

Content-based Audio Processing with MPEG-7

DAFx-03 Tutorial II
8 September 2003

Dr Michael A. Casey, City University
Adam T. Lindsay, Lancaster University

Overview

- ◆ Introduction
 - ◆ MPEG context & MPEG-7 Overview
- ◆ Applying the descriptions for synthesis
 - ◆ Descriptors in detail
- ◆ XML Processing
 - ◆ Best practices & XSLT
- ◆ Tools
 - ◆ Databases, Java & Reference software

MPEG as a standard

- ◆ Working group of ISO/IEC
 - ◆ Set up by international treaty (Geneva)
 - ◆ National bodies are members of MPEG
 - ◆ Companies send people as representatives of the national bodies
- ◆ Other bodies compare
 - ◆ ITU, SMPTE, AES
 - ◆ Compete and complement

MPEG in the past

- ◆ 1993 MPEG-1
 - ISO/IEC 11172
 - ◆ Target CD ROM distribution of digital video
 - ◆ Parts 1, 2, 3 followed by enhancements
- ◆ 1996 MPEG-2
 - ISO/IEC 13818
 - ◆ Target Digital Video Broadcast
 - ◆ Parts 1, 2, 3 followed by enhancements
- ◆ 1999 MPEG-4
 - ISO/IEC 14496
 - ◆ Target standardised production distribution and access to digital television interactive graphics and multimedia
 - ◆ Enhancements ongoing

MPEG-7 (from 35 000 ft)

- ◆ Description, not compression
- ◆ Not in traditional MPEG realm
 - ◆ new community
- ◆ A kind of coding, in a sense
- ◆ Prompted by MPEG's previous successes
- ◆ Multimedia Content Description Interface

Multimedia Content Description Interface

- ◆ Multimedia
 - ◆ Pictures and/or sound
 - ◆ (with or without associated information)
- ◆ Content
 - ◆ What's inside the multimedia?
- ◆ Description
 - ◆ Representation using some other means than what is being described
- ◆ Interface
 - ◆ The join between two or more entities

The Parts of MPEG-7

ISO/IEC 15938

- ◆ Part 1 • Systems
- ◆ Part 2 • Description Definition Language
- ◆ Part 3 • Visual
- ◆ Part 4 • Audio
- ◆ Part 5 • Multimedia Description Schemes
- ◆ Part 6 • Reference Software
- ◆ Part 7 • Conformance
- ◆ Part 8 • Extraction and use of description

MPEG-7 Overview

Definitions

DDL

Description Definition Language, the XML Schema-based declarative language that the majority of the syntax of the standard is written in

D

Descriptor: a basic unit of description

DS

Description Scheme: a combination of DSs and/or Ds, written in the DDL, and expressing a relationship between these description units

Description

An instantiated Description Scheme, which is natively expressed as an XML document (but may also exist in Binary form)

Definitions

BiM

Binary format for MPEG-7 descriptions: a DS-aware compression format

MDS

Multimedia Description Schemes: a new sub-group in MPEG, with a corresponding Part (part 5, easily the largest) of the standard.

WD, CD, FCD, FDIS, IS

Working Draft, Committee Draft, Final Committee Draft, Final Draft International Standard, and International Standard: all stages (listed in order) in the development of ISO standards

XML Basics

- ◆ W3C standard
- ◆ Simplification/popularisation of SGML
- ◆ Standard (meta-) syntax, to ease parsing
- ◆ Used as basis for MPEG-7's DDL
- ◆ Both definitions (XML Schema) and instantiations are in XML syntax

XML Basics

- ◆ `<element attribute="here">
content</element>`
- ◆ `<outer>
 <inner empty="true"/>
</outer>`
- ◆ `<namespace:element/>`

DDL Basics

```
<complexType name="SoundModelStatePathType">  
  <complexContent>  
    <extension base="mpeg7:AudioLLDScalarType">  
      <sequence>  
        <element name="SoundModelRef"  
          type="anyURI"/>  
      </sequence>  
    </extension>  
  </complexContent>  
</complexType>
```

- ◆ MPEG-7 defines a library of datatypes

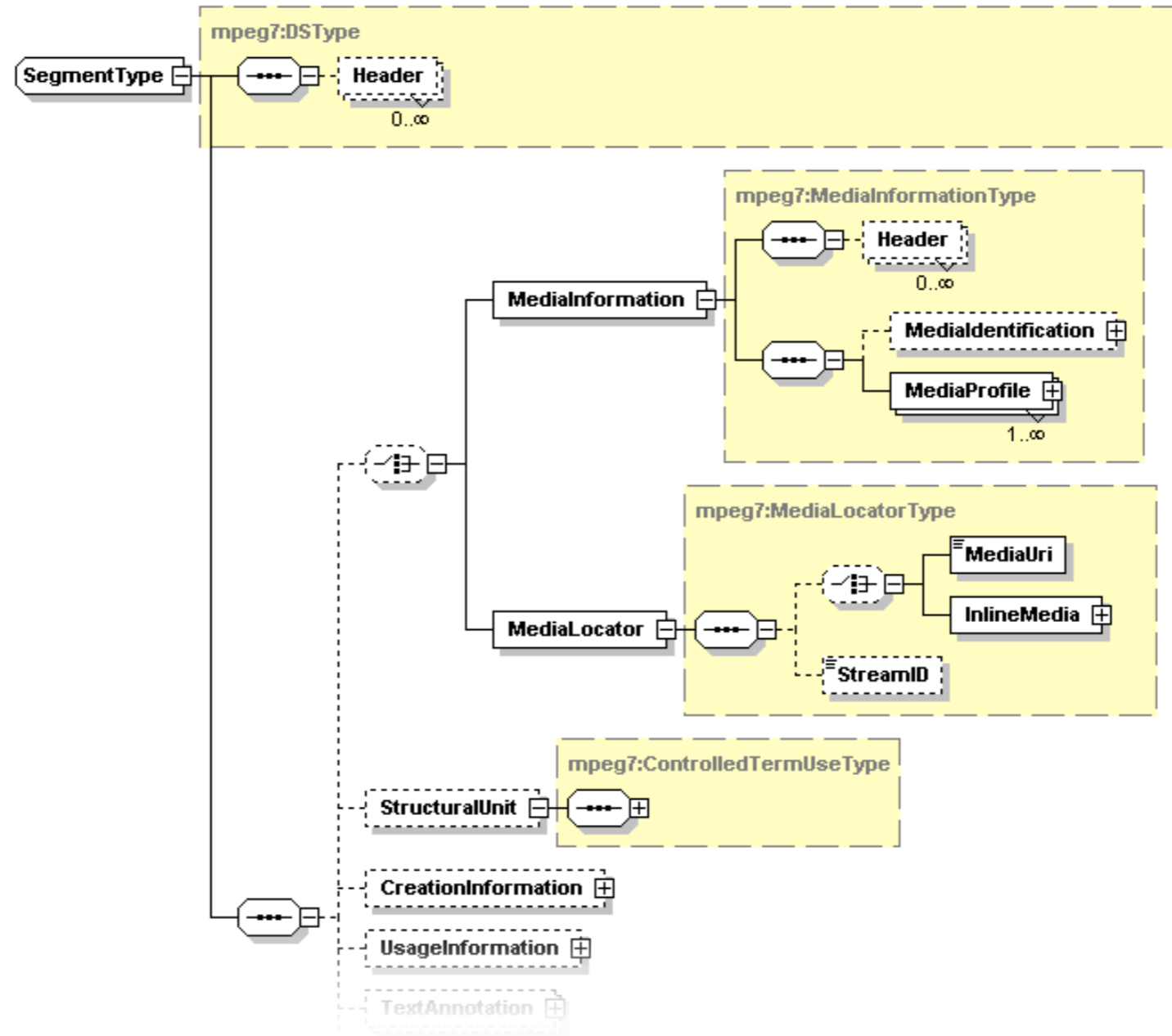
MPEG-7 Conventions

```
<SeriesOfVector totalNumOfSamples="42"  
  vectorSize="29" hopSize="PT10N1000F">  
  <Scaling ratio="1" numOfElements="42"/>  
  <Raw mpeg7:dim="42 29">  
0.00021216 0.00024503 0.0033497 0.0053583...  
  </Raw>  
</SeriesOfVector>
```

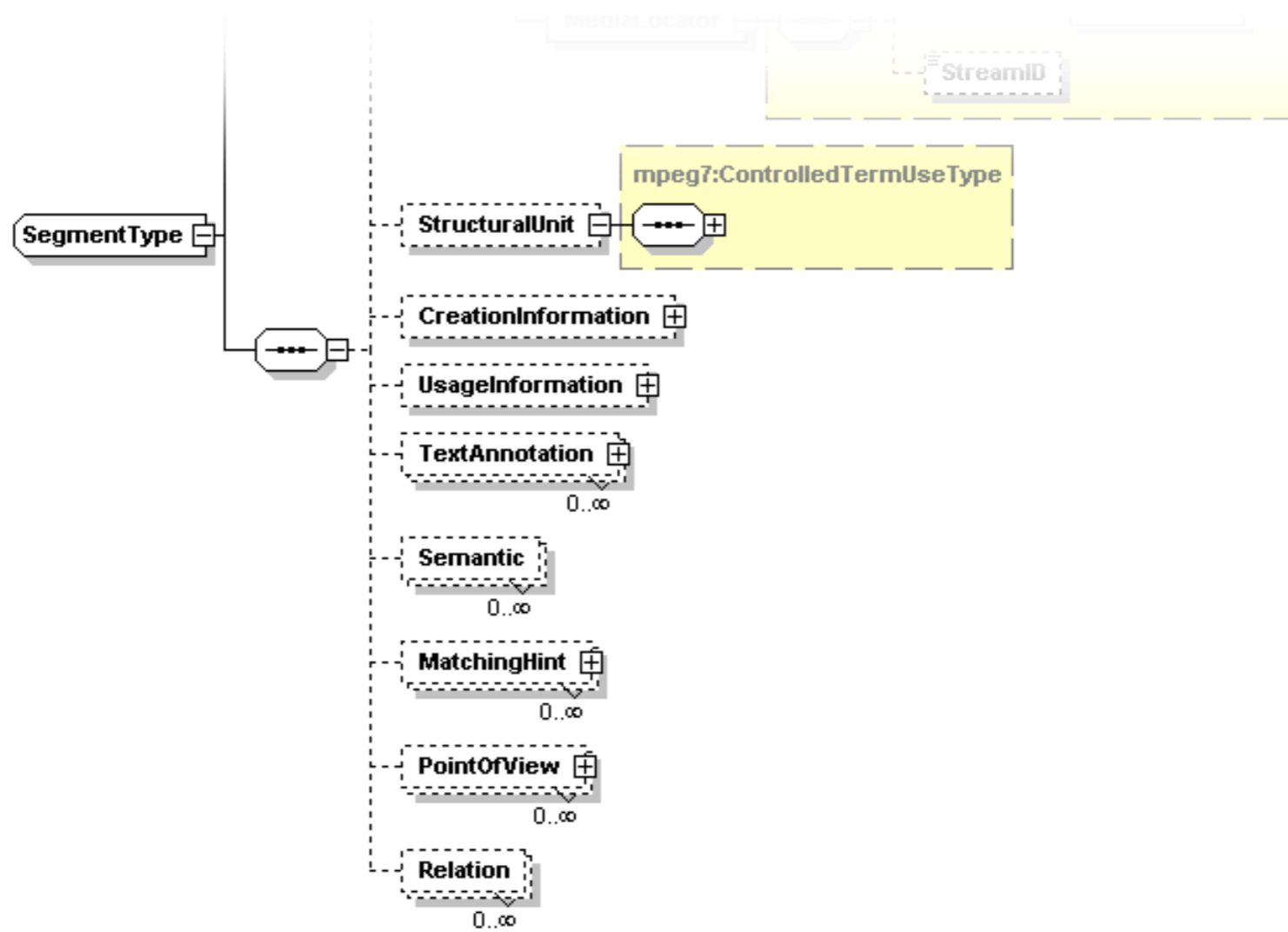
MDS Basic Constructs

- ◆ Datatypes
 - ◆ Constrained integers
 - ◆ Vectors, Matrices, Histograms
- ◆ Reference
 - ◆ Links
 - ◆ Unique IDs
 - ◆ Time elements
 - ◆ Media locators

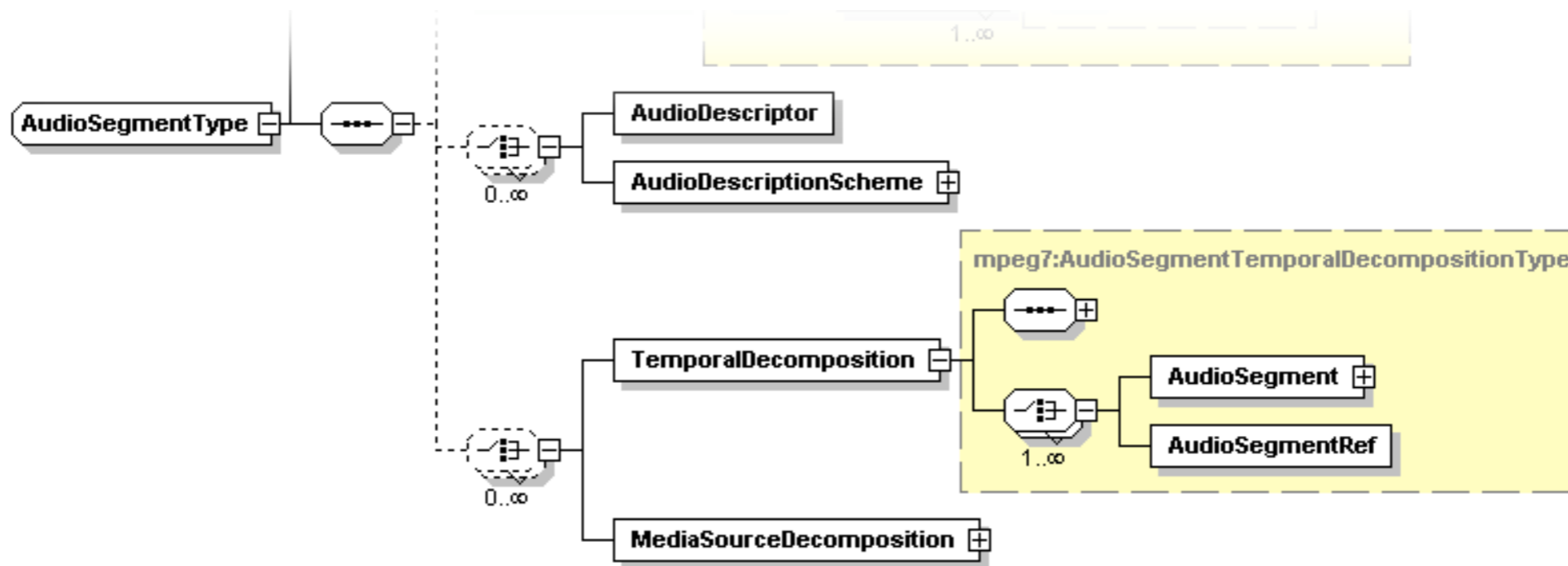
The Segment



The Segment



Segment Decomposition



- ◆ One of the elements of a Segment is a SegmentDecomposition
- ◆ Each N-Segment Decomposition may in turn contain an N-Segment decomposition

MPEG-7 Audio Basics

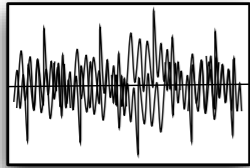
Describing an AudioSegment

- ◆ A Segment contains any number of Audio descriptors
- ◆ An Audio Low-level descriptor (LLD) contains a single value or a series of values
 - ◆ Scalable series normally is sampled at 100Hz
 - ◆ Can be set arbitrarily and non-uniformly
- ◆ Segments can contain other segments

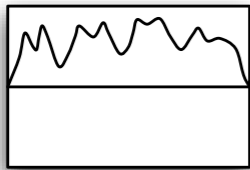
Low-level Descriptors

- ◆ Six loose categories:
 - ◆ Basic
 - ◆ Basic spectral
 - ◆ Signal parameters
 - ◆ Timbral temporal
 - ◆ Timbral spectral
 - ◆ Spectral basis representations

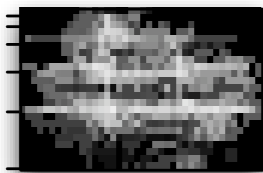
Basics



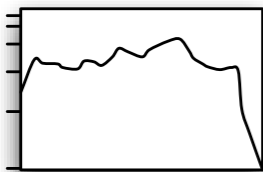
AudioWaveform



AudioPower

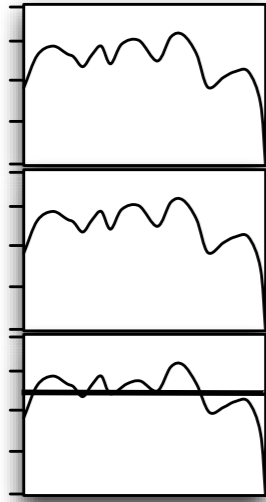


AudioSpectrumEnvelope



AudioSpectrumCentroid, -Spread,
-Flatness

Parameters and Timbre

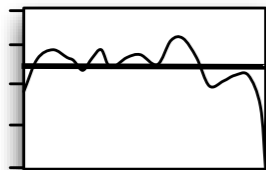


AudioFundamentalFrequency

AudioHarmonicity

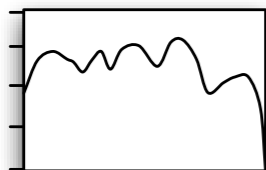
Timbral Temporal

◆ LogAttackTime, TemporalCentroid



Timbral Spectral

◆ HarmonicSpectralCentroid, -Spread,
-Deviation, -Variation



SpectralCentroid

Applications

- ◆ Spoken content
 - ◆ Lattices
- ◆ Musical instrument timbre
 - ◆ Set of LLDs
- ◆ Melody description
 - ◆ Intervallic & contour
- ◆ Sound (effect) recognition and indexing

Why are we here?

- ◆ To tell you about MPEG-7
- ◆ Share some techniques and tools we like
- ◆ Talk about MPEG-7 outside the context of information retrieval:

Content-based retrieval is an integral part of the production chain

Descriptions can be used for more than search; they can contribute to processing