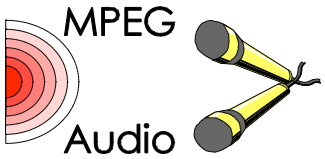


Workshop W14 - Audio Gets Smart:

Semantic Audio Analysis & Metadata Standards

Jürgen Herre
Fraunhofer Institut for Integrated Circuits (FhG-IIS)
Erlangen, Germany

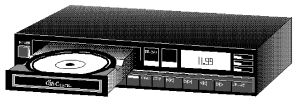
Overview

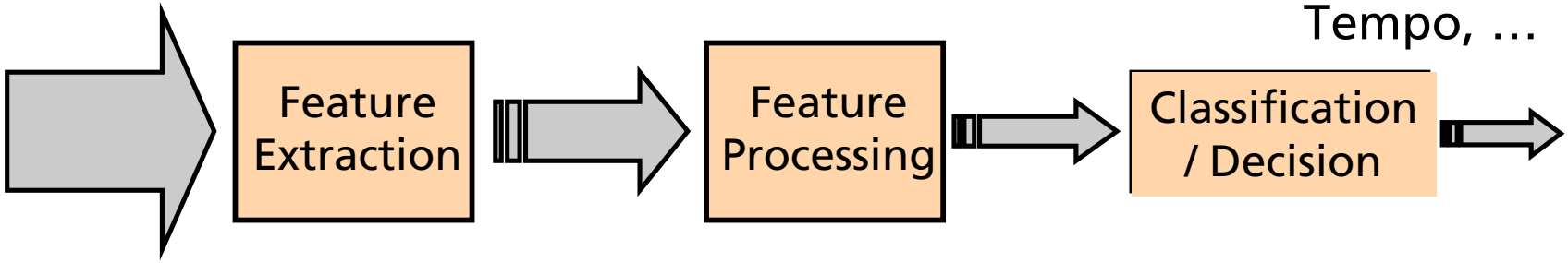
- Extracting meaning from audio
- Why semantic description / metadata?
- Why standardization?
- The MPEG-7 Standard
- MPEG-7 Audio
  - Tools
  - Some Applications
- Conclusions

Fraunhofer Institut
Integrierte Schaltungen

Extracting Meaning From Audio

Example: Generic pattern matching system - data flow



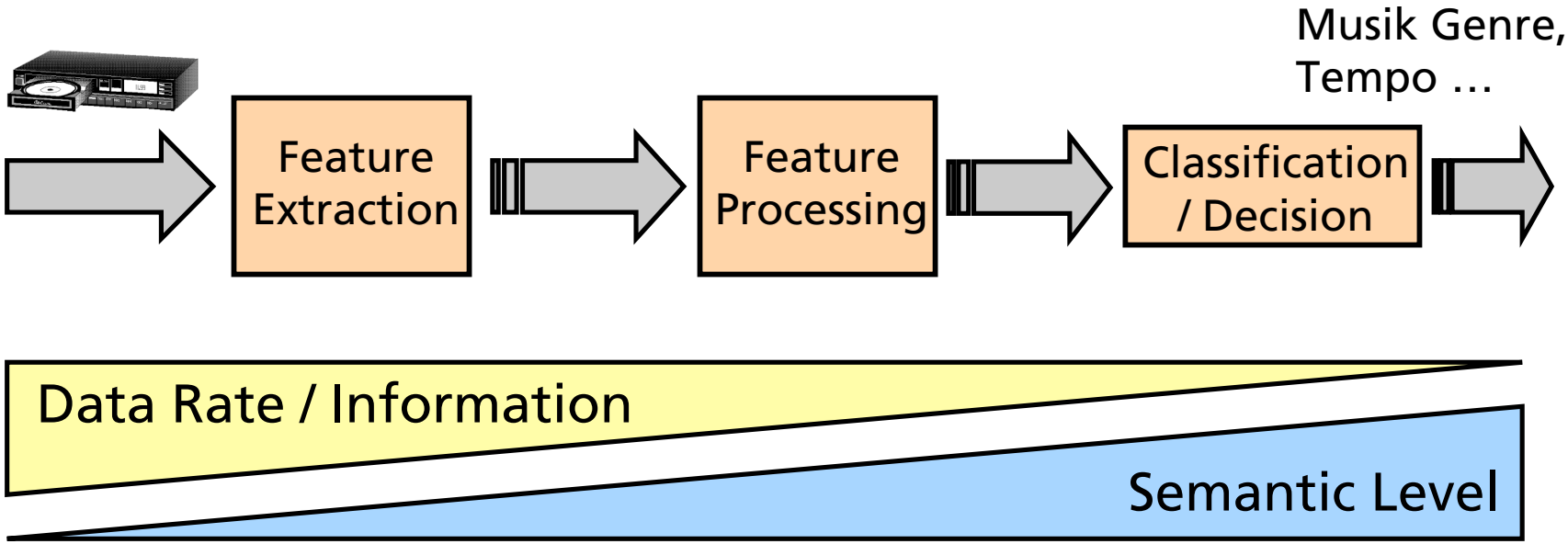


Audio data (PCM)		Features		Clustered Fs.		Output
1.4 Mbit/s	□	10kbit/s	□	100bit/s	□	1bit/s

Fraunhofer Institut
Integrierte Schaltungen

Extracting Meaning From Audio (2)

Data reduction & semantic level perspective



What is the “optimum” type of description?

Fraunhofer Institut
Integrierte Schaltungen

Why Semantic Description / Metadata?

Challenges of the "Multimedia Age"

- How to enable efficient search, filtering, management of multimedia content ?
- Today: Text search (e.g. popular search engines, such as Google, Yahoo, Lycos, ...)
- "Find audio / images / movies just as easily by simple intuitive queries"!

Use semantic analysis directly?

- Extraction / semantic analysis of content
  - is usually complex & costly
  - requires access to raw A/V data (not always possible; high bandwidth demand)
- Not everything can be "extracted" from the content itself (e.g. labeling "metadata")

Fraunhofer Institut
Integrierte Schaltungen

The Semantic Description / Metadata Idea

Approach

- Supplement A/V data with Content Description (Metadata = "data about data")
- Holds semantic descriptions of any level
- Search metadata rather A/V content itself
  - Key to efficient content handling!

Descriptions

- enable search & retrieval / filtering
- enable intelligent navigation
- summarization
- ...

2 Key Areas of Development

- Efficient automatic production and usage of descriptive metadata
- Standardization of metadata formats

Fraunhofer Institut Integrierte Schaltungen

Why Standardization?

Imagine there is …

- metadata available for the existing A/V resources (e.g. on the web)
- a common way of representing semantics
  - which key parameters to use
  - how to represent & "package" them

Then …

- *any* multimedia application could use & exploit these resources!! (incl. search & recommendation engines, user agents, …)

Key Aspect

- Interoperability between metadata DBs and applications on a world wide scale

⇒ Standardization!

Fraunhofer Institut
Integrierte Schaltungen

MPEG-7 or: “This is not about source coding!”

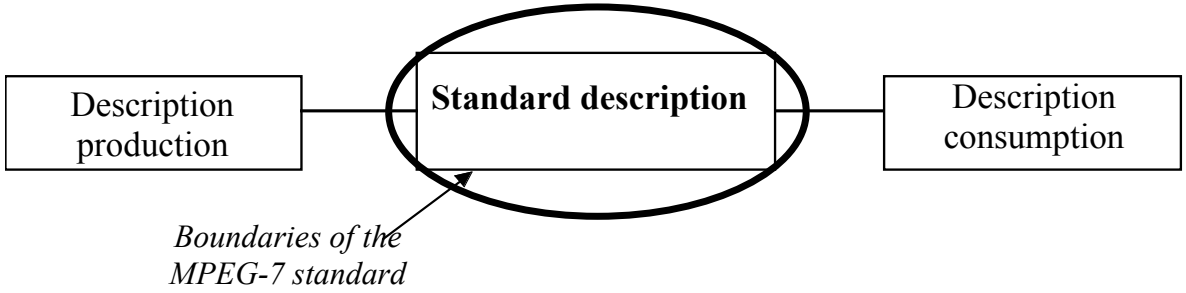
MPEG	“Moving Pictures Expert Group”, ISO/IEC JTC1/SC29/WG11
MPEG-1 (1992)	• First generic audio coding standard, Layers 1-3, (DAB, Worldspace, DVB, Internet Audio,”MP3”)
MPEG-2 (1994)	• Extending MPEG-1 coders towards lower sampling rates & multi-channel...
MPEG-2 AAC (1997)	• More powerful mono ... multi-channel coding
MPEG-4 (1999+)	• New functionalities (scalability, object oriented representation, interactivity ...)
MPEG-7 (2001+)	• “Multimedia Content Description Interface” Metadata standard (not compression!)

Fraunhofer Institut
Integrierte Schaltungen

The MPEG-7 Standard

Scope

- Standardization of “description language”:



- Description language based on XML / XML Schema (and binary versions of it)
- Provides elaborated concepts for describing
  - Generic properties of A/V data (“metadata”)
  - Visual data (content-derived)
  - Audio data (content-derived)

unique!

Fraunhofer Institut
Integrierte Schaltungen

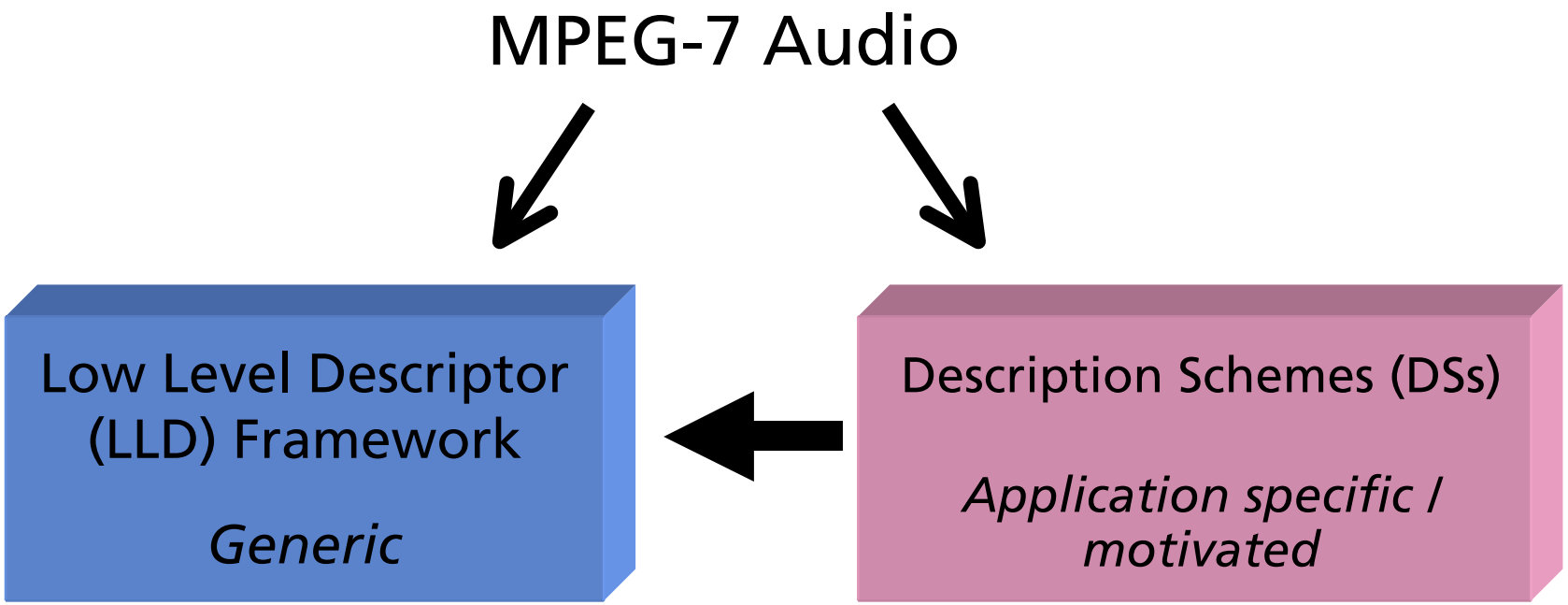
MPEG-7 & MPEG-7 Audio

ISO/IEC JTC1/SC29/WG11 IS 15938
"Multimedia Content Description Interface":

Part 1	• Systems
Part 2	• Description Definition Language (DDL)
Part 3	• Visual
Part 4	• Audio
Part 5	• Multimedia Description Schemes
Part 6	• Reference Software
Part 7	• Conformance Testing
Part 8	• Extraction & Use of MPEG-7 Descriptions

Fraunhofer Institut
Integrierte Schaltungen

MPEG-7 Audio: Basic Architecture



Fraunhofer Institut
Integrierte Schaltungen

Low Level Descriptor Framework (1)

- Collection of 17 Descriptors
- Usefulness shown in Core Experiments (CEs) for specific applications

Generality

- Generic signal characteristics
  - applicable to a wide range of signals
  - useful for many purposes
- Toolbox: Future applications can make use of them in flexible ways (form DS according to their specific needs)
- Provides basic layer of interoperability

Fraunhofer Institut
Integrierte Schaltungen

Low Level Descriptor Framework (2)

Basic	• Instantaneous waveform, power, silence
Basic Spectral	• Power spectrum, spectral centroid, spectral spread, spectral flatness …
Signal Parameters	• Fundamental frequency, harmonicity
Spectral Basis Representations	• Used primarily for sound recognition, projections into low-dimensional space
Timbral Temporal	• Log attack time, temporal centroid of a monophonic sound
Timbral Spectral	• Features specific to the harmonic portions of signals (harmonic spectral centroid, spectral deviation, spectral spread, …) etc.

Fraunhofer Institut
Integrierte Schaltungen

MPEG-7 Audio Version 2 Additions (2003)

Additions

- Audio signal quality description:
  - Background Noise, channel cross-correlation, relative delay, balance, DC offset, bandwidth, transmission technology
  - Errors in recordings (clicks, clipping, drop outs, …)
- Musical tempo [bpm]

Extensions

- Description of stereo / multi-channel signals
- Extensions for spoken content description

Fraunhofer Institut
Integrierte Schaltungen

Some MPEG-7 Audio Applications

- Spoken Content Search
- General Sound Recognition & Indexing
- Audio Archiving and Restoration
- Instrument Timbre Search
- Melody Search / Query by Humming
- Robust Audio Identification / Fingerprinting

Fraunhofer Institut
Integrierte Schaltungen

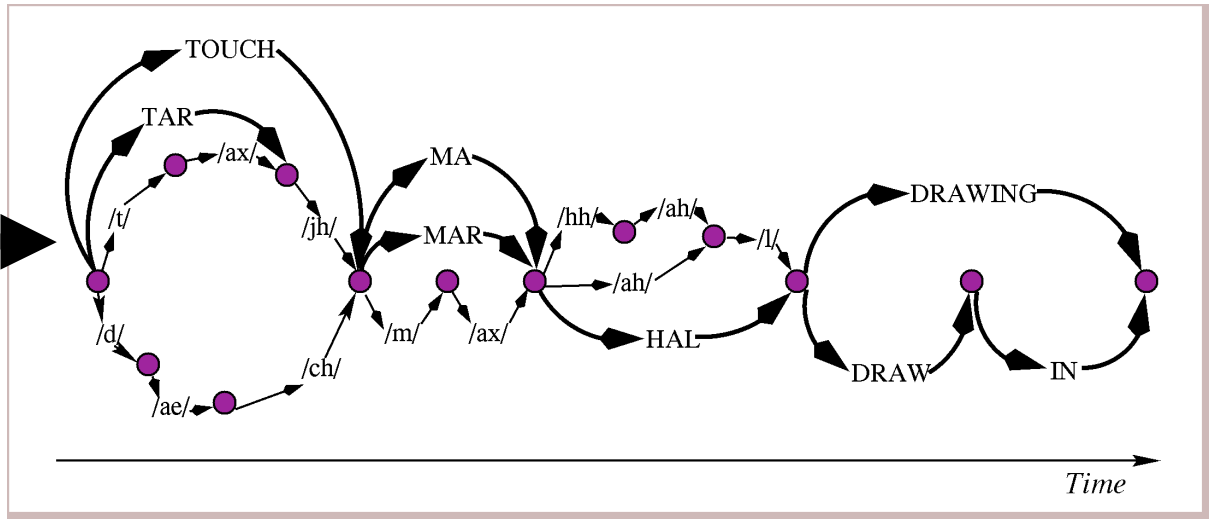
Application: Spoken Content Search

Context

- Speech is most important means of communication for humans
- Today's automatic speech recognizers represent speech internally as *word & phone lattices*

"Taj Mahal Drawing"

(example by Phil Garner)



Fraunhofer Institut
Integrierte Schaltungen

Application: Spoken Content Search (2)

Approach

- Store lattices as descriptive data (rather than plain text)
- Use for querying/matching
  - Allows handling of ambiguities (“recognize speech” vs. “wreck a nice beach”)
  - Allows even retrieval of words unknown at annotation time

Applications

- Spoken document retrieval
- Annotation & retrieval using speech (e.g.: personal photo database with spoken comments a la “the kids at the beach”)

Fraunhofer Institut
Integrierte Schaltungen

Application: General Sound Recognition/Indexing

Context

- How can various sounds be recognized, even within complex sound mixtures ?
- How to address sound similarity problem?

Approach

- Independent Component Analysis of spectral components (ICA), HMM

Applications

- A/V indexing & search by audio (index/find scenes with animals, gun fights, laughter, explosions, …)
- Determine scene content/context from audio portion:
  Gun shot / explosion action scene

Fraunhofer Institut
Integrierte Schaltungen

Application: Archiving and Restoration

Context

- Preservation of cultural heritage (audio)
- Archivists always like to keep *raw* digitized recording for later processing & restoration

Approach

- Audio quality description encodes:
  - General audio quality
  - Technical recording parameters
  - Location & type of defects

Applications

- Search for archived audio material
- Restoration of audio material (may use latest restoration technology)

Fraunhofer Institut
Integrierte Schaltungen

Application: Instrument Timbre Search

Context

- Human perception of sound includes timbre (besides pitch, loudness, duration)

Approach

- Use perceptually relevant features to describe monophonic isolated instrument sounds enables *comparison of sounds*
- Considers sustained harmonic & percussive instrument sounds

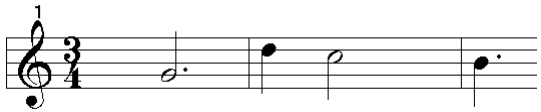
Applications

- Navigation through / search in instrument sound databases (“Giga-samplers”)
- Content-based editing

Fraunhofer Institut
Integrierte Schaltungen

Application: Melody Search

Context



- How to search for melodies?
  - Needs to be both efficient and tolerant to pitch & timing imprecisions

Approach

- Melody description encodes:
  - Pitch
  - Timing/rhythmic information
  - Other optional information (key …)

Applications

- “Query by Humming / Playing / Singing”
- Search for themes in music catalogs
- Musicology

Fraunhofer Institut
Integrierte Schaltungen

Application: Audio Identification/Fingerprinting

Context

- How to automatically recognize/identify a recording?

Approach

- Store unique MPEG-7 signature / fingerprint of original items in DB (robust, compact)
- Identification of unknown audio material by *matching* its signature with the DB signatures

Audio Identification/Fingerprinting (2)

Applications

- Music recognition on handheld devices (PDA, cell phone), …
- Content search on the Internet
- Broadcast monitoring
- How to find metadata for a given piece of legacy content (CD, CC, VCR Tape, ... no metadata attached!)?

Fraunhofer Institut
Integrierte Schaltungen

Conclusions

Notion of *"Metadata"*

- Essential to allow efficient handling of A/V data for real-world applications
  - Search&retrieval, summarization, navigation …
- Container for semantic audio analysis output
- Interoperability is crucial

MPEG-7 Audio

- First international standard providing content-derived audio metadata unique!
  - Generic descriptor toolbox
  - Support for a number of attractive applications
- Large potential for more, still unconceived applications …

Fraunhofer Institut
Integrierte Schaltungen