Robust Matching of Audio Material

An MPEG-7 Audio Functionality

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

Basic Idea

- “Oh - I know this tune - wait a second ...”

Fraunhofer Institut
Integrierte Schaltungen

Content-based Identification of Audio Material

Idea

- Mimicking human recognition capability
- Automatic identification of audio material by *robust matching* with previously registered audio content

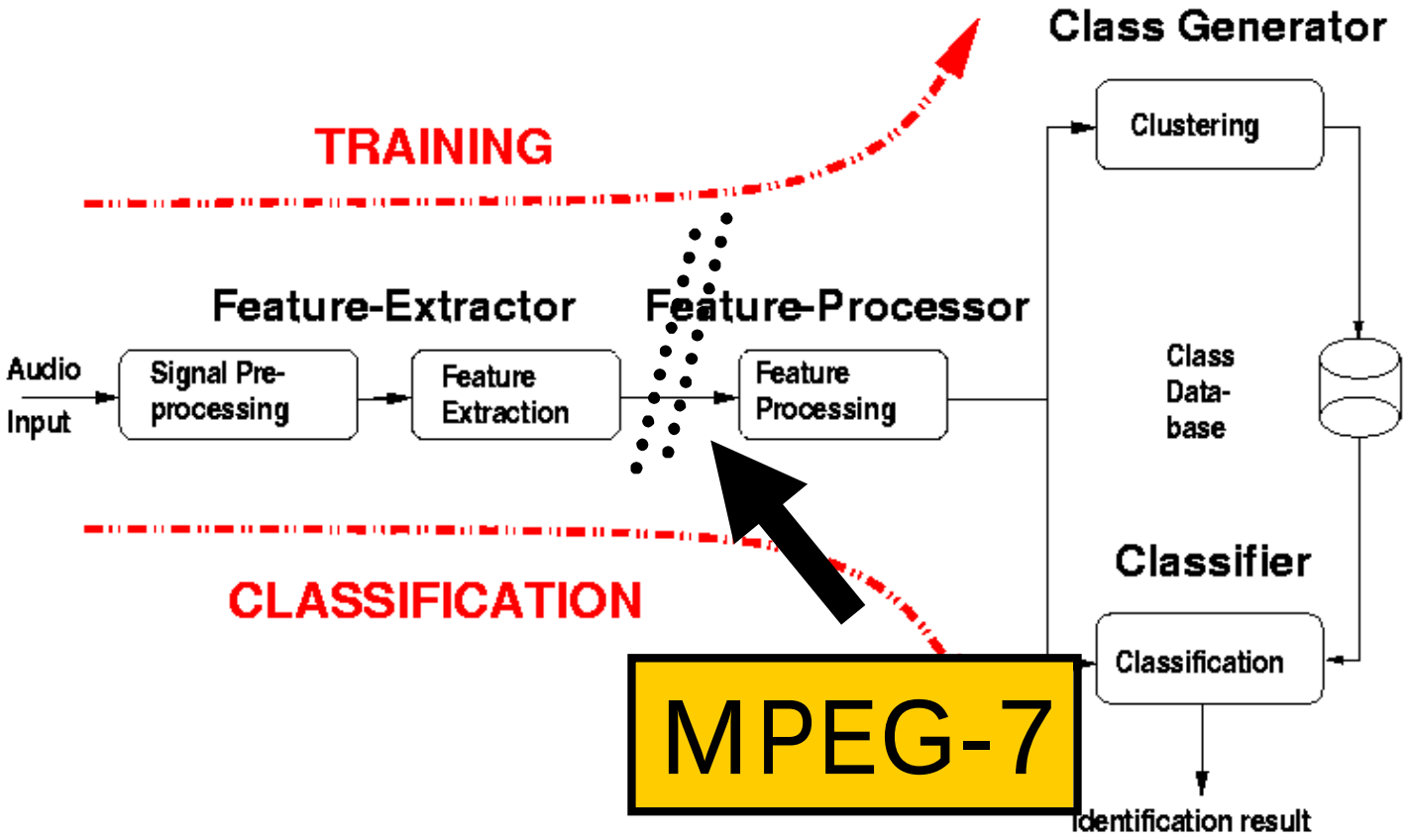
Why "robust" matching ?

"Real-world distortions" happen:

- Linear distortion (level change, filtering, EQ, band limiting, ...)
- Non-linear distortion, incl. compression (MP3, ...)
- Availability of excerpt only (cropping)

Fraunhofer Institut
Integrierte Schaltungen

System Overview (Example)



Fraunhofer Institut
Integrierte Schaltungen

Robust Audio Matching in MPEG-7

MPEG-7 Element

- Low Level Descriptor (LLD): `AudioSpectrumFlatness()`

Background

- Describes spectral flatness properties in several frequency bands (tonal ↔ noise)
- Very robust w.r.t. distortions
  → see Paper M-6, Tuesday 10:30
- Compact: e.g. 4 values/s @ 8bit/value
  - Note: Compact binary representation of descriptor values not (yet) part of MPEG-7
- Basis for *audio fingerprint*: Interoperability & openness ensured by ISO standard

Fraunhofer Institut
Integrierte Schaltungen

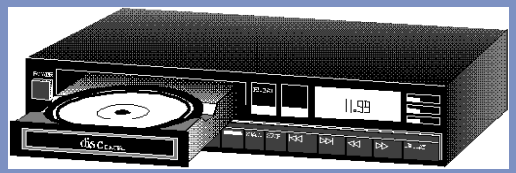
Applications

- Search for specific audio content (e.g. on the Internet)
- Broadcast monitoring
  - Transmission protocols
  - Charts analysis
- Music Sales (find your favorite songs)
- „Audio Fingerprinting”
- Linking music to metadata
  - How to find metadata associated to a piece of audio content ? (today: via CDDB, ID3v2, ...)

Fraunhofer Institut Integrierte Schaltungen

Linking MPEG-7 To the Legacy Audio World

Audio Content
(legacy format)



"Matching"

Descriptions
(MPEG-7)

Title, Artist, Composer
Melody, Lyrics, . . .

Audiospectrum
Flatness
"Fingerprint"

Fraunhofer Institut
Integrierte Schaltungen

Some Performance Results

Current recognition performance

- 15,000 items pop/rock database (30s each)
  - No distortion 100.0% (100.0%)
  - Cropping (20s) 99.9% / (100.0%)
  - MP3 (96kbps st.) 100.0% / (100.0%)
  - MP3 & Cropping 99.7% / (100.0%)
  - Loudsp./Microph. 99.3% / (99.7%)

  (Performance in Top1% (Top10%))

Recognition speed

- Ca. 80x faster than real-time on Pentium class PC @ 500Mhz (0.25s)

Signature size

- 15 MByte (1kbyte/item 30s)

Fraunhofer Institut
Integrierte Schaltungen

Real-Time Demonstration

- Standard Laptop (Pentium II @ 333 MHz)
- Database of trained items:
  - 1000 audio items
  - Mostly from Rock / Pop genre
  - 60s excerpts each
- Robustness against acoustic transmission (D/A → Loudspeaker → Room/Noise → Microphone → A/D)
- Average classification time for 60s features: 0.4s