**INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO**

Source:	**Requirements Group**
Status:	**Approved**
Title:	**MPEG-7: Context and Objectives (version - 10 Atlantic City)**

MPEG-7 Context and Objectives

1. Context

More and more audio-visual information is available in digital form, in various places around the world. Along with the information, people appear that want to use it. Before one can use any information, however, it will have to be located first. At the same time, the increasing availability of potentially interesting material makes this search harder. Currently, solutions exist that allow searching for textual information. Many text-based search engines are available on the World Wide Web, and they are among the most visited sites - indicating they foresee a real demand. Identifying information is, however, not possible for audio-visual content, as no generally recognised description of this material exists. In general, it is not possible to efficiently search the web for, say, a picture of 'the Motorbike from Terminator II', or to search a sequence where "King Lear congratulates his assistants on the night after the battle," or to search for "twenty minutes of video according to my preferences of today". In specific cases, solutions do exist. Multimedia databases on the market today allow

searching for pictures using characteristics like colour, texture and information about the shape of objects in the picture. One could envisage a similar example for audio, in which one can whistle a melody to find a song.

The question of finding content is not restricted to database retrieval applications; also in other areas similar questions exist. For instance, there is an increasing amount of (digital) broadcast channels available, and this makes it harder to select the broadcast channel (radio or TV) that is potentially interesting.

2. MPEG-7 Objectives

In October 1996, MPEG started a new work item to provide a solution to the questions described above. The new member of the MPEG family, called "Multimedia Content Description Interface" (in short 'MPEG-7'), will extend the limited capabilities of proprietary solutions in identifying content that exist today, notably by including more data types. In other words: MPEG-7 will specify a standard set of descriptors that can be used to describe various types of multimedia information. MPEG-7 will also standardise ways to define other descriptors as well as structures (Description Schemes) for the descriptors and their relationships (see also 2.1 What is the Scope of the Standard). This description (i.e. the combination of descriptors and description schemes) shall be associated with the content itself, to allow fast and efficient searching for material of a user's interest. MPEG-7 will also standardise a language to specify description schemes, i.e. a Description Definition Language (DDL). AV material that has MPEG-7 data associated with it, can be indexed and searched for. This 'material' may include: still pictures, graphics, 3D models, audio, speech, video, and information about how these elements are combined in a multimedia presentation ('scenarios', composition information). Special cases of these general data types may include facial expressions and personal characteristics.

MPEG-7, like the other members of the MPEG family, is a standard representation of audio-visual information satisfying particular requirements. The MPEG-7 standard builds on other (standard) representations such as analogue, PCM, MPEG-1, -2 and -4. One functionality of the standard is to provide references to suitable portions of them. For example, perhaps a shape descriptor used in MPEG-4 is useful in an MPEG-7 context as well, and the same may apply to motion vector fields used in MPEG-1 and MPEG-2.

MPEG-7 descriptors do, however, not depend on the ways the described content is coded or stored. It is possible to attach an MPEG-7 description to an analogue movie or to a picture that is printed on paper. Even though the MPEG-7 description does not depend on the (coded) representation of the material, the standard in a way builds on MPEG-4, which provides the means to encode audio-visual material as objects having certain relations in time (synchronisation) and space (on the screen for video, or in the room for audio). Using MPEG-4 encoding, it will be possible to attach descriptions to elements (objects) *within* the scene, such as audio and visual objects.. MPEG-7 will allow different granularity in its descriptions, offering the possibility to have different levels of discrimination.

Because the descriptive features must be meaningful in the context of the application,

they will be different for different user domains and different applications.

This implies that the same material can be described using different types of features, tuned to the area of application. To take the example of visual material: a lower abstraction level would be a description of e.g. shape, size, texture, colour, movement (trajectory) and position ('where in the scene can the object be found?). And for audio: key, mood, tempo, tempo changes, position in sound space. The highest level would give semantic information: 'This is a scene with a barking brown dog on the left and a blue ball that falls down on the right, with the sound of passing cars in the background.' All these descriptions are of course coded in an efficient way - efficient for search that is. Intermediate levels of abstraction may also exist.

The level of abstraction is related to the way the features can be extracted: many low-level features can be extracted in fully automatic ways, whereas high level features need (much) more human interaction.

Next to having a description of the content, it may also be required to include other types of information about the multimedia data:

- *The form* - An example of the form is the coding scheme used (e.g. JPEG, MPEG-2), or the overall data size. This information helps determining whether the material can be 'read' by the user.
- *Conditions for accessing the material* - This could include copyright information, and price;
- *Classification* - This could include parental rating, and content classification into a number of pre-defined categories;
- *Links to other relevant material* - The information may help the user speeding up the search.
- *The context* - In the case of recorded non-fiction content, it is very important to know the occasion of the recording (e.g. Olympic Games 1996, final of 200 meter hurdles, men)

In many cases, it will be desirable to use textual information for the descriptions. Care must be taken, however, that the usefulness of the descriptions is as independent from the language area as possible. (A very clear example where text comes in handy is in giving names of authors, film, places.)

MPEG-7 data may be physically located with the associated AV material, in the same data stream or on the same storage system, but the descriptions could also live somewhere else on the globe. When the content and its descriptions are not co-located, mechanisms that link AV material and their MPEG-7 descriptions are useful; these links should work in both directions.

2.1 What is the Scope of the Standard?

MPEG-7 will address applications that can be stored (on-line or off-line) or streamed (e.g. broadcast, push models on the Internet), and can operate in both real-time and non real-time environments. A 'real-time environment' means that information is associated with the content while it is being captured.

Figure 1 below shows a highly abstract block diagram of a possible MPEG-7 processing chain, included here to explain the scope of the MPEG-7 standard. This chain includes feature extraction (analysis), the description itself, and the search engine (application). To fully exploit the possibilities of MPEG-7 descriptions, automatic extraction of features (or 'descriptors') will be extremely useful. It is also clear that automatic extraction is not always possible, however. As was noted above, the higher the level of abstraction, the more difficult automatic extraction is, and interactive extraction tools will be of good use. However useful they are, neither automatic nor semi-automatic feature extraction algorithms will be inside the scope of the standard. The main reason is that their standardisation is not required to allow interoperability, while leaving space for industry competition. Another reason not to standardise analysis is to allow making good use of the expected improvements in these technical areas.

Also the search engines will not be specified within the scope of MPEG-7; again this is not necessary, and here too, competition will produce the best results.
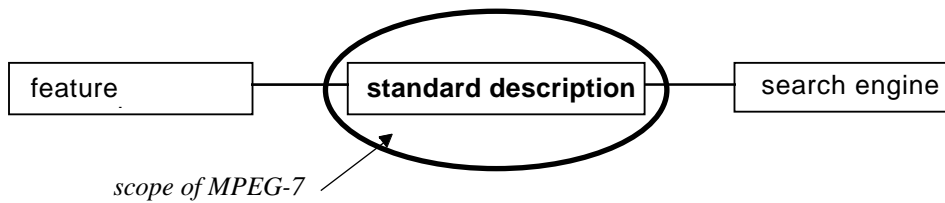
Figure 1: Scope of MPEG-7

To provide a better understanding of the introduced terminology, i.e. Descriptor, Description Scheme, and DDL, please find below Figures 2 – 4. The dotted boxes in the figures encompasses the normative elements of the MPEG-7 standard. Note that the presence of a box or ellipse in one of this drawings does not imply that the corresponding element shall be present in all MPEG-7 applications.

Figure 2 shows the extensibility of the above concepts. Note, the arrows from DDL to DS signify that the DSs are generated using DDL. Furthermore, the drawing reveals the fact that you can build a new DS using an existing DS.
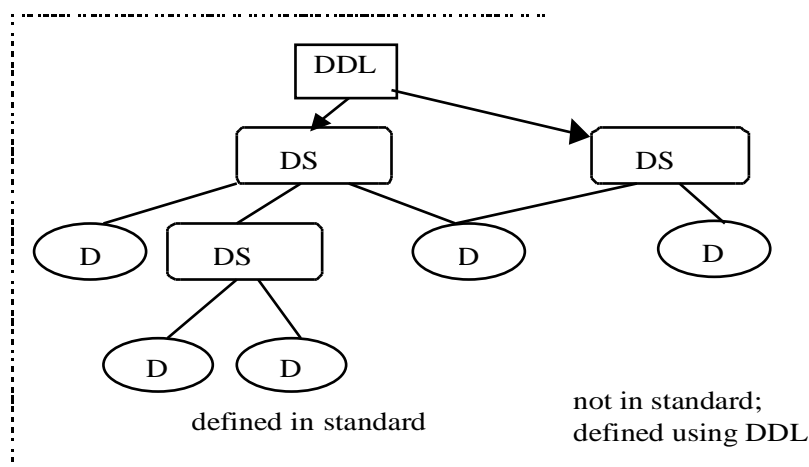
Figure 2: An abstract representation of possible relations between Ds and DSs.

Figure 3 that the DDL provides the mechanism to built a description scheme which in turn forms the basis for the generation of a description. The instantiation of the DS is

4

described as part of Figure 4.

.
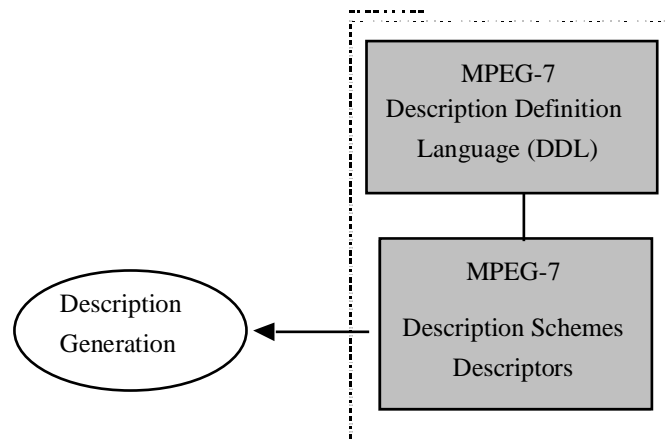
Figure 3: The role of Ds and DSs for the generation of descriptions

Figure 4 explains how MPEG-7 would work in practice. Note: There can be other streams from content to user; these are not depicted here. Furthermore, the use for the encoder and decoder is optional.
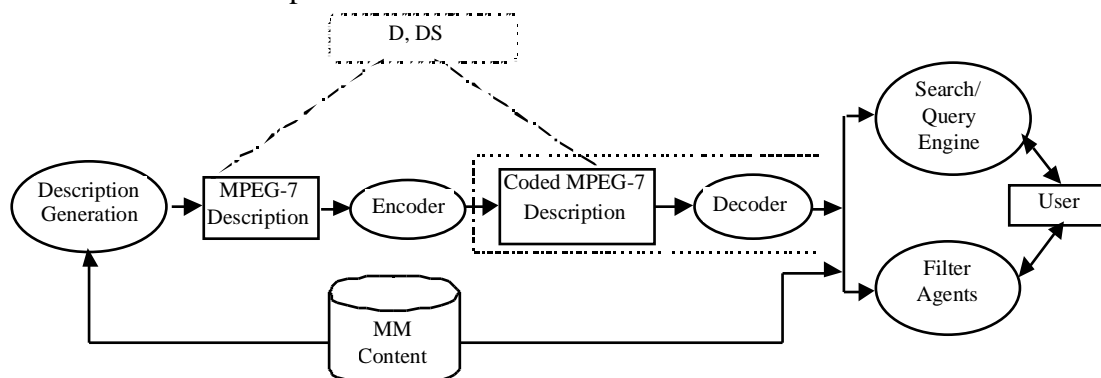
Figure 4: An abstract representation of possible applications using MPEG-7.

The emphasis of MPEG-7 will be the provision of novel solutions for audio-visual content description. Thus, addressing text-only documents will not be among the goals of MPEG-7. However, audio-visual content may include or refer to text in addition to its audio-visual information. MPEG-7, , therefore, will consider existing solutions developed by other standardisation organisations for text only documents and support them as appropriate.

Besides the descriptors themselves, the database structure plays a crucial role in the final retrieval's performance. To allow the desired fast judgement about whether the material is of interest, the indexing information will have to be structured, e.g. in a hierarchical or associative way.

More detailed descriptions of requirements can be found in the 'MPEG-7 Requirements Document' [1].

3. Areas of Interest

There are many applications and application domains which will benefit from the MPEG-7 standard. A few application examples are:

- Digital libraries (image catalogue, musical dictionary,…)
- Multimedia directory services (e.g. yellow pages)
- Broadcast media selection (radio channel, TV channel,…)
- Multimedia editing (personalised electronic news service, media authoring)

The potential applications are spread over the following application domains:
- Education,
- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face),
- Tourist information,
- Cultural services (history museums, art galleries, etc.),
- Entertainment (e.g. searching a game, karaoke),
- Investigation services (human characteristics recognition, forensics),
- Geographical information systems,
- Remote sensing (cartography, ecology, natural resources management, etc.),
- Surveillance (traffic control, surface transportation, non-destructive testing in hostile environments, etc.),
- Bio-medical applications,
- Shopping (e.g. searching for clothes that you like),
- Architecture, real estate, and interior design,
- Social (e.g. dating services), and
- Film, Video and Radio archives.

The way MPEG-7 data will be used to answer user queries is outside the scope of the standard. In principle, any type of AV material may be retrieved by means of any type of query material. This means, for example, that video material may be queried using video, music, speech, etc. It is to the search engine to match the query data and the MPEG-7 AV description. A few query examples are:

1. *Music*
Play a few notes on a keyboard and get in return a list of musical pieces containing (or close to) the required tune or images somehow matching the notes, e.g. in terms of emotions.

2. *Graphics*
Draw a few lines on a screen and get in return a set of images containing similar graphics, logos, ideograms,...

3. *Image*
Define objects, including colour patches or textures and get in return examples among which you select the interesting objects to compose your image.

4. *Movement*
On a given set of objects, describe movements and relations between objects and get in return a list of animations fulfilling the described temporal and spatial relations.

5. *Scenario*

On a given content, describe actions and get a list of scenarios where similar actions happen.

5. *Voice*

Using an excerpt of Pavarotti's voice, and getting a list of Pavarotti's records, video clips where Pavarotti is singing or video clips where Pavarotti is present.

More detailed descriptions of applications can be found in the 'MPEG-7 Applications Document' [2].

4. Method of Work and Work Plan

The method of development is comparable to that of the previous MPEG standards. After defining the requirements (this process has already started), an open Call for Proposals will be issued. The Call will ask for relevant technology fitting the requirements, and after an evaluation of the technology that was received, a choice will be made and development will continue with the most promising submission(s). In the course of developing the standard, additional calls can be issued when not enough technology is present within MPEG to meet the requirements, and there is a reasonable belief that the technology does indeed exist.

As this new MPEG work item will require technology available in technological areas not yet sufficiently represented in the MPEG community, it shall be necessary to seek the collaboration of new experts in the relevant areas. As always, MPEG is open to anyone interested to participate and contribute.

The preliminary work plan for MPEG-7 foresees:
Call for Proposals	October 1998
Working Draft	December1999
Committee Draft	October 2000
Final Committee Draft	February2001
Draft International Standard	July 2001
International Standard	September 2001

More detailed regarding the call for proposals can be found in the 'MPEG-7 Evaluation Document' [3] and the 'MPEG-7 Proposal Package Description (PPD)' [4].

5. Frequently Asked Questions

1. **What is MPEG-7?**

   MPEG-7 will be a standardised description of various types of multimedia information. This description will be associated with the content itself, to allow fast and efficient searching for material that is of interest to the user. MPEG-7 is formally called 'Multimedia Content Description Interface'.

The standard does not comprise the (automatic) extraction of descriptions/features. Nor does it specify the search engine (or any other program) that can make use of the description.

2. From whom or where did the demand for MPEG-7 come?

The demand logically follows the increasing availability of digital audio-visual content. MPEG members recognised this demand, and initiated a new work item. The work on the definition of MPEG-7 has already started to attract new people to MPEG.

3. Why is MPEG-7 needed?

Nowadays, more and more audio-visual information is available, from many sources around the world. Also, there are people who want to use this audio-visual information for various purposes. However, before the information can be used, it must be located. At the same time, the increasing availability of potentially interesting material makes this search more difficult. This challenging situation led to the need of a solution to the problem of quickly and efficiently searching for various types of multimedia material interesting to the user. MPEG-7 wants to answer to this need, providing this solution.

4. Who is currently participating in the development of the MPEG-7 standard?

The people taking part in defining MPEG-7 represent broadcasters, equipment manufacturers, digital content creators and managers, transmission providers, publishers and intellectual property rights managers, as well as university researchers.

5. Where are you in the process of specifying the MPEG-7 standard?

We are in the phase of defining the scope of the standard and its requirements, and the ideas are likely to evolve considerably. Much is still open to input from interested parties, and MPEG is aware that useful work has already been carried out in several areas. The work plan is as follows:

Call for Proposals	*October 1998*
Working Draft	*December1999*
Committee Draft	*October 2000*
Final Committee Draft	*February2001*
Draft International Standard	*July 2001*
International Standard	*September 2001*

6. Will MPEG-7 include audio or video content recognition?

The standardisation of audio-visual content recognition tools is beyond the scope of MPEG-7. Following its principle 'specifying the minimum for maximum usability, MPEG-7 will concentrate on standardising a representation that can be used for description. Development of audio-visual content recognition tools will be a task for industries which will build and sell MPEG-7 enabled products.

In developing the standard, however, MPEG might build some coding tools, just as it did with the predecessors of MPEG-7, namely MPEG-1, -2 and -4. Also for

these standards, coding tools were built for research purposes, but they did not become part of the standard itself.

7. Will MPEG-7 support audio or video content retrieval?

In the same way that MPEG will not standardise the tools to generate the description, MPEG-7 will also not standardise the tools that use the description. It might however be necessary to address the interface between the description and the search engine.

8. What form will the "descriptions" of multimedia content in MPEG-7 take?

The words 'descriptions' or 'features' represent a rich concept, that can be related to several levels of abstraction. Descriptions vary according to the types of data. Furthermore, different types of descriptions are necessary for different purposes of the categorisation.

9. Will the standard allow automatic extraction of descriptions as well as manual entry?

The descriptions that conform to the MPEG-7 standard could be entered by hand, but they could also be automatically extracted. Some features can be best extracted automatically (colour, texture), but for some other features ('this scene contains three shoes and that music was recorded in 1995') this is very hard or even impossible.

10. A 'Call for Proposals', how does that work?

A Call for Proposals (CfP) asks for technology for inclusion in the standard. It is addressed at all interested parties, no matter whether they participate or have participated in MPEG.

MPEG work is usually carried out in two stages, a competitive and a collaborative one. In the competitive stage, participants work on their technology by themselves. In answer to the CfP, people submit their technology to MPEG, after which MPEG makes a fair comparison between the submissions. In MPEG-2 and -4 this was done using subjective tests and additional expert evaluation. How such evaluations will be carried out for MPEG-7 is not yet known, but this will be described in the CfP when it is published in 1998.

Based on the outcome of the evaluation, MPEG will decide which proposals to use for the collaborative stage. In this stage, members of the Experts Group work together on improving and expanding the standard under construction, building on the selected proposals.

Before the final CfP in November 1998, preliminary versions may be published. This is comparable to what happened for MPEG-4.

11. What is the relationship between MPEG-7 and other MPEG activities?

MPEG-7 can be used independently of the other MPEG standards - the description might even be attached to an analog movie. The representation that is defined within MPEG-4, i.e. the representation of audio-visual data in terms of objects, is however very well suited to what will be built on the MPEG-7 standard. This

representation is basic to the process of categorisation. In addition, MPEG-7 descriptions could be used to improve the functionality of previous MPEG standards.

12. If I want to get involved in MPEG-7, what do I need to know about the other MPEG standards?

In principle, knowledge about the other three MPEG standards is not required for taking part in the MPEG-7 work. However, since some of MPEG-7's tools may be close to those of MPEG-4, some knowledge about them could be useful.

13. If I want to know more about the other MPEG standards, where do I look?

You can start by taking a look at MPEG's home page (http://www.cselt.it/mpeg/) which contains many useful references, including more lists with "Frequently Asked Questions" about MPEG activities.

14. So what happened to MPEG-5 and -6? (And how about 3?)

MPEG-3 existed once upon a time, but its goal, enabling HDTV, could be accomplished using the tools of MPEG-2, and hence the work item was abandoned. So after 1,2 and 4, there was much speculation about the next number. Should it be 5 (the next) or 8 (creating an obvious binary pattern)? MPEG, however, decided not to follow either logical expansion of the sequence, but chose the number of 7 instead. So MPEG-5 and MPEG-6 are, just like MPEG-3, not defined.

15. When will MPEG-7 replace the existing MPEG-1 and MPEG-2 standards?

MPEG-7 will not replace MPEG-1 MPEG-2 or in fact MPEG-4 it is intended to provide complementary functionality to these other MPEG standards: representing information about the content, not the content itself ("the bits about the bits") This functionality is the standardisation of multimedia content descriptions.

16. If I want to know more about, be involved in, or give an input to the MPEG-7 development process, whom should I contact?

You can contact any of the people listed below with their email addresses and telephone numbers. To visit MPEG meetings you need to be on your national delegation, but the people listed below can explain how this works.

Rob Koenen (KPN Research - the Netherlands / chairman MPEG Requirements)
r.h.koenen@research.kpn.com   +31 70 332 5310

Sylvie Jeannin (Philips Research – US / evaluation contact)
sjn@philabs.research.philips.com

Fernando Pereira (Instituto Superior Técnico - Portugal)
fp@lx.it.pt   +351 1 8418460

Ibrahim Sezan (Sharp Labs - USA)
sezan@sharplabs.com   +1 360 817 8401

Adam Lindsay (Riverland, Belgium/ audio contact)
adam@riv.be   + 32 2 721 5454

Frank Nack (GMD-IPSI – Germany / requirements contact)
nack@darmstadt.gmd.de   + 49 6151 869833

Seungyup Paek (Columbia University – US / test material contact)
Syp@ee.columbia.edu   +1 212 854 7447

V.V.Vinod (Kent Ridge digital Labs – Singapore / PPD contact)
vinod@krdl.org.sg   +65 874 5225

References

[*1*]   *MPEG Requirements Group, "MPEG-7Requirements Document", Doc. ISO/MPEG N2461, MPEG Atlantic City Meeting, October 1998*

[*2*] *MPEG Requirements Group, "Applications for MPEG-7", Doc. ISO/MPEG N2462, MPEG Atlantic City Meeting, October 1998*

[*3*]   *MPEG Requirements Group, " MPEG-7Evaluation Procedure", Doc. ISO/MPEG N2463, MPEG Atlantic City Meeting, October 1998*

[*4*]   *MPEG Requirements Group, "*MPEG-7 Proposal Package Description (PPD)*", Doc. ISO/MPEG N2464, MPEG Atlantic City Meeting, October 1998*