MPEG Audio FAQ

MPEG-7: description of meta-information on sound

Overview of MPEG-7 Audio

What does MPEG-7 standardize?

It is officially called "Multimedia Content Description Interface", a means of providing meta-data for multimedia. MPEG-7 focusses on two different work items in this context:

- standardisation of a language to specify descriptions for multimedia data (the Definition Description Language, or "DDL") and
- standardisation of a set of obviously required description components for multimedia data (the Descriptors, "Ds" and Description Schemes "DSs").

For the second work item, MPEG-7 is not restricting itself to traditional "card catalogue" descriptions (e.g. Dublin Core) of video and/or audio material but also develops descriptions that require (human or automatic) content analysis of multimedia material and are able to grasp information on structural and semantic content. MPEG-7 does not standardize extraction algorithms but only focusses on the specifications of outcomes of such algorithms. However, in developing the standard, MPEG will develop tools for creating descriptions to prove their effectiveness, and these tools may be extraction algorithms. They will not be part of the standard, or "normative," however.

What is the relationship to previous MPEG efforts?

Where MPEG-1 and -2 concentrated almost entirely on compression, MPEG-4 moved to a higher level of abstraction in coding objects and using content-specific techniques for coding content. MPEG-7 moves to an even higher level of abstraction, a cognitive coding, some might say.
In principle, MPEG-1, -2, and -4 are designed to represent the information itself, while MPEG-7 is meant to represent information about the information (although there are areas common between MPEG-4 and -7). Another way of looking at it is that MPEG-1, -2, and -4 made content available. MPEG-7 allows you to describe and thus find the content you need.

What are the general applications for MPEG-7?

MPEG-7 aims at making the web more searchable for multimedia content than it is for text today. This also applies to making large content archives accessible to the public (or to enable people to identify content to buy). A big issue here is a standardized set of descriptors for the same type of data within every content archive such that the same tools are able to access a lot of different collections

in the same way. The information used for content retrieval may also be used by agents for selection and filtering of broadcast material. Additionally, the meta-data may be used for more advanced access to the underlying data, by enabling automatic or semi-automatic multimedia presentation or editing.

In summary, example application areas of MPEG-7 audio are:

- Setting-up audio archives (radio, movies, tv)
- Retrieving audio files from the Internet/from an archive
- Filtering audio broadcasts
- Music education
- Surveillance

What are specific functionalities foreseen for MPEG-7 audio? (or, what can I tell my organization MPEG-7 audio is good for?)

Although still an expanding list, we can envision indexing music, sound effects, and spoken-word content in the audio-only arena. MPEG-7 will enable query-by-example such as query-by-humming. In addition, audio tools play a large role in typical audio-visual content in terms of indexing film soundtracks and the like. If someone wants to manage a large amount of audio content, whether selling it, managing it internally, or making it openly available to the world, MPEG-7 is potentially the solution.

One specific application area supported by MPEG-7 is that of automatic speech recognition (ASR). A spoken content description scheme will embody the output of any state of the art ASR system, at a host of different semantic levels. This will allow searching on spoken and visual events: "Find me the part where Romeo says `It is the East and Juliet is the sun'".

For more details, please see the MPEG-7 Applications Document (N2426) for a non-exhaustive list of representative examples.

MPEG-7 language and abbreviations

What do the abbreviations used within MPEG-7 audio stand for?

DDL = Description Definition Language
D = Descriptor
DS = Description Scheme
XM = Experimentation Model
CE = Core Experiment

Technicalities of MPEG-7

What are the foreseen elements of MPEG-7?

MPEG-7 work is currently seen as being in three parts: Descriptors (D's), Description Schemes (DS's), and a Description Definition Language (DDL). Each is equally crucial to the entire MPEG-7 effort.

1. **Descriptors** are the representations of <u>low-level features,</u> the fundamental qualities of audiovisual content which may range from statistical models of signal amplitude, to fundamental frequency of a signal, to an estimate of the number of sources present in a signal, to spectral tilt, to emotional content, to an explicit sound-effect model, to any number of concrete or abstract features. This is the place where the most involvement from the signal processing community is foreseen. Note that not all of the descriptors need to be automatically extracted--the essential part of the standard is to establish a normalized representation and interpretation of the Descriptor.

2. **Description Schemes** are <u>structured combinations of Descriptors.</u> This structure may be used to annotate a document, to directly express the structure of a document, or to create combinations of features which form a richer expression of a higher-level concept. For example, a radio segment DS may note the recording date, the broadcast date, the producer, the talent, and include pointers to a transcript. A classical music DS may encode the musical structures (and allow for exceptions) of a Sonata form. Various spectral and temporal Descriptors may be combined to form a DS appropriate for describing timbre or short sound effects.

3. The **Description Definition Language** is to be the mechanism which allows a great degreed flexibility to be included in MPEG-7. Not all documents will fit into a prescribed structure. There are fields (e.g. biomedical imagery) which would find the MPEG-7 framework very useful, but which lie outside of MPEG's scope. A solution provider may have a better method for combining MPEG-7 Descriptors than a normative description scheme. The DDL is to address all of these situations.

All of these elements combine to create a **Description**, or an instantiated DS, which is represented using the DDL, and which incorporates Ds as appropriate to the data model.

How will MPEG-7 descriptions originate? (Aka "Aren't media producers too busy to spend time annotating content?")

We see descriptors as coming from the following possible places:

- Preservation of existing descriptive data (e.g. scripts) through the production/delivery chain
- Generated automatically by capture devices (e.g. time or GPS location in a camera)
- Extracted automatically & semiautomatically (i.e. with some human assistance)
- Manually produced (e.g. for legacy material such as existing film archives)

Ultimately, we recognise that there is more than one way to accomplish the feat of content description, and expect users to adopt the solutions matching their own needs.

What are Core Experiments?

Core experiments (CEs) are the procedure by which a description proposal makes its way into the standards document and reference software for the evolving standard. It is where multiple parties both vie for position and collaborate to find the best technology for inclusion in the standard. There will be a set of procedures that have to be followed for execution of a CE.

What is the XM? What is the development environment?

The eXperimentation Model (XM) is the development base for the Core Experiments and is the reference software for the evolving standard. The development environment for Windows is Visual Studio v5.0. *The development environment for Linux has yet to be fully defined.*

What audio description tools are currently in the Working Draft or are being worked on?

The only audio description tool currently included in the WD are the speech recognition description tools. They consist of combined word and phoneme lattices and may additionally store topic labels, predefined taxonomy or textual summaries related to the ASR output stored in the lattice. Other audio description tools being at CE status are:

- **Instrument description tools**, which will enable "query by example" search based on instrument timbre similarity.
- **Sound effects description tools**, which will enable sound classification based on a parameter space.

In addition, there are generic description tools in the WD which are highly relevant to audio descriptions. Examples are:

- the SegmentDS with an AudioSegmentDS as subclass; it describes the physical structure and the signal properties of an audio document and can form segment trees and define the structure of an audio document.
- the SemanticDS which can define a semantic index table to an audio document. It may consist of EventDSs which are segments with a temporal dimension like e.g. music segments, speech segments, sound effects segments, news items etc.
- the MediaIdentificationDS which contains descriptions of the storage format of the media in question.

Relation to other standards / methods

What are potential connections between MPEG-7 tools and MPEG-4 tools?

There are many possible connections between MPEG-4 tools and MPEG-7. Most of the content-

specific tools contained in MPEG-4 have great potential because a model for the content is already specified: by choosing a method of coding, one selects the features that are important to the material. For example, if one encodes a sound by using sinusoidal tracks, then MPEG-7 asks which of those tracks are most significant in distinguishing the sound. It is a matter of abstraction up to the point of measuring similarity.

The Structured Audio tools also have a strong relationship to MPEG-7. They synthesize a sound from an already-existing description. The challenge in this case is to reach a suitable level of abstraction. There can be many, very different, descriptions which can be synthesized into perceptually indistinguishable sounds. It is clear that the models (e.g. of a musical instrument or an acoustic space) used directly within structured audio will not be sufficiently abstract or constrained. Therefore, it is an open research question as to how to identify, select, and build a structured audio model for MPEG-7.

MPEG-4 also talks about "description" of media. What is the relation to MPEG-7?

MPEG-4 is concerned with generative descriptions - ie the descriptions consist of a more or less complete set of instructions from which a system may generate a piece of content. In other words, an MPEG-4 description is essentially a compressed form of the content.

By contrast, MPEG-7 descriptions are (in general) non-generative - they are designed for other purposes (search, etc) and do not provide complete desriptions from which the media experience may be created. MPEG-7 descriptions exist alongside the media and provide information about it ("bits about the bits").

It is planned that there will be some object content identifiers and other simple meta-data included in MPEG-4. However, their capabilities will be severely limited compared to MPEG-7.

Bibliographic references

Can you propose more detailed information in the literature?

- ISO/IEC 15938 Working Draft, 2000.
- F. Nack, A. Lindsay: "Everything you wanted to know about MPEG-7", Part 1; in IEEE Multimedia Vol. 6, No. 3, 1999, pp.65-77.
- F. Nack, A. Lindsay: "Everything you wanted to know about MPEG-7", Part 2; in IEEE Multimedia Vol. 6, No. 4, 1999, pp.64-73.
- joint paper in Image Comm J. in press
- J. New Music Research in press

Organisational details

What is the current time frame for MPEG-7?

- **December, 1999:** MPEG-7 Working Draft
- **October, 2000:** MPEG-7 Committee Draft
- **February, 2001:** MPEG-7 Final Committee Draft
- **July, 2001:** MPEG-7 Final Draft International Standard
- **September, 2001:** MPEG-7 International Standard

How do I join the MPEG-7 Audio Issues AHG to keep track of current development?

Send a message with "subscribe" in the Subject line to: mpeg-7-aud-request@starlab.net

Where can I find out more about MPEG-7?

- Official MPEG site: http://www.cselt.it/mpeg/
  has all of the latest public MPEG documents.
- GMD-Darmstadt: http://www.darmstadt.gmd.de/mobile/MPEG7/index.html
  has done an excellent job of organizing and presenting the relevant MPEG-7 documents.
- MPEG-7 Audio Home Page: http://www.mpeg-7.com/audio (old location)
- StarLab: http://www.mpeg-7.com/
  has some official MPEG-7 documents in Portable Document Format.

Where do I go if I have more questions?

If you have any questions, please do not hesitate to join the MPEG-7 Audio Issues AHG and direct your questions to the reflector. Chances are, there are others who share your concerns, and the AHG exists specifically to identify those concerns. If you must reach an individual, Adam Lindsay <adam@starlab.net> will be happy to answer any questions regarding MPEG-7.

Heiko Purnhagen 10-Mar-2000