# 7 Multimedia Content Description

7.1 Metadata: Concepts and Overview

7.2 MPEG-7: Generic Metadata Standard

7.3 Selected Metadata-Relevant Standards

7.4 Automated Extraction of Multimedia Metadata

Literature:

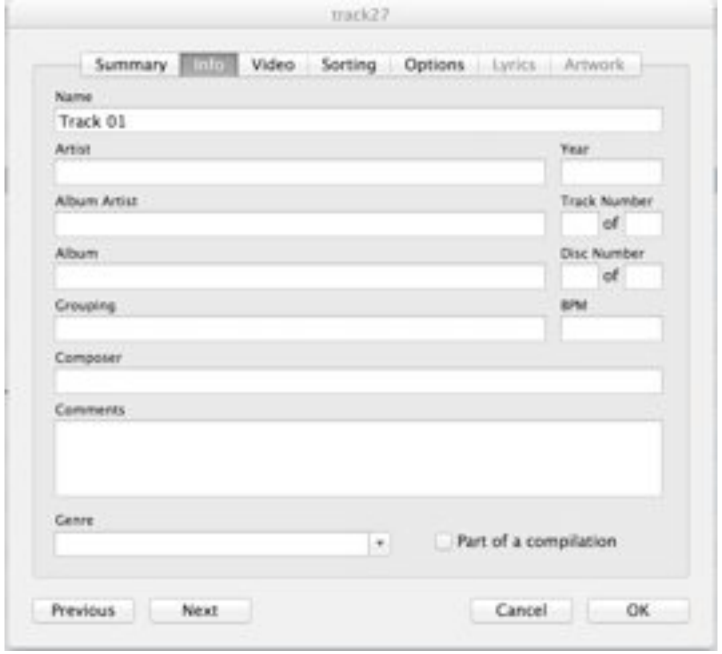
Rosenblatt/Trippe/Mooney, Digital Rights Management, Chapter 6

Troncy/Huet/Schenk, Multimedia Semantics - Metadata, Analysis and Interaction, Wiley 2011

# Unlabelled Stuff

playitagainvideo.com

- The Unlabeled Video Tape Problem
  - Even worse with digital media: Various formats, variants
- Digital media production:
  - Labeling of parts to be composed
    - » Date, time, format, ...
  - Representing the composition
- Digital media on the Internet
  - Identifying digital media
    - » Title, author, genre, ...
  - Searching for specific media, e.g. audio, video content
  - Fine-grained search within media
    - » e.g. person search within video content
  - Bringing together related media (e.g. text news and photos)
    - » (Automated) syndication



track27

Name: Track 01

# Content, Essence, Metadata

- Content
  - consists of *essence* data and *metadata*
- Essence
  - parts of content that directly represent program material such as audio, video, graphic, still-image, text, or sensor-data
- Metadata
  - parts of content that contain data used
    - » to *describe* essence or
    - » to provide information on its *use*
  - metadata objects sometimes called “mobs”
- Metadata may be
  - stored separately from the essence data
  - combined with the essence data (“embedded metadata”)

Source: AAF Developer Overview

# Types of Multimedia Metadata

- Technical Metadata:
  - Form (data format, representation parameters like resolution, color depth...)
  - For live captured material: Time, date, location of original occurrence
  - Technical parameters of capture (e.g. aperture, exposure etc. for images)
- Content Description Metadata:
  - High level, structured:
    - » Title, author, composer, artist, cast, ....
  - High level, unstructured:
    - » Summary, textual description, thumbnail, ...
  - Low level:
    - » Objects and time positions
    - » Audio and video attributes: Key, mood, tempo ...
- Additional information:
  - Digital rights, classification, context, further links, ...

# Types of Origin for Metadata

- Automatic creation or derivation:
  - All technical metadata
  - Some low level metadata (e.g. average brightness, musical tempo)
- Retrieval from external databases:
  - High-level metadata
  - Retrieval may be based on identifier or analysis of media content
  - Example: GraceNote database for music
- Manual addition
  - Archival, indexing, annotation, ...

# Metadata Problems

- Creation metadata
  - During the creation of media essence, metadata is created but often ignored
  - Example: EXIF data in JPEG
- Manually added metadata
  - Users notoriously ignore the administration of metadata
- Metadata incompatibility
  - Metadata exists in various formats specific for media types, applications, product vendors, ...
  - Exchange of metadata is difficult
- Broad range of metadata
  - Metadata exists on various levels, covering all is expensive
- Metadata economy
  - How much of the metadata will be used?
  - When to create metadata?

# 7 Multimedia Content Description

7.1 Metadata: Concepts and Overview

7.2 MPEG-7: Generic Metadata Standard

7.3 Selected Metadata-Relevant Standards

7.4 Automated Extraction of Multimedia Metadata

Literature:

B. S. Manjunath, Philippe Salembier, Thomas Sikora:
Introduction to MPEG-7, Wiley 2002

www.chiariglione.org
mpeg-7.joanneum.at
www.multimedia-metadata.info

# MPEG-7

- Moving Picture Experts Group (MPEG)
  - = ISO/IEC JTC1/SC29/WG11 “Moving Pictures and Audio”
  - Main Web presence now: www.chiariglione.org/mpeg
- MPEG-7 “Multimedia Content Description Interface” (since 1996)
  - “ ... a standard for 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. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes support as broad a range of applications as possible.”
- ISO/IEC 15938 standard since 2002, parts still being added
  - MPEG 7 Audiovisual Description Profile (AVDP): 2012
- Industrial uptake very slow
  - Ambitious standard
- Some research and open source prototypes available
  - See e.g.
    http://mpeg7.joanneum.at,
    http://www.multimedia-metadata.info

# Parts of the MPEG-7 Standard

- MPEG-7 Systems
- MPEG-7 Description Definition Language (DDL)
  - Descriptors (D) and description schemes (DS) specify the syntax and semantics of each *feature* (metadata element)
  - DDL allows the creation of Ds and DSs
    - » XML-based language with some small extensions to XML Schema
- MPEG-7 Visual
- MPEG-7 Audio
- MPEG-7 Multimedia Description Schemes
- MPEG-7 Reference Software
  - eXperimentation Model XM
- MPEG-7 Conformance (rules for conformance checking)
- Extraction and use of MPEG-7 descriptors
- MPEG-7 Profiles and Levels (Profile Schemas, Schema Definition)
- MPEG-7 Query Format

# MPEG-7 Profiles

Part 9 of MPEG-7 (2005):

- Simple Metadata Profile (SMP)
  - Single document or simple collection, similar to EXIF or ID3
- User Description Profile
  - Tools for describing personal preferences and usage patterns
  - Adopted by TV-Anytime standard
- Core Description Profile
  - Collections of multimedia content, description of relationships

Later profiles:

- Audiovisual Description Profile (AVDP)
  - Targeted at requirements of audiovisual media production
  - Mainly driven by European Broadcasting Union (EBU)

# Application Areas of MPEG-7

- Architecture, real estate, and interior design (e.g., searching for ideas).
- Broadcast media selection (e.g., radio channel, TV channel).
- Cultural services (e.g., virtual museums).
- Digital libraries (e.g., image catalogue, musical dictionary).
- Education (e.g., repositories of multimedia courses).
- Home Entertainment (e.g., home video management).
- Investigation services (e.g., human characteristics recognition, forensics).
- Journalism (e.g. searching for video footage of political event).
- Multimedia directory services (e.g. yellow pages, tourist information).
- Multimedia editing (e.g., personalized electronic news service, media authoring).
- Remote sensing (e.g., cartography, ecology, natural resources management).
- Shopping (e.g., searching for clothes that you like).
- Surveillance (e.g., traffic control, surface transportation).
- ...

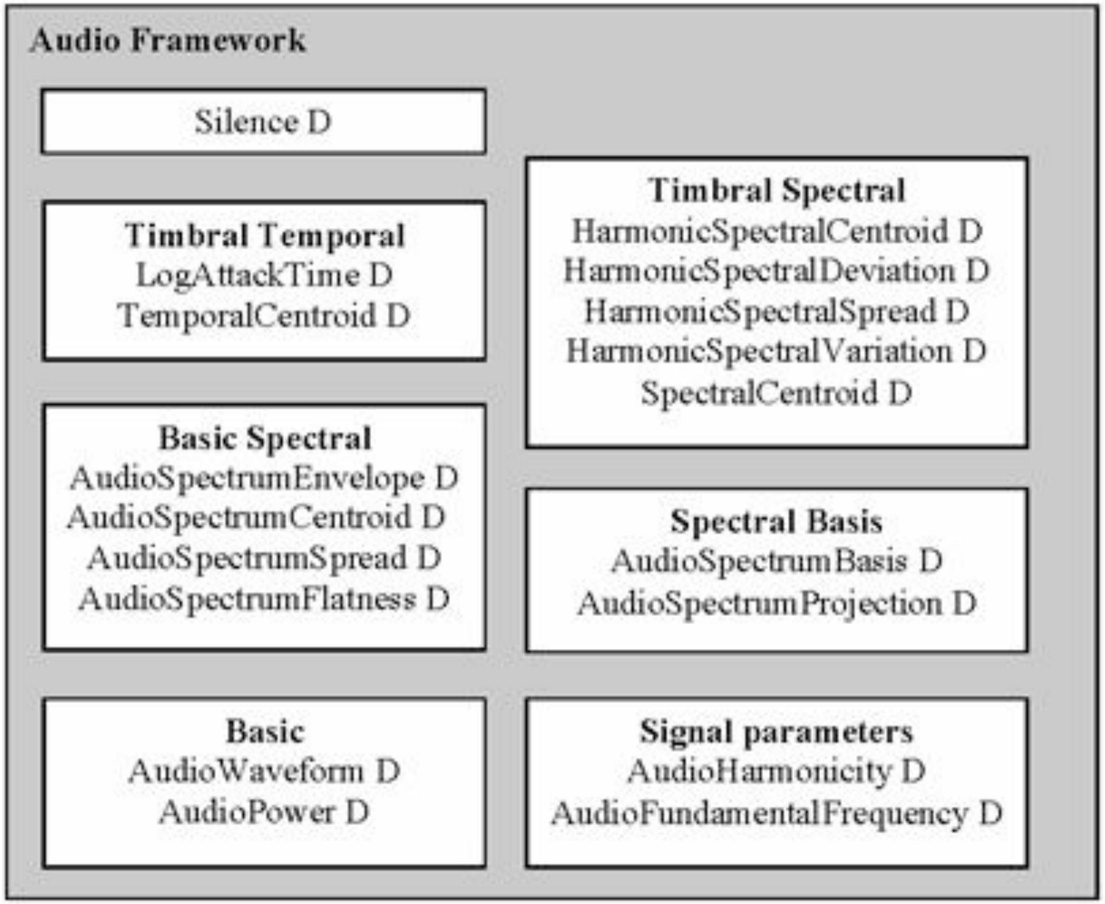
# Examples of Advanced Queries

- 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.
- Draw a few lines on a screen and find a set of images containing similar graphics, logos, ideograms,...
- Define objects, including color patches or textures and retrieve examples among which you select the interesting objects to compose your design.
- 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.
- Describe actions and get a list of scenarios containing such actions.
- Using an excerpt of Pavarotti's voice, obtaining a list of Pavarotti's records, video clips where Pavarotti is singing and photographic material portraying Pavarotti.

From: MPEG-7 Overview

# MPEG-7 Audio Low-Level Descriptors

- Structures:
  - Single scalar value
  - Series of sampled values
- Features:
  - See figure
- MPEG-7 descriptions may contain features described using different (external) methods and algorithms



**Audio Framework**

Silence D

**Timbral Temporal**
LogAttackTime D
TemporalCentroid D

**Timbral Spectral**
HarmonicSpectralCentroid D
HarmonicSpectralDeviation D
HarmonicSpectralSpread D
HarmonicSpectralVariation D
SpectralCentroid D

**Basic Spectral**
AudioSpectrumEnvelope D
AudioSpectrumCentroid D
AudioSpectrumSpread D
AudioSpectrumFlatness D

**Spectral Basis**
AudioSpectrumBasis D
AudioSpectrumProjection D

**Basic**
AudioWaveform D
AudioPower D

**Signal parameters**
AudioHarmonicity D
AudioFundamentalFrequency D

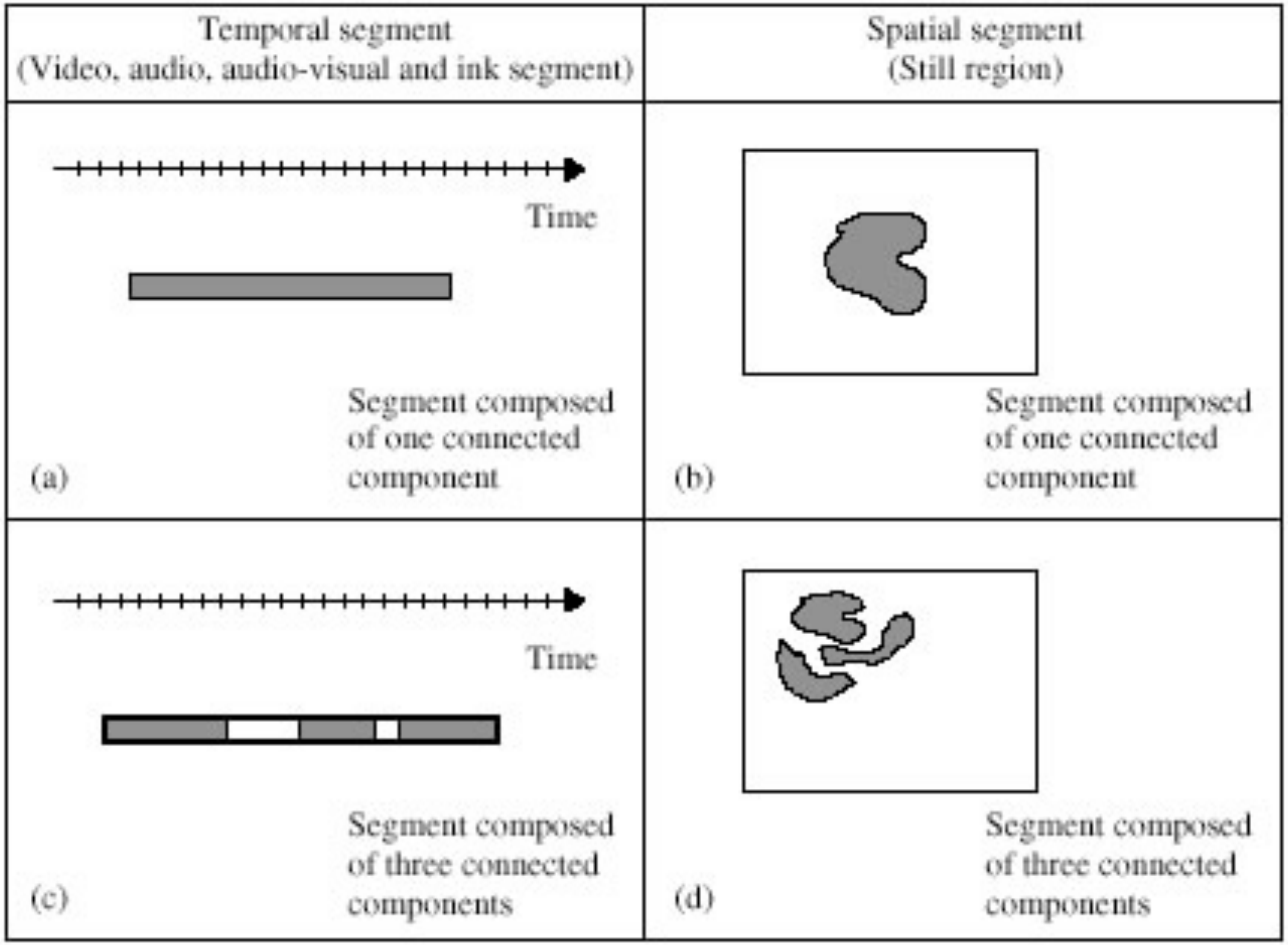
# MPEG-7 Audio High-Level Descriptors

- Audio signature
  - Statistical summary of spectral flatness descriptor
  - Fingerprinting, identification of audio content
- Musical instrument timbre
- Melody description
  - MelodyContour (terse, efficient)
  - MelodySequence
    - » Query by Humming
    - » Example: http://www.musicline.de/de/melodiesuche
- General sound recognition and indexing
  - Probabilistic classifiers for sound classes
- Spoken content
  - Output and intermediate results of Automatic Speech Recognition (ASR)

# Structural Content Description: Segments

- A segment represents a section of an audio-visual content item.
- The Segment Description Scheme (DS) is an abstract class (in the sense of object-oriented programming).
- It has nine major subclasses:
  - Still Region DS (spatial)
  - Video Segment DS (temporal)
  - Moving Region DS (spatiotemporal)
  - Audio Segment DS (temporal)
  - AudioVisual Segment DS (temporal)
  - AudioVisual Region DS (spatiotemporal)
  - Still Region 3D DS (3D spatial)
  - Ink Segment DS (electronic ink from pen, smartboard etc. )
  - Multimedia Segment DS (composite of segments)

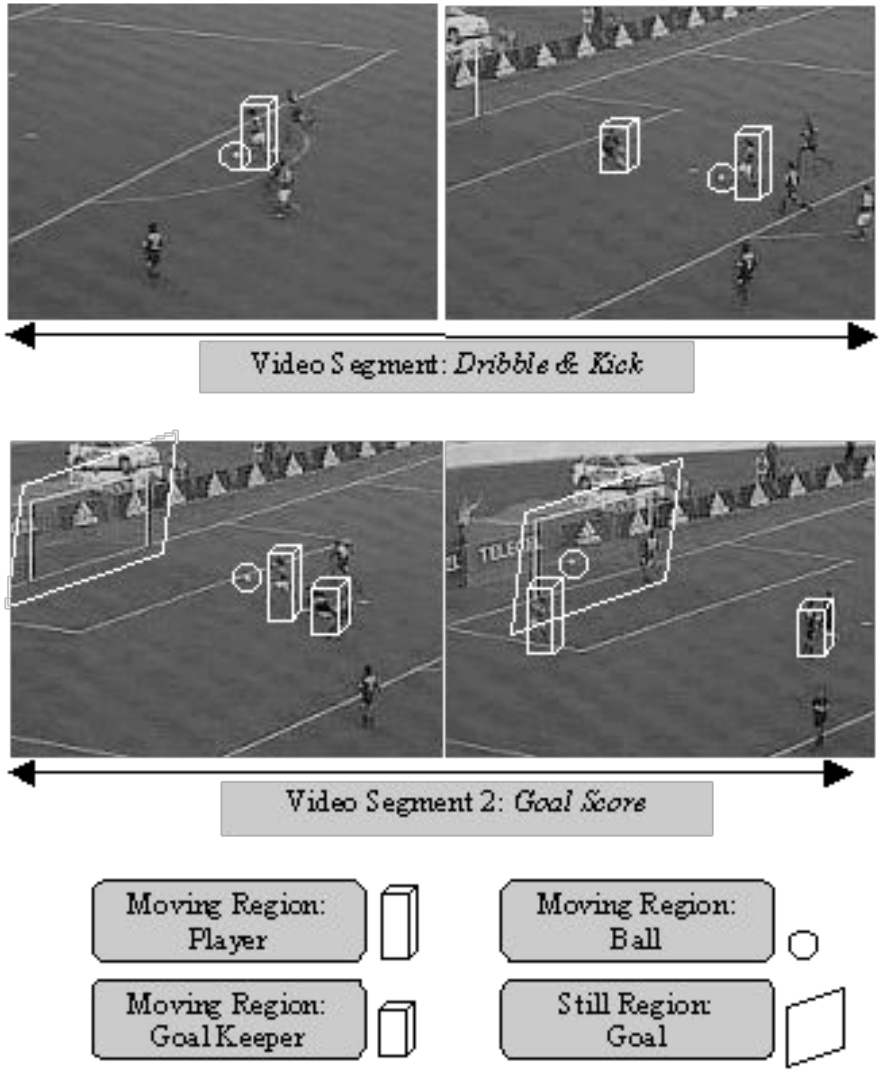
# Examples of Segments

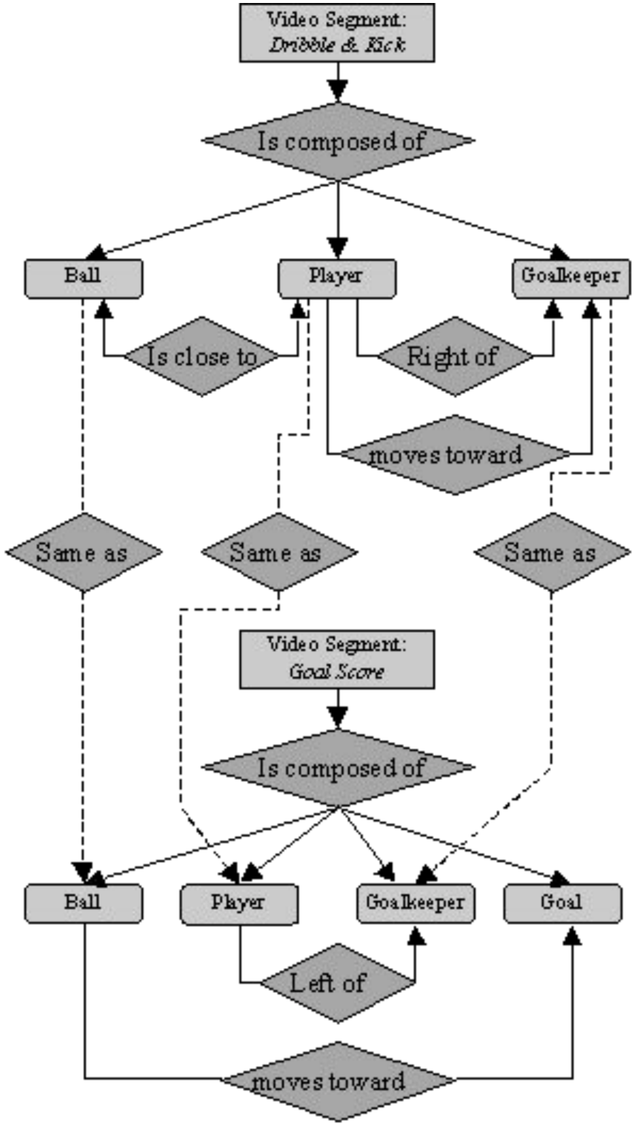

Temporal segment
(Video, audio, audio-visual and ink segment)
Spatial segment
(Still region)
Time
Segment composed of one connected component
(a)
Segment composed of one connected component
(b)
Time
Segment composed of three connected components
(c)
Segment composed of three connected components
(d)

# Structural Relations of Segments

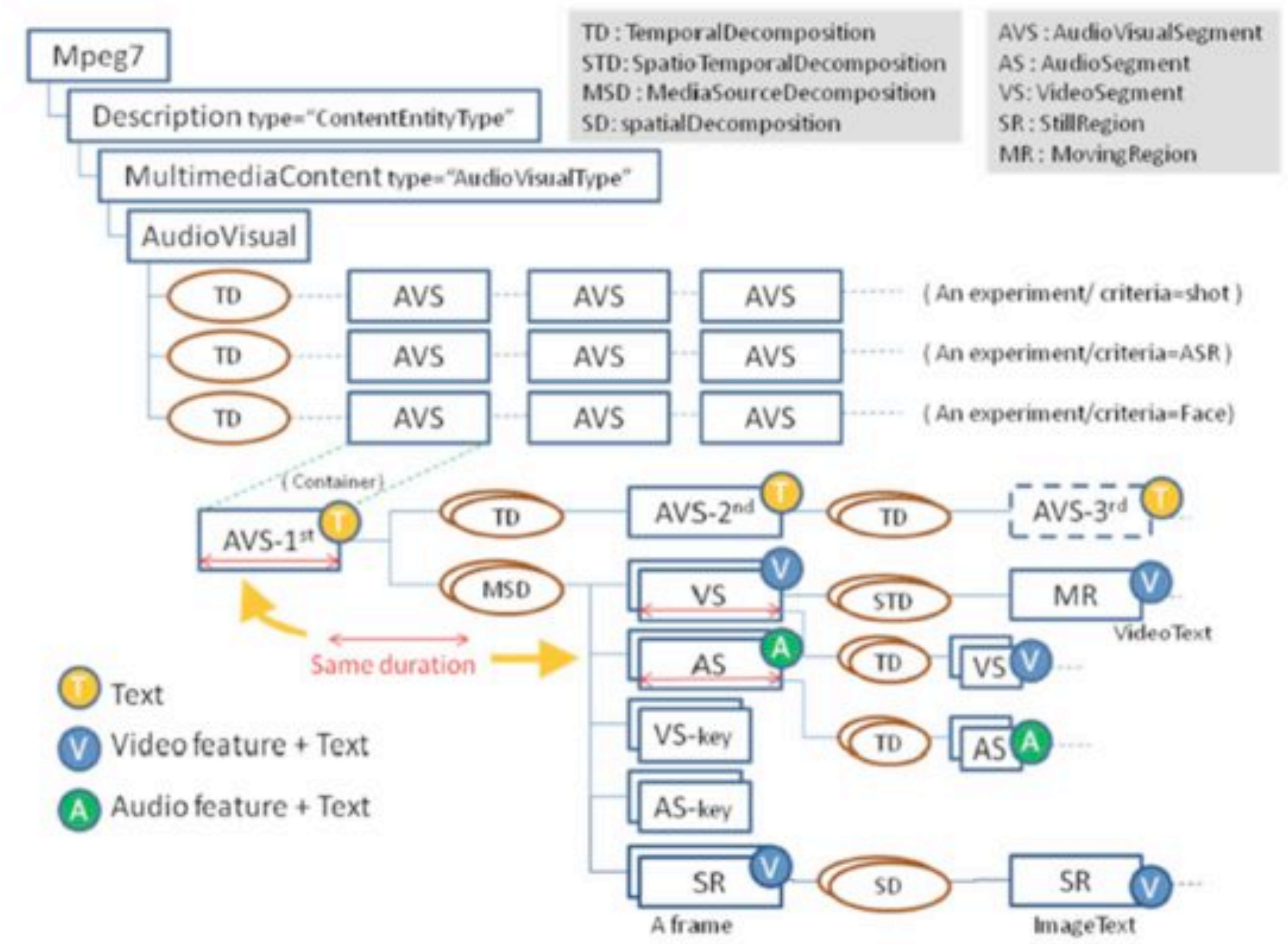
- Content structure:
  - Either hierarchical segment decomposition
  - Or general segment relationship graph
- Predefined structural relations in MPEG-7 (can be extended):
  - Generic:
    - Identical, union, disjoint
  - Spatial:
    - South, north, west, east, northwest, northeast, southwest, southeast, left, right, below, above, over, under
  - Temporal:
    - Precedes, follows, meets, metBy, overlaps, overlappedBy, contains, during, strictContains, strictDuring, starts, startedBy, finishes, finishedBy, coOccurs, contiguous, sequential, coBegin, coEnd, parallel, overlapping
- For each relation, the inverse relation is implicitly defined.

# Video Segmentation with Moving Regions


Video Segment: *Dribble & Kick*
Video Segment 2: *Goal Score*
Moving Region: Player
Moving Region: Ball
Moving Region: Goal Keeper
Still Region: Goal



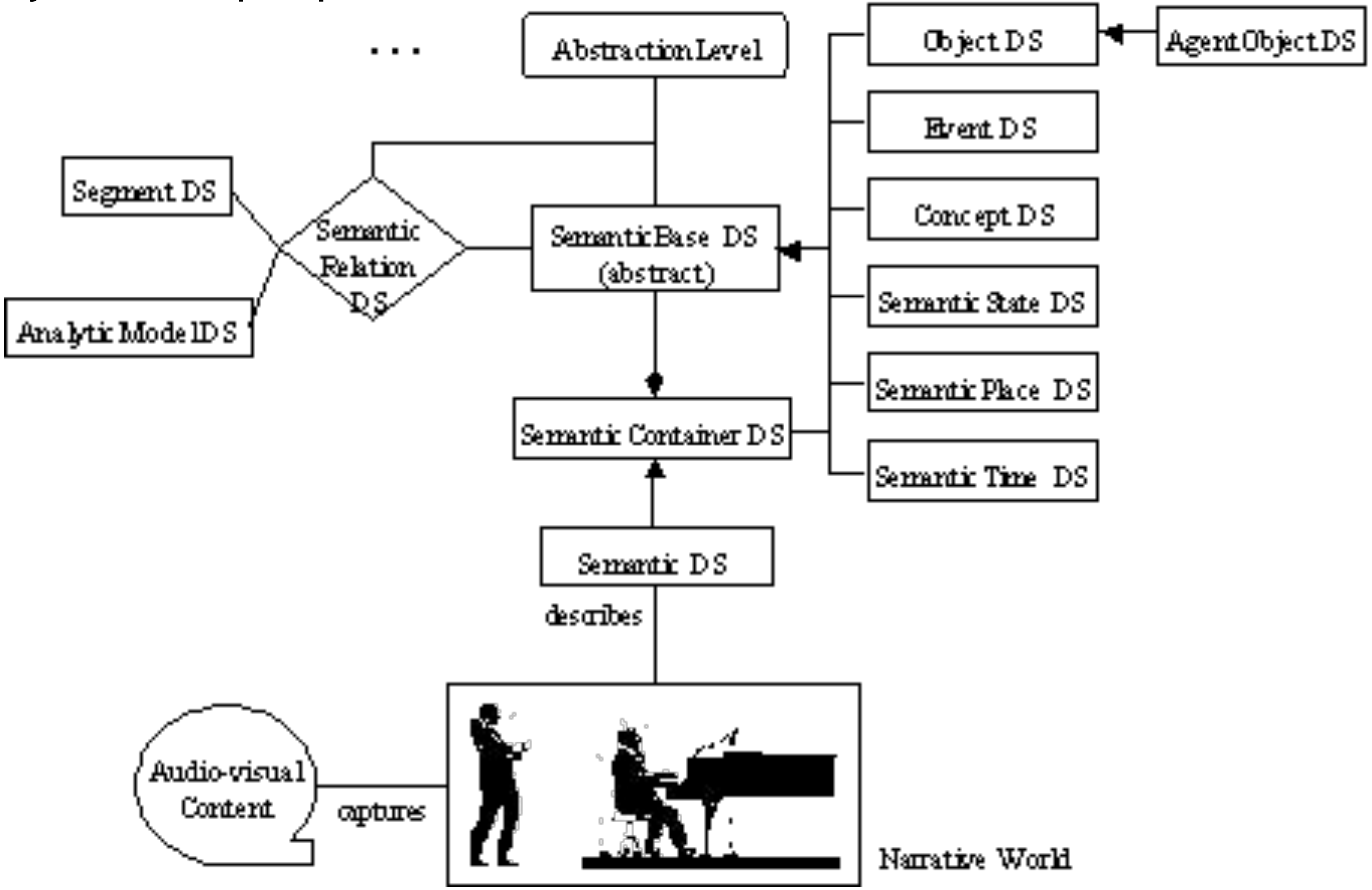
# Audiovisual Description Profile Structure



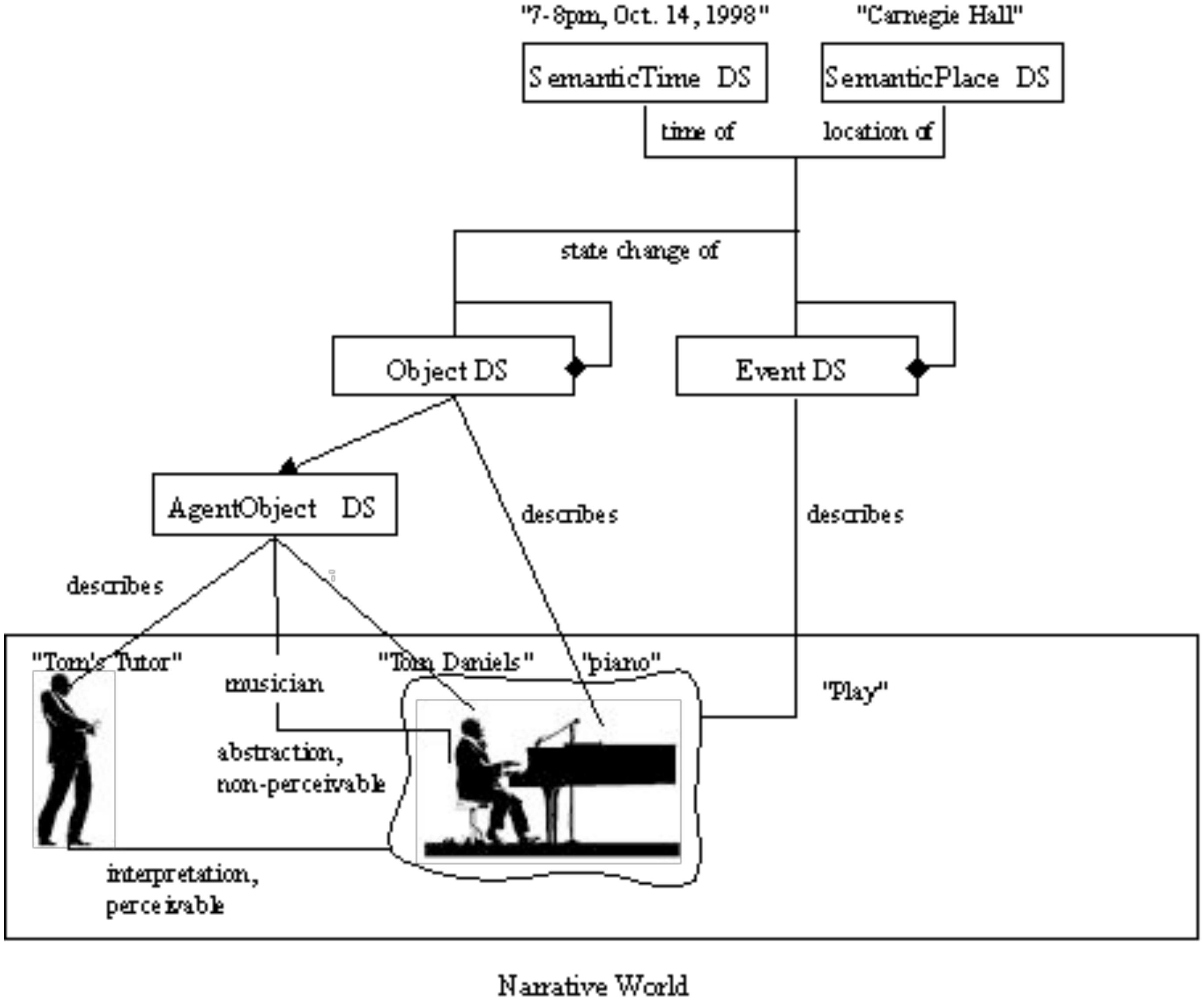
johanneum.at

# Content Semantics in MPEG-7

- Event: Occasion when something happens
  - Occurs at some time and place
  - Populated by objects and people
- “Narrative world” for a piece of content



# Content Semantics in MPEG-7: Example



# 7 Multimedia Content Description

7.1 Metadata: Concepts and Overview

7.2 MPEG-7: Generic Metadata Standard

7.3 Selected Metadata-Relevant Standards

7.4 Automated Extraction of Multimedia Metadata

# Metadata in Classic Multimedia Formats

- EXIF (Exchangeable Image File Format)
  - EXIF header for captured image or sound
  - Massively used in still-picture cameras
- ID3 for MP3
  - ID3 tag: association of information frames (each for specific metadata)
  - Predefined frames: identification, technical metadata, rights, lyrics, …
  - Extensible by new frames
- News/G2
  - Developed by IPTC (International Press Telecommunications Council)
  - XML-based specific languages: NewsML-G2, SportsML-G2, EventsML-G2, ProgramGuideML-G2, WeatherML
  - Controlled vocabularies, e.g. IPTC News Codes

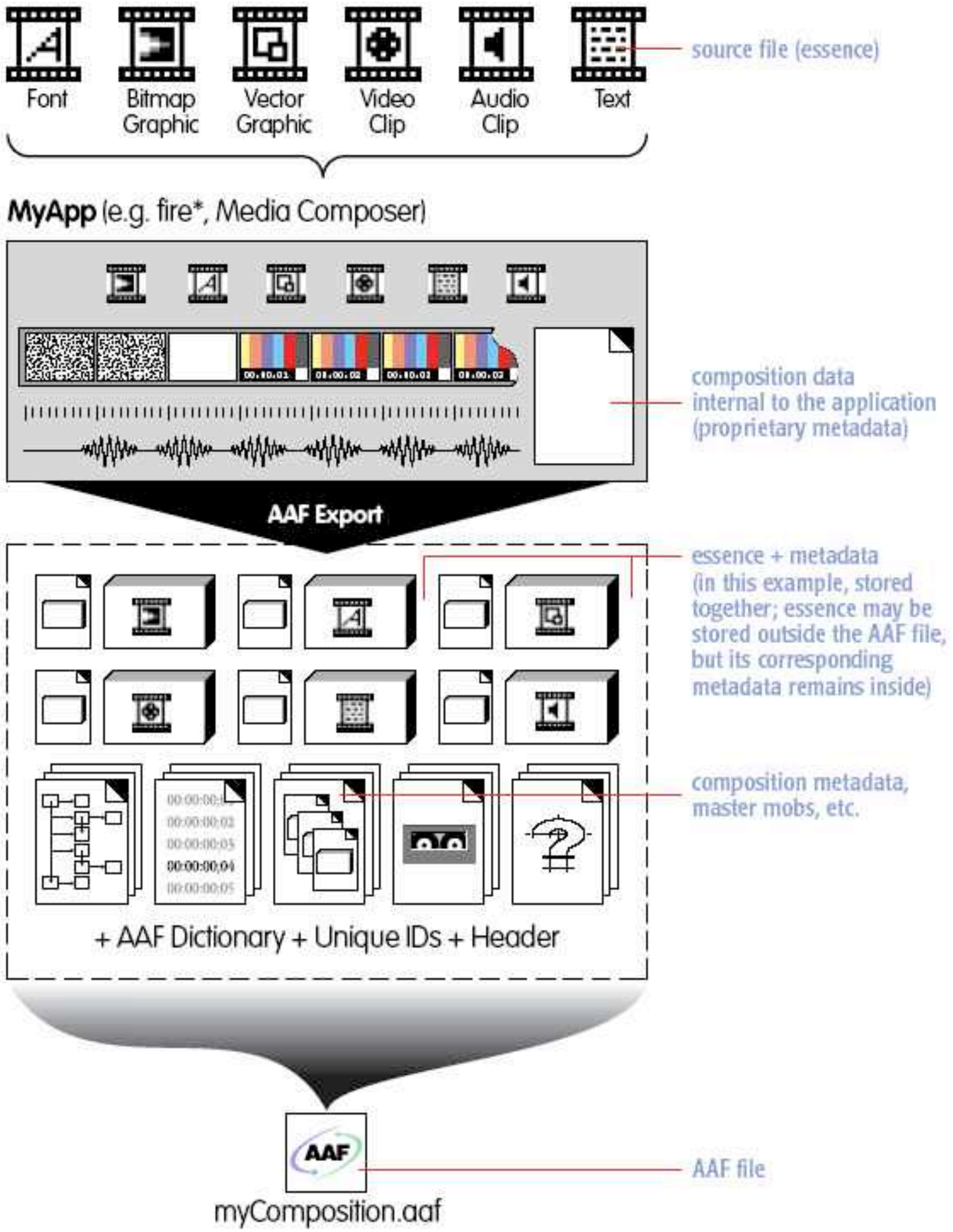
# Selected Media Metadata Standards

- Dublin Core Metadata Initiative (DCMI) &
  PRISM (Publishing Requirements for Industry Standard Metadata)
  - Oriented towards books, magazines, journals etc.
  - Uses XML, RDF, Dublin Core
  - dublincore.org, www.prismstandard.org
- TV Anytime (www.tv-anytime.org)
  - Devoted to audio-visual services making use of local mass-storage
  - Focus on Electronic Program Guide and user profiles
- EBU P/Meta
  - Devoted to material exchange between broadcasting stations
  - Vocabulary for program structure and metadata
- SMPTE Metadata Dictionary
  - Structured list of 1500 metadata elements, used e.e. in MXF format
- Commercial solutions e.g. by Rovi (www.rovicorp.com), *ex* Macrovision

# Integration of Digital Media in Video Production

- Example: Putting together all audio elements for a film soundtrack
  - Music tracks, ambient sound tracks, performer’s synchronized sound, ...
  - Metadata related to creation process need homogeneous treatment
- Standards in the broadcasting industry
  - SMPTE (Society of Motion Picture and Television Engineers)
  - EBU (European Broadcasting Union)
  - Working on hardware-based standards for a long time
- EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Program Material as Bit Streams (1996-1999)
  - Results further developed into Advanced Authoring Format (AAF)
  - AAF: Industry-driven, cross-platform, multimedia file format
  - "Advanced Media Workflow Association" (AMWA)
    - » see http://www.amwa.tv/

# Interchanging Compositions with AAF


Font
Bitmap Graphic
Vector Graphic
Video Clip
Audio Clip
Text
source file (essence)
MyApp (e.g. fire*, Media Composer)
composition data internal to the application (proprietary metadata)
AAF Export
essence + metadata (in this example, stored together; essence may be stored outside the AAF file, but its corresponding metadata remains inside)
composition metadata, master mobs, etc.
+ AAF Dictionary + Unique IDs + Header
AAF
AAF file
myComposition.aaf

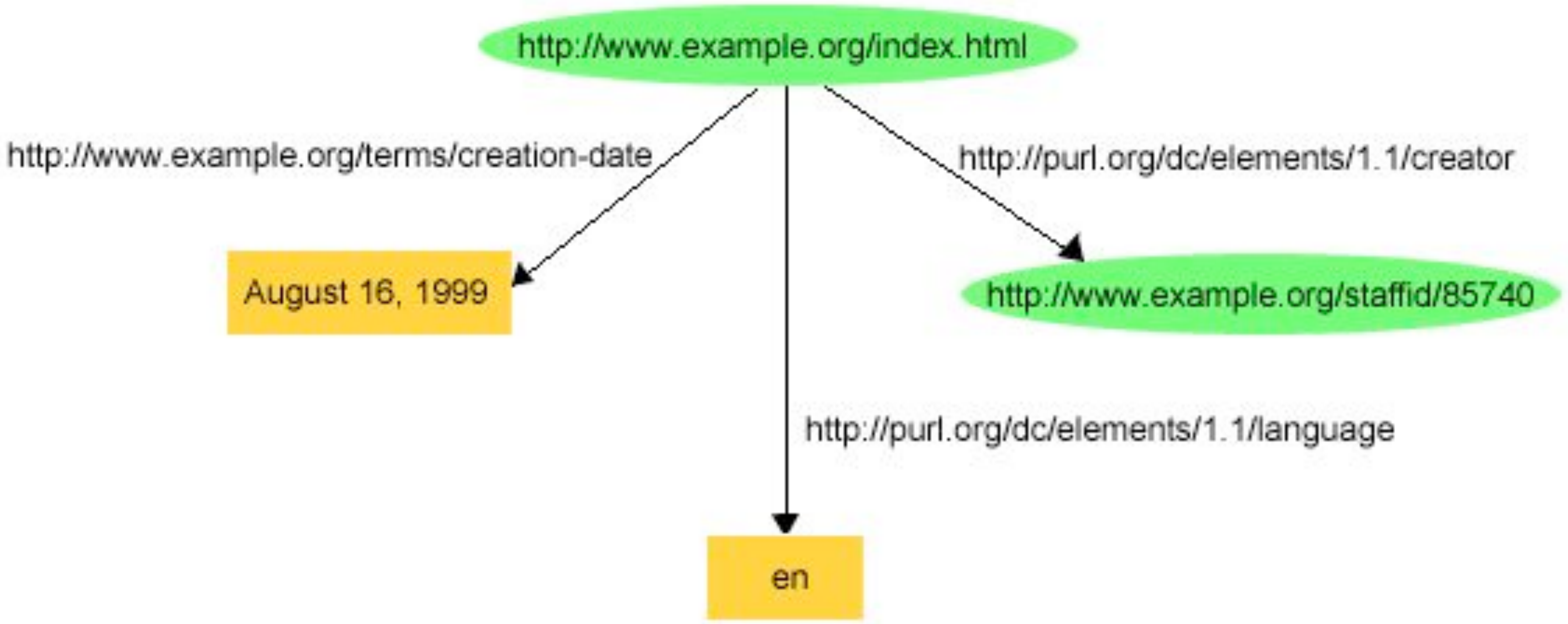
# Resource Description Framework RDF

- Language for representing information about resources in the WWW
  - W3C's Semantic Web activity

- *Resource:* Anything that can be identified by a URI (e.g. all Web pages)
- *Property:* An attribute of a described thing which can take on specific values
- *Statement:* A triple consisting of
  - *Subject:* Some resource to be described
  - *Predicate:* A property of the subject
  - *Object:* A specified value
- Properties, values and statements are resources themselves,
  - i.e. can be identified by a URI
  - i.e. can be subject to further description

# RDF Example

- **http://www.example.org/index.html** has a **creator** whose value is **John Smith**
- **http://www.example.org/index.html** has a **creation-date** whose value is **August 16, 1999**
- **http://www.example.org/index.html** has a **language** whose value is **English**


http://www.example.org/index.html
http://www.example.org/terms/creation-date
http://purl.org/dc/elements/1.1/creator
August 16, 1999
http://www.example.org/staffid/85740
http://purl.org/dc/elements/1.1/language
en

# Example: Audio Metadata in DC-based RDF/XML

RDF/XML is an XML language for representing descriptions

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description
     rdf:about="http:www.medien.ifi.lmu.de/team/
     heinrich.hussmann/files/mmn8a.m4b>
    <dc:creator>Heinrich Hussmann</dc:creator>
    <dc:title>Multimedia Content Description I</dc:title>
    <dc:description>Discusses multimedia metadata
     standards.</dc:description>
    <dc:date>2009-12-10</dc:date>
    <dc:format>audio/mp4</dc:format>
  </rdf:Description>

</rdf:RDF>
```

# Adobe XMP

- Defined by Adobe 2001, since 2007 under BSD license
- Embedding of metadata into distributed files
  - In particular into PDF
- Data model and XML-Based storage model
  - Following the RDF description principle
- Formal schema definitions for metadata properties
- Application:
  - Adobe products (e.g. Photoshop, In-Design)
  - International Press and Telecommunications Council (IPTC) has integrated XMP into its Image Metadata specifications

# 7 Multimedia Content Description

7.1 Metadata: Concepts and Overview

7.2 MPEG-7: Generic Metadata Standard

7.3 Selected Metadata-Relevant Standards

7.4 Automated Extraction of Multimedia Metadata

- Music, Images, Video

Literature:

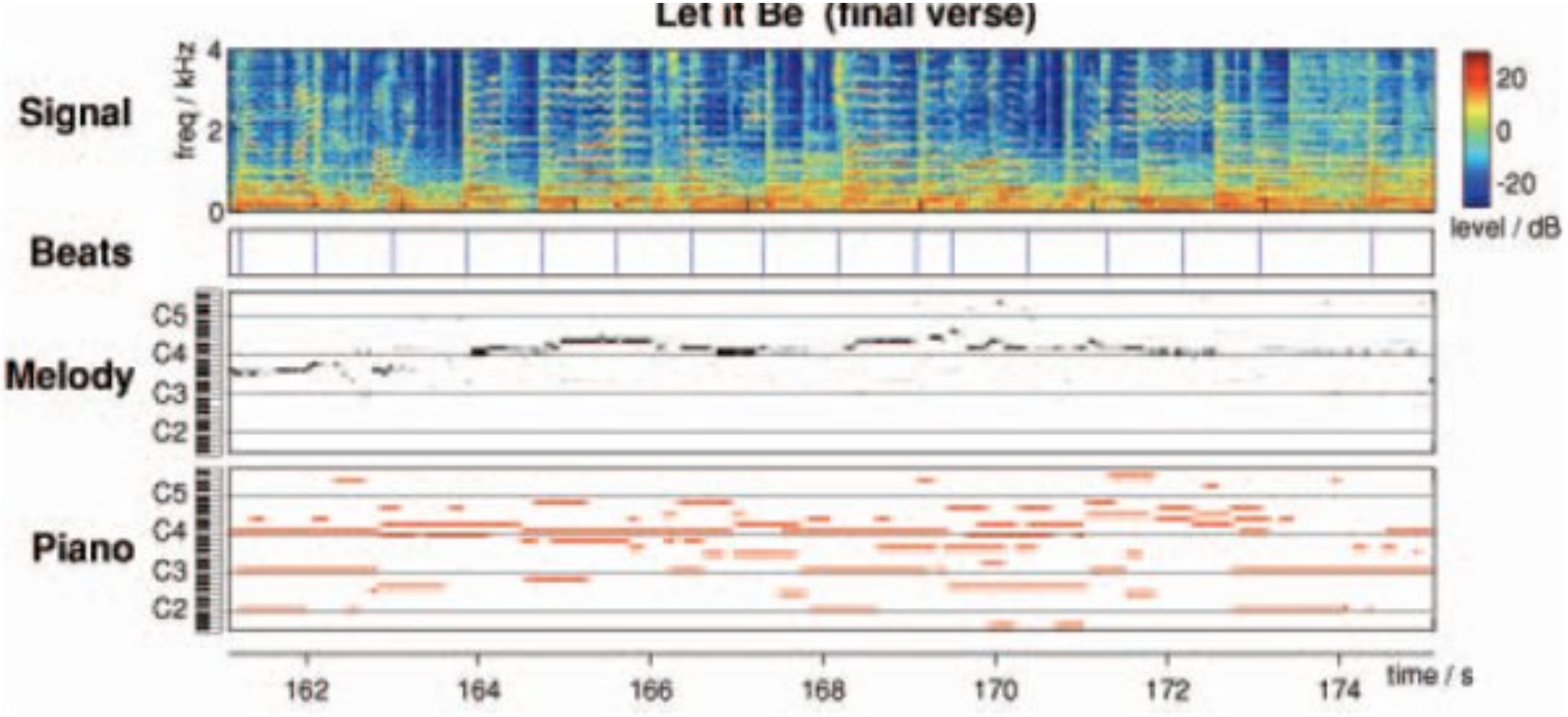
*Communications of the ACM* 49(8), August 2006,
Special section on Music Information Retrieval, pp. 28-60

S. Dasiopoulou, E. Giannakidou, G. Litos, P. Malasioti, and Y. Kompatsiaris. 2011. A survey of semantic image and video annotation tools. In *Knowledge-driven multimedia information extraction and ontology evolution*, G. Paliouras, C. D. Spyropoulos, and G. Tsatsaronis (Eds.). Springer, 196-239.

# Timescales of Musical Information

- Individual music note events
  - Extraction of the music score
  - Identification of instrument playing
- Chords (simultaneous notes)
  - Identification of chords
- Phrase level
  - Tempo extraction
  - Identification of phrases (based on repetition/alternation of segments) e.g. identification of chorus
- Piece level
  - Genre identification (“rock”, “jazz”, “classical”)

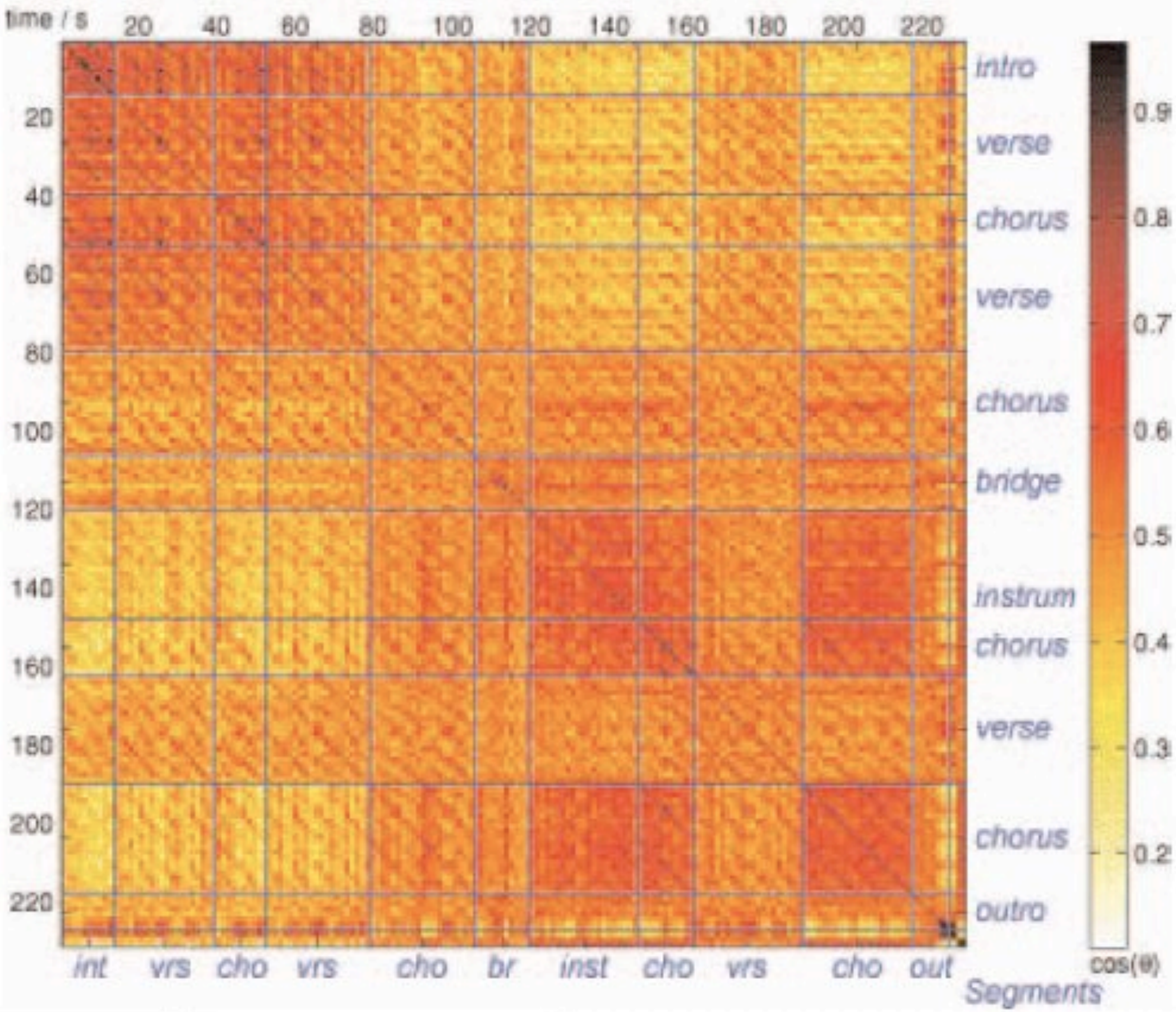
# Automatic Score Transcription



- Beats determined by tempo-smoothed event detector
- Melody recognized by general-purpose support-vector classifier
  - Trained to recognize spectral slices to be labelled with pitch values

# Automatic Phrase Detection

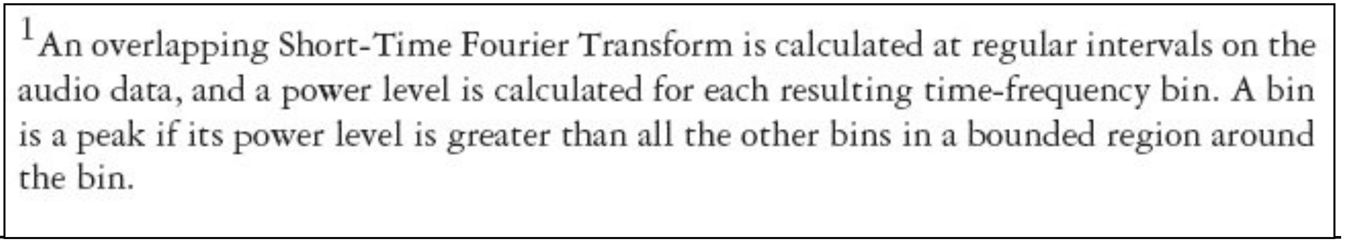
- Self-similarity matrix
  - Values represent acoustic similarity
  - Looking for diagonal ridges off the main diagonal
  - Blue lines are manually inserted for comparison



See also: http://www.fxpal.com/publications/FXPAL-PR-99-093.pdf

# Example: Shazam Music Tagging (1)

- Commercial service for mobile phones:
  Identify music from a short audio sample *(query by example)*
  - See http://www.shazam.com (London, founded 2000)
  - A. Wang: The Shazam Music Recognition Service, *Comm. ACM* Aug. 2006
- Challenges:
  - Distinguishing music from noise
  - Dealing with distortions
  - Keeping fingerprints small (in order to deal with millions of songs)
- Basic idea:
  - Spectrogram peaks (energy distribution in time and frequency)[1]
  - Few "anchor" peaks are combined with peaks in a certain surrounding zone (time and frequency offsets)
    - » Combinatorial hashing creates 32b fingerprint hash token



[1] An overlapping Short-Time Fourier Transform is calculated at regular intervals on the audio data, and a power level is calculated for each resulting time-frequency bin. A bin is a peak if its power level is greater than all the other bins in a bounded region around the bin.

# Example: Shazam Music Tagging (2)

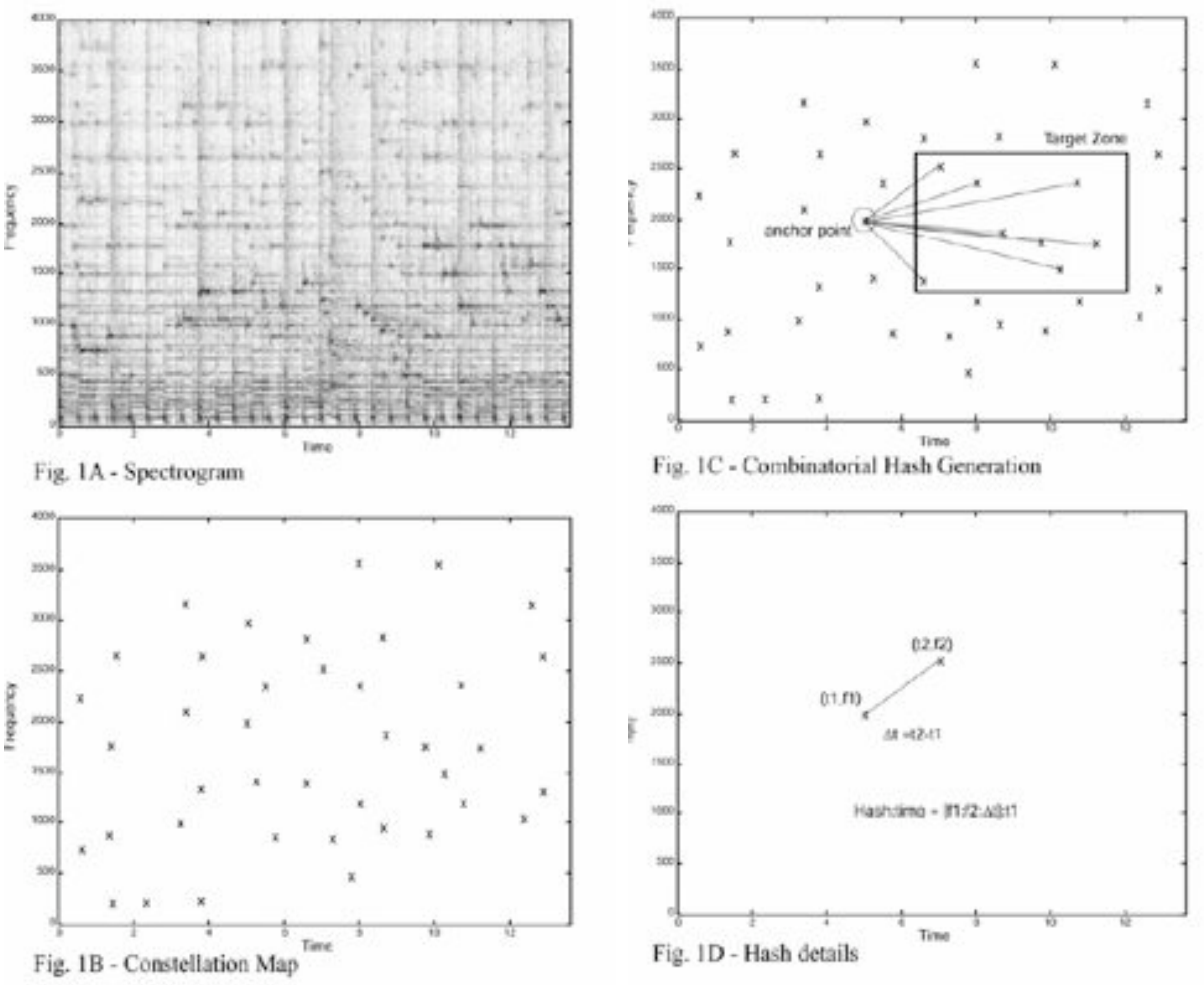


Fig. 1A - Spectrogram

Fig. 1C - Combinatorial Hash Generation

Fig. 1B - Constellation Map

Fig. 1D - Hash details

http://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

# Fingerprint Complexity Tradeoff

- Computing a more complex fingerprint:
  - Increases search time (more tokens to inspect)
  - Improves entropy
    - » Better descriptiveness distinguishes more clearly between items
- Shazam example:
  - Combinatorial expansion increases token number by factor 10 (roughly)
  - Combinatorial expansion accelerates index search by a factor of more than a million!

# Example: Shazam Music Tagging (3)

- Comparing tokens from sample and database:
  - Only tokens having peaks from target signal are relevant
  - Even presence of a few well matching tokens is significant
- Temporal alignment of fingerprint features:
  - Matching set of features must have identical relative positions in time
  - Find linear time correspondence
    - » By searching a histogram of relative time differences for peaks

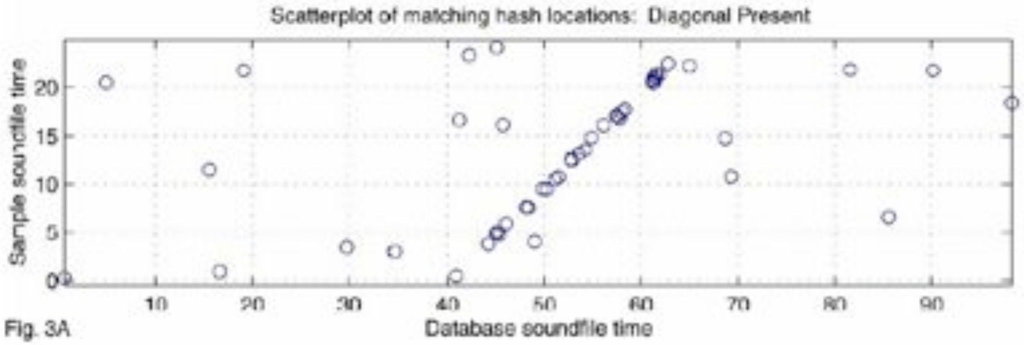


Fig. 3A

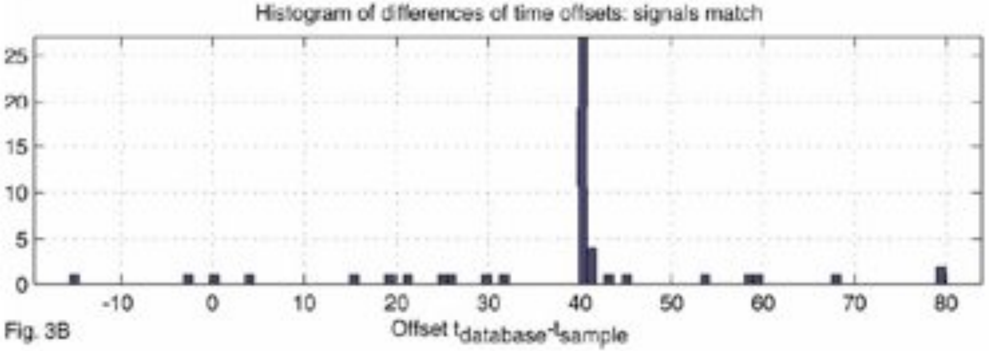
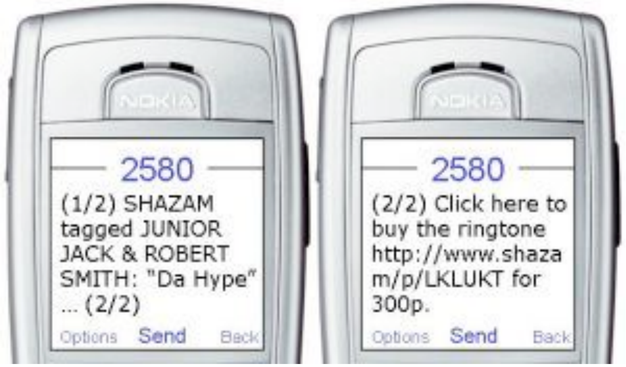


Fig. 3B

# Example: Shazam Music Tagging (4)

- Commercial situation:
  - 2012: Shazam app used 10 million times per day
  - More than 11 million tracks in database
- Without Internet connectivity (1999/2000):
  - Query via speech channel, result via text message
- Smartphone apps (Shazam/Encore)
  - Require Internet connectivity
  - Query and result via Internet
- Steady changes
  in business model:
  - Secondary content for TV
  - Music retail
  - Social music network



2580
(1/2) SHAZAM tagged JUNIOR JACK & ROBERT SMITH: "Da Hype" ... (2/2)
Options Send Back

2580
(2/2) Click here to buy the ringtone http://www.shazam/p/LKLUKT for 300p.
Options Send Back

“2580” service

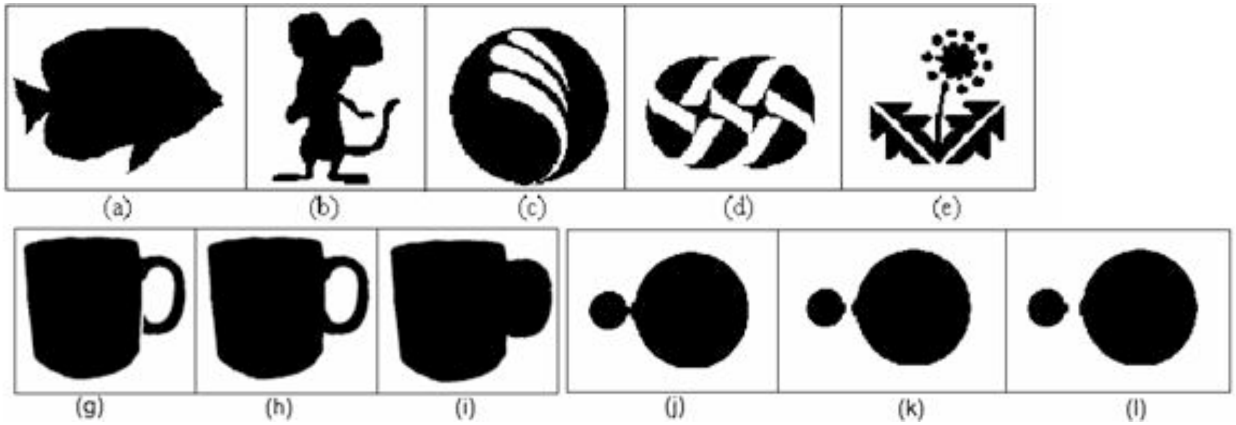
shazam.org

# MPEG-7's Sophisticated Shape Descriptors

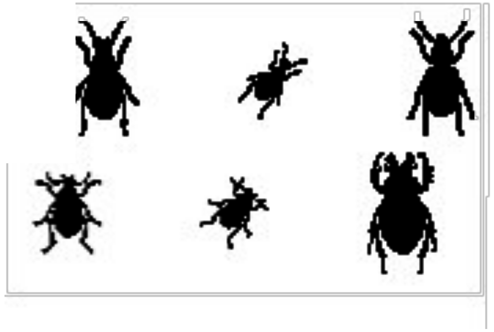
- Region shapes
  - Pixel distribution, using both boundary and internal pixels
  - Can describe complex objects with multiple disconnected regions
  - Shape analysis based on moments
    - Angular Radial Transformation (ART)
- Contour shapes
  - Based on Curvature Scale-Space (CSS) representation of contour
  - Recognized characteristic contour shapes
  - Similar to human perception
- Desirable properties of extraction methods
  - Able to handle complex shapes
  - Robust to minor deformations, perspective transformations, movement, splits, occlusions etc.
  - Compact and efficient

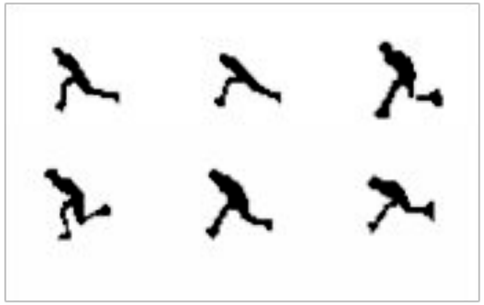
# Examples for Shape Descriptors

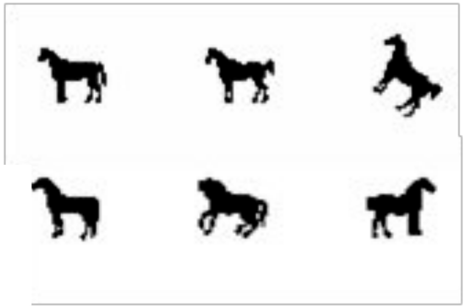
Region shapes:



Contour shapes:

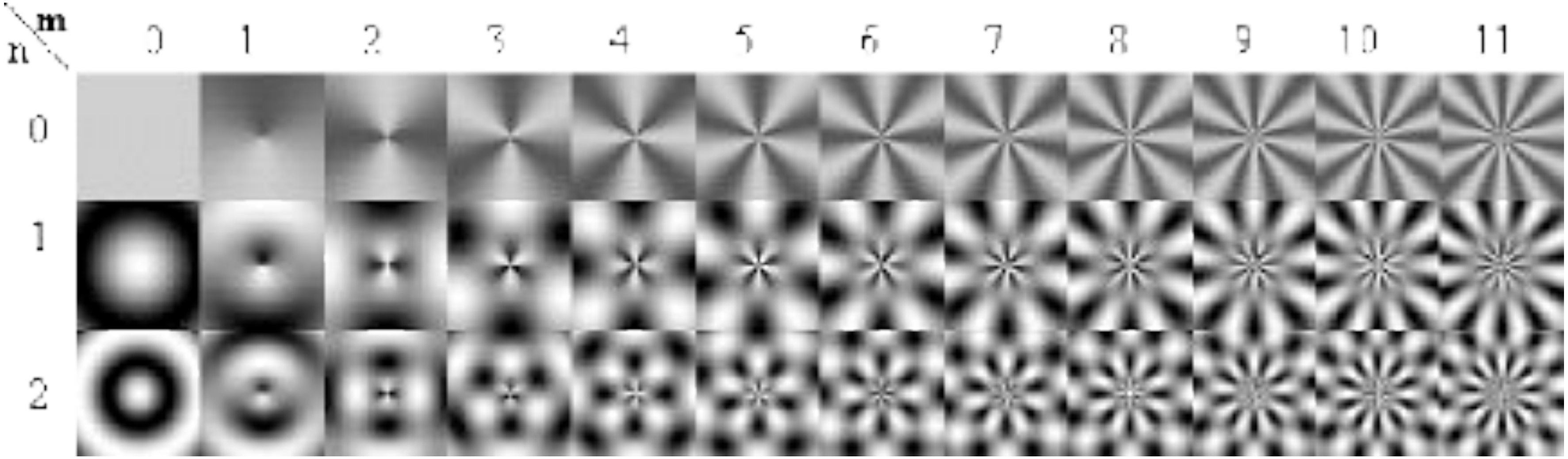






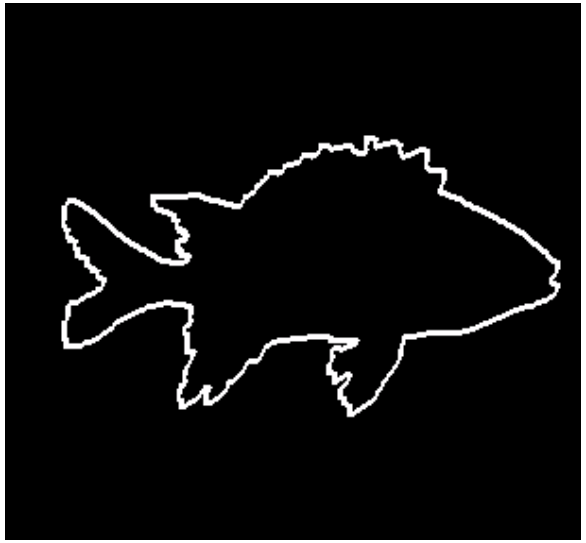
# Angular Radial Transformation (ART)

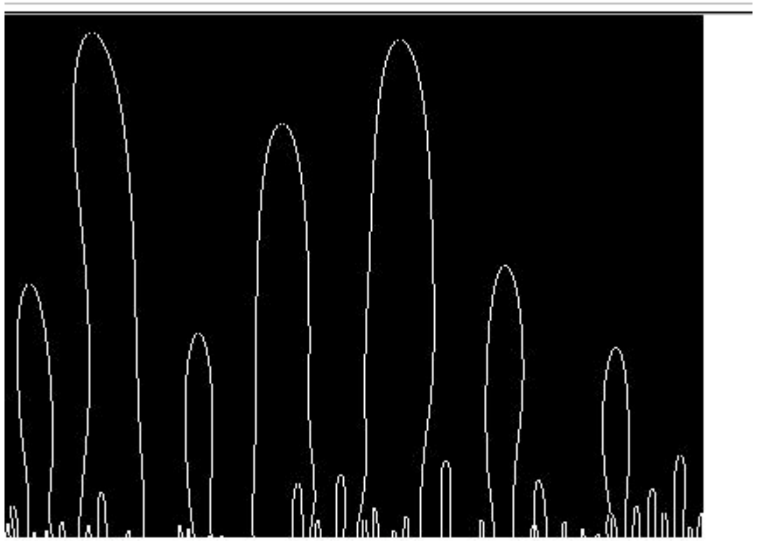
- Convert image information into angular and radial parts
- Represent image as coefficients of basis functions
- First 36 basis functions:



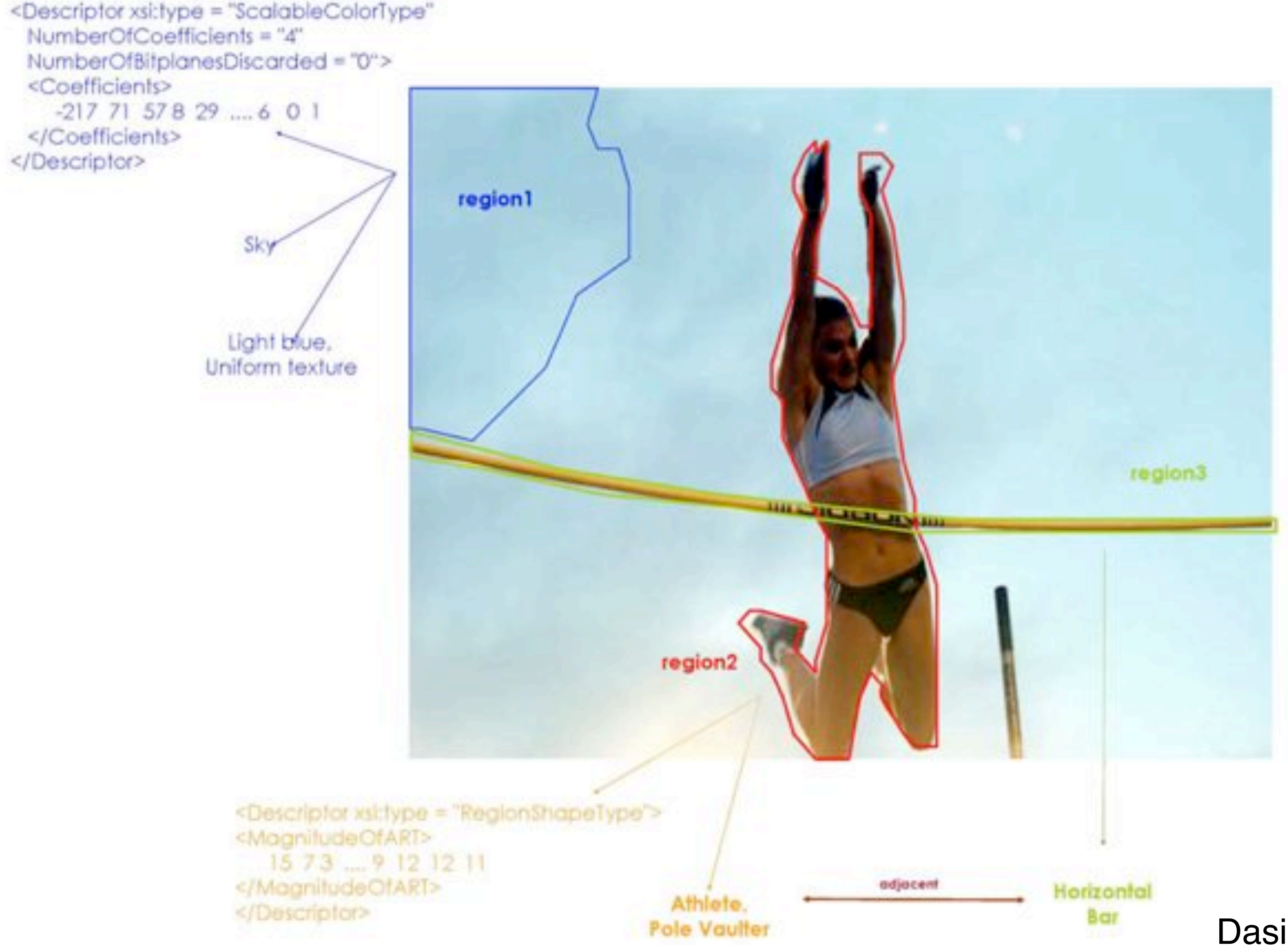
# Curvature-Scale Space Computation

- Curvature is a local measure of how fast a curvature is turning
  - Curvature zero crossing points are essential for contours
  - Contour is sampled with increasing precision and smoothed stepwise to retrieve curvature zero-crossings of various scales
- Mokhtarian, Abbasi et al., University of Surrey, UK http://www.ee.surrey.ac.uk/CVSSP/demos/css/demo.html





# Semantic Segmentation/Annotation of Images

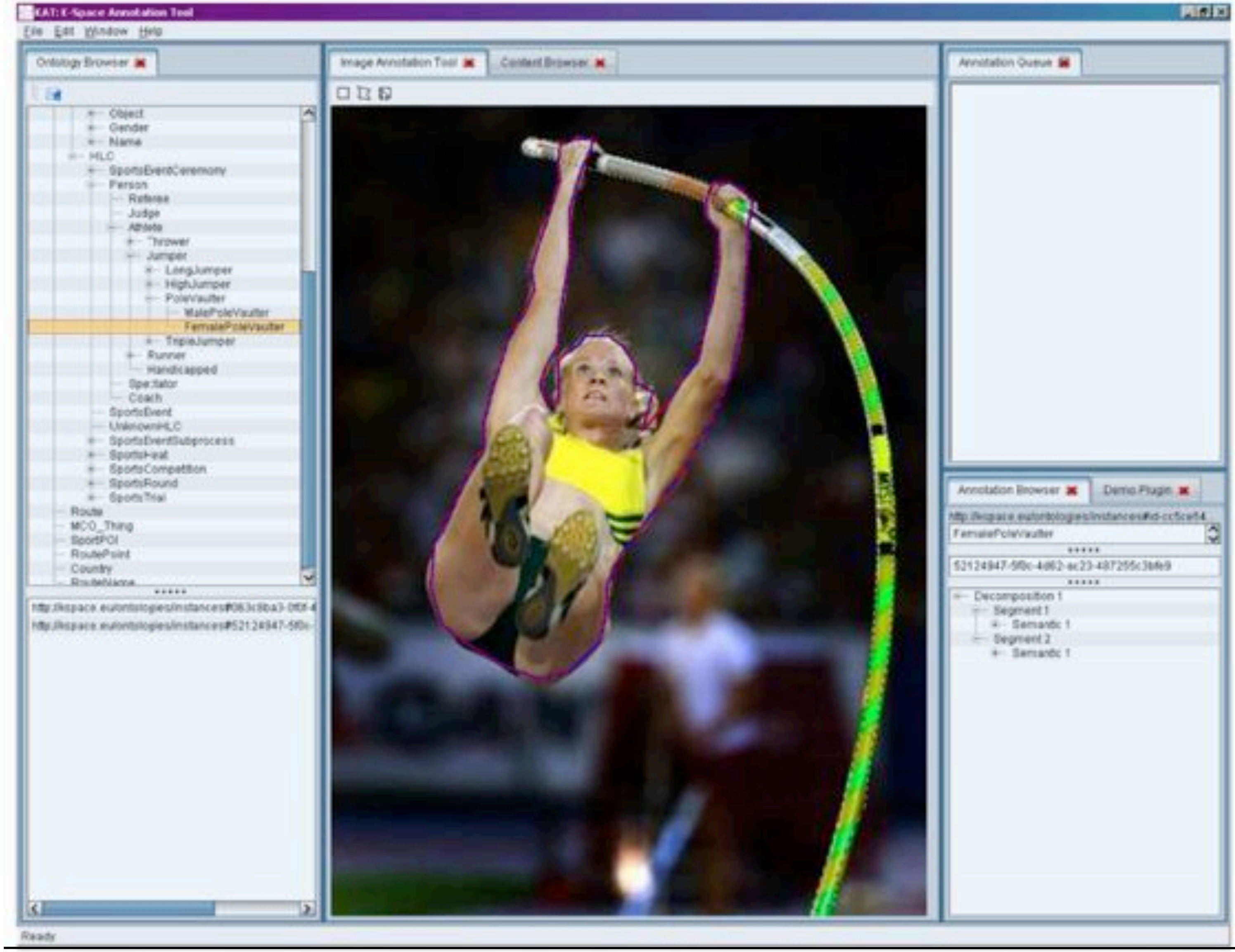


```
<Descriptor xsi:type = "ScalableColorType"
  NumberOfCoefficients = "4"
  NumberOfBitplanesDiscarded = "0">
  <Coefficients>
     -217 71 57 8 29 .... 6 0 1
  </Coefficients>
</Descriptor>
```

```
<Descriptor xsi:type = "RegionShapeType">
<MagnitudeOfART>
    15 7 3 .... 9 12 12 11
</MagnitudeOfART>
</Descriptor>
```

Dasiopoulou et al.

# Example of Visual Annotation Tool



http://kspace.eu/ontologies/instances#063c8ba3-0f0f-4

http://kspace.eu/ontologies/instances#52124947-5f0c-

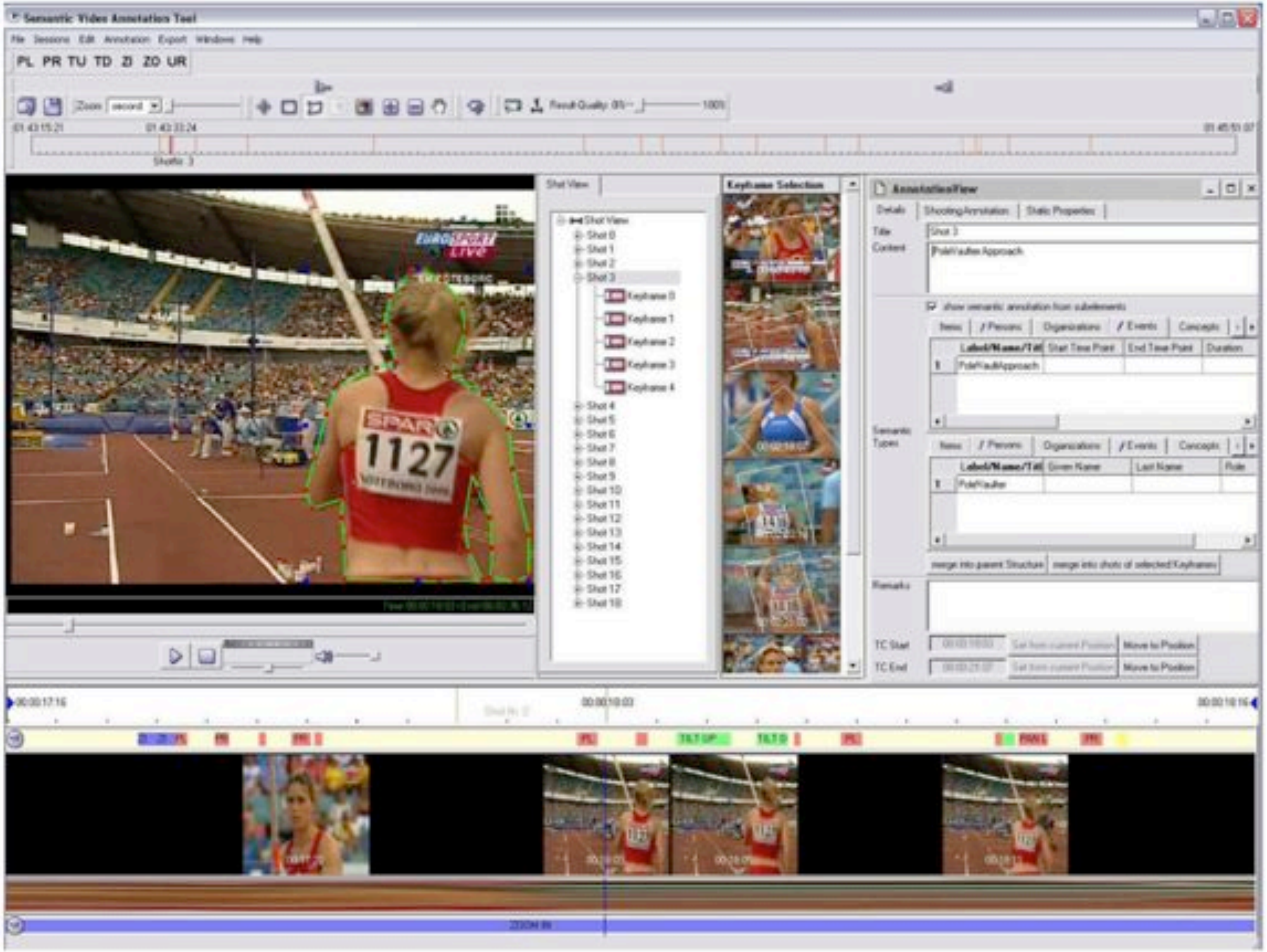
http://kspace.eu/ontologies/instances#id-cc5ce54

FemalePoleVaulter

52124947-5f0c-4d62-ac23-487255c3bfe9

KAT tool
K-Space EU project

# Example of Video Annotation Tool



Title: Shot 3

Content: PoleVaulter Approach

☑ show semantic annotation from subelements

Semantic Types

| | Label/Name/Title | Start Time Point | End Time Point | Duration |
|---|---|---|---|---|
| 1 | PoleVaultApproach | | | |

| | Label/Name/Title | Given Name | Last Name | Role |
|---|---|---|---|---|
| 1 | PoleVaulter | | | |

Remarks:

TC Start: 00:00:18:03

TC End: 00:00:21:07

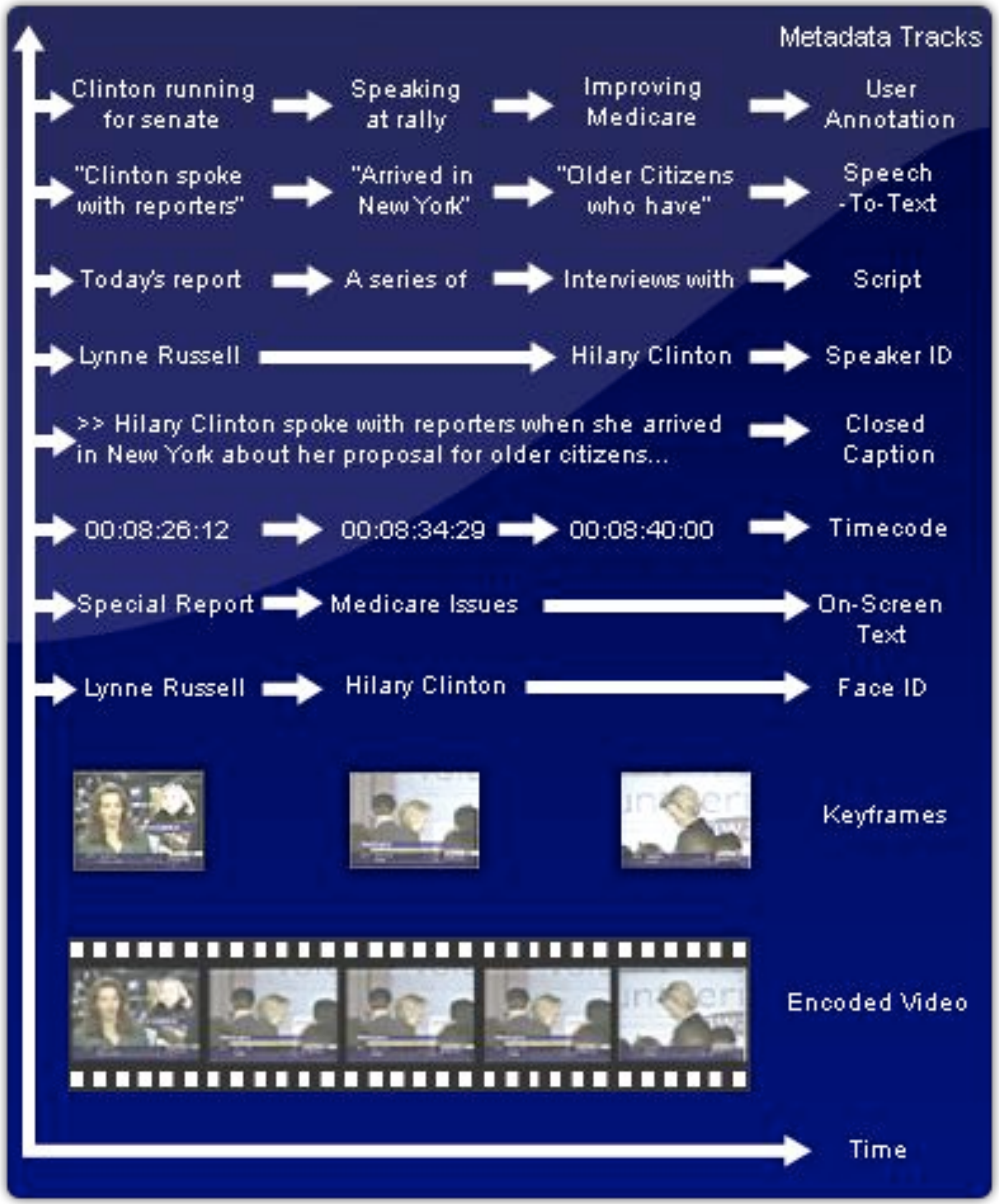
Shot View
- Shot 0
- Shot 1
- Shot 2
- Shot 3
  - Keyframe 0
  - Keyframe 1
  - Keyframe 2
  - Keyframe 3
  - Keyframe 4
- Shot 4
- Shot 5
- Shot 6
- Shot 7
- Shot 8
- Shot 9
- Shot 10
- Shot 11
- Shot 12
- Shot 13
- Shot 14
- Shot 15
- Shot 16
- Shot 17
- Shot 18

Semantic Video Annotation Suite (joanneum.at)

# Example: Autonomy Virage

www.virage.com:

"Using advanced image and audio analysis engines that watch, listen to and read a video signal in real-time, Autonomy Virage delves into the video file itself to extract the meaning of the information contained within."



Metadata Tracks

Clinton running for senate → Speaking at rally → Improving Medicare → User Annotation

"Clinton spoke with reporters" → "Arrived in New York" → "Older Citizens who have" → Speech -To-Text

Today's report → A series of → Interviews with → Script

Lynne Russell → Hilary Clinton → Speaker ID

>> Hilary Clinton spoke with reporters when she arrived in New York about her proposal for older citizens... → Closed Caption

00:08:26:12 → 00:08:34:29 → 00:08:40:00 → Timecode

Special Report → Medicare Issues → On-Screen Text

Lynne Russell → Hilary Clinton → Face ID

Keyframes

Encoded Video

Time

# Example: Autonomy Virage / DVI / IDOL

- DVI = Deep Video Indexing
- IDOL = Intelligent Data Operating Layer

"Openly configurable, the DVI functionality uses numerous proprietary approaches in a configurable lattice which can be weighted within the DVI fingerprint as separate entities or as a whole unit.

These technological approaches include:

- Texture trajectory analysis
- Advanced Optical Character Recognition (OCR)
- Spectrum trajectory analysis
- Advanced scene analysis

Autonomy Virage can extract a comprehensive range of data from multimedia resources, including full transcripts of audio streams, on-screen character recognition, keyframes, facial recognition and speaker information, all of which is linked to the original video file, allowing users to locate content with pinpoint accuracy. Utilizing the power of IDOL, Autonomy Virage provides users with an unrivaled range of video analysis tools such as scene detection, 'find similar' functions and conceptual analysis such as automatic hyperlinking of related content, categorization and clustering."

www.virage.com

# Automated Media Composition and Repurposing?

- Sufficient metadata annotation enables automated semantic decisions
  - Combination of media elements
  - Repurposing, e.g. creating adequate still picture from video
  - Nack, F. (2004) The Future in Digital Media Computing is Meta. *IEEE MultiMedia, Vol 11, No. 2, pp. 10-13*
- Serious open research issue: *How to express the overall structure of a composed piece of media?*
  - Grammar-like system to express the semantic and aesthetic composition
    - » For movies
    - » For general multimedia productions
  - *"Applied media aesthetics"* (H. Zettl)
  - M. Davis: Media Streams - An iconic language for video representation, 1995