network new data to get a response about what the new data is.

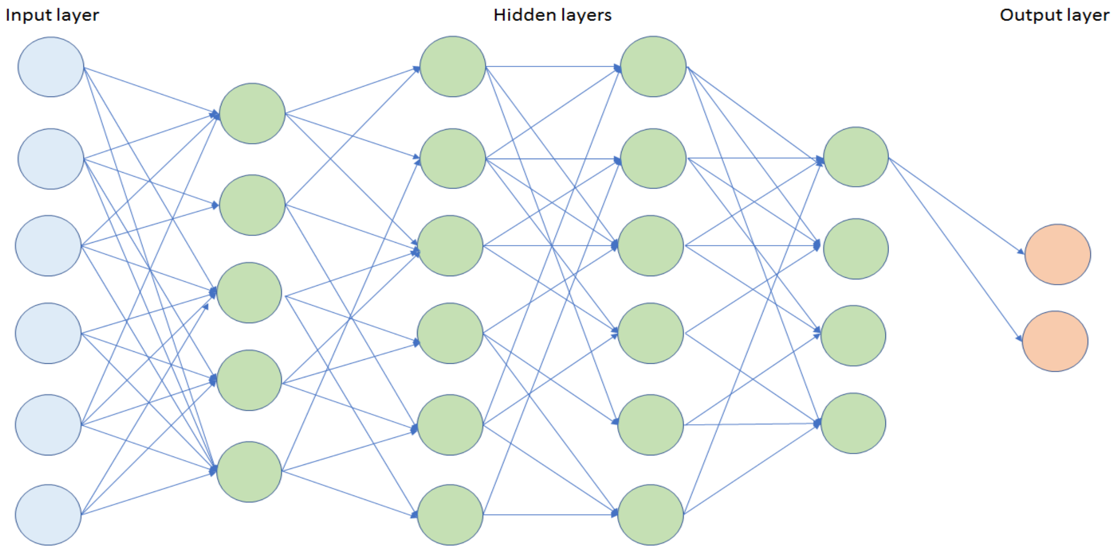


Figure 44: An example of artificial neural network

When is NN compression useful?
Compression is useful whenever there is a need to distribute NNs to remotely located devices. Depending on the specific use case, compression should be accompanied by other features. In the following two major use cases will be analysed.

Public surveillance
In 2009 MPEG developed the Surveillance Application Format. This is a standard that specifies the package (file format) containing audio, video and metadata to be transmitted to a surveillance centre. Today, however, it is possible to ask the surveillance network to do more intelligent things by distributing intelligence even down to the level of visual and audio sensors.
For this more advanced scenarios MPEG is developing a suite of specifications under the title of Internet of Media Things (IoMT) where Media Things (MThing) are the media "versions" of IoT's Things. Parts 2 (IoMT Discovery and Communication API) and 3 (IoMT Media Data Formats and API) of the IoMT standard (ISO/IEC 23093) has reached FDIS level in March 2019. Part 1 (Architecture) will reach FDIS level in October 2019.
The IoMT reference model is represented in *Figure 45*

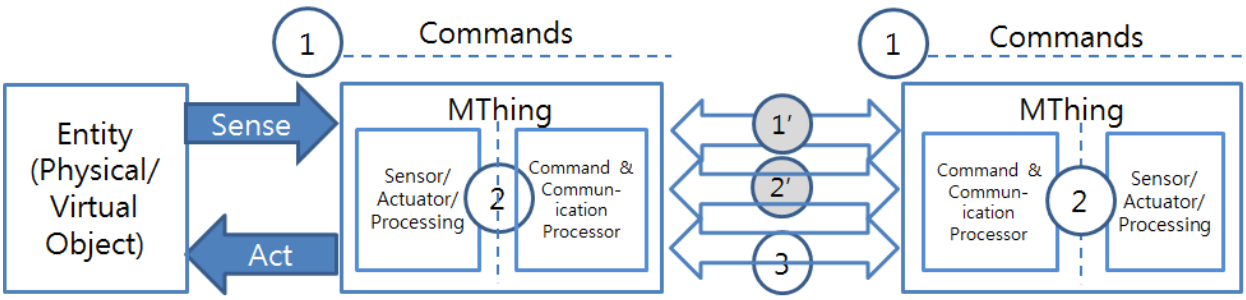


Figure 45: IoT in MPEG stands for "media" – IoMT

IoMT standardises the following interfaces:

1	User commands (setup info.) between a system manager and an MThing
1’	User commands forwarded by an MThing to another MThing, possibly in a modified form (e.g., subset of 1)
2	Sensed data (Raw or processed data in the form of just compressed data or resulting from a semantic extraction) and actuation information
2’	Wrapped interface 2 (e.g. for transmission)
3	MThing characteristics, discovery

IoMT is neutral as to the type of semantic extraction or, more generally, to nature of intelligence actually present in the cameras. However, as NNs networks are demonstrating better and better results for visual pattern recognition, such as object detection, object tracking and action recognition, cameras can be equipped with NNs capable to process the information captured to achieve a level of understanding and transmit that understanding through interface 2.

Therefore, one can imagine that re-trained or brand new NNs can be regularly uploaded to a server that distributes NNs to surveillance cameras. Distribution need not be uniform since different areas may need different NNs, depending on the tasks that given areas need to specifically carry out.

NN compression is a vitally important technology to make the described scenarios real because automatic surveillance system may use many cameras (e.g. thousands and even million units) and because, as the technology to create NNs matures, the time between NN updates will progressively become shorter.

Distribution of NN-based apps to devices

There are many cases where compression is useful to efficiently distribute heavy NN-based apps to a large number of devices, in particular mobile. Here 3 case are considered.

- Visual apps. Updating a NN-based camera app in one’s mobile handset will soon become common place. Ditto for the many conceivable application where the smart phone understands some of the objects in the world around. Both will happen at an accelerated frequency.
- Machine translation (speech-to-text, translation, text-to-speech). NN-based translation apps already exist and their number, efficiency, and language support can only increase.
- Adaptive streaming. As AI-based methods can improve the QoE, the coded representation of NNs can initially be made available to clients prior to streaming while updates can be made during streaming to enable better adaptation decisions, i.e. better QoE.

Requirements

MPEG has identified a number of requirements for compressing NNs. Even though not all applications need the support of all requirements, the NN compression algorithm must eventually be able to support all the identified requirements.

1. Compression shall have a lossless mode, i.e. the performance of the compressed NN is exactly the same as the uncompressed NN
2. Compression shall have a lossy mode, i.e. the performance of the decompressed NN can be different than the performance of the uncompressed NN of course in exchange for more compression
3. Compression shall be scalable, i.e. even if only a subset of the compressed NN is used, there is still a level of performance
4. Compression shall support incremental updates, i.e. as more data are received the performance of NN improves
5. Decompression shall be possible with limited resources, i.e. with limited processing performance and memory
6. Compression shall be error resilient, i.e. if an error occurs during transmission, the file is not lost

7. Compression shall be robust to interference, i.e. it is possible to detect that the compressed NN has been tampered with
8. Compression shall be possible even if there is no access to the original training data
9. Inference shall be possible using compressed NN
10. Compression shall support incremental updates from multiple providers to improve performance of a NN

Conclusions

The standard Compression of neural networks for multimedia content description and analysis (part 17 of MPEG-7) that MPEG is currently developing will initially produce a base layer standard that will help the industry move its first steps in this exciting field that will certainly shape the way intelligence is added to things near to all of us.

12.3 MPEG standards for genomics

Introduction

The well-known double helix carries the DNA of living beings. The DNA of humans contains about 3.2 billion nucleotide base pairs represented by the quaternary symbols (A, G, C, T). Today, with high-speed sequencing machines it is possible to "read" the DNA. The resulting file contains millions of “reads”, short segments of symbols, typically all of the same length. The size of the file is an unwieldy few Terabytes.

The MPEG-G standard parts 1, 2 and 3, developed jointly by MPEG and ISO TC 276 Biotechnology, will reduce the size of the file, without loss of information, by exploiting the inherent redundancy of the reads and make at the same time the information in the file more easily accessible.

This chapter provides some context and explains the basic ideas of the standard and the benefits it can yield to those who need to access genomic information.

Reading the DNA

There are two main obstacles that prevent a direct use of files from sequencing machines: the position of a read on the DNA sample is unknown and the value of each symbol of the read is not entirely reliable.

Figure 46 represents a case where there are 17 reads each with a read length of 15 nucleotides. These have already been aligned to a reference genome (first line). Reads with a higher number start further down in the reference genome.

Ref Gen	G	C	T	A	T	C	A	G	G	C	T	A	G	G	T	T	A	C	A	G	T	G	C	A	T	G	C
Read #1	G	C	T	A	T	C	A	G	G	C	T	A	G	A	T												
Read #2		C	T	A	T	C	A	G	G	C	T	A	G	A	T	T											
Read #3		C	T	A	A	C	A	G	G	C	T	A	G	A	T	T											
Read #4			T	A	T	C	A	G	G	C	T	A	T	A	T	T	A										
Read #5				A	T	C	A	G	G	C	T	A	G	A	T	T	A	C									
Read #6				A	T	C	A	G	G	C	T	A	G	G	T	T	A	C									
Read #7					T	C	A	G	G	C	T	A	G	A	T	T	A	C	A								
Read #8					T	C	A	G	G	C	T	A	G	A	T	T	A	C	A								
Read #9						C	A	G	G	C	T	A	G	A	T	T	A	C	A	G							
Read #10							A	G	G	C	T	A	G	A	T	T	A	C	A	G	T						
Read #11								G	G	C	T	A	G	A	T	T	A	C	A	G	T	G					
Read #12									G	C	T	A	G	A	T	T	A	C	A	G	T	G	C				
Read #13										C	T	A	G	A	T	T	A	C	A	G	T	G	C	A			
Read #14										C	T	A	G	G	T	T	A	C	A	G	T	G	C	A			
Read #15											T	A	G	A	T	T	A	C	A	G	T	G	C	A	T		
Read #16												A	G	A	T	T	A	C	A	G	T	G	C	A	T	G	
Read #17													G	A	T	T	A	C	A	G	T	G	C	A	T	G	C

Figure 46 – A group of reads aligned to a reference genome

Reading column-wise, we see that in most cases the values have exactly the value of the reference genome. A single difference (represented by isolated red symbols) may be caused by read errors while a quasi-completely different column (most symbols in red) may be caused by the fact that 1) a given DNA is unlikely to be exactly equal to a reference genome or 2) the person with this particular DNA may have health problems.

Use of genomics today

Genomics is already used in the clinical practice. An example of genomic workflow is depicted in *Figure 46* which could very well represent a blood test workflow if “DNA” were replaced by “blood”. Patients go to a hospital where a sample of their DNA is taken and read by a sequencing machine. The files are analysed by experts who produce reports which are read and analysed by doctors who decide actions.

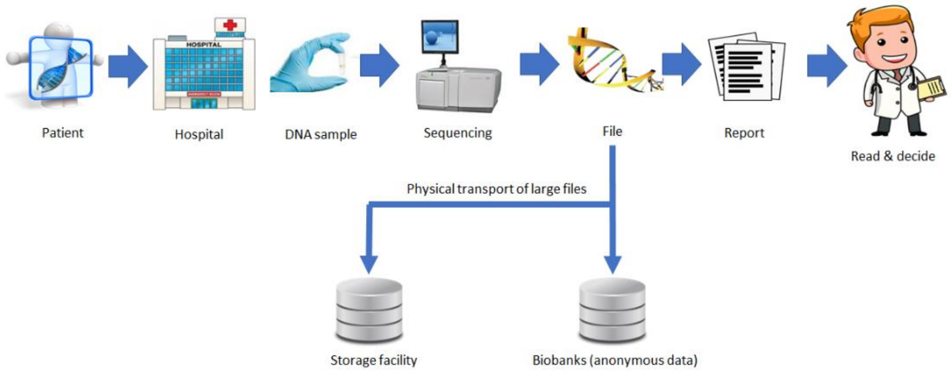


Figure 47 – A typical genome workflow today

Use of genomics tomorrow

Today, genomic workflows take time – even months – and may be costly – thousands of USD per DNA sample. While there is not much room to cut the time it takes to obtain a DNA sample, sequencing cost has been decreasing and are expected to continue doing so.

Big savings could be achieved by acting on data transport and processing. If the size of a 3 Terabytes file is reduced by, say, a factor of 100, the transport of the resulting 30 Gigabytes would be compatible with today’s internet access speeds of 1 Gbit/s (~4 min). Faster data access, a by-product of compression, would allow doctors to get the information they are searching, locally or from remote, in a fraction of a second.

The new possible scenario is depicted in *Figure 48*.

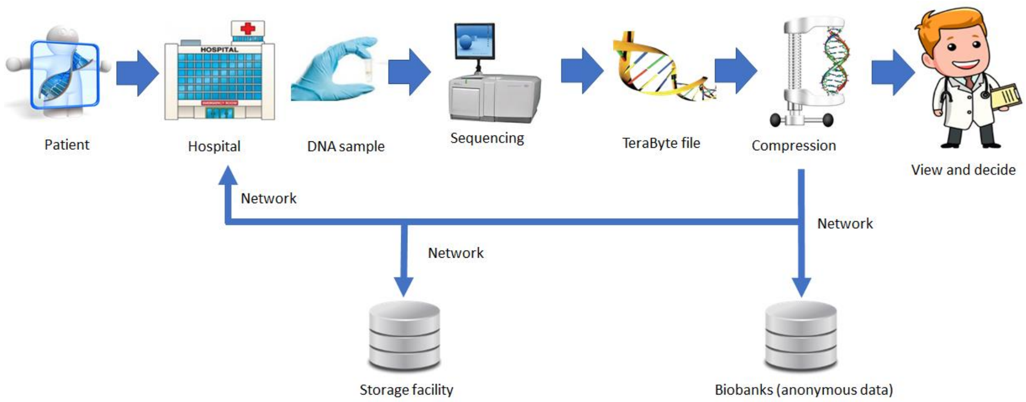


Figure 48 – How genomic workflows can change with compression

MPEG makes genome compression real

Not much had been done to make the scenario above real (zip is the oft-used compression technology today) until the time (April 2013) MPEG received a proposal to develop a standard to losslessly compress files from DNA sequencing machines.

The MPEG-G standard – titled *Genomic Information Representation* – has 5-parts: Parts 1, 2 and 3 have been approved as FDIS and the other parts are expected to follow suit shortly after.

MPEG-G is an excellent example of how MPEG could apply its expertise to a different field than media. Part 1, an adaptation of the MP4 File Format present in all smartphones/tablets/PCs, specifies how to make and transport compressed files. Part 2 specifies how to compress reads and Part 3 how to invoke the APIs to access specific compressed portions of a file. Part 4 and 5 are Conformance and Reference Software, respectively.

Figure 49 depicts the very sophisticated operation specified in Part 2 in a simplified way.

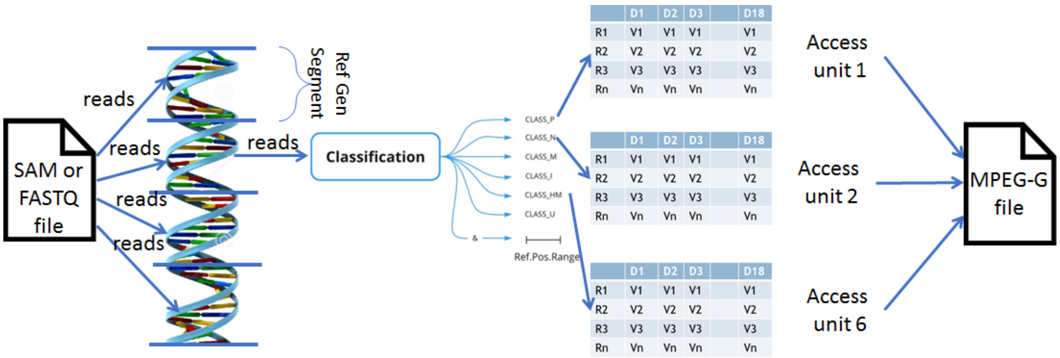


Figure 49 – The MPEG compression

An MPEG-G file can be created with the following sequence of operations:

1. Put the reads in the input file (aligned or unaligned) in bins corresponding to segments of the reference genome
2. Classify the reads in each bin in 6 classes: P (perfect match with the reference genome), M (reads with variants), etc.
3. Convert the reads of each bin to a subset of 18 descriptors specific of the class: e.g., a class P descriptor is the start position of the read etc.
4. Put the descriptors in the columns of a matrix
5. Compress each descriptor column (MPEG-G uses the very efficient CABAC compressor already present in several video coding standards)
6. Put compressed descriptors of a class of a bin in an Access Unit (AU) for a maximum of 6 AUs per bin

Therefore, an MPEG-G file contains all AUs of all bins corresponding to all segments of the reference genome. A file may contain the compressed reads of more than one DNA sample.

The benefits of MPEG-G

Compression is beneficial but is not necessarily the only or primary benefit. More important is the fact that while designing compression, MPEG has given a structure to the information. In MPEG-G the structure is provided by Part 1 (File and transport) and by Part 2 (Compression).

In MPEG-G most information relevant to applications is immediately accessible, locally and, more importantly, also from remote without the need to download the entire file to be able to access the information of interest. Part 3 (Application Programming Interfaces) makes this fast access even more convenient because it facilitates the work of developers of genomics applications who may not have in-depth information of the – certainly complex – MPEG-G standard.

Conclusions

In the best MPEG tradition, MPEG-G is a generic standard, i.e. a standard that can be employed in a wide variety of applications that require small footprint of and fast access to genomic information.

A certainly incomplete list includes: Assistance to medical doctors' decisions; Lifetime Genetic Testing; Personal DNA mapping on demand; Personal design of pharmaceuticals; Analysis of immune repertoire; Characterisation of micro-organisms living in the human host; Mapping of micro-organisms in the environment (e.g. biodiversity).

Standards are living beings, but MPEG standards have a DNA that allows them to grow and evolve to cope with the manifold needs of its ever-growing number of users.

12.4 Compression of other data

In Compression standards for the data industries I reported the proposal made by the Italian ISO member body to establish a Technical Committee on Data Compression Technologies. The proposal was rejected on the ground that Data Compression is part of Information Technology.

It was a big mistake because it has stopped the coordinated development of standards that would have fostered the move of different industries to the digital world. The article identified a few such as Automotive, Industry Automation, Geographic information and more.

MPEG has done some exploratory work and found that there quite a few of its existing standards could be extended to serve new application areas. One example is the conversion of MPEG-21 Contracts to Smart Contracts. An area of potential interest is data generated by machine tools in industry automation.

13 The MPEG standards

So far MPEG has developed, is completing or is planning to develop 22 standards for a total of 201 specifications. For those not in MPEG, and even for some active in MPEG, there is natural question: what is the purpose of all these standards? Assuming that the answer to this question is given, a second one pops up: is there a logic in all these MPEG standards?

Depending on the amount of understanding of the MPEG phenomenon, you can receive different answers ranging from

"There is no logic. MPEG started its first standard with a vision of giving the telco and CE industries a single format. Later it exploited the opportunities that that its growing expertise allowed."

to

"There is a logic. The driver of MPEG work was to extend its vision to more industries leveraging its assets while covering more functionalities."

The reader is encouraged to decide where to place their decision on this continuum of possibilities after reading this chapter.

Figure 50 presents a view of all MPEG standards, completed, under development or planned. Yellow indicates that the standard has been dormant for some time and light brown indicates that the standard is still active.

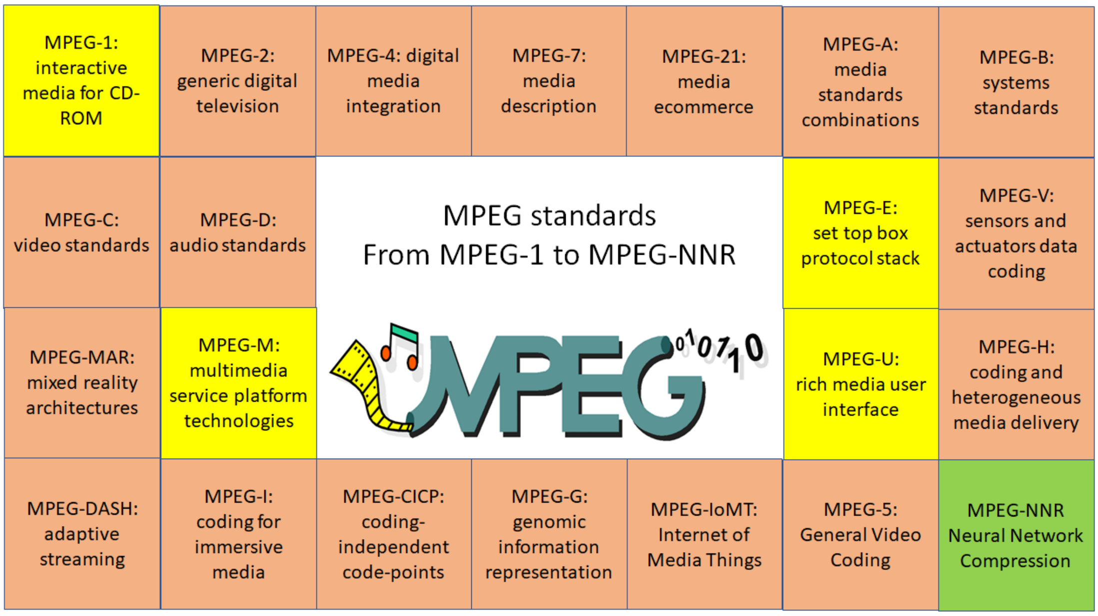


Figure 50: The 22 MPEG standards

13.1 MPEG-1

The goal of MPEG-1 was to leverage the manufacturing power of the Consumer Electronics (CE) industry to develop the basic audio and video compression technology for an application that was considered particularly attractive when MPEG was established (1988), namely interactive audio and video on CD-ROM. This was the logic of the telco industry who thought that their future would be "real time audio-visual communication" but did not have a friendly industry to ask to develop the terminal equipment.

The bitrate of 1.5 Mbit/s mentioned in the official title of MPEG-1 *Coding of moving pictures and associated audio at up to about 1,5 Mbit/s* was an excellent common point for the telecom industry with their ADSL technology whose first generation targeted that bitrate and for the CE industry whose Compact Disc had a throughput of 1.44 Mbit/s (1.2 for the CD-ROM). With that bitrate, compression technology of the late 1980's could only deal with a rather low, but still acceptable resolution (1/2 the horizontal and 1/2 the vertical resolution obtained by subsampling every other field, so that the input video is progressive), Considering that audio had to be impeccable (that is what humans want), at least 200 kbit/s had to be assigned to audio.

Figure 51 depicts the model of an MPEG-1 decoder

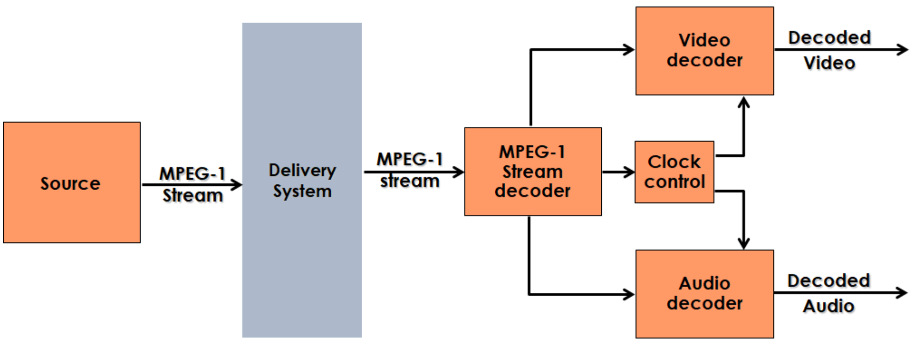


Figure 51 – Model of the MPEG-1 standard

The structure adopted for MPEG-1 set the pattern for most MPEG standards:

1. *Part 1 – Systems* specifies how to combine one or more audio and video data streams with timing information to form a single stream (link)
2. *Part 2 – Video* specifies the video coding algorithm applied to so-called SIF video of ¼ the standard definition TV (link)
3. *Part 3 – Audio* specifies the audio compression. Audio is stereo and can be compressed with 3 different performance "layers": layer 1 is for an entry level digital audio, layer 2 for digital broadcasting and layer 3, aka MP3, for digital music. The MPEG-1 Audio layers were the predecessors of MPEG-2 profiles (and of most subsequent MPEG standards) (link)
4. *Part 4 – Compliance testing* (link)
5. *Part 5 – Software simulation* (link).

13.2 MPEG-2

MPEG-2 was a more complex beast to deal with. A digitised TV channel can yield 20-24 Mbit/s, depending on the delivery system (terrestrial/satellite broadcasting or cable TV). Digital stereo audio can take 0.2 Mbit/s and standard resolution 4 Mbit/s (say a little less with more compression). Audio could be multichannel (say, 5.1) and hopefully consume less bitrate for a total bitrate of a TV program of 4 Mbit/s. Hence the bandwidth taken by an analogue TV program can be used for 5-6 digital TV programs.

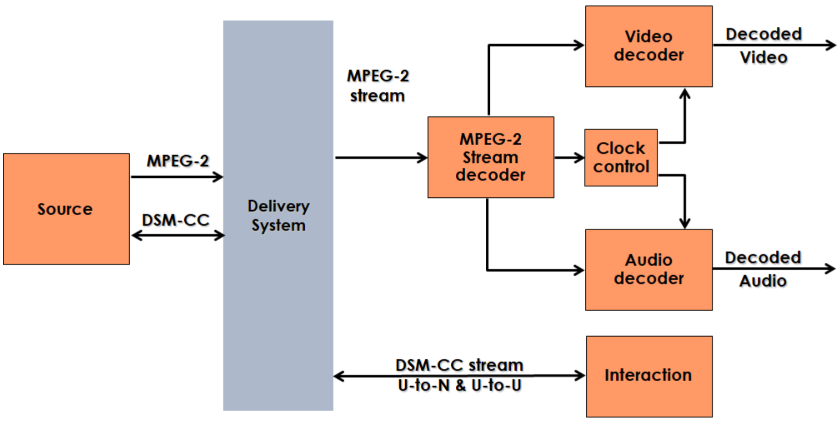


Figure 52 – Model of the MPEG-2 standard

The fact that digital TV programs part of a multiplex may come from independent sources and that digital channels in the real world are subject to errors force the design of an entirely different Systems layer for MPEG-2. The fact that users need to access other data sent in a carousel, that in an interactive scenario (with a return channel) there is a need for session management and that a user may interact with a server forced MPEG to add a new stream for user-to-network and user-to-user protocols.

In conclusion the MPEG-2 model is a natural extension of the MPEG-1 model (superficially, the DSM-CC line, but the impact is more pervasive).

The official title of MPEG-2 is *Generic coding of moving pictures and associated audio information*. It was originally intended for coding of standard definition television (MPEG-3 was expected to deal with coding of High Definition Television). As the work progressed, however, it became clear that a single format for both standard and high definition was not only desirable but possible. Therefore, the MPEG-3 project never took off.

The standard is not specific of a video resolution (this was already the case for MPEG-1 Video) but rationalises the notion of profiles, i.e. assemblies of coding tools and levels a notion that applies to, say, resolution, bitrate etc. Profiles and levels have subsequently adopted in most MPEG standardisation areas.

The standard is composed of 10 parts, some of which are

- *Part 1 – Systems* specifies the Systems layer to enable the transport of a multichannel digital TV stream on a variety of delivery media (link)
- *Part 2 – Video* specifies the video coding algorithm. Video is interlaced and may have a wide range of resolutions with support to scalability and multiview in appropriate profiles (link)
- *Part 3 – Audio* specifies a MPEG-1 Audio backward-compatible multichannel audio coding algorithm. This means that an MPEG-1 Audio decoder is capable of extracting and decoding an MPEG-1 Audio bitstream (link)
- *Part 6 – Extensions for DSM-CC* specifies User-to-User and User-to-Network protocols for both broadcasting and interactive applications. For instance DSM-CC can be used to enable such functionalities as carousel or session set up (link)
- *Part 7 – Advanced Audio Coding (AAC)* specifies a non-backward compatible multichannel audio coding algorithm. This was done because backward compatibility imposes too big a penalty for some applications, e.g. those that do not need backward compatibility (link), the first time MPEG was forced to develop two standards for apparently the same applications.

13.3 MPEG-4

MPEG-4 had the ambition of bringing interactive 3D spaces to every home. Media objects such as audio, video, 2D graphics were an enticing notion in the mid-1990's. The WWW had shown that it was possible to implement interactivity inexpensively and the extension to media interactivity looked like it would be the next step. Hence the official title of MPEG-4 *Coding of audio-visual objects*. *Figure 53* shows the MPEG-4 model.

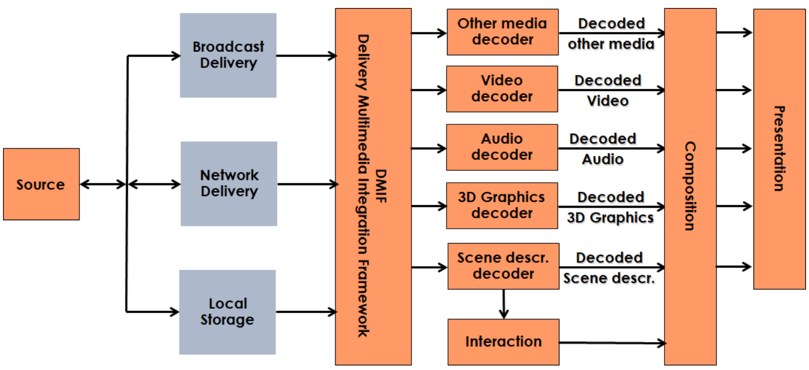


Figure 53 – Model of the MPEG-4 standard

This vision did not become true and one could say that even today it is not entirely clear what is interactivity and what is the interactive media experience a user is seeking, assuming that just one exists.

Is this a signal that MPEG-4 was a failure?

- *Yes*, it was a failure, and so what? MPEG operates like a company. Its "audio-visual objects" product looked like a great idea, but the market thought differently.
- *No*, it was a success, because 6 years after MPEG-2, MPEG-4 Visual yielded some 30% improvement in terms of compression.
- *Yes*, it was a failure because a patent pool dealt a fatal blow with their "content fee" (i.e. "you pay royalties by the amount of time you stream").
- *No* it was a success because MPEG-4 has 34 parts, the largest number ever achieved by MPEG in its standards, that include some of the most foundational and successful standards such as

the AAC audio coding format, the MP4 File Format, the Open Font Format and, of course the still ubiquitous Advanced Video Coding AVC video coding format whose success was not dictated so much by the 20% more compression that it delivers compared to MPEG-4 Visual (always nice to have), but to the industry-friendly licence released by a patent pool. Most important, the development of most MPEG standards is driven by a vision. Therefore, users have available a packaged solution, but they can also take the pieces that they need.

An overview of the entire MPEG-4 standard is available here. The standard is composed of 34 parts, some of which are

1. *Part 1 – Systems* specifies the means to interactively and synchronously represent and deliver audio-visual content composed of various objects (link)
2. *Part 2 – Visual* specifies the coded representation of visual information in the form of natural objects (video sequences of rectangular or arbitrarily shaped pictures) and synthetic visual objects (moving 2D meshes, animated 3D face and body models, and texture) (link).
3. *Part 3 – Audio* specifies a multi-channel perceptual audio coder with transparent quality compression of Compact Disc music coded at 128 kb/s that made it the standard of choice for many streaming and downloading applications (link)
4. *Part 6 – Delivery Multimedia Integration Framework (DMIF)* specifies interfaces to virtualise the network
5. *Part 9 – Reference hardware description* specifies the VHDL representation of MPEG-4 Visual (link)
6. *Part 10 – Advanced Video Coding* adds another 20% of performance to part 2 (link)
1. *Part 11 – Scene description and application engine* provides a time dependent interactive 3D environment building on VRML (link)
7. *Part 12 – ISO base media file format* specifies a file format that has been enriched with many functionalities over the years to satisfy the needs of the multiple MPEG client industries (link)
8. *Part 16 – Animation Framework eXtension (AFX)* specifies a range of 3D Graphics technologies, including 3D mesh compression (link)
9. *Part 22 – Open Font Format* (OFF) is the result of the MPEG effort that took over an industry initiative (OpenType font format specification), brought it under the folds of international standardisation and expanded/maintained it in response to evolving industry needs (link)
10. *Part 29 – Web video coding (WebVC)* specifies the Constrained Baseline Profile of AVC in a separate document
11. *Part 30 – Timed text and other visual overlays in ISO base media file format* supports applications that need to overlay other media to video (link)
12. *Part 31 – Video coding for browsers (VCB)* specifies a video compression format (unpublished)
13. *Part 33 – Internet Video Coding (IVC)* specifies a video compression format (link).

Parts 29, 31 and 33 are the results of 3 attempts made by MPEG to develop Option 1 Video Coding standards with a good performance. All did not reach the goal because ISO rules allow a company to make a patent declaration without specifying which is the patented technology that the declaring company alleges to be affected by a standard. The patented technologies could not be removed because MPEG did not have a clue about which were the allegedly infringing technologies.

13.4 MPEG-7

In the late 1990's the industry had been captured by the vision of "500 hundred channels" and telcos thought they could offer interactive media services. With the then being deployed MPEG-1 and MPEG-2, and, with MPEG-4 under development, MPEG expected that users would have zillions of media items.

MPEG-7 started with the idea of providing a standard that would enable users to find the media content of their interest in a sea of media content. Definitely, MPEG-7 deviates from the logic of the previous two standards and the technologies used reflect that because it provides formats for data (called metadata) extracted from multimedia content to facilitate searching in multimedia items. As shown in the figure, metadata can be classified as Descriptions (metadata extracted from the media items, especially audio and video) and Description Schemes (compositions of descriptions). *Figure 54* also shows two additional key MPEG-7 technologies. The first is the Description Definition Language (DDL) used to define new Descriptors and the second id XML Compression. With Descriptions and Description Schemes represented in verbose XML, it is clear that MPEG needed a technology to effectively compress XML.

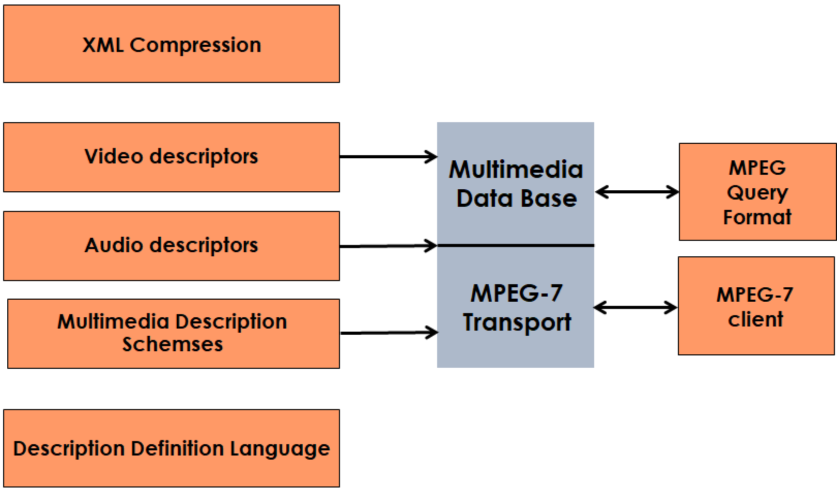


Figure 54 – Components of the MPEG-7 standard

An overview of the entire MPEG-7 standard is available here. The official title of MPEG-7 is *Multimedia content description interface* and the standard is composed of 16 parts, some of which are:

1. Part 1 – *Systems* has similar functions as the parts 1 of previous standards. In addition, it specifies a compression method for XML schemas used to represent MPEG-7 Descriptions and Description Schemes.
2. Part 2 – *Description definition language* breaks the Systems-Video-Audio traditional sequences of previous standards to provide a language to describe descriptions (link)
3. Part 3 – *Visual* specifies a wide variety of visual descriptors such as colour, texture, shape, motion etc. (link)
4. Part 4 – *Audio* specifies a wide variety of audio descriptors such as signature, instrument timber, melody description, spoken content description etc. (link)
5. Part 5 – *Multimedia description schemes* specifies description tools that are not visual and audio ones, i.e., generic and multimedia description tools such as description of the content structural aspects (link)
6. Part 8 – *Extraction and use of MPEG-7 descriptions* explains how MPEG-7 descriptions can be practically extracted and used
7. Part 12 – *Query format* defines format to query multimedia repositories (link)
8. Part 13 – *Compact descriptors for visual search* specifies a format that can be used to search images (link)
9. Part 15 – *Compact descriptors for video analysis* specifies a format that can be used to analyse video clips (link).

13.5 MPEG-21

In the year 1999 MPEG understood that its technologies were having a disruptive impact on the media business. MPEG thought that the industry should not fend of a new threat with old repressive

tools. The industry should convert the threat into an opportunity, but there were no standard tools to do that.
MPEG-21 is the standard resulting from the effort by MPEG to create a framework that would facilitate electronic commerce of digital media. It is a suite of specifications for end-to-end multimedia creation, delivery and consumption that can be used to enable open media markets.
This is represented in *Figure 55*. The basic MPEG-21 element is the Digital Item, a structured digital object with a standard representation, identification and metadata, around which a set of specifications were developed. MPEG-21 also includes specifications of Rights and Contracts and basic technologies such as the file format.

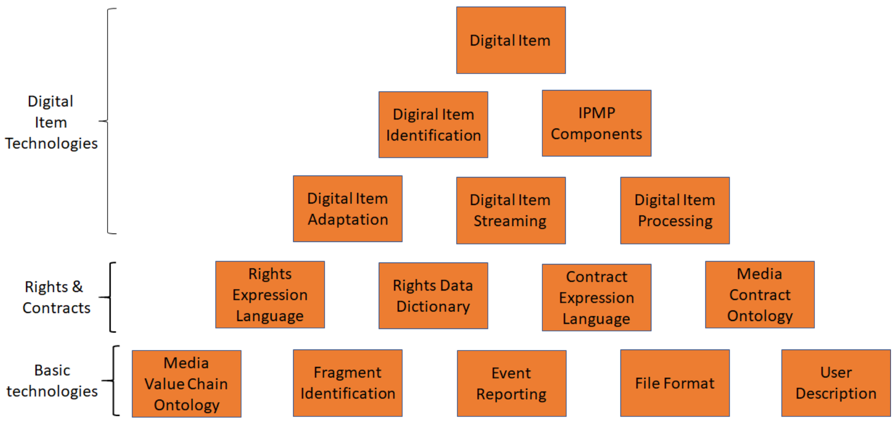


Figure 55 – Components of the MPEG-21 standard

An overview of the entire MPEG-21 standard, whose official title of MPEG-21 is *Multimedia Framework*, is available here. Some of the 21 MPEG-21 parts are briefly described below:
1. Part 2 – *Digital Item Declaration* specifies Digital Item (link)
2. Part 3 – *Digital Item Identification* specifies identification methods for Digital Items and their components (link)
3. Part 4 – *Intellectual Property Management and Protection (IPMP) Components* specifies how to include management and protection information and protected parts in a Digital Item (link)
4. Part 5 – *Rights Expression Language* specifies a language to express rights (link)
5. Part 6 – *Rights Data Dictionary* specifies a dictionary of rights-related data (link)
6. Part 7 – *Digital Item Adaptation* specifies description tools to enable optimised adaptation of multimedia content (link)
7. Part 15 – *Event Reporting* specifies a format to report events (links)
8. Part 17 – *Fragment Identification of MPEG Resources* specifies a syntax for URI Fragment Identifiers (link)
9. Part 19 – *Media Value Chain Ontology* specifies an ontology for Media Value Chains (link)
10. Part 20 – *Contract Expression Language* specifies a language to express digital contracts (link)
11. Part 21 – *Media Contract Ontology* specifies an ontology for media-related digital contracts (link).

13.6 MPEG-A

The official title of MPEG-A is *Multimedia Application Formats*. The idea behind MPEG-A is kind of obvious: we have standards for media elements (audio, video 3D graphics, metadata etc.), but what should one do to be interoperable when combining different media elements? Therefore MPEG-A is a suite of specifications that define application formats integrating existing MPEG technologies to provide interoperability for specific applications. Unlike the preceding standards that provided generic technologies for specific contexts, the link that unites MPEG-A specifications is the task of combing MPEG and, when necessary, other technologies for specific needs.

An overview of the MPEG-A standard is available here. Some of the 20 MPEG-A specifications are briefly described below:

1. *Part 2 – MPEG music player application format* specifies an “extended MP3 format” to enable augmented sound experiences (link)
2. *Part 3 – MPEG photo player application format* specifies additional information to a JPEG file to enable augmented photo experiences (link)
3. *Part 4 – Musical slide show application format* is a superset of the Music and Photo Player Application Formats enabling slide shows accompanied by music
4. *Part 6 – Professional archival application format* specifies a format for carriage of content, metadata and logical structure of stored content and related data protection, integrity, governance, and compression tools (link)
5. *Part 10 – Surveillance application format* specifies a format for storage and exchange of surveillance data that include compression video and audio, file format and metadata (link)
6. *Part 13 – Augmented reality application format* specifies a format to enable consumption of 2D/3D multimedia content including both stored and real time, and both natural and synthetic content (link)
7. *Part 15 – Multimedia Preservation Application Format* specifies the Multimedia Preservation Description Information (MPDI) that enables a user to discover, access and deliver multimedia resources (link)
8. *Part 18 – Media Linking Application Format* specifies a data format called "bridget", a link from a media item to another media item that includes source, destination, metadata etc. (link)
9. *Part 19 – Common Media Application Format* combines and restricts different technologies to deliver and combine CMAF Media Objects in a flexible way to form multimedia presentations adapted to specific users, devices, and networks (link)
10. *Part 22 – Multi-Image Application Format* enables precise interoperability points for creating, reading, parsing, and decoding images embedded in a High Efficiency Image File (HEIF).

13.7 MPEG-B

The official title of MPEG-B is *MPEG systems technologies.* After developing MPEG-1, -2, -4 and -7, MPEG realised that there were specific systems technologies that did not fit naturally into any part 1 of those standards. Thus, after using the letter A in MPEG-A, MPEG decided to use the letter B for this new family of specifications.MPEG-B is composed of 13 parts, some of which are

1. *Part 1 – Binary MPEG format for XML*, also called Binary MPEG format for XML or BiM, specifies a set of generic technologies for encoding XML documents adding to and integrating the specifications developed in MPEG-7 Part 1 and MPEG-21 part 16 (link)
2. *Part 4 – Codec configuration representation* specifies a framework that enables a terminal to build a new video or 3D Graphics decoder by assembling standardised tools expressed in the RVC CAL language (link)
3. *Part 5 – Bitstream Syntax Description Language (BSDL)* specifies a language to describe the syntax of a bitstream
4. Part 7 – *Common encryption format for ISO base media file format files* specifies elementary stream encryption and encryption parameter storage to enable a single MP4 file to be used on different devices supporting different content protection systems (link)
5. *Part 9 – Common Encryption for MPEG-2 Transport Streams* is a similar specification as Part 7 for MPEG-2 Transport Stream
6. *Part 11 – Green metadata* specifies metadata to enable a decoder to consume less energy while still providing a good quality video (link)

7. *Part 12 – Sample Variants* defines a Sample Variant framework to identify the content protection system used in the client (link)
8. *Part 13 – Media Orchestration* contains tools for orchestrating in time (synchronisation) and space the automated combination of multiple media sources (i.e. cameras, microphones) into a coherent multimedia experience rendered on multiple devices simultaneously
9. Part 14 – *Partial File Format* enables the description of an MP4 file partially received over lossy communication channels by providing tools to describe reception data, the received data and document transmission information
10. Part 15 – *Carriage of Web Resource* in MP4 FF specifies how to use MP4 File Format tools to enrich audio/video content, as well as audio-only content, with synchronised, animated, interactive web data, including overlays.

13.8 MPEG-C

The official title of MPEG-C *is MPEG video technologies*. As for systems, MPEG realised that there were specific video technologies supplemental to video compression that did not fit naturally into any part 2 of the MPEG-1, -2, -4 and -7 standards.
MPEG-C is composed of 6 parts, two of which are

1. Part 1 – *Accuracy requirements for implementation of integer-output 8x8 inverse discrete cosine transform*, was created after IEEE had discontinued a similar standard which is at the basis of important video coding standards
2. Part 4 – *Media tool library* contains modules called Functional Units expressed in the RVC-CAL language. These can be used to assemble some of the main MPEG Video coding standards, including HEVC and 3D Graphics compression standards, including 3DMC.

13.9 MPEG-D

The official title of MPEG-D is MPEG audio technologies. Unlike MPEG-C, MPEG-D parts 1, 2 and 3 actually specify audio codecs that are not generic as MPEG-1, MPEG-2 and MPEG-4 but intended to address specific application targets.
MPEG-D is composed of 5 parts, the first 4 of which are

1. Part 1 – *MPEG Surround* specifies an extremely efficient method for coding of multi-channel sound via the transmission of a 1) compressed stereo- or mono-audio program and 2) a low-rate side-information channel with the advantage of retaining backward compatibility to now ubiquitous stereo playback systems while giving the possibility to next-generation players to present a high-quality multi-channel surround experience
2. Part 2 – *Spatial Audio Object Coding* (SAOC) specifies an audio coding algorithm capable to efficiently handle individual audio objects (e.g. voices, instruments, ambience, ..) in an audio mix and to allow the listener to adjust the mix based on their personal taste, e.g. by changing the rendering configuration of the audio scene from stereo over surround to possibly binaural reproduction
3. Part 3 – *Unified Speech and Audio Coding* (USAC) specifies an audio coding algorithm capable to provide consistent quality for mixed speech and music content with a quality that is better than codecs that are optimized for either speech content or music content
4. Part 4 – *Dynamic Range Control* (DRC) specifies a unified and flexible format supporting comprehensive dynamic range and loudness control, addressing a wide range of use cases including media streaming and broadcast applications. The DRC metadata attached to the audio content can be applied during playback to enhance the user experience in scenarios such as ‘in a crowded room’ or ‘late at night’.

13.10 MPEG-E

This standard is the result of an entirely new direction of MPEG standardisation. Starting from the need to define API that applications can call to access key MPEG technologies, MPEG developed a Call for Proposal to which several responses were received. MPEG reviewed the responses and developed the ISO/IEC standard called Multimedia Middleware.

MPEG-E is composed of 8 parts

1. Part 1 – *Architecture* specifies the MPEG Multimedia Middleware (M3W) architecture that allows applications to execute multimedia functions without requiring detailed knowledge of the middleware and to update, upgrade and extend the M3W
2. Part 2 – *Multimedia application programming interface (API)* specifies the M3W API that provide media functions suitable for products with different capabilities for use in multiple domains
3. Part 3 – *Component model* specifies the M3W component model and the support API for instantiating and interacting with components and services
4. Part 4 – *Resource and quality management,* Part 5 – *Component download,* Part 6 – *Fault management* and Part 7 – *System integrity management* specify the support API and the technology used for M3W Component Download Fault Management Integrity Management and Resource Management, respectively

13.11 MPEG-V

The development of the MPEG-V standard Media context and control started in 2006 from the consideration that MPEG media – audio, video, 3D graphics etc. – offer virtual experiences that may be a digital replica of a real world, a digital instance of a virtual world or a combination of natural and virtual worlds. At that time, however, MPEG could not offer users any means to interact with those worlds.

MPEG undertook the task to provide standard interactivity technologies that allow a user to

1. Map their real-world sensor and actuator context to a virtual-world sensor and actuator context, and vice-versa and
2. Achieve communication between virtual worlds.

All data streams indicated are specified in one or more of the 7 MPEG-V parts

1. Part 1 – *Architecture* expands of the figure above
2. Part 2 – *Control information* specifies control devices interoperability (actuators and sensors) in real and virtual worlds
3. Part 3 – *Sensory information* specifies the XML Schema-based Sensory Effect Description Language to describe actuator commands such as light, wind, fog, vibration, etc. that trigger human senses
4. Part 4 – *Virtual world object characteristics* defines a base type of attributes and characteristics of the virtual world objects shared by avatars and generic virtual objects
5. Part 5 – *Data formats for interaction devices* specifies syntax and semantics of data formats for interaction devices – Actuator Commands and Sensed Information – required to achieve interoperability in controlling interaction devices (actuators) and in sensing information from interaction devices (sensors) in real and virtual worlds
6. Part 6 – *Common types and tools* specifies syntax and semantics of data types and tools used across MPEG-V parts.

13.12 MPEG-MAR

When MPEG began the development of the Augmented Reality Application Format (ARAF) it also started a specification called Augmented Reality Reference Model. Later it became aware that SC 24 *Computer graphics, image processing and environmental data representation* was doing a similar work and joined forces to develop a standard called Mixed and Augmented Reality Reference Model with them.

In the Mixed and Augmented Reality (MAR) paradigm, representations of physical and computer mediated virtual objects are combined in various modalities and the. MAR standard has been developed to enable

1. *The design of MAR applications or services*. The designer may refer and select the needed components from those specified in the MAR model architecture taking into account the given application/service requirements.
2. *The development of a MAR business model*. Value chain and actors are identified in the Reference Model and implementors may map them to their business models or invent new ones.
3. *The extension of existing or creation of new MAR standards*. MAR is interdisciplinary and creates ample opportunities for extending existing technology solutions and standards.

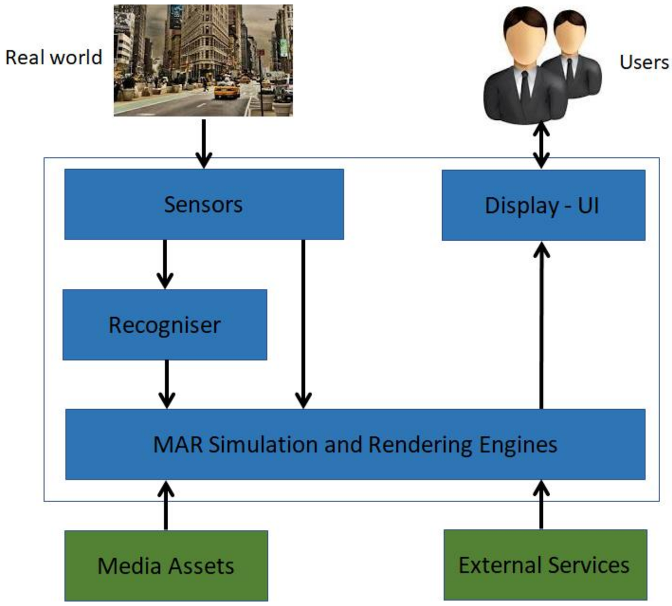


Figure 56 – MAR Reference System Architecture

MAR-RM and ARAF paradigmatically express the differences between MPEG standardisation and regular IT standardisation. MPEG defines interfaces and technologies while IT standards typically concern architectures and reference models. This explains why most patent declarations received by ISO relate to MPEG standards. It is also worth noting that in the 6 years it took to develop the standard, MPEG developed 3 editions of its ARAF standard.

The Reference architecture of the MAR standard is depicted in *Figure 56*. Information from the real world is sensed and enters the MAR engine either directly or after being "understood". The engine can also access media assets or external services. All information is processed by the engine which outputs the result of its processing and manages the interaction with the user.

Based on this the standard elaborates the *Enterprise Viewpoint* with classes of actors, roles, business model, successful criteria, the Computational Viewpoint with functionalities at the component level and the Informational Viewpoint with data communication between components.

MM-RM is a one-part standard.

13.13 MPEG-M

Multimedia service platform technologies (MPEG-M) specifies two main components of a multi-media device, called peer in MPEG-M.

The first component is API: High-Level API for applications and Low-Level API for network, energy and security. The second components is a middleware called MXM that relies specifically on MPEG multimedia technologies.
The Middleware is composed of two types of engine. *Technology Engines* are used to call functionalities defined by MPEG standards such as creating or interpreting a licence attached to a content item. *Protocol Engines* are used to communicate with other peer, e.g. in case a peer does not have a particular Technology Engine that another peer has. For instance, a peer can use a Protocol Engine to call a licence server to get a licence to attach to a multimedia content item. The MPEG-M middleware can create chains of Technology Engines (Orchestration) or Protocol Engines (Aggregation).
MPEG-M is a 5-part standard
1. Part 1 – *Architecture* specifies the architecture, and High- and Low-Level API
2. Part 2 – *MPEG extensible m*iddleware *(MXM) API* specifies the API
3. Part 3 – *Conformance and re*ference software
4. Part 4 – *Elementary services* specifies the elementary services provided by the Protocol Engines
5. Part 5 – *Service aggregation* specifies how elementary services can be aggregated.

13.14 MPEG-U

The development of the MPEG-U standards was motivated by the evolution of User Interfaces that integrate advanced rich media content such as 2D/3D, animations and video/audio clips and aggregate dedicated small applications called widgets. These are standalone applications embedded in a Web page and rely on Web technologies (HTML, CSS, JS) or equivalent.
With its MPEG-U standard, MPEG sought to have a common UI on different devices, e.g. TV, Phone, Desktop and Web page.
Therefore MPEG-U extends W3C recommendations to
1. Cover non-Web domains (Home network, Mobile, Broadcast)
2. Support MPEG media types (BIFS and LASeR) and transports (MP4 FF and MPEG-2 TS)
3. Enable Widget Communications with restricted profiles (without scripting)

The MPEG-U architecture is depicted in *Figure 57*.

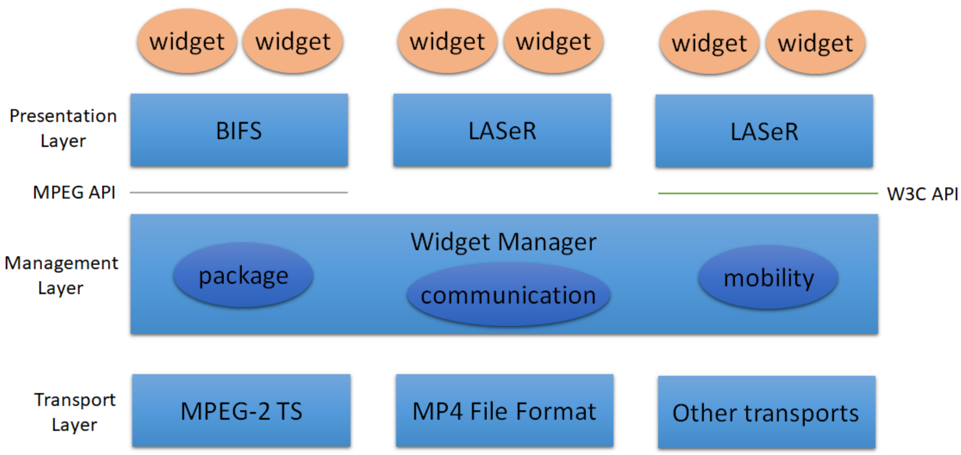


Figure 57 – MPEG-U Architecture

The normative behaviour of the Widget Manager includes the following elements of a widget
1. Packaging formats
2. Representation format (manifest)
3. Life Cycle handling

4. Communication handling
5. Context and Mobility management
6. Individual rendering (i.e. scene description normative behaviour)

Figure 58 depicts the operation of an MPEG-U widget for TV in a DLNA environment.

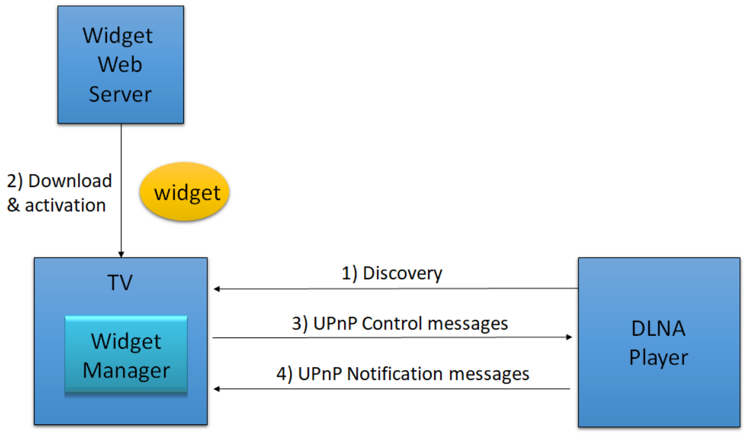


Figure 58 – MPEG-U for TV in a DLNA environment

MPEG-U is a 3-part standard
1. Part 1 – *Widgets*
2. Part 2 – *Additional gestures and multimodal interaction*
3. Part 3 – *Conformance and reference software*

13.15 MPEG-H

High efficiency coding and media delivery in heterogeneous environments (MPEG-H) is an integrated standard that resumes the original MPEG "one and trine" Systems-Video-Audio standards approach. In the wake of those standards, the 3 parts can be and are actually used independently, e.g. in video streaming applications. On the other hand, ATSC have adopted the full Systems-Video-Audio triad with extensions of their own.

MPEG-H has 15 parts, as follows
1. Part 1 – *MPEG Media Transport* (MMT) is the solution for the new world of broadcasting where delivery of content can take place over different channels each with different characteristics, e.g. one-way (traditional broadcasting) and two-way (the ever more pervasive broadband network). MMT assumes that the Internet Protocol is common to all channels.
2. Part 2 – *High Efficiency Video Coding* (HEVC) is the latest approved MPEG video coding standard supporting a range of functionalities: scalability, multiview, from 4:2:0 to 4:4:4, up to 16 bits, Wider Colour Gamut and High Dynamic Range and Screen Content Coding
3. Part 3 – *3D Audio* is the latest approved audio coding standards supporting enhanced 3D audio experiences
4. Parts 4, 5 and 6 *Reference software* for MMT, HEVC and 3D Audio
5. Parts 7, 8, 9 *Conformance testing* for MMT, HEVC and 3D Audio
6. Part 10 – *MPEG Media Transport FEC Codes* specifies several Forward Error Correcting Codes for use by MMT.
7. Part 11 – *MPEG Composition Information* specifies an extension to HTML 5 for use with MMT
8. Part 12 - *Image File Format* specifies a file format for individual images and image sequences
9. Part 13 – *MMT Implementation Guidelines* collects useful guidelines for MMT use

10. Parts 14 – Conversion and coding practices for high-dynamic-range and wide-colour-gamut video and 15 – *Signalling, backward compatibility and display adaptation for HDR/WCG video* are technical reports to guide users in supporting HDR/WCC,

13.16 MPEG-DASH

Dynamic adaptive streaming over HTTP (DASH) is a suite of standards providing a standard solution for the efficient and easy streaming of multimedia using existing available HTTP infrastructure (particularly servers and CDNs, but also proxies, caches, etc.). DASH was motivated by the popularity of HTTP streaming and the existence of different protocols used in different streaming platforms, e.g. different manifest and segment formats.

By developing the DASH standard for HTTP streaming of multimedia content, MPEG has enabled a standard-based client to stream content from any standard-based server, thereby enabling interoperability between servers and clients of different vendors.

As depicted in *Figure 59*, the multimedia content is stored on an HTTP server in two components: 1) Media Presentation Description (MPD) which describes a manifest of the available content, its various alternatives, their URL addresses and other characteristics, and 2) Segments which contain the actual multimedia bitstreams in form of chunks, in single or multiple files.

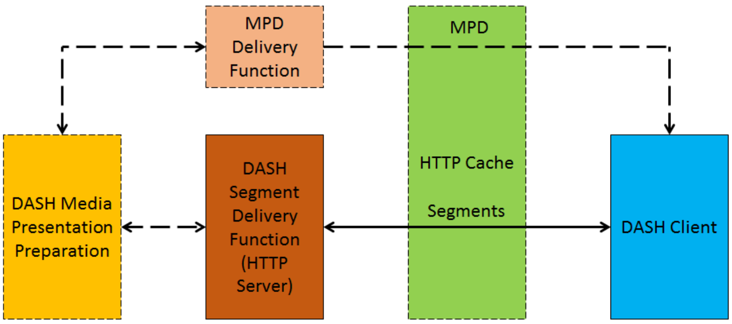


Figure 59 – DASH model

Currently DASH is composed of 8 parts

1. Part 1 – *Media presentation description and segment formats* specifies 1) the Media Presentation Description (MPD) which provides sufficient information for a DASH client to adaptive stream the content by downloading the media segments from a HTTP server, and 2) the segment formats which specify the formats of the entity body of the request response when issuing a HTTP GET request or a partial HTTP GET.
2. Part 2 – *Conformance and reference software* the regular component of an MPEG standard
3. Part 3 – *Implementation guidelines* provides guidance to implementors
4. Part 4 – *Segment encryption and authentication* specifies encryption and authentication of DASH segments
5. Part 5 – *Server and Network Assisted DASH* specifies asynchronous network-to-client and network-to-network communication of quality-related assisting information
6. Part 6 – *DASH with Server Push and WebSockets* specified the carriage of MPEG-DASH media presentations over full duplex HTTP-compatible protocols, particularly HTTP/2 and WebSockets
7. Part 7 – *Delivery of CMAF content with DASH* specifies how the content specified by the Common Media Application Format can be carried by DASH
8. Part 8 – *Session based DASH operation* will specify a method for MPD to manage DASH sessions for the server to instruct the client about some operation continuously applied during the session.

13.17 MPEG-I

Coded representation of immersive media (MPEG-I) represents the current MPEG effort to develop a suite of standards to support immersive media products, services and applications.

Currently MPEG-I has 11 parts and more parts are being added.

1. Part 1 – *Immersive Media Architectures* outlines possible architectures for immersive media services.
2. Part 2 – *Omnidirectional MediA Format* specifies an application format that enables consumption of omnidirectional video, aka Video 360. Version 2 is under development
3. Part 3 – *Immersive Video Coding* will specify the emerging Versatile Video Coding standard
4. Part 4 – *Immersive Audio Coding* will specify metadata to enable enhanced immersive audio experiences compared to what is possible today with MPEG-H 3D Audio
5. Part 5 – *Video-based Point Cloud Compression* will specify a standard to compress dense static and dynamic point clouds
6. Part 6 – *Immersive Media Metrics* will specify different parameters useful for immersive media services and their measurability
7. Part 7 – *Immersive Media Metadata* will specify systems, video and audio metadata for immersive experiences. One example is the current 3DoF+ Video activity
8. Part 8 – *Network-Based Media Processing* will specify APIs to access remote media processing services
9. Part 9 – *Geometry-based Point Cloud Compression* will specify a standard to compress sparse static and dynamic point clouds
10. Part 10 – *Carriage of Point Cloud Data* will specify how to accommodate compressed point clouds in the MP4 File Format
11. Part 11 – *Implementation Guidelines for Network-based Media Processing* is the usual collection of guidelines

13.18 MPEG-CICP

Coding-Independent Code-Points (MPEG-CICP) is a collection of code points that have been assembled in single documents because they are not standard-specific

Part 1 – *Systems*, Part 2 – *Video* and Part 3 – *Audio* collect the respective code points and Part 4 – *Usage of video signal type code points* contains guidelines for their use

13.19 MPEG-G

Genomic Information Representation (MPEG-G) is a suite of specifications developed jointly with TC 276 Biotechnology that allows to reduce the amount of information required to losslessly store and transmit DNA reads from high speed sequencing machines.

An MPEG-G file can be created with the following sequence of operations:

1. Put the reads in the input file (aligned or unaligned) in bins corresponding to segments of the reference genome
2. Classify the reads in each bin in 6 classes: P (perfect match with the reference genome), M (reads with variants), etc.
3. Convert the reads of each bin to a subset of 18 descriptors specific of the class: e.g., a class P descriptor is the start position of the read etc.
4. Put the descriptors in the columns of a matrix
5. Compress each descriptor column (MPEG-G uses the very efficient CABAC compressor already present in several video coding standards)
6. Put compressed descriptors of a class of a bin in an Access Unit (AU) for a maximum of 6 AUs per bin

MPEG-G currently includes 6 parts

1. Part 1 – *Transport and Storage of Genomic Information* specifies the file and streaming formats
2. Part 2 – *Genomic Information Representation* specified the algorithm to compress DNA reads from high speed sequencing machines
3. Part 3 – *Genomic information metadata and application programming interfaces (APIs)* specifies metadata and API to access an MPEG-G file
4. Part 4 – *Reference Software* and Part 5 – *Conformance* are the usual components of a standard
5. Part 6 – *Genomic Annotation Representation* will specify how to compress annotations.

13.20 MPEG-IoMT

Internet of Media Things (MPEG-IoMT) is a suite of specifications:
1. API to discover Media Things,
2. Data formats and API to enable communication between Media Things.

A Media Thing (MThing) is the media "version" of IoT's Things and *Figure 60* represents the IoMT reference model.

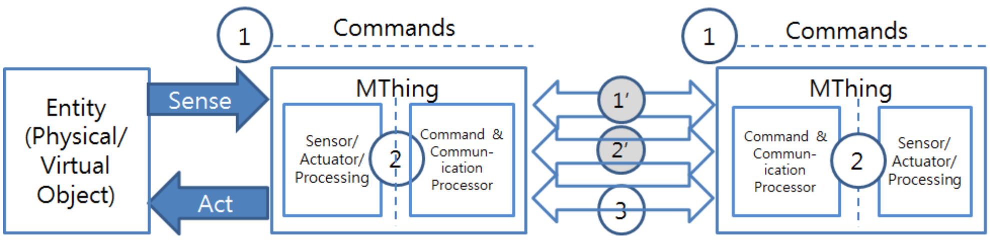


Figure 60: IoT in MPEG is for media – IoMT

Currently MPEG-IoMT includes 4 parts
1. Part 1 – *IoMT Architecture* will specify the architecture
2. Part 2 – *IoMT Discovery and Communication API* specifies Discovery and Communication API
3. Part 3 – *IoMT Media Data Formats and API* specifies Media Data Formats and API
4. Part 4 – *Reference Software and Conformance* is the usual part of MPEG standards

13.21 MPEG-5

General Video Coding (MPEG-5) will contain new video coding specifications.
1. Part 1 – *Essential Video Coding* will specify a video codec with two layers. The first layer will provide a significant improvement over AVC but significantly less than HEVC and the second layer will provide a significant improvement over HEVC but significantly less than to VVC.
2. Part 2 – *Low Complexity Video Coding Enhancements* will specify a data stream structure defined by two component streams, a base stream decodable by a hardware decoder, and an enhancement stream suitable for software processing implementation with sustainable power consumption. The enhancement stream will provide new features such as compression capability extension to existing codecs, lower encoding and decoding complexity, for on demand and live streaming applications. *Figure 61* sketches an LCEVC decoder.

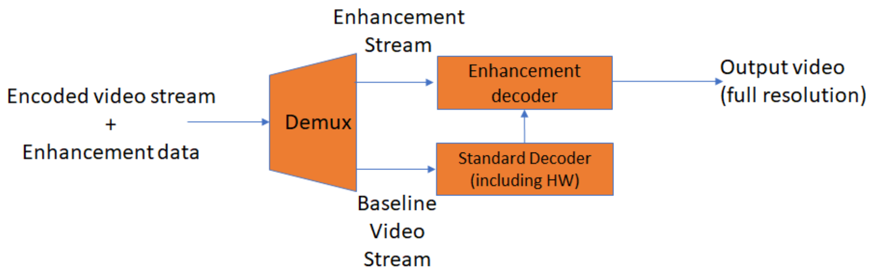


Figure 61: Low Complexity Enhancement Video Coding

14 The MPEG work plan

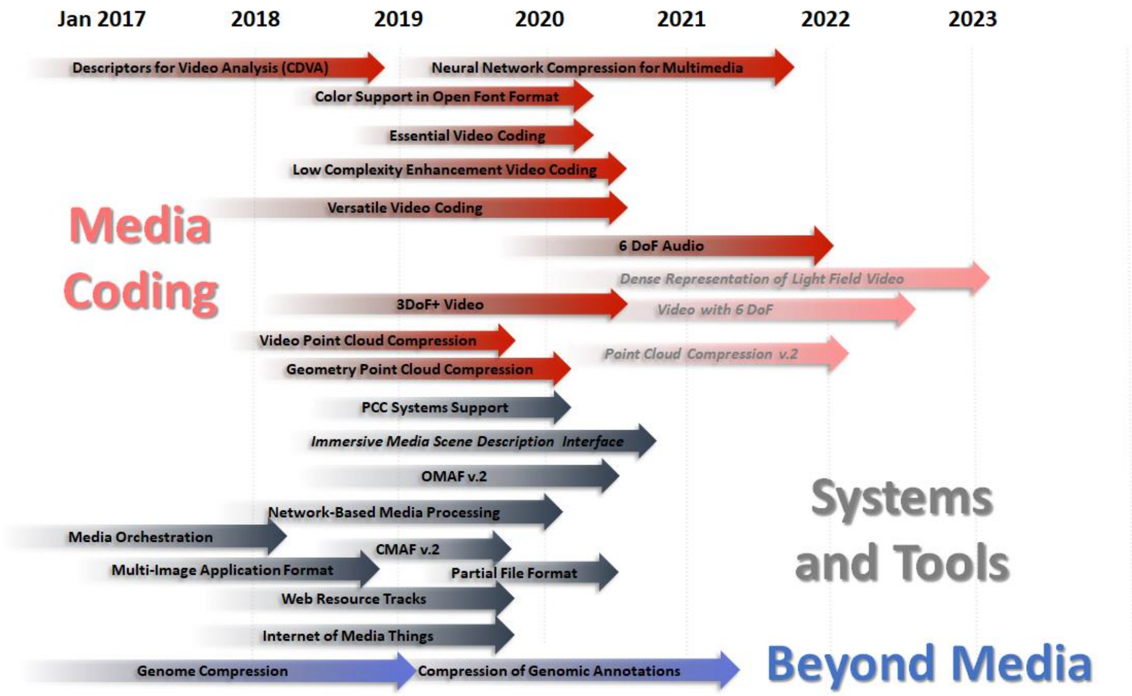


Figure 62 – The MPEG work plan (March 2019)

The MPEG work plan at a glance

Figure 62 shows the *main* standards that MPEG has developed or is developing in the 2017-2023 period. It is organised in 3 main sections:

- Media Coding (e.g. AAC and AVC)
- Systems and Tools (e.g. MPEG-2 TS and File Format)
- Beyond Media (currently Genome Compression).

Disclaimer: in *Figure 62* and in the following all dates are planned dates.

Navigating the areas of the MPEG work plan

The 1st column in *Figure 63* gives the currently active MPEG standardisation areas. The first row gives the currently active MPEG standards. The non-empty white cells give the number of "deliverables" (Standards, Amendments and Technical Reports) currently identified in the work plan.

Video coding

In the Video coding area MPEG is currently developing specifications for 4 standards: MPEG-H, -I, -5 and -CICP) and is exploring advanced technologies for immersive visual experiences.

	2	4	5	7	A	B	CICG	D	DASH	G	H	I	IoMT	V
Video Coding			2S				1T				1S	1S		
Audio Coding								1S				1S		
3D Graphics Coding												2S		
Font Coding		1S												
Genome Coding										1S				
Neural Network Coding				1S										
Media Description				1S								1S		
Systems Support		1S									1A	1S		
Transport	2A	2A				2A			2S2A2T			1S		
Application Formats					1A									
API												1S		
Media Systems													1S	
Reference Implementation		1S		1S	1A				1S	1S		3S	1S	1S
Conformance				1S	2A				1S	1S		2S	1S	

Figure 63 – Standards (S), Amendments (A) and Technical Reports (T) in the MPEG work plan (March 2019)

MPEG-H
Part 3 – *High Efficiency Video Coding* 4th edition specifies a profile of HEVC with an encoding of a single (i.e. monochrome) colour plane and will be restricted to a maximum of 10 bits per sample, as done in past HEVC range extensions profiles, and additional Supplemental Enhancement Information (SEI) messages, e.g. fisheye video, SEI manifest, and SEI prefix messages.

MPEG-I
Part 3 – *Versatile Video Coding*, currently being developed jointly with VCEG, MPEG is working on the new video compression standard after HEVC. VVC is expected to reach FDIS stage in July 2020 for the core compression engine. Other parts, such as high-level syntax and SEI messages will follow later.

MPEG-CICP
Part 4 – *Usage of video signal type code points* 2nd edition will document additional combinations of commonly used code points and baseband signalling.

MPEG-5
MPEG has already obtained all technologies necessary to develop standards with the intended functionalities and performance from the Calls for Proposals (CfP) for the two parts.

1. Part 1 – *Essential Video Coding* will specify a video codec with two layers: layer 1 significantly improves over AVC but performs significantly less than HEVC and layer 2 significantly improves over HEVC but performs significantly less than VVC.
2. Part 2 – *Low Complexity Video Coding Enhancements* will specify a data stream structure defined by two component streams: stream 1 is decodable by a hardware decoder, stream 2 can be decoded in software with sustainable power consumption. Stream 2 provides new features such as compression capability extension to existing codecs, lower encoding and decoding complexity, for on demand and live streaming applications.

Explorations
MPEG experts are collaborating in the development of support tools, acquisition of test sequences and understanding of technologies required for 6DoF and light field.

1. *Compression of 6DoF visual* will enable a user to move more freely than in 3DoF+, eventually, allowing any translation and rotation in space.

2. *Compression of dense representation of light fields* is stimulated by new devices that capture light field with both spatial and angular light information. As the size of data is large and different from traditional images, effective compression schemes are required.

Audio coding
In the Audio coding area MPEG is working on 2 standards (MPEG-D, and -I).

MPEG-D
In Part 5 – *Uncompressed Audio in MP4 File Format*, MPEG extends MP4 to enable carriage of uncompressed audio (e.g. PCM). At the moment, MP4 only carries compressed audio.

MPEG-I
Part 3 *Immersive Audio*. As MPEG-H 3D Audio already supports a 3DoF user experience, MPEG-I builds upon it to provide a 6DoF immersive audio experience. A Call for Proposal will be issued in October 2019. Submissions are expected in October 2021 and FDIS stage is expected to be reached in April 2022. Even though this standard will not be about compression, but about metadata as for 3DoF+ Visual, we have kept this activity under Audio Coding.

3D Graphics Coding
In the 3D Graphics Coding area MPEG is developing two parts of MPEG-I.

- Part 5 – *Video-based Point Cloud Compression* (V-PCC) for which FDIS stage is planned to be reached in October 2019.
- Part 9 – *Geometry-based Point Cloud Compression* (G-PCC) for which FDIS stage is planned to be reached in January 2020.

The two PCC standards employ different technologies and target different application areas, generally speaking, entertainment and automotive/unmanned aerial vehicles.

Font Coding
In the Font coding area MPEG is working on a new edition of MPEG-4 part 22.
Part 22 – Open Font Format. 4th edition specifies support of complex layouts and additional support for new layout features. FDIS stage will be reached in April 2020.

Genome Coding
In the Genome coding area MPEG has achieved FDIS level for the 3 foundational parts of the MPEG-G standard:

- Part 1 – *Transport and Storage of Genomic Information*
- Part 2 – *Genomic Information Representation*
- Part 3 – *Genomic information metadata and application programming interfaces (APIs)*.

In October 2019 MPEG will complete Part 4 – *Reference Software* and Part 5 – *Conformance.* In July 2019 MPEG will issue a Call for Proposals for Part 6 – *Genomic Annotation Representation*.

Neural Network Coding
Compression of this type of data is motivated by the increasing use of neural networks in many applications that require the deployment of a particularly trained network instance to a potentially large number of devices, which may have limited processing power and memory.
MPEG has restricted the general field to neural networks trained with media data, e.g. for the object identification and content description, and is therefore developing the standard in MPEG-7 which already contains two standards – CDVS and CDVA – which offer similar functionalities

achieved with different technologies (and therefore the standard should be classified under Media description).

MPEG-7
Part 17 – *Compression of neural networks for multimedia content description and analysis* MPEG is developing a standard that enable compression of artificial neural networks trained with audio and video data. FDIS is expected in January 2021.

Media Description
Media description is the goal of the MPEG-7 standard which contains technologies for describing media, e.g. for the purpose of searching media.
In the Media Description area MPEG has completed *Part 15 Compact descriptors for video analysis* (CDVA) in October 2018 and is now working on 3DoF+ visual.

MPEG-I
Part 7 – *Immersive Media Metadata* will specify a set of metadata that enable a decoder to provide a more realistic user experience in OMAF v2. The FDIS is planned for July 2021.

System support
In the System support area MPEG is working on MPEG-4 and -I.

MPEG-4
Part 34 - *Registration Authorities* aligns the existing MPEG-4 Registration Authorities to current ISO practice.

MPEG-H
In MPEG-H MPEG is working on
Part 10 – *MPEG Media Transport FEC Codes.* This is being enhanced with the Window-based FEC code. FDAM is expected to be reached in January 2020.

MPEG-I
Part 6 – *Immersive Media Metrics* specifies the metrics and measurement framework in support of immersive media experiences. FDIS stage is planned to be reached in July 2020.

Transport
In the Transport area MPEG is working on MPEG-2, -4, -B, -H, -DASH, -I and Explorations.

MPEG-2
Part 2 – *Systems* continues to be a lively area of work 25 years after MPEG-2 Systems reached FDIS. After producing Edition 7, MPEG is working on two amendments to carry two different types of content

- Carriage of JPEG XS in MPEG-2 TS JPEG XS
- Carriage of associated CMAF boxes for audio-visual elementary streams in MPEG-2 TS

MPEG-4
Part 12 – *ISO Based Media File Format Systems* continues to be a lively area of work 20 years after MP4 File Format reached FDIS. MPEG is working on two amendments

- *Corrected audio handling*, expected to reach FDAM in July 2019
- *Compact movie fragment* is expected to reach FDAM stage in January 2020

MPEG-B

In MPEG-B MPEG is working on two new standards

- Part 14 – *Partial File Format* provides a standard mechanism to store HTTP entities and the partial file in broadcast applications for later cache population. The standard is planned to reach FDIS stage in July 2020.
- Part 15 – *Carriage of Web Resources in ISOBMFF* will make it possible to enrich audio/video content, as well as audio-only content, with synchronised, animated, interactive web data, including overlays. The standard is planned to reach FDIS stage in January 2020.

MPEG-DASH

In MPEG-DASH MPEG is working on

- Part 1 – *Media presentation description and segment formats* will see a new edition in July 2019 and will be enhanced with an Amendment on Client event and timed metadata processing. FDAM is planned to be reached in January 2020.
- Part 3 – *MPEG-DASH Implementation Guidelines* 2nd edition will become TR in July 2019
- Part 5 – Server and network assisted DASH (SAND) will be enriched by an Amendment on Improvements on SAND messages. FDAM to be reached in July 2019.
- Part 7 – *Delivery of CMAF content with DASH* a Technical Report with guidelines on the use of the most popular delivery schemes for CMAF content using DASH. TR is planned to be reached in March 2019
- Part 8 – *Session based DASH operation* will reach FDIS in July 2020.

MPEG-I

Part 2 – *Omnidirectional Media Format* (OMAF) released in October 2017 is the first standard format for delivery of omnidirectional content. With OMAF 2nd Edition *Interactivity support for OMAF*, planned to reach FDIS in July 2020, MPEG is extending OMAF with 3DoF+ functionalities.

Application Formats

MPEG-A ISO/IEC 23000 Multimedia Application Formats is a suite of standards for combinations of MPEG and other standards (only if there are no suitable MPEG standard for the purpose). MPEG is working on

Part 19 – *Common Media Application Format* 2nd edition with support of new formats

Application Programming Interfaces

The Application Programming Interfaces area comprises standards that make possible effective use of some MPEG standards.

MPEG-I

Part 8 – *Network-based Media Processing* (NBMP), a framework that will allow users to describe media processing operations to be performed by the network. The standard is expected to reach FDIS stage in January 2020.

Media Systems

Media Systems includes standards or Technical Reports targeting architectures and frameworks.

IoMT

Part 1 – *IoMT Architecture*, expected to reach FDIS stage in October 2019. The architecture used in this standard is compatible with the IoT architecture developed by JTC 1/SC 41.

Reference Implementation
MPEG is working on the development of standards for reference software of MPEG-4, -7, A, -B, -V, -H, -DASH, -G, -IoMT

Conformance
MPEG is working on the development of standards for conformance of MPEG-4, -7, A, -B, -V, -H, -DASH, -G, -IoMT.

15 ISO numbers of MPEG standards

The 3 columns of *Table 15* give the commonly used MPEG acronym (column 1), the 5-digit ISO number identifying the standard (column 2) and the official title of the standard (column 3). The standards are listed in ascending ISO numbers.

Table 15 –Acronyms, ISO numbers and titles of MPEG standards

Acronym	ISO #	Title
MPEG-1	11172	Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s
MPEG-2	13818	Generic coding of moving pictures and associated audio information
MPEG-4	14496	Coding of audio-visual objects
MPEG-7	15938	Multimedia content description interface
MPEG-21	21000	Multimedia Framework
MPEG-A	23000	Multimedia Application Formats
MPEG-B	23001	MPEG systems technologies
MPEG-C	23002	MPEG video technologies
MPEG-D	23003	MPEG audio technologies
MPEG-E	23004	Multimedia Middleware
MPEG-V	23005	Media context and control
MPEG-M	23006	Multimedia service platform technologies
MPEG-U	23007	Rich media user interfaces
MPEG-H	23008	High efficiency coding and media delivery in heterogeneous environments
MPEG-DASH	23009	Dynamic adaptive streaming over HTTP (DASH)
MPEG-I	23090	Coded representation of immersive media
MPEG-CICP	23091	Coding-Independent Code-Points
MPEG-G	23092	Genomic Information Representation
MPEG-IoMT	23093	Internet of Media Things
MPEG-5	23094	General Video Coding

16 Conclusions

Thirty years ago, MPEG was an experts group in an unknown working group in a subcommittee developing standards for character coding.
Today MPEG is the group behind global media digitisation, media industry convergence, an installed base of media devices (tablets, smartphones, set top boxes and digital TVs) worth 2.8 T$ and a global yearly turnover in excess of 1 T$.

As this book tried to demonstrate, this did not happen by chance and I claim that will continue in the future. If mindless parties do not get in the way, I mean.