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

ISO/IEC JTC1/SC29/WG11 **N4675**

Jeju, March 2002

Title:	**Introduction to MPEG-7 (v4.0)**
Source:	**Requirements**
Status:	**Approved**
Editors:	**Neil Day (Bluemetrix, JP), José M. Martínez (UPM-GTI, ES)**

1. Introduction to MPEG-7

How many times have you seen science fiction movies such as 2001: A Space Odyssey and thought, "Wow, we're so far away from having any of the fancy gadgets depicted in these movies!" In 2001, Hal, the talking computer intelligently navigates and retrieves information or runs complex operations instigated by spoken input. Or how about using an image-based query, say an image of the motorbike used by Arnold Schwartzenegger in the movie T2, to find images of similarly looking motorbikes. Dreams or reality?

As more and more audiovisual information becomes available from many sources around the world, many people would like to use this information for various purposes. This challenging situation led to the need for a solution that quickly and efficiently searches for and/or filters various types of multimedia material that's interesting to the user.

For example, finding information by rich-spoken queries, hand-drawn images, and humming improves the user-friendliness of computer systems and finally addresses what most people have been expecting from computers. For professionals, a new generation of applications will enable high-quality information search and retrieval. For example, TV program producers can search with "laser-like precision" for occurrences of famous events or references to certain people, stored in thousands of hours of audiovisual records, in order to collect material for a program. This will reduce program production time and increase the quality of its content.

MPEG-7 is a multimedia content description standard, (), that addresses how humans and computer systems expect to interact with computer systems, since it develops rich descriptions that reflect those expectations. This document gives an introductory overview of the MPEG-7 standard. More information about MPEG-7 can be found at the MPEG website (http://mpeg.telecomitalialab.com/) and the MPEG-7 Alliance website **http://www.mpeg-alliance.com** These web pages contain links to a wealth of information about MPEG, including many publicly available documents, several lists of 'Frequently Asked Questions' and links to other MPEG-7 web pages.

Formatted

1.1 What Are the MPEG Standards?

The Moving Picture Coding Experts Group (MPEG) is a working group of the Geneva-based ISO/IEC standards organization, (International Standards Organization/International Electro-technical Committee) in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio, and a combination of the two. MPEG-7 then is an ISO/IEC standard being developed by MPEG, the committee that also developed the Emmy Award-winning standards known as MPEG-1 and MPEG-2, and the 1999 MPEG-4 standard.

- *MPEG-1:* For the storage and retrieval of moving pictures and audio on storage media.
- *MPEG-2:* For digital television, it's the timely response for the satellite broadcasting and cable television industries in their transition from analog to digital formats.
- *MPEG-4:* Codes content as objects and enables those objects to be manipulated individually or collectively on an audiovisual scene.

MPEG-1, -2, and -4 make content available. MPEG-7 lets you to find the content you need.

Besides these standards, MPEG is currently also working in MPEG-21 a Technical Report about Multimedia Framework.

1.2 Defining MPEG-7

MPEG-7 is a standard for describing features of multimedia content.

1.2.1 Qualifying MPEG-7

MPEG-7 provides the world's richest set of audio-visual descriptions.

These descriptions are based on catalogue (e.g., title, creator, rights), semantic (e.g., the who, what, when, where information about objects and events) and structural (e.g., the colour histogram - measurement of the amount of colour associated with an image or the timbre of an recorded instrument) features of the AV content and leverages on AV data representation defined by MPEG-1, 2 and 4.

Comprehensive Scope of Data Interoperability.

MPEG-7 uses XML Schema as the language of choice for content description MPEG-7 will be interoperable with other leading standards such as, SMPTE Metadata Dictionary, Dublin Core, EBU P/Meta, and TV Anytime.

1.3 The Key Role of MPEG-7

MPEG-7, formally named "Multimedia Content Description Inter-face," is the standard that describes multimedia content so users can search, browse, and retrieve that content more efficiently and effectively than they could using today's mainly text-based search engines. It's a standard for describing the features of multimedia content. *However…*

MPEG-7 will not standardize the (automatic) extraction of AV descriptions/features. Nor will it specify the search engine (or any other program) that can make use of the description. It will be left to the creativity and innovation of search engine companies, for example, to manipulate and massage the MPEG-7-described content into search indices that can be used by their browser and retrieval tools, (*see figure 1*).

2. MPEG-7 Technical Activities

It is important to note that MPEG-7 addresses many different applications in many different environments, which means that it needs to provide a flexible and extensible framework for describing audio-visual data. Therefore, MPEG-7 will define a multimedia library of methods and tools. It will standardize:

• *A set of descriptors:* A descriptor (D) is a representation of a feature that defines the syntax and semantics of the feature representation.

• *A set of description schemes:* A description scheme (DS) specifies the structure and semantics of the relationships between its components, which may be both descriptors and description schemes.

• *A language that specifies description schemes, the Description Definition Language (DDL):* It also allows for the extension and modification of existing description schemes. MPEG-7 adopted XML Schema Language as the MPEG-7 DDL. However, the DDL requires some specific extensions to XML Schema Language to satisfy all the requirements of MPEG-7. These extensions are currently being discussed through liaison activities between MPEG and W3C, the group standardizing XML.

• *One or more ways (textual, binary) to encode descriptions:* A coded description is a description that's been encoded to fulfill relevant requirements such as compression efficiency, error resilience, and random access.
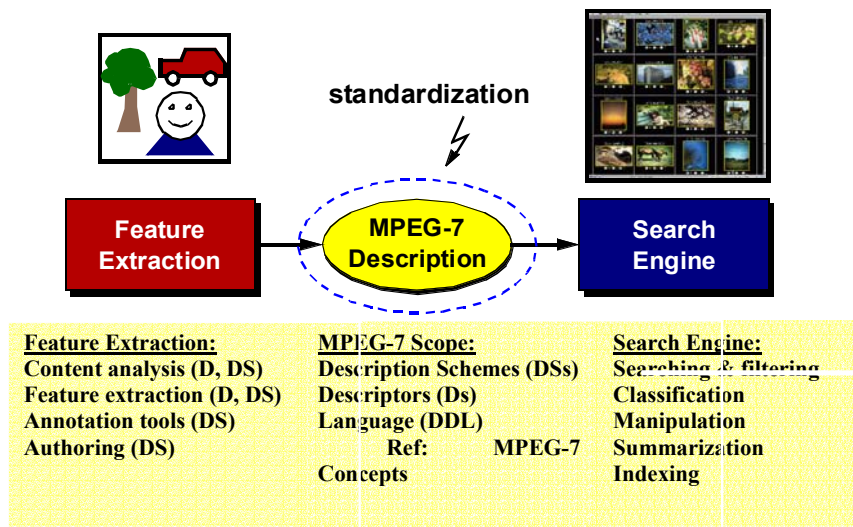

standardization

| Feature Extraction | MPEG-7 Description | Search Engine |

Feature Extraction:	**MPEG-7 Scope:**	**Search Engine:**
Content analysis (D, DS)	Description Schemes (DSs)	Searching & filtering
Feature extraction (D, DS)	Descriptors (Ds)	Classification
Annotation tools (DS)	Language (DDL)	Manipulation
Authoring (DS)	Ref: MPEG-7	Summarization
	Concepts	Indexing

Figure 1: The Scope of MPEG-7

2.1 Organization of MPEG-7 Description Tools

Over 100 MPEG-7 Description Tools have been developed and refined. The relationships between the MPEG-7 Description Tools are outlined in Figure 2.

The basic elements, at the lower level, deal with basic data types, mathematical structures, schema tools, linking and media localization tools, as well as basic DSs, which are elementary components of more complex DSs. The Schema tools section specifies elements for creating valid MPEG-7 schema instance documents and description fragments.

In addition, this section specifies tools for managing and organizing the elements and datatypes of the schema.

3

Based on this lower level, content description and management elements can be defined. These elements describe the content from several viewpoints. Currently five viewpoints are defined: creation and production, media, usage, structural aspects, and conceptual aspects. The first three elements primarily address information that's related to the management of the content (content management), whereas the last two are mainly devoted to the description of perceivable information (content description).
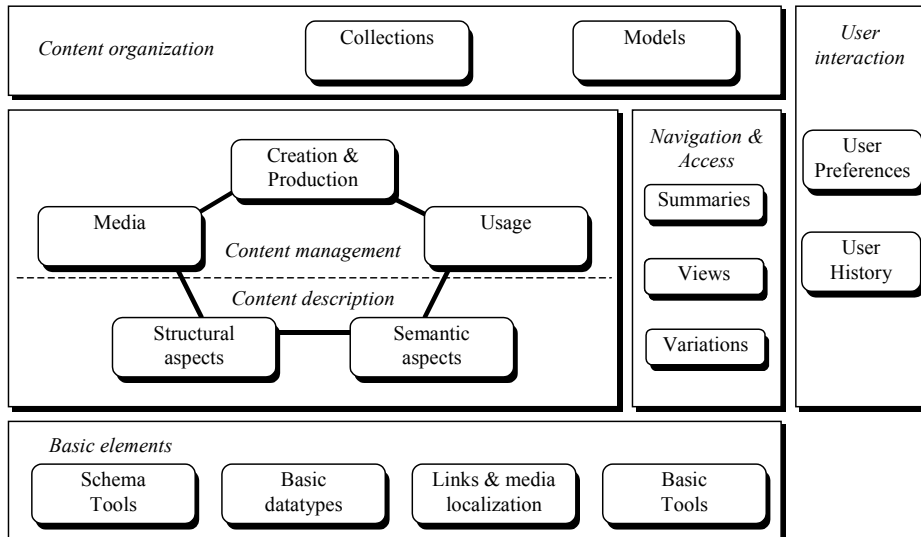

Figure 2: Overview of MPEG-7 Multimedia Description Schemes (DSs)

• **Creation and Production:** Contains meta information that describes the creation and production of the content; typical features include title, creator, classification, and purpose of the creation. Most of the time this information is author-generated since it can't be extracted from the content.

• **Usage:** Contains meta information that's related to the usage of the content; typical features involve rights holders, access rights, publication, and financial information. This information may be subject to change during the lifetime of the AV content.

• **Media:** Contains the description of the storage media; typical features include the storage format, the encoding of the AV content, and elements for the identification of the media. *Note:* Several instances of storage media for the same AV content can be described.

• **Structural aspects:** Contains the description of the AV content from the viewpoint of its structure. The description is structured around segments that represent physical, spatial, temporal, or spatio-temporal components of the AV content. Each segment may be described by signal-based features (color, texture, shape, motion, audio) and some elementary semantic information.

• **Conceptual Aspects:** Contains a description of the AV content from the viewpoint of its conceptual notions.

The five sets of Description Tools are presented here as separate entities, however, they are interrelated and may be partially included in each other. For example, Media, Usage or Creation & Production elements can be attached to individual segments involved in the structural description of the content. Tools are also defined for

navigation and access and there is another set of tools for *Content organization* which addresses the organization of content by classification, by the definition of collections and by modeling. Finally, the last set of tools is *User Interaction* which describes user's preferences for the consumption of multimedia content and usage history.

2.2 MPEG-7 Working Groups

Currently MPEG-7 concentrates on the specification of description tools (Descriptors and Description Schemes), together with the development of the MPEG-7 reference software, known as XM (eXperimentation Model). The XML Schema Language was chosen as the base for the Description Definition Language (DDL).

The **MPEG-7 Audio group** develops a range of Description Tools, from generic audio descriptors (e.g., waveform and spectrum envelopes, fundamental frequency) to more sophisticated description tools like Spoken Content and Timbre. Generic Audio Description tools will allow the search for similar voices, by searching similar envelopes and fundamental frequencies of a voice sample against a database of voices. The Spoken Content Description Scheme (DS) is designed to represent the output of a great number of state of the art Automatic Speech Recognition systems, containing both words and phonemes representations and most likely transitions. This alleviates the problem of out-of-vocabulary words, allowing retrieval even when the original word was wrongly decoded. The Timbre descriptors (Ds) describe the perceptual features of instrument sound, that make two sounds having the same pitch and loudness appear different to the human ear. These descriptors allow searching for melodies independently of the instruments.

The **MPEG-7 Visual group** is developing four groups of description tools: Color, Texture, Shape and Motion. Color and Texture Description Tools will allow the search and filtering of visual content (images, graphics, video) by dominant color or textures in some (arbitrarily shaped) regions or the whole image. Shape Description Tools will facilitate "query by sketch" or by contour similarity in image databases, or, for example, searching trademarks in registration databases. Motion Description Tools will allow searching of videos with similar motion patterns that can be applicable to news (e.g. similar movements in a soccer or football game) or to surveillance applications (e.g., detect intrusion as a movement towards the safe zone).

The **MPEG-7 Multimedia Description Schemes group** is developing the description tools dealing with generic and audiovisual and archival features. Its central tools deal with content management and content description as outlined in section 2.1.

The **MPEG-7 Implementation Studies group** is designing and implementing the MPEG-7 Reference Software known as XM.

The **MPEG-7 Systems group** is developing the DDL and the binary format (known as BiM), besides working in the definition of the terminal architecture and access units.

3. MPEG-7 Application Domains

The elements that MPEG-7 standardizes will support a broad a range of applications (for example, multimedia digital libraries, broadcast media selection, multimedia editing, home entertainment devices, etc.). MPEG-7 will also make the web as searchable for multimedia content as it is searchable for text today. This would apply especially to large content archives, which are being made accessible to the public, as well as to multimedia catalogues enabling people to identify content for purchase. The information used for content retrieval may also be used by *agents*, for the selection and filtering of broadcasted "push" material or for personalized advertising. Additionally, MPEG-7 descriptions will allow fast and cost-effective usage of the underlying data, by enabling

semi-automatic multimedia presentation and editing. All domains making use of multimedia will benefit from MPEG-7 including,

- Digital libraries, Education (image catalogue, musical dictionary, Bio-medical imaging catalogues…)
- Multimedia editing (personalised electronic news service, media authoring)
- Cultural services (history museums, art galleries, etc.),
- Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems)
- Broadcast media selection (radio channel, TV channel,…)
- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face),
- E-Commerce (personalised advertising, on-line catalogues, directories of e-shops,…)
- Surveillance (traffic control, surface transportation, non-destructive testing in hostile environments, etc.),
- Investigation services (human characteristics recognition, forensics),
- Home Entertainment (systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke,…)
- Social (e.g. dating services),

3.1 Typical applications enabled by MPEG-7 technology include:

• *Audio:* I want to search for songs by humming or whistling a tune or, using an excerpt of Pavarotti's voice, get a list of Pavarotti's records and video clips in which Pavarotti sings or simply makes an appearance. Or, play a few notes on a keyboard and retrieve a list of musical pieces similar to the required tune, or images matching the notes in a certain way, e.g. in terms of emotions.

• *Graphic*s**:** Sketch a few lines on a screen and get a set of images containing similar graphics, logos, and ideograms.

• *Image:* Define objects, including color patches or textures, and get examples from which you select items to compose your image. Or check if your company logo was advertised on a TV channel as contracted.

• *Visual:* Allow mobile phone access to video clips of goals scored in a soccer game, or automatically search and retrieve any unusual movements from surveillance videos.

• *Multimedia:* On a given set of multimedia objects, describe movements and relations between objects and so search for animations fulfilling the described temporal and spatial relations. Or, describe actions and get a list of scenarios containing such actions.

3.2 Example of an MPEG-7 Application

The following application is an examples of the type of solutions that MPEG-7 can solve. There are many more applications being developed around the world, (see the MPEG-7 Alliance website, www.mpeg-industry.com.

3.2.1 Content Retrieval using Image as the Query

Figure 3 shows possible ways to search for visual content using the inherent structural features of an image. In this example there are four image features detailed. The color histogram feature (1) of an image allows me to search for images that have the same color. Note, the position of the colors is not important but rather the amount of similar color in the image is important. The next feature, spatial color distribution (2) allows me to search for

images where the location of the same color is important. You can see that the added object in the right-bottom flag does not affect this type of search. You can additionally search for images that have a similar edge or contour profile as in the spatial edge distribution (3) search technique. Note, color does not make a difference to this type of search. Finally, you can see an example of searching by object shape (4). Here, the color and edge profiles are not important.
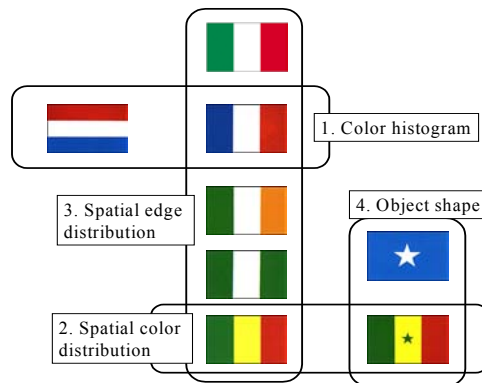

Figure 3: Search using image features

4. Validation Tools in MPEG-7

4.1 XSV (TBD)

4.2 XERCES (TBD)

Validation of the MDS tools can be with the SAX2 library from the Apache's Xerces-J package, which supports the W3C Schema 1.0 released in May 2, 2001.

4.3 MPEG-7 Validation Service: Interoperability Test Bed (http://m7itb.nist.gov/M7Validation.html)

In order to achieve true interoperability between MPEG-7 standard based applications, an MPEG-7 Interoperability Test Bed (M7ITB) is needed. The M7ITB serves: a) to validate MPEG-7 DSs and Ds of Multimedia Description Scheme and Description Definition Language (DDL) from the syntactic and semantic structure point of view, b) to conduct conformance and interoperability testing between MPEG-7 encoded metadata instances, and c) to verify applicability of MPEG-7 technology in a real application or production environment.

5. MPEG-7 in the 21st Century Media Landscape

MPEG-7 is about the future of media in the 21st century. This is not an overstatement. MPEG-7 provides a comprehensive and flexible framework for describing the content of multimedia. To describe content implies knowledge of elements it consists of, as well as, knowledge of interrelations between those elements. The most straightforward application is multimedia management, where such knowledge is prerequisite for efficiency and accuracy. However, there are other serious implications. Knowledge of the structural features of multimedia

information as well as its semantic features will help generate solutions that will provide more comprehensive and accurate indexing and search applications, (leading to greater ability for content manipulation, content reuse - and thus new content creation). Many issues, it is true, remain including copyrights issues and interoperability between applications and systems that wish to adhere to the MPEG-7 standard. But such issues are balanced by incredible economical, educational, and ergonomic benefits that will be brought by MPEG-7 technology. Potential concerns will be resolved, and after some years hence, we will not be able to imagine media without MPEG-7 technologies.

6.Advantages of MPEG-7 – A Summary

1. MPEG Standards have been Successful in the Marketplace.

MPEG standards, to date, have been extremely successful in the marketplace due to MPEG's unique process of sharing leading-edge technology while protecting intellectual property. MPEG-7 builds on the demonstrated success of this process.

2.   Taking Advantage of MPEG-7 Expertise

The contributors to MPEG-7 include experts in every portion of the content value chain: production, post-production, delivery, and consumption. Through this process MPEG-7 has standardized description schemes for content description, management, and organization, as well as navigation, access, user preferences and usage history.

3.   Interoperability: rapid uptake of MPEG-7, as it is built on enabling technologies and standards.

MPEG-7 is harmonizing with has employed (or has otherwise harmonized with) other standards that have demonstrated success and acceptance in both traditional media and new media businesses, e.g., W3C (XML, XML Schema), IETF (URI, URN, URL), Dublin Core, ISO/ANSI Thesaurus guidelines, SMPTE Metadata Dictionary, TV-Anytime, etc. This will allow rapid integration into your company's products

ATSC example: *MPEG-7 brings ubiquitous MPEG-2 platforms new EPG functionality*

Various MPEG-2 solutions are looking to MPEG-7 as the standard for specifying additional program metadata. For example, ATSC recently announced an "RFP for Metadata for Advance Electronic Program Guide (EPG) Functionality". This request for proposal applies to a protocol for the carriage of metadata in an ATSC MPEG-2-based digital television broadcast stream to support advanced EPG functionality in a DTV receiver. The RFP explicitly states that "it is highly desirable that any ATSC standard for enhanced metadata to support advanced EPG features should be harmonized with other standards efforts, such as MPEG-7…"

(*Request for Proposal For Potential Revisions to ATSC Standards in the Area of Metadata for Advanced EPG Functionality*, ATSC T3/S8 Doc 427, 29 January 2001, http://www.atsc.org/T3S8_Adv_EPG_RFP.pdf)

4.   Data Exchange between Subsidiaries

MPEG-7 will enable the content management system at one subsidiary to leverage the content of another subsidiary. For example, broadcasters and content producers could categorize, exchange, process and manage assets across boundaries or along an entire supply chain, e.g., from production house to advertising agency to a broadcaster's network.

5.   Market Potential for MPEG-7 Applications

According to a Goldman Sachs projection, the market for content management tools will grow from US$378Million in 2000 to US$4.5Billion by 2005. Interoperable tools sell better than non-interoperable tools. MPEG-7 is the gold standard for content management interoperability, not just entertainment companies - but every company, every industry, everywhere.

6. MPEG-7 will enable a New Generation of Multimedia Applications

MPEG-7 uniquely provides comprehensive standardised multimedia description tools for content. Descriptions for the catalogue level (e.g. title), the semantic level (who, what, when, where) and the structural level (spatio-temporal region, color histogram, timbre, texture) will provide tools for creative developers to generate new waves of multimedia applications. Standardized MPEG-7 description tools, then, are a key enabler of the following application domains:

- Search Engines, Digital Libraries, Broadcast Networks, Entertainment and News Distributors, Streaming Businesses
- Dynamic start-up companies, searching for cutting edge technologies.
- Governmental, Educational, Law, Medical & Remedial Services, and Non-profit organizations looking for digital media solutions. For example, the U.S. Library of Congress receives over 10,000 multimedia items each week, and is committed to a) the long term preservation of these multimedia items in digital format, and b) making much of their collection accessible to U.S. citizens in digital format.
- XML, Metadata, Modeling/Simulation, & Surveillance Industries
- AI Practitioners, Content Creators and Providers.

7. MPEG-4 and MPEG-7 Tools for Killer Applications

With MPEG-7's sister standard, MPEG-4, an ideal combination is made for solutions that require efficient streaming of content, content manipulation, and indexing and retrieval of that content. In particular mobile application developers have already begun to use these two standards and the trend is set to hugely increase as the demand for visual and audio information services continues to grow.

8. MPEG-7 Intellectual Property and Management Protection

MPEG works closely with representatives of the creative industries to ensure that the best possible protection of the rights of stakeholders is maintained both in content and in metadata

9. MPEG-7 Makes Content More Valuable

Stored audio-visual content, gathered over the years, by broadcasters, libraries, and publishers becomes more valuable because, with MPEG-7 indexing technology, more comprehensive methods are available for users to access and retrieve more detailed descriptions of that content.

10. MPEG-7 provides a seamless path towards increasingly intelligent content management systems

We live in the age of convergence, from the level of production through to distribution and consumption. The technical hardware and communication infrastructure is evolving and will soon reach the point where computing and communications will become embedded in everyday objects and environments. Media will also then become ubiquitous. Ubiquitous media will create a huge demand for new content, and meeting this demand must involve fundamental changes to all stages of media production, management and delivery.

Media archives will become vast and interconnected pools of content, too large to be managed manually. Customization of content within programs, e.g. substitution of structural elements (characters, music, voices) according to viewer desires, content scaling for PDA, cell phones, will be not only possible, but easy and

pleasant. MPEG-7 will enable the creation of tools, (through its structured combination of low level features and high-level meta-data), for coping with this "outbreak" of generic content.

7. REFERENCES

MPEG-7 Documents and Software Reference (Experimental Model, XM)

There are a number of public documents, describing MPEG-7 technology in much greater detail, available at the MPEG website (Information more focused to industry, including a publicly accessible news mailing list, is also available at the MPEG-7 Alliance website, (**http://www.mpeg-industry.com**). At the latter site, tutorial overviews on MPEG-7 audio, visual, systems and the DDL technologies are available. Links to the MPEG-7 XM are accessible from the web sites allowing review and testing of the XM, (both Linux and Windows NT versions of the software exist).

Related MPEG-7 summary documents include,

> MPEG-7 (Technical) Overview, W: Jeju, Korea, March, 2002
>
> MPEG-7 Projects and Demos, W: Jeju, Korea, March, 2002

MPEG-7 Awareness Events (www.mpeg-industry.com)

A number of MPEG-7 Awareness Events have been held (Paris, Oct. 2000; Singapore, March 2001) highlighting MPEG-7 technology and demos. The aim of the Awareness Events is to encourage industry to interact with the MPEG-7 development community so that a standard is developed that closely reflects the needs of the multimedia content market. The next MPEG-7 Awareness Event will be held in the USA, May 3/4, 2002.

MPEG-7 Personal Contacts

MPEG Convenor: Leonardo.Chiariglione (Leonardo.Chiariglione@CSELT.IT)

For Technical Issues:

> *Requirements*:   Fernando Pereira (fp@lx.it.pt)
>
> *Audio*:   Juergen Herre (hrr@iis.fhg.de)
>
> *Visual*:   Miroslaw Bober (miroslaw.bober@vil.ite.mee.com)
>
> *Multimedia DS*:   John Smith (jsmith@us.ibm.com)
>
> *Systems*:   Olivier Avaro (olivier.avaro@francetelecom.fr)
>
> *XM Software Implementation*: Stephan Herrmann (stephanh@lis.e-technik.tu-muenchen.de)

MPEG-7Alliance: Neil Day (neil@mpeg-industry.com),

Editors: Neil Day, (neil@mpeg-industry.com), José M. Martínez (jms@gti.ssr.upm.es)

Deleted: http://www.cselt.it/mpeg/).