25th International AES Conference
Metadata for Audio
17th June 2004
MPEG-7 WORKSHOP

Coordinator: Geoffroy Peeters (Ircam, SemanticHIFI, Paris)

Co-organized with the TC on Semantic Audio Analysis
http://www.aes.org/technical/saa/

Speakers:
Max Jacob (Ircam, SemanticHIFI, Paris)
Jurgen Herre et al. (FHG IIS, Erlangen)
Michael Casey (City University, London)
Emilia Gomez et al. (IUA/UPF,Barcelona)

Introduction

- Objectives of this workshop:
  - there have been a number of tutorials describing the goal and possibilities of MPEG-7 in the audio domain.
  - the practical use of this specification remains unclear to many application designers.
  - present the MPEG-7 standard from a practical point of view
  - providing information for its step-by-step use in audio applications
    - what forms does the MPEG-7 standard take ?
    - where to find/buy the standard ?
    - how to read the standard ?
    - who should use it ?
    - how to implement it in a database context ?
    - how to connect databases to automatic extractors ?

Introduction

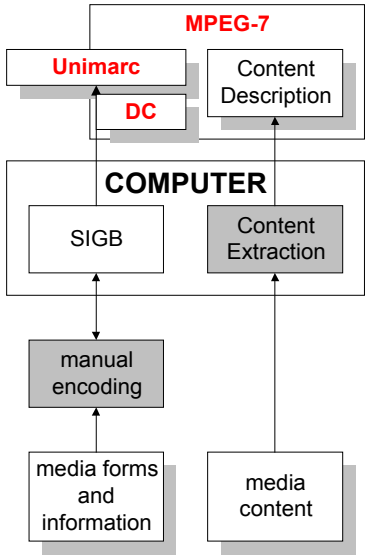
- 1. What is MPEG-7 ?
- 2. How to get into MPEG-7 ?
- 3. MPEG-7 Audio How to use it ?
  - Jürgen Herre : Using MPEG-7 Audio Low-Level Scalability: A guided Tour
- 4. MPEG-7 MDS How to use it ?
  - Max Jacob : Managing Large Sound Databases Using MPEG-7
- 5. How to implement MPEG-7 in an application ?
  - Max Jacob : Managing Large Sound Databases Using MPEG-7
- 6. Storing and searching MPEG-7 Audio ?
  - Michael Casey : Integrating Low-Level Metadata in Multimedia Database Management Systems
- 7. Example of applications using MPEG-7 ?
  - Emilia Gomez : Tools for Content-Based Retrieval and Transformation using MPEG7: The SPOff and the MDTools

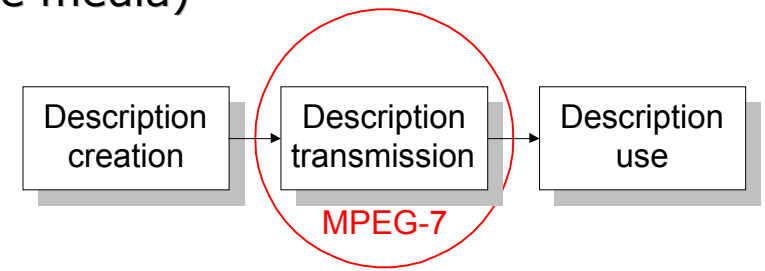
- 1. What is MPEG-7
  - MPEG-7 Objectives
  - MPEG-7 Parts
  - What does it describes ?
  - MPEG-7 Components
  - MPEG-7 Standard / Profiles
- 2. How to get into MPEG-7
  - Where to get the standard from ?
    - Text document
    - XML Schema
    - Reference software
    - Conformance
- 3. MPEG-7 Audio
  - Version1, 2, 3
  - Reference software

- 1. What is MPEG-7
  - MPEG-7 Objectives
  - MPEG-7 Parts
  - What does it describes ?
  - MPEG-7 Components
  - MPEG-7 Standard / Profiles
- 2. How to get into MPEG-7
  - Where to get the standard from ?
    - Text document
    - XML Schema
    - Reference software
    - Conformance
- 3. MPEG-7 Audio
  - Version1, 2, 3
  - Reference software

1. What is MPEG-7: Objectives

- Objectives of MPEG-7:
  - MPEG-7 = Multimedia Content Description Interface
  - A metadata standard (not a compression standard)
  - Goal: "describing the multimedia content data that supports some degree of interpretation of the information's meaning, which can be passed onto, or accessed by, a device or a computer code"
  - Differences with other standard (MARK, DublinCore, …)
    - include numerical description (signal based)
  - description linked to the media (not in the media)
  - description format = XML/DDL format or Binary format (BIM)
  - MPEG-7: standardization of description not of extraction methods





1. What is MPEG-7: Parts

- MPEG-7 = International Standard since September 2001
- MPEG-7 = ISO/IEC 15938

- MPEG-7 Parts:

Part 1. MPEG-7 System	15938-1
Part 2. MPEG-7 Description Definition Language	15938-2
Part 3. MPEG-7 Visual	15938-3
Part 4. MPEG-7 **Audio**	**15938-4**
Amendment1	**15938-4/A1**
Part 5. MPEG-7 MDS	15938-5
Part 6. MPEG-7 Reference Software	15938-6
Part 7. MPEG-7 Conformance testing	15938-7
Part 8. MPEG-7 Extraction and use of MPEG-7 descriptions	15938-8
Part 9. MPEG-7 Profiles	15938-9
Part10. MPEG-7 Schema definition	15938-10

1. What is MPEG-7: Parts

- MPEG-7 Parts:
  - Part 1. MPEG-7 Systems:
    The tools that are needed to prepare MPEG-7 Descriptions for efficient transport and storage, and to allow synchronization between content en descriptions. Tools related to managing and protecting intellectual property
  - Part 2. MPEG-7 Description Definition Language:
    The language for defining new Description Schemes and perhaps eventually also for new Descriptors.
  - Part 3. MPEG-7 Visual:
    The Descriptors and Description Schemes dealing with (only) Visual descriptions
  - Part 4. MPEG-7 Audio:
    The Descriptors and Description Schemes dealing with (only) Audio descriptions
  - Part 5. MPEG-7 Multimedia Description Schemes:
    The Descriptors and Description Schemes dealing with generic features and multimedia descriptions
  - Part 6. MPEG-7 Reference Software:
    A software implementation of relevant parts of the standard.
  - Part 7. MPEG-7 Conformance:
    Guidelines and procedures for testing conformance of MPEG-7 implementations.

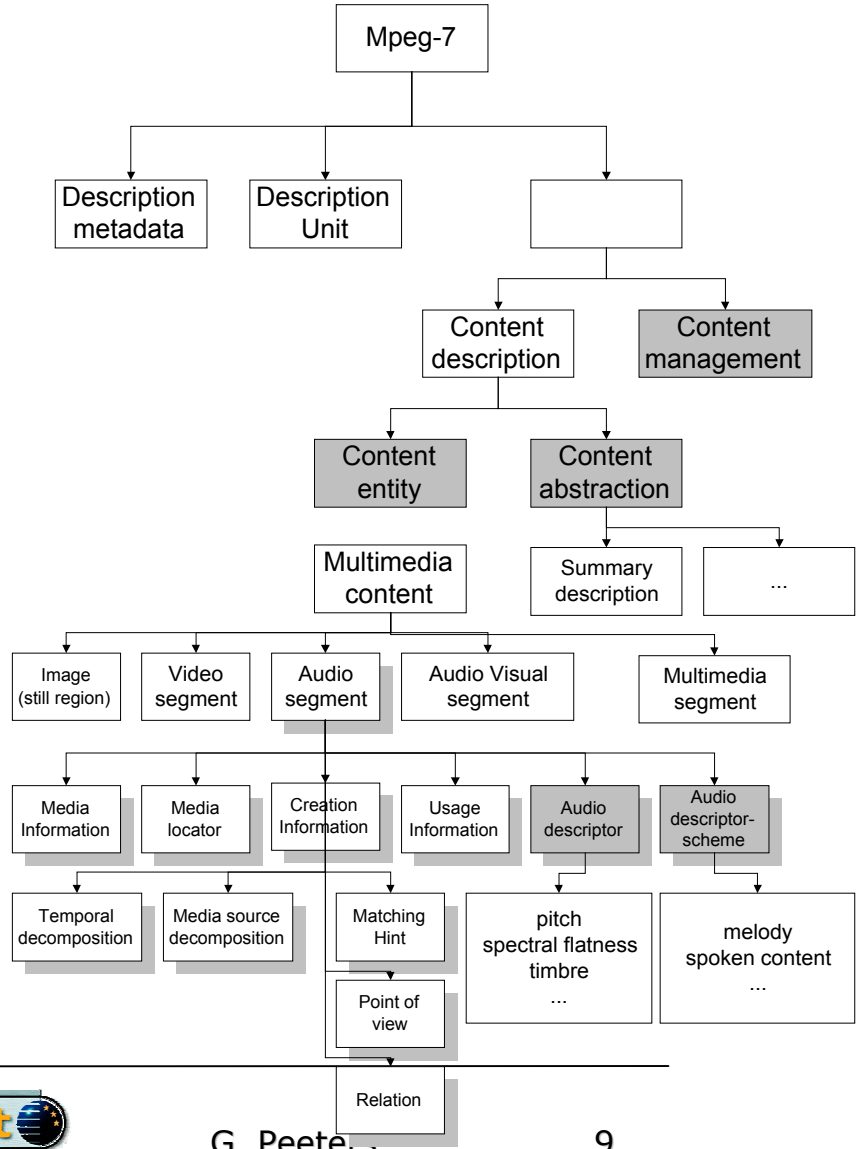
What is MPEG-7: What does it describe ?
bibliographical AND content description

MPEG-7

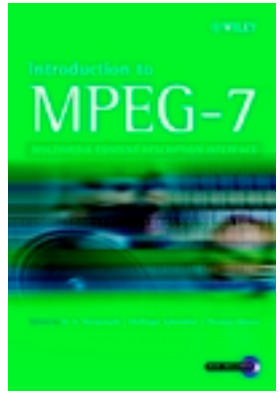
- Descriptions:
  - Bibliographical information
    - media
    - meta (creation/production)
  - Content description
  - Structure description (table of content)
  - Semantic description (index)
  - Navigation and access control
    - summary

- D: descriptors
- DS: description schemes

- B

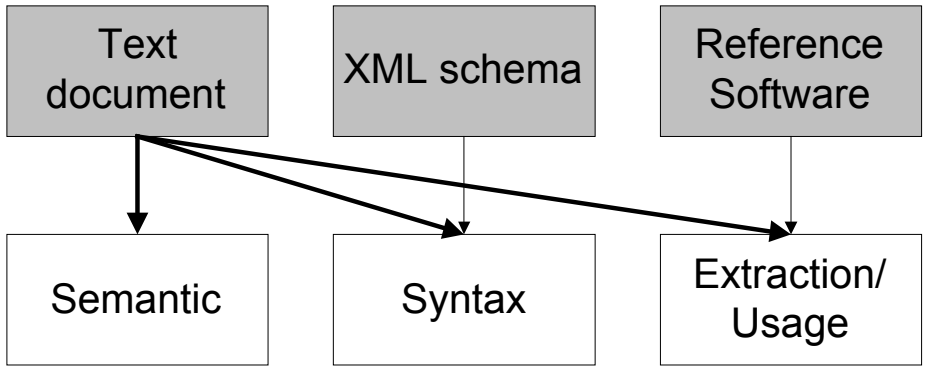


Introduction to MPEG-7: Multimedia Content Description Interface
B. S. Manjunath (Editor), Philippe Salembier (Editor), Thomas Sikora (Editor)
ISBN: 0-471-48678-7



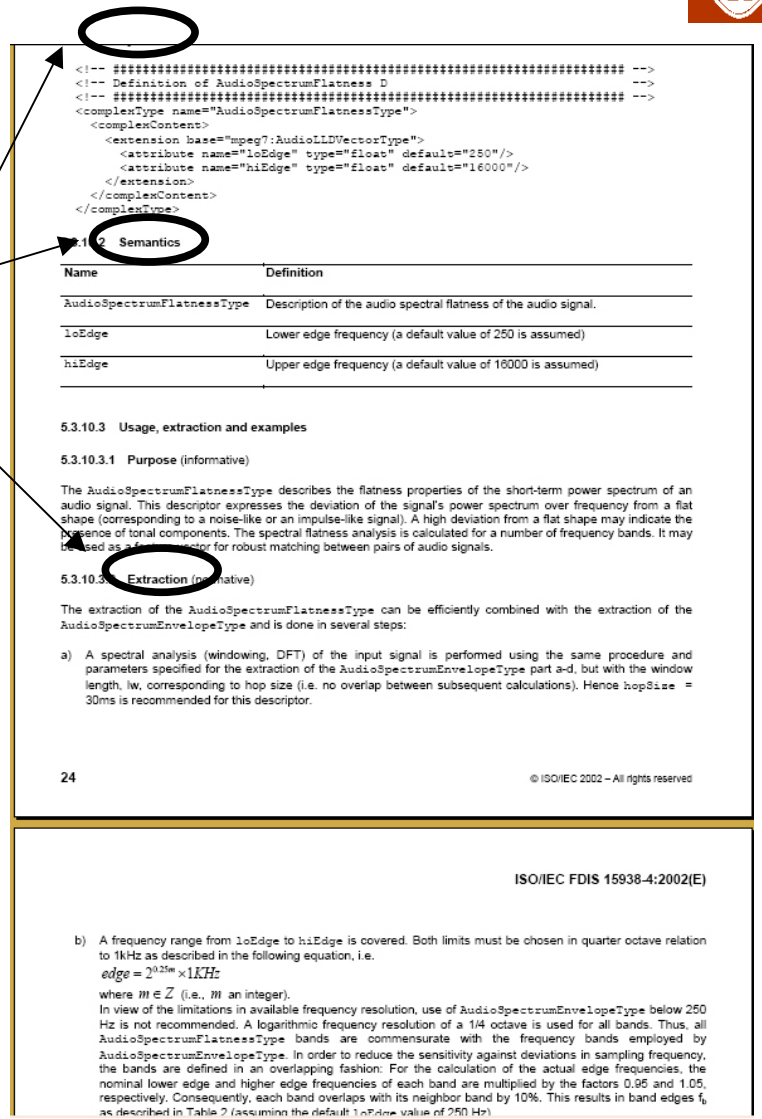
1. What is MPEG-7: Components

- What is practically MPEG-7 ?
  - 1) Semantic of the description
    (text document: semantic part)
  - 2) Syntax of the description
    (text document:syntax part / xml schema)
  - 3) Example of extraction methods for the description
    (text document: usage and extraction part normative or informative / reference software)



1. What is MPEG-7: Components

- What is practically MPEG-7
  - 1) Text document
    - defines the semantic of the descriptions
    - defines the syntax of the description
    - provide explanation for descriptor extraction and usage (normative of informative)
  - Audio: 15938-4.pdf
  - MDS: 15938-5.pdf



```
<!-- ################################################################################ -->
<!-- Definition of AudioSpectrumFlatness D                                            -->
<!-- ################################################################################ -->
<complexType name="AudioSpectrumFlatnessType">
  <complexContent>
    <extension base="mpeg7:AudioLLDVectorType">
      <attribute name="loEdge" type="float" default="250"/>
      <attribute name="hiEdge" type="float" default="16000"/>
    </extension>
  </complexContent>
</complexType>
```

5.3.10.2 Semantics

Name	Definition
AudioSpectrumFlatnessType	Description of the audio spectral flatness of the audio signal.
loEdge	Lower edge frequency (a default value of 250 is assumed)
hiEdge	Upper edge frequency (a default value of 16000 is assumed)

5.3.10.3 Usage, extraction and examples

5.3.10.3.1 Purpose (informative)

The AudioSpectrumFlatnessType describes the flatness properties of the short-term power spectrum of an audio signal. This descriptor expresses the deviation of the signal's power spectrum over frequency from a flat shape (corresponding to a noise-like or an impulse-like signal). A high deviation from a flat shape may indicate the presence of tonal components. The spectral flatness analysis is calculated for a number of frequency bands. It may be used as a feature vector for robust matching between pairs of audio signals.

5.3.10.3.2 Extraction (normative)

The extraction of the AudioSpectrumFlatnessType can be efficiently combined with the extraction of the AudioSpectrumEnvelopeType and is done in several steps:

a) A spectral analysis (windowing, DFT) of the input signal is performed using the same procedure and parameters specified for the extraction of the AudioSpectrumEnvelopeType part a-d, but with the window length, lw, corresponding to hop size (i.e. no overlap between subsequent calculations). Hence hopSize = 30ms is recommended for this descriptor.

24 © ISO/IEC 2002 – All rights reserved

ISO/IEC FDIS 15938-4:2002(E)

b) A frequency range from loEdge to hiEdge is covered. Both limits must be chosen in quarter octave relation to 1kHz as described in the following equation, i.e.

$$edge = 2^{0.25m} \times 1KHz$$

where $m \in Z$ (i.e., m an integer).

In view of the limitations in available frequency resolution, use of AudioSpectrumEnvelopeType below 250 Hz is not recommended. A logarithmic frequency resolution of a 1/4 octave is used for all bands. Thus, all AudioSpectrumFlatnessType bands are commensurate with the frequency bands employed by AudioSpectrumEnvelopeType. In order to reduce the sensitivity against deviations in sampling frequency, the bands are defined in an overlapping fashion: For the calculation of the actual edge frequencies, the nominal lower edge and higher edge frequencies of each band are multiplied by the factors 0.95 and 1.05, respectively. Consequently, each band overlaps with its neighbor band by 10%. This results in band edges f_b as described in Table 2 (assuming the default loEdge value of 250 Hz).

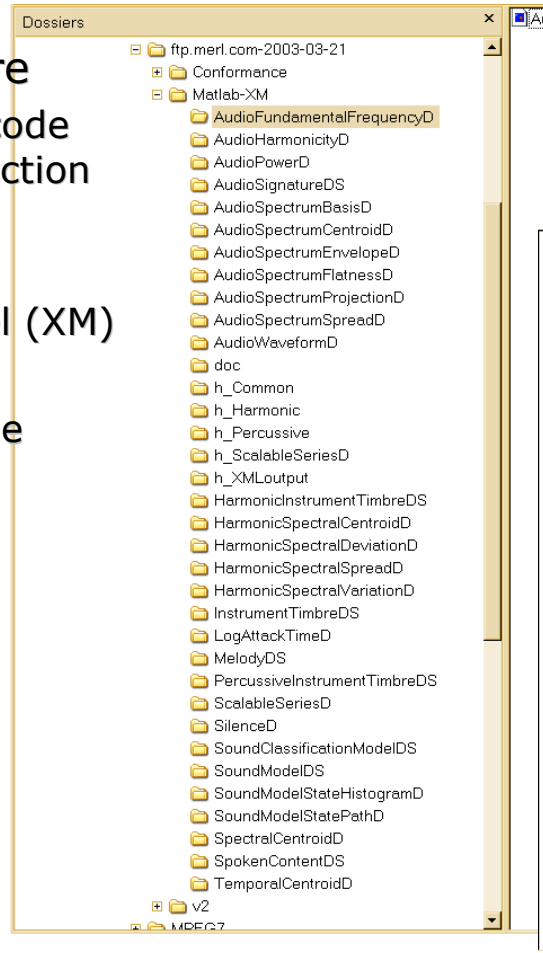
1. What is MPEG-7: Components

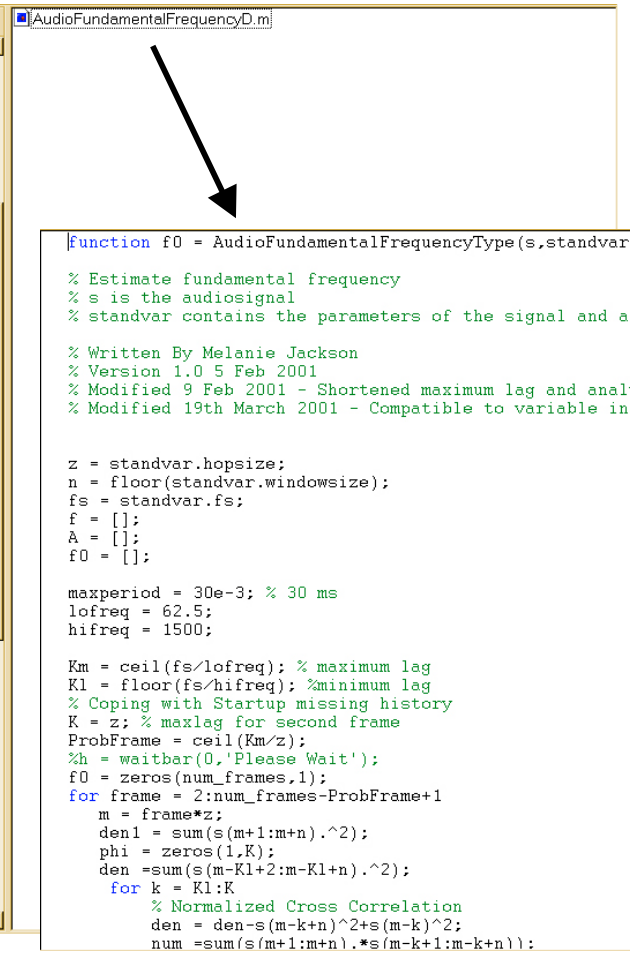
- What is practically MPEG-7
  - 2) Xml schema
    - defines the syntax of the description
    - used to validate the description (see conformance)
    - Mpeg7-2001.xsd.xml
    - ddl-2001.xsd.xml
    - visual-2001.xsd.xml
    - audio-2001.xsd.xml
    - mds-2001.xsd.xml
    - xml-1998.xsd.xml
  - 2b) Binary XML format (BIM) coder/decoder

```
<!-- Definition of AudioWaveform D -->
<!-- ################################################################ -->
+ <complexType name="AudioWaveformType">
<!-- ################################################################ -->
<!-- Definition of audioSpectrumAttributeGrp -->
<!-- ################################################################ -->
+ <attributeGroup name="audioSpectrumAttributeGrp">
<!-- ################################################################ -->
<!-- Definition of AudioSpectrumEnvelope D -->
<!-- ################################################################ -->
+ <complexType name="AudioSpectrumEnvelopeType">
<!-- ################################################################ -->
<!-- Definition of AudioPower D -->
<!-- ################################################################ -->
+ <complexType name="AudioPowerType">
<!-- ################################################################ -->
<!-- Definition of AudioSpectrumCentroid D -->
<!-- ################################################################ -->
- <complexType name="AudioSpectrumCentroidType">
- <complexContent>
<extension base="mpeg7:AudioLLDScalarType" />
</complexContent>
</complexType>
<!-- ################################################################ -->
<!-- Definition of AudioSpectrumSpread D -->
<!-- ################################################################ -->
- <complexType name="AudioSpectrumSpreadType">
- <complexContent>
<extension base="mpeg7:AudioLLDScalarType" />
</complexContent>
</complexType>
<!-- ################################################################ -->
<!-- Definition of AudioFundamentalFrequency D -->
<!-- ################################################################ -->
- <complexType name="AudioFundamentalFrequencyType">
- <complexContent>
- <extension base="mpeg7:AudioLLDScalarType">
<attribute name="loLimit" type="float" default="25" />
<attribute name="hiLimit" type="float" use="optional" />
</extension>
```

1. What is MPEG-7: Components

- What is practically MPEG-7
  - 3) Reference software
    - provides example code for descriptor extraction and matching
    - eXperimental Model (XM)
    - C++ code
    - Audio: Matlab© code





1. What is MPEG-7: Standard / Profiles

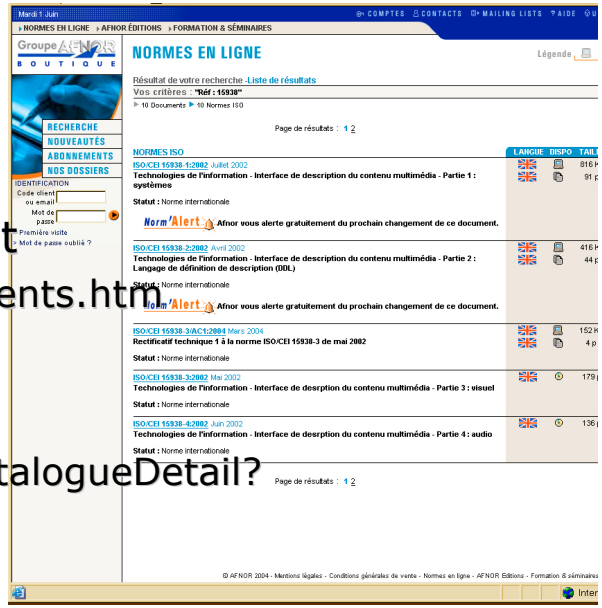
- MPEG-7 standard and MPEG-7 profiles
  - Profiles ?
    - A profile is a subset of tools defined in ISO/IEC 15938, providing a particular set of functionalities for one or more classes of applications
    - Level: A level is a defined set of constraints on a profile to limit the complexity of the profile
    - New schema: combination of ISO/IEC 15938
    - Description profiles /System profiles
      - description: profiles define subsets of description tools across the different parts of ISO/IEC 15938
      - system: capable of constraining systems-related issues, such as transport, access units and binary encoding
  - Current description profiles ?
    - Simple Metadata Profile (SMP): simple metadata tagging for single instances of multimedia clip. This profile can be used in the areas such as music, images, and mobile applications, just to name a few
      - ID3 -> MPEG-7: Mapping ID3 V1.1 tags into MPEG-7 tools to describe song title, album title, artist, year of recording, genre, and user comment for MP3
      - 3GPP: ...
      - EXIF: …
    - User Description Profile (UDP): describe the personal preferences and usage patterns of users of multimedia content
    - Core Description Profile (CDP): describe general multimedia content such as images, videos, audio, and collections

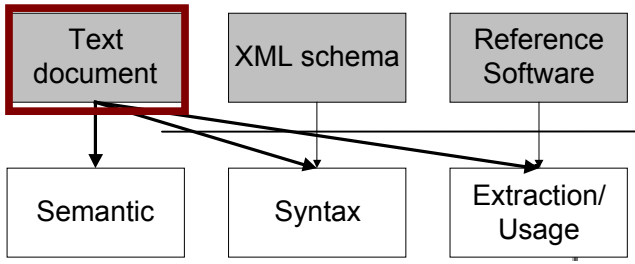
- 1. What is MPEG-7
  - MPEG-7 Objectives
  - MPEG-7 Parts
  - What does it describes ?
  - MPEG-7 Components
  - MPEG-7 Standard / Profiles
- 2. How to get into MPEG-7
  - Where to get the standard from ?
    - Text document
    - XML Schema
    - Reference software
    - Conformance
- 3. MPEG-7 Audio
  - Version1, 2, 3
  - Reference software

2. How to get into MPEG-7

- Where to get the standard from ?

- 1) Text document
  - Current working documents : can be download at
    - http://www.chiariglione.org/mpeg/working_documents.htm
  - IS at ISO
    - http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?
  - IS at BSI
    - http://bsonline.techindex.co.uk/...
  - IS at AFNOR
    - http://www.boutique.afnor.fr/Boutique.asp?url=...
- WD (Working Draft)
  - CD (Committee) - FCD (Final) - FDIS (International Standard)- IS
  - PDAM (Proposed Draft Amendment) - FPDAM (Final) - FDAM - AMD



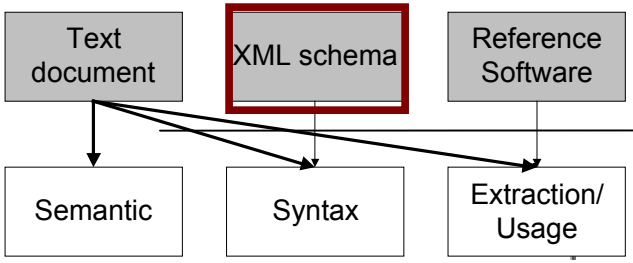


2. How to get into MPEG-7

- Where to get the standard from ?

- 2) XML Schema
  - provided with the CD version of the standard

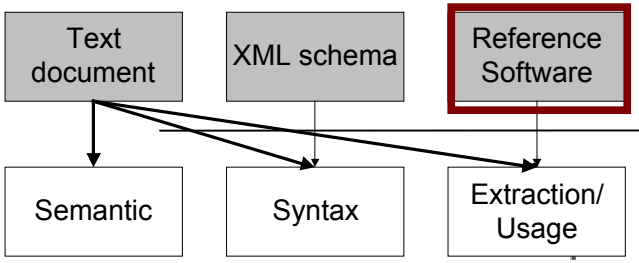
  - NIST MPEG-7 Validation Site:
    http://m7itb.nist.gov/M7Validation.html
  - MPEG-7 Audio/Multimedia Software and Resources:
    http://ccc.soi.city.ac.uk/mpeg7/



2. How to get into MPEG-7

- Where to get the standard from ?

- 3) Reference software: XM (eXperimental Model)
  - Separate document:
    - 15938-6
  - Audio (Matlab)
    - FTP: ftp.merl.com
    - HTTP: http://ccc.soi.city.ac.uk/mpeg7/mirror/
  - MDS/Video (C/C++)
    - HTTP: http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html
    - CVS server



2. How to get into MPEG-7

- Where to get the standard from ?

- 4) Conformance:
  - Distinct document:
    - 15938-7

  - Audio conformance ?
    - Mandatory: audio schema conformance
      -> essentially a syntactic conformance
      -> range of numerical descriptors
    - Recommended (2002): semantic validation in audio
      (see informative/normative part of text document)

    - Audio use: do not generate an error

- 1. What is MPEG-7
  - MPEG-7 Objectives
  - MPEG-7 Parts
  - What does it describes ?
  - MPEG-7 Components
  - MPEG-7 Standard / Profiles
- 2. How to get into MPEG-7
  - Where to get the standard from ?
    - Text document
    - XML Schema
    - Reference software
    - Conformance
- 3. MPEG-7 Audio
  - Version1, 2, 3
  - Reference software

3. MPEG-7 Audio How to use it

- MPEG-7 Audio 15938-4
- MPEG-7 Audio version2=Amendment1 15938-4/A1
- *MPEG-7 Audio version3=Amendment2* *in work ...*

3. MPEG-7 Audio How to use it

AudioFramework

SeriesOfScalar
SeriesOfVector
AudioLLDScalar
AudioLLDVector

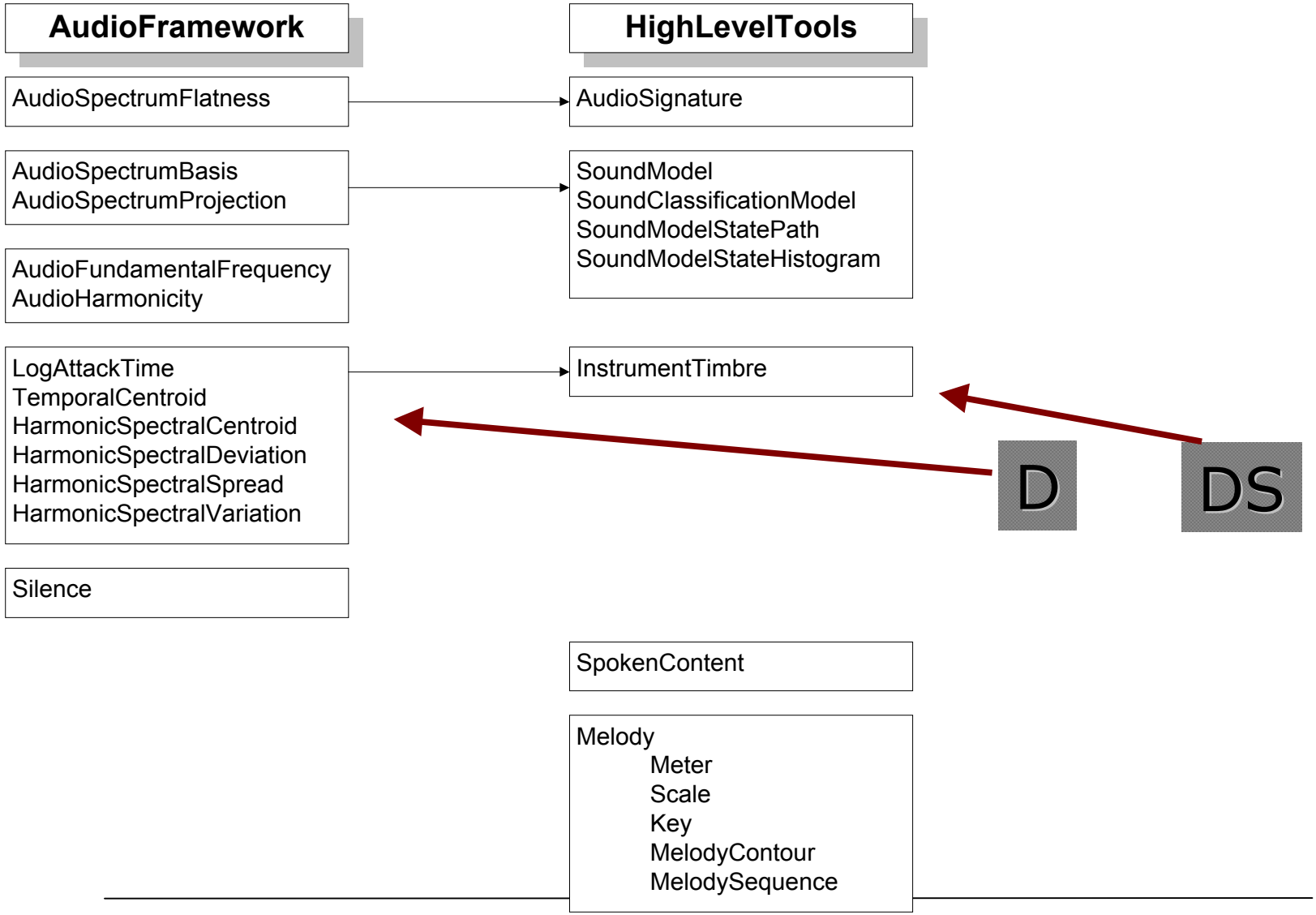
AudioWaveform
AudioPower
AudioSpectrumEnvelope
AudioSpectrumCentroid
AudioSpectrumSpread

HighLevelTools

- data types => scalable series
- future-proof descriptors

3. MPEG-7 Audio How to use it

MPEG-7 Audio version 1



3. MPEG-7 Audio How to use it

High-level audio description tools (descriptors Scheme)

- **MPEG-7 Audio version 1**

- **AudioSignature**

  The AudioSignatureDS is a condensed representation of an audio signal designed to provide a unique content identifier for the purpose of robust automatic identification of audio signals. The AudioSignatureDS uses statistical data summarization on a series of values of the AudioSpectrumFlatnessType to determine the signature.

- **General Sound Recognition and Indexing**

  The sound recognition descriptors and description schemes are a collection of tools for indexing and categorization of general sounds, with immediate application to sound effects. Support for automatic sound identification and indexing is included

- **Musical Instrument Timbre description tools**

  Timbre descriptors aim at describing perceptual features of instrument sounds. Timbre is currently defined in the literature as the perceptual features that make two sounds having the same pitch and loudness sound different.

- **Spoken Content description tools**

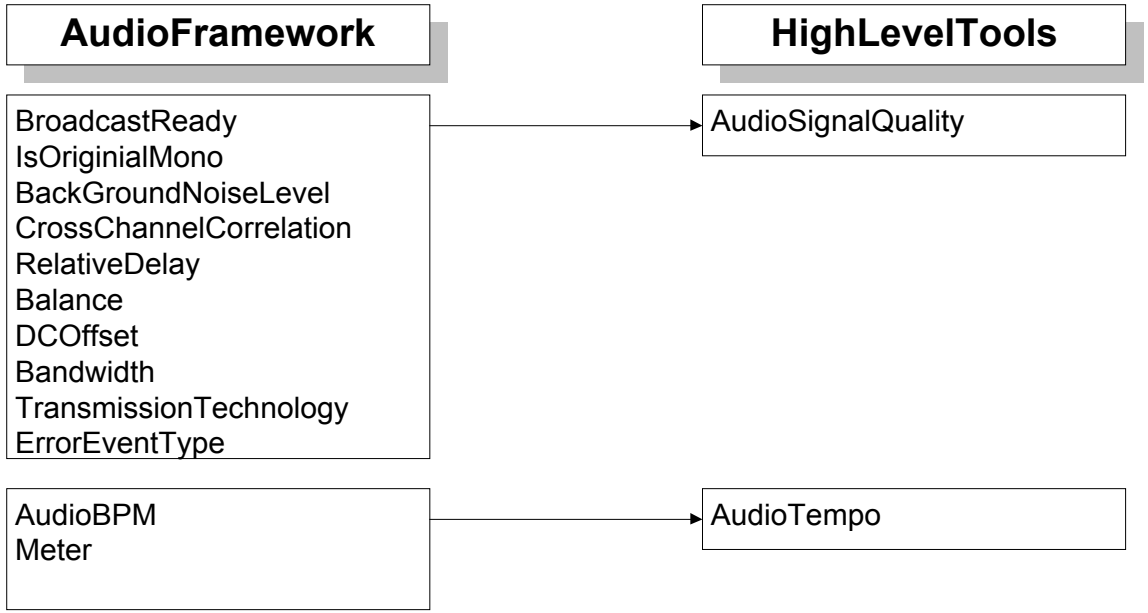
  The Spoken Content description tools allow detailed description of words spoken within an audio stream. The tools can be used for two broad classes of retrieval scenario: indexing into and retrieval of an audio stream, and indexing of multimedia objects annotated with speech.

- **Melody description tools (Melody, Melody contour)**

  The Melody Contour DS is a compact representation for melodic information, which allows for efficient and robust melodic similarity matching, for example, in query-by-humming. The Melody Contour DS uses a 5-step contour (representing the interval difference between adjacent notes), in which intervals are quantized. The Melody Contour DS also represents basic rhythmic information by storing the number of the nearest whole beat of each note, which can dramatically increase the accuracy of matches to a query.

3. MPEG-7 Audio How to use it

MPEG-7 Audio version 2



- **MPEG-7 Audio version 2 (Amendment 1)**
- **Handling of multi-channel signals**
- **WordLexiconType**
- **Audio Signal Quality tools**

  If an AudioSegmentDS contains a piece of music, several features describing the signal's quality can be computed to describe the quality attributes

  - *Operator, UsedTool, BackgroundNoiseLevel, RelativeDelay, Balance, DcOffset, CrossChannelCorrelation, Bandwidth, TransmissionTechnology, ErrorEventList, IsOriginalMono, BroadcastReady*
- **Tempo / Bpm description tools**

  characterize the underlying temporal structure of musical material. Musical tempo information may be used as an efficient search criterion to find musical content for various purposes (e.g. dancing) or belonging to certain musical genres

  - *AudioBPM, Meter*

- **MPEG-7 Audio version 3 (Amendment 2)**

- **work in progress**

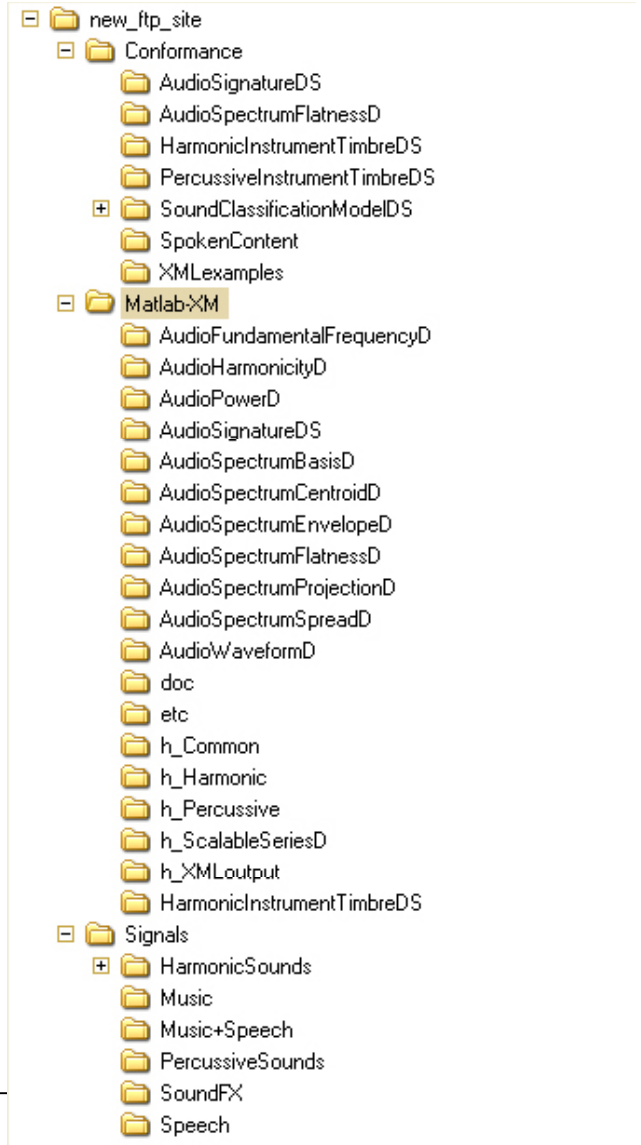
3. MPEG-7 Audio How to use it

MPEG-7 Audio Reference Software

- **'Matlab-XM'** folder : contains the source code for the extraction of the audio descriptors.
  - Each descriptor (and description scheme) has its own folder
  - Secondary functions : 'h_' folders
  - XML reading and writing functions: h_XMLoutput.
  - ScalableSeries/Vector, ...: h_ScalableSeriesD.
  - Documentation: 'doc' folder

- **'Conformance' folder** :
  - -scripts (called top-scripts) allowing to instantiate an xml/bim description from a '.wav' audiofile. -xml/bim description example of instantiations

- The **'Signal' folder;**
  - 'test material' in order to provide audio-signal for the various kind of description: music, speech, musical instruments (harmonic/percussive), sound effects (SoundFX).



- Invited speakers

Jürgen Herre

Biography: Jürgen Herre joined the Fraunhofer Institute for Integrated Circuits (IIS) in Erlangen, Germany, in 1989. Since then he has been involved in the development of perceptual coding algorithms for high quality audio, including the well-known ISO/MPEG-Audio Layer III coder (aka "MP3"). In 1995, Dr. Herre joined Bell Laboratories for a PostDoc term working on the development of MPEG-2 Advanced Audio Coding (AAC). Since the end of '96 he is back at Fraunhofer working on the development of advanced multimedia technology including MPEG-4, MPEG-7 and secure delivery of audiovisual content, currently as the Chief Scientist for the Audio/Multimedia activities at Fraunhofer IIS, Erlangen.

Dr. Herre is a fellow of the Audio Engineering Society, co-chair of the AES Technical Committee on Coding of Audio Signals and vice chair of the AES Technical Council. He also is an active member of the MPEG audio subgroup.

"Using MPEG-7 Audio low level scalability - a guided tour"

- The need for scalability on metadata
- What does metadata scalability mean?
- How is it implemented in MPEG-7 Audio? -> scalable series
- The possibilities offered by ScalableSeries
- Some examples of application of scalability

Max Jacob

Biography: Max Jacob received his degree in musicology in Florence in 1998 specialized in computer music. Autodidact as computer scientist, he has worked for several software and multimedia production companies in Italy between 1996 and 2001. Autodidact as composer he has made the music for movies, video clips and art installations since 1990. Autodidact as sound engineer, he has worked for some recording studio in Florence between 1991 and 1998. Since 2001 at IRCAM, he has been involved in the French project ECRINS and the European projects CUIDADO and Semantic HiFi.

"Managing large sound databases using Mpeg7"

Sound databases are widely used for scientific, commercial and artistic purposes. Nevertheless there is yet no standard way to manage them. This is due to the complexity of describing and indexing audio content and to the variety of purposes a sound database might address. Recently there appeared Mpeg7, a standard for audio/visual content meta-data that could be a good starting point. Mpeg7 not only defines a set of description tools, but is more generally an open framework allowing to host specific extensions for specific needs in a common environment. This is crucial since there would be no way to freeze in a monolithic definition all the possible needs of a sound database. This paper tries to line out how the Mpeg7 framework can be used, how it can be extended and how all this can t into an extensible database design.

6. Storing and searching MPEG-7 Audio

Michael Casey:

Biography: Michael A. Casey obtained his Ph.D. from the Massachusetts Institute of Technology (MIT) Media Laboratory in 1998. Since leaving MIT, his career has focused on research for broadcast and multimedia industries working for News Digital Systems (News Corp. International) and Mitsubishi Electric Research Laboratories. Now a lecturer in the Department of Computing at City University in London, his Multimedia Informatics research group focuses on large-scale systems for multimedia retrieval, general audio retrieval, music information retrieval and intelligent performance systems. Michael is supported by the Engineering and Physical Sciences Research Council (Grant GR/S84750/01). He is a member of the AES, IEEE, British Standards Institute and the International Standards Organization's Moving Pictures Experts Group (MPEG).

"Integrating Low-Level Metadata in Multimedia Database Management Systems"

Employing MPEG-7 low-level descriptors for indexing and searching in large-scale multimedia databases requires strategies for metadata integration. This requirement arises because XML or binary low-level descriptor data is semi-structured and databases require structured table-based data representations. Furthermore, these descriptors are generally incompatible with the built-in datatypes of most database systems, rendering standard SQL scripts unwieldy and inefficient. This talk will give an overview of some practical techniques for efficiently representing and retrieving low-level metadata, and associated media assets, using common database systems such as Oracle9i. Applications to automatic indexing of heterogeneous media catalogues will be presented.

7. Example of applications using MPEG-7

Emilia Gomez

Biography: Emilia Gómez received a MSc degree in Telecommunication Engineering from the University of Seville in 1999. Then, she received a master degree in Acoustics, Signal Processing and Computer Science Applied to Music (ATIAM) from IRCAM, Paris, France. In 2000, she joined the Music Technology Group (IUA/UPF), where she is a researcher and PhD candidate. In 2003, she was a visiting researcher at the Music Acoustics Group, TMH-Royal Institute of Technology, Stockholm. She is also a professor at the Higher Music School of Catalonia, where she teaches Music Acoustics and Sound Synthesis and Processing.
Her research interests are in content-based description and transformation of audio. Specifically, she has been working on melodic and tonal description of audio recordings for music content processing. She has been involved in the CUIDADO IST European Project and the TABASCO Spanish National Project Expressive Transformations of Audio). She is now involved in the SIMAC IST European project (Semantic Interaction with Music Audio Contents).

Oscar Celma

Biography:Òscar Celma was born in Barcelona in 1976. He studied Computer Science at the Universitat Politècnica de Catalunya (Barcelona). In 2000, he joined the Music Technology Group (Universitat Pompeu Fabra) where he is an Associate Professor of Computer Science at the Department of Technology.
His main research field is Music Information Retrieval: metadata description and ontology management, and web-crawlers focused on musical information. From 2001 to 2003, he has been involved in OpenDrama IST European Project and, currently he is involved in Simac IST European Project.

"Tools for content-based retrieval and transformation of audio using MPEG-7: the SPOffline and the MDTools"

In this paper we present a set of applications for content-based retrieval and transformations of audio recordings. They illustrate diverse aspects of a common framework for music content description and structuring implemented using the MPEG-7 standard. MPEG-7 descriptions can be generated either manually or automatically, and are stored in a XML database. Retrieval services are implemented in the database. A set of musical transformations are defined directly at the level of musically meaningful MPEG-7 descriptors and are automatically mapped onto low-level audio signal transformations.