DEIUNDERCOVER

Music Information Retrieval
Overview & Current Trends

Rui Pedro Paiva

DEI

10 de Dezembro de 2008

Contents

- Introduction
- Applications
- Short MIR Tale
- Disciplines
- Generic MIR System
- Techniques

Introduction

- **Music and Man**
  - **Music expresses “that which cannot be put into words and that which cannot remain silent” (Victor Hugo)**
  - **We associate music with the most unique moments of our lives and music is part of our individual and social imaginary**
    - By listening to music, emotions and memories, thoughts and reactions, are awakened
  - **“Life has a soundtrack” (Ana Gomes, “Festivais de Verão”, Público)**
  - **“The history of a people is found in its songs” (George Jellinek)**

Introduction

- **Music and World economy**
  - **Music industry runs, only in the USA an amount of money in the order of several billion US dollars per year.**
  - **Explosion of the Electronic Music Industry (EMD)**
    - Widespread access to the Internet
    - Bandwidth increasing in domestic and mobile accesses
    - Compact audio formats with near CD quality (mp3, wma)
    - Portable music devices (iPod, mp3 readers)
    - Peer-to-peer networks (Napster, Kazaa, eMule)
    - Online music stores (iTunes, Calabash Music, Sapo Music) → resolution is the song, not the CD
    - Music identification platforms (Shazam, 411-Song, Gracenote MusicID / TrackID)
    - Music recommendation systems (MusicSurfer)

Introduction

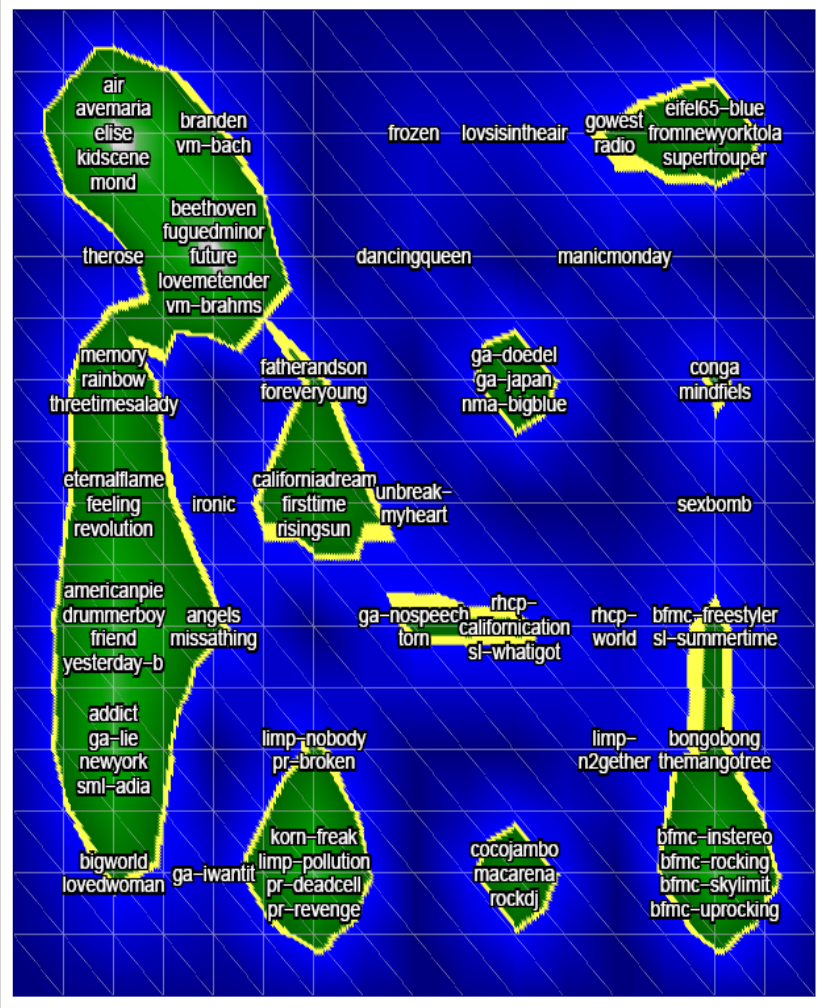
- **Music and World economy (cont.)**
  - **By 2005, Apple iTunes was selling ≈ 1.25 million songs each day**
    - Since the service was launched and until thee beginning of 2005 ≈ 250 million songs had been sold in total
  - **Number and dimension of digital music archives continuously growing**
    - Database size (these days, over 2 million songs)
    - Genres covered
  - **Challenges to music providers and music librarians**
    - Organization, maintenance, labeling, user interaction
    - Any large music database is only really useful if users can find what they are looking for in an efficient manner!

Introduction

- **Database Organization and Music Retrieval**
  - **Presently, databases are manually annotated → search and retrieval is mostly textual (artist, title, album, genre)**
    - Service providers
      - Difficulties regarding manual song labeling: subjective and time-consuming,
    - Customers
      - Difficulties in performing “content-based” queries
        - “Music’s preeminent functions are social and psychological”, and so “the most useful retrieval indexes are those that facilitate searching in conformity with such social and psychological functions. Typically, such indexes will focus on stylistic, mood, and similarity information” [Huron, 2000].
  - **→ Music Information Retrieval (MIR) emerges from the necessity to manage huge collections of digital music for “preservation, access, research and other uses” [Futrelle and Downie, 2003].**

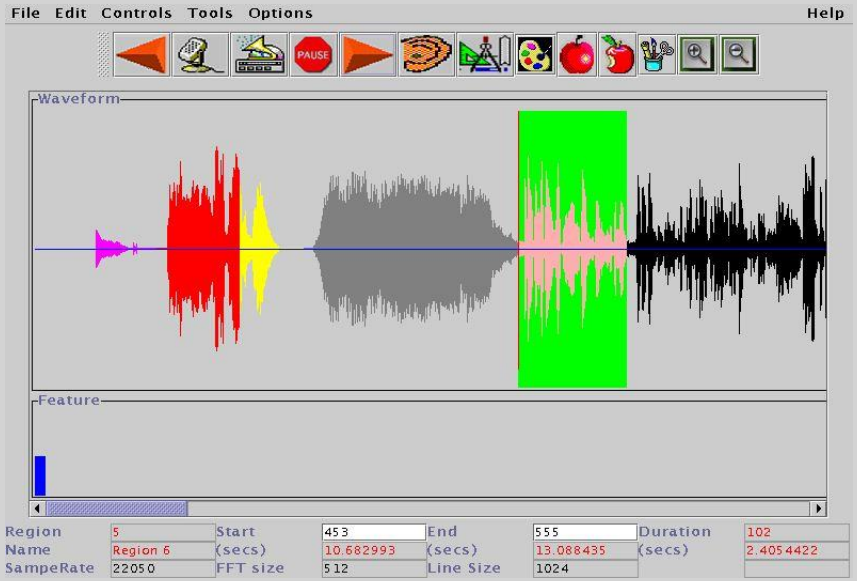
Applications

- **Platforms for EMD**
  - **Similarity-based retrieval tools**
    - Query-by-example
      - Music identification (trackID, Tunatic)
      - Music recommendation
        - Discovery of new music
      - Islands of music
        - Metaphor of geographic maps: similar genres are “physically” close together
    - Automatic playlist generation
    - Query-by-melody (query-by-humming, query-by-singing)
    - Plagiarism detection
    - Music web crawlers



Applications

- **Music education and training**
  - **Automatic music transcription**
    - →Music composition, analysis, performance evaluation
- **Audio software**
  - **Intelligent audio (music) editors → automatic indexing**
- **Multimedia databases and operating systems**



© George Tzanetakis, 2002

Applications

- **Video indexing and searching**
  - Segmentation based on audio (music) content → detection of scene transitions
- **Advertisement and cinema**
  - Tools for mood-based retrieval
- **Sports**
  - Music to induce a certain cardiac frequency

Short MIR Tale

- Precursors of computer-based MIR: incipit and theme indexes, e.g., Harold Barlow and Sam Morganstern's dictionary of musical themes
- 1966: potential of applying automatic information retrieval techniques to music was recognized (Kassler)
- 1970s and 1980s: automatic music transcription systems
- 1990s: surge of interest, mostly in topics such as query-by-humming (impulse from research on digital libraries)
- 2000: 1st International Symposium on Music Information Retrieval (ISMIR)

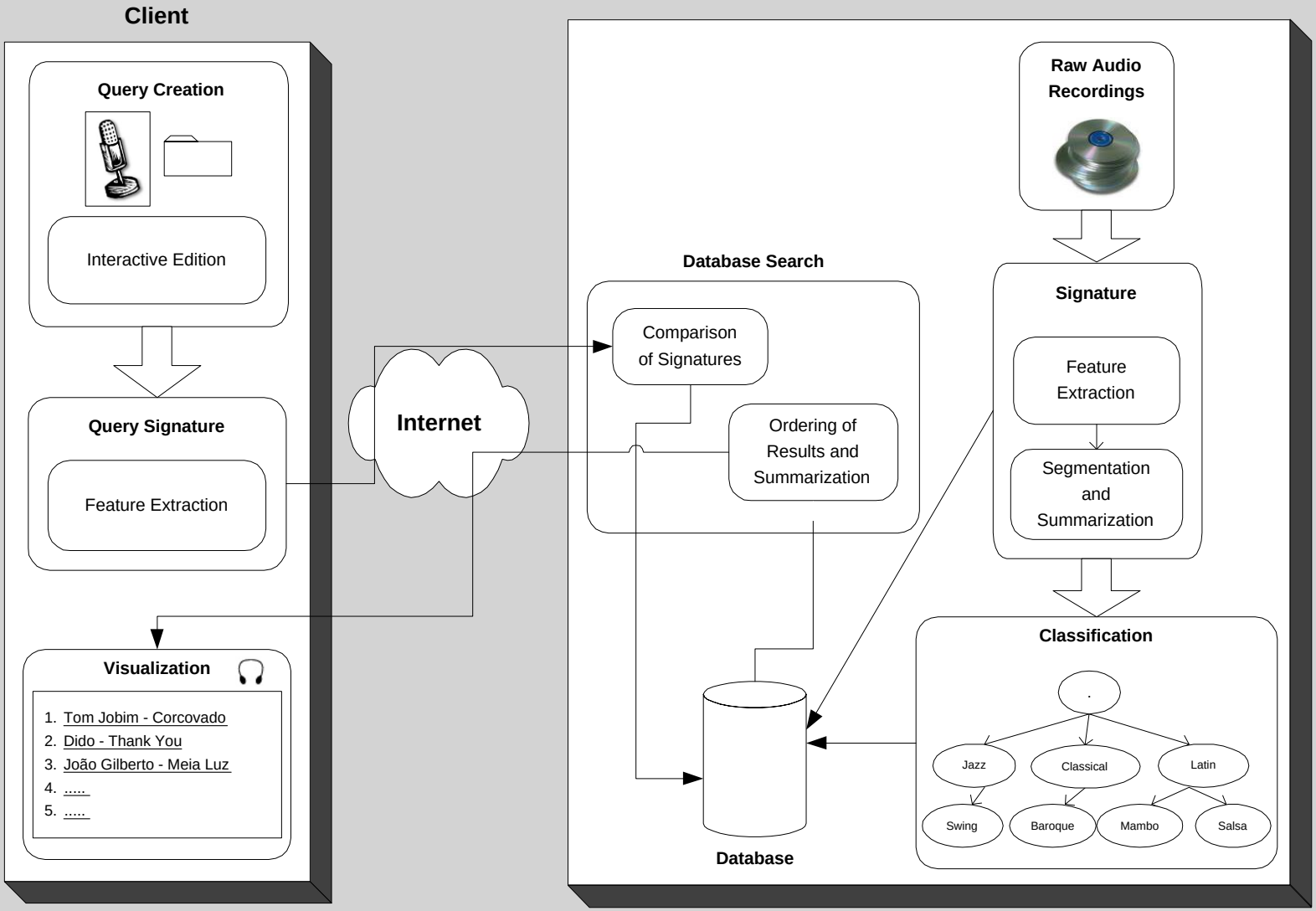
Short MIR Tale

- **Presently: strongly cross-disciplinary field**

Community	Type of Institution(s)	Typical Research Areas
Computer Science, Information Retrieval	Academic, Commercial	Representation, Indexing, Retrieval, Machine Learning, User Interface Design
Audio Engineering, Digital Signal Processing	Academic, Commercial	Compression, Feature Detection, Pitch Tracking, Machine Learning, Classification, Playlist Generation, Musical Analysis
Musicology, Music Theory	Academic	Representation, Musical Analysis
Library Science	Libraries, Academic	Representation, Metadata, User Studies, Classification, Intellectual Property Rights, User Interface Design
Cognitive Science, Psychology, Philosophy	Academic	Representation, Perception, User Studies, Ontology
Law	Government, Legal Profession, Academic	Intellectual Property Rights

  - **My interests: content analysis and similarity assessment and retrieval in audio song databases**
    - Metrics of similarity, music identification, music recommendation, audio fingerprinting, music classification and feature extraction, tempo and melody detection, music summarization

Generic MIR System

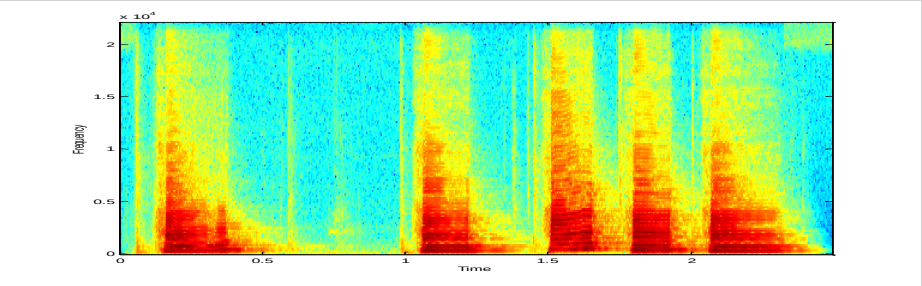


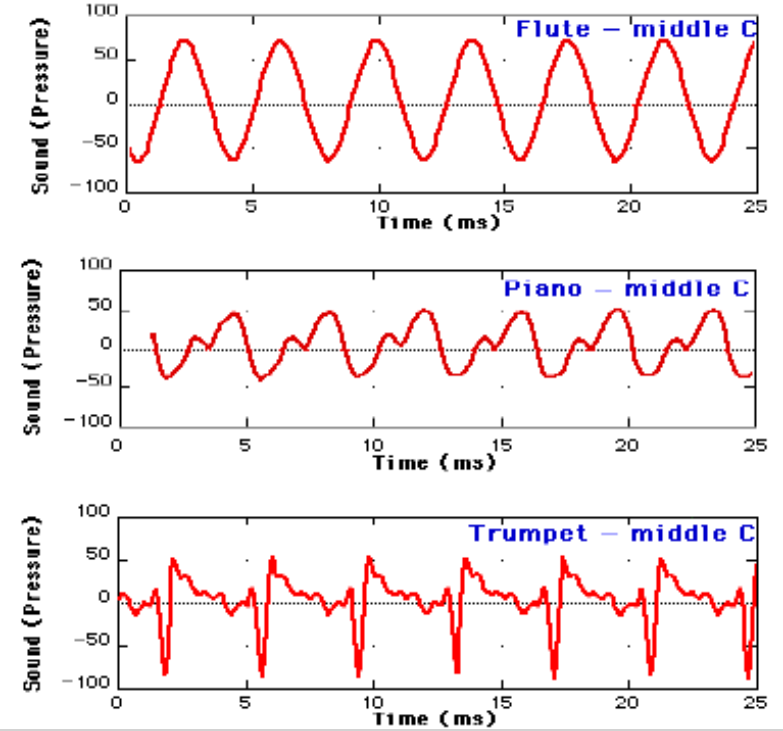
Techniques

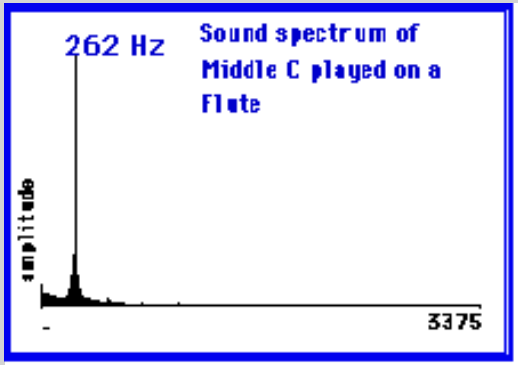
- **Idea**
  - **Extract semantic information from low-level data**
  - **Feature extraction**
    - Physical: F0, intensity, centroid, uniformity, rolloff, flux
    - Perceptual: pitch, loudness, timbre, beat
    - Musicological: notes, melodies, measures, motives, themes
    - Higher-level (semantic) features: emotion, genre, instruments, artist

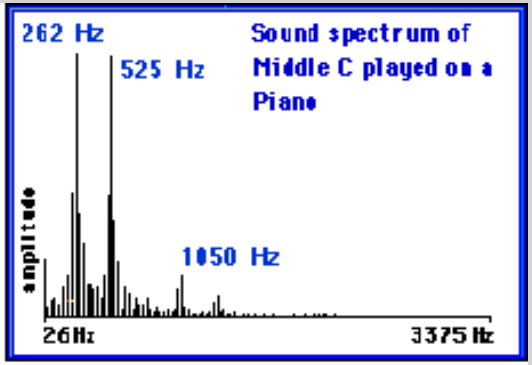
Techniques

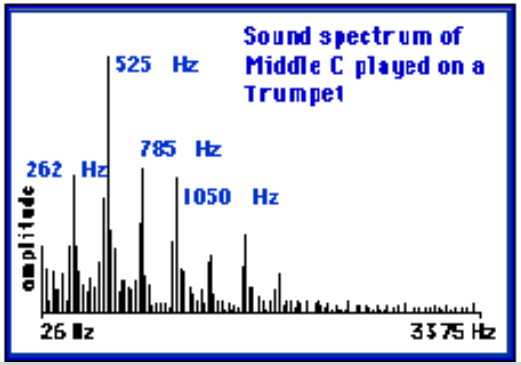
- **Physical features**
  - **Frequency domain**
    - DFT, STFT, spectrogram
    - F0 detectors







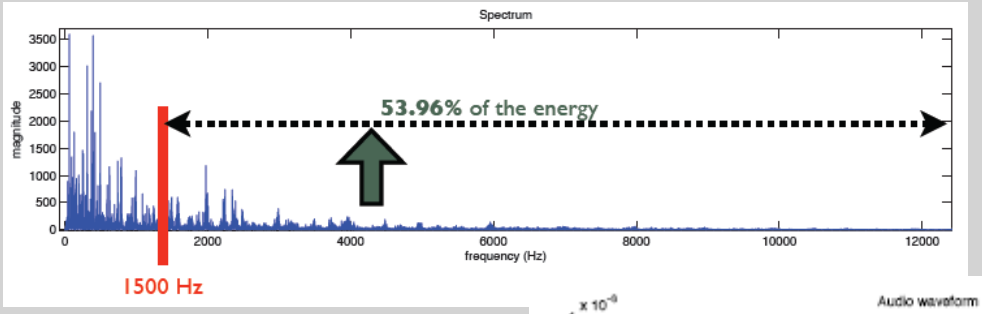




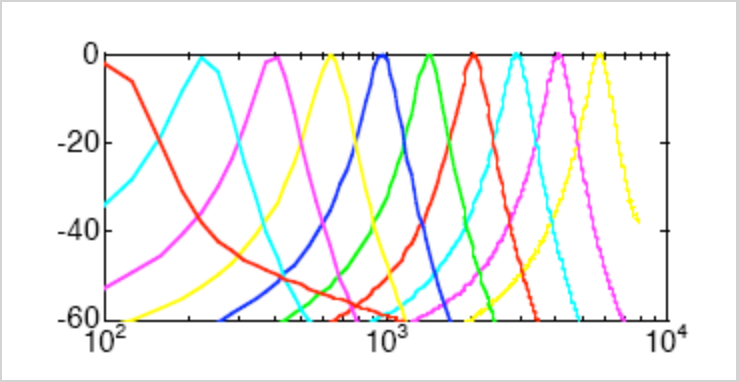
Techniques

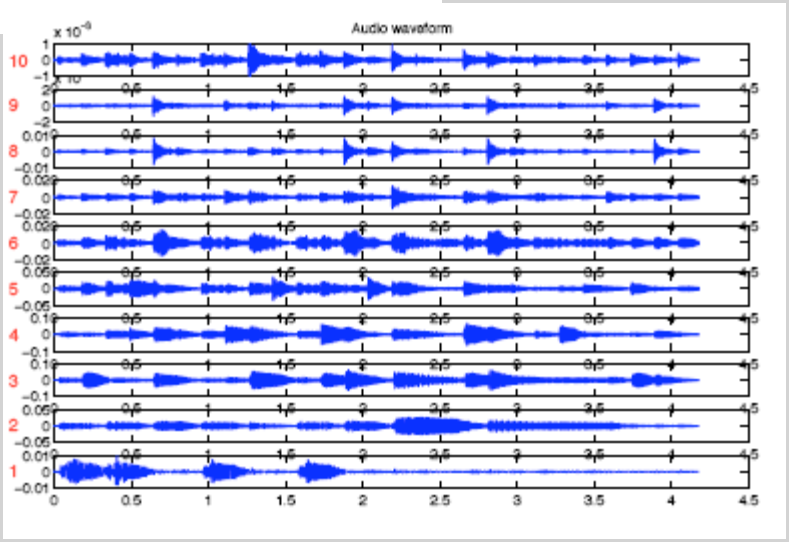
- **Physical features (cont.)**
  - **Spectral features**
    - MFCCs, centroid, rolloff, flux, harmonicity, high-frequency content, …

© MIR Toolbox, 2008



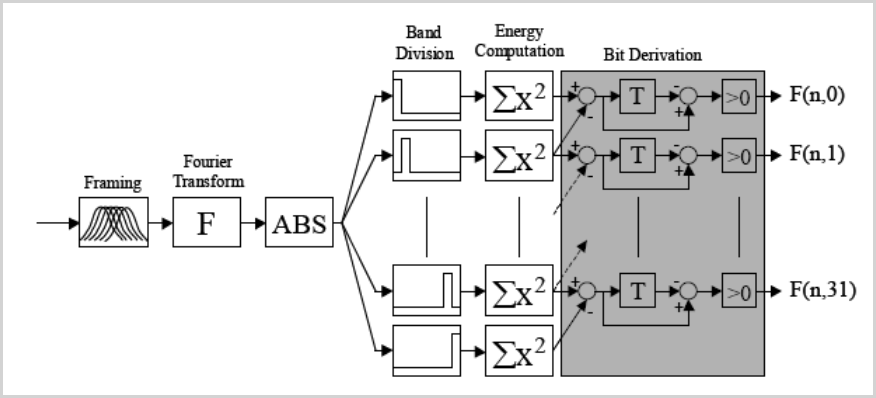
  - **Sub-band features**





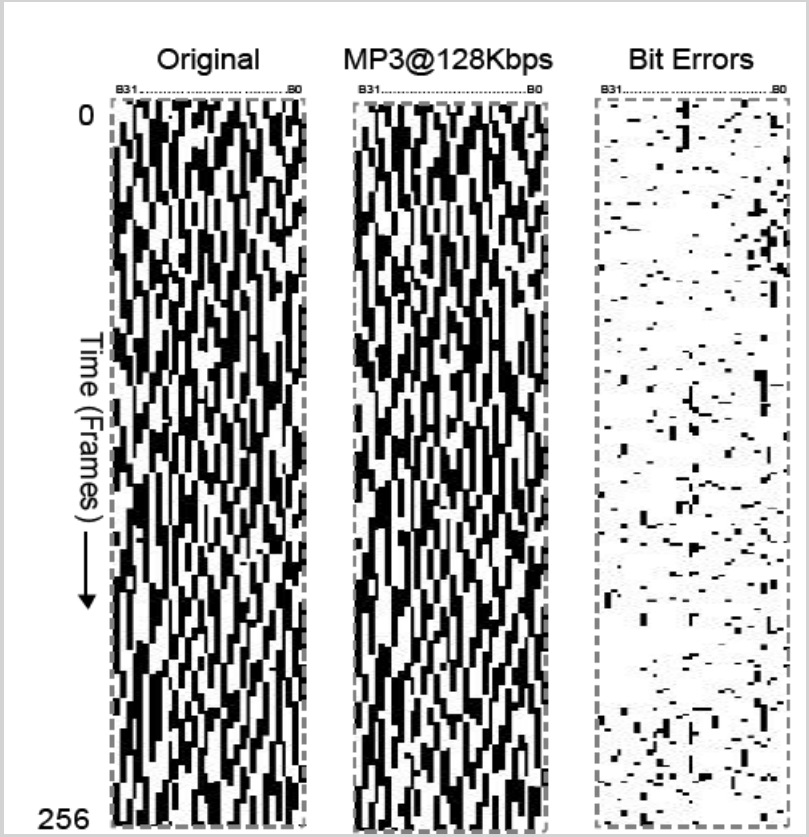
Techniques

- **Physical features (cont.)**
  - **Spectral features + sub-band features (e.g., audio fingerprinting)**



$$F(n,m)=\begin{cases}1 & \text{if } E(n,m)-E(n,m+1)-\big(E(n-1,m)-E(n-1,m+1)\big)>0\\ 0 & \text{if } E(n,m)-E(n,m+1)-\big(E(n-1,m)-E(n-1,m+1)\big)\le 0\end{cases}$$

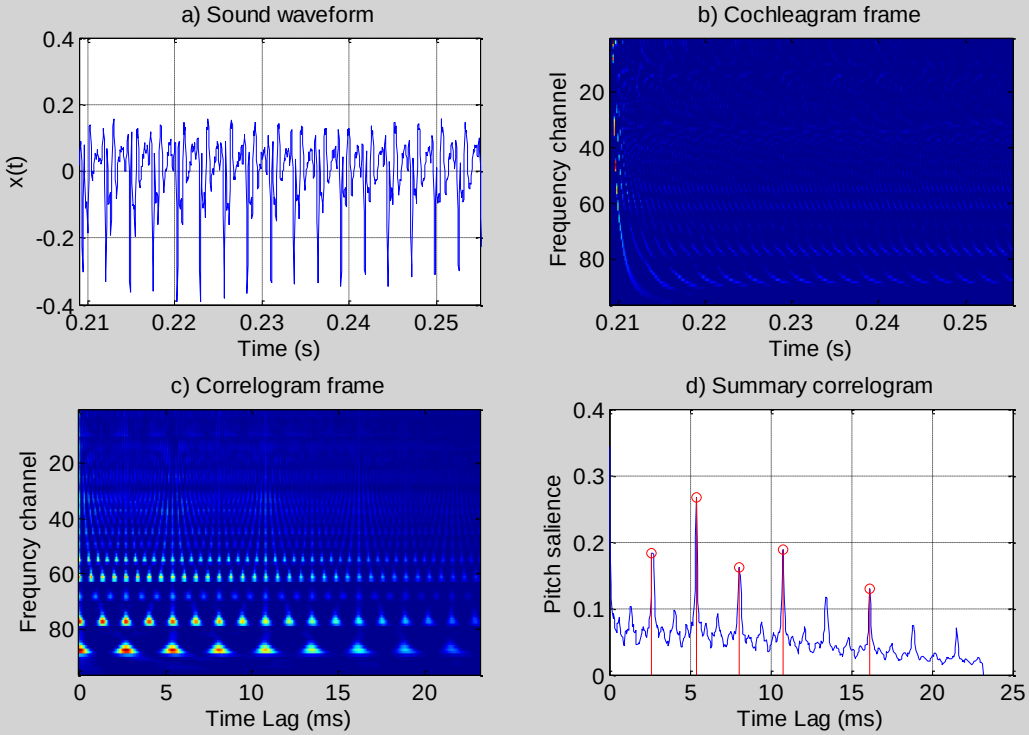
© Haitsma & Kalker, 2002

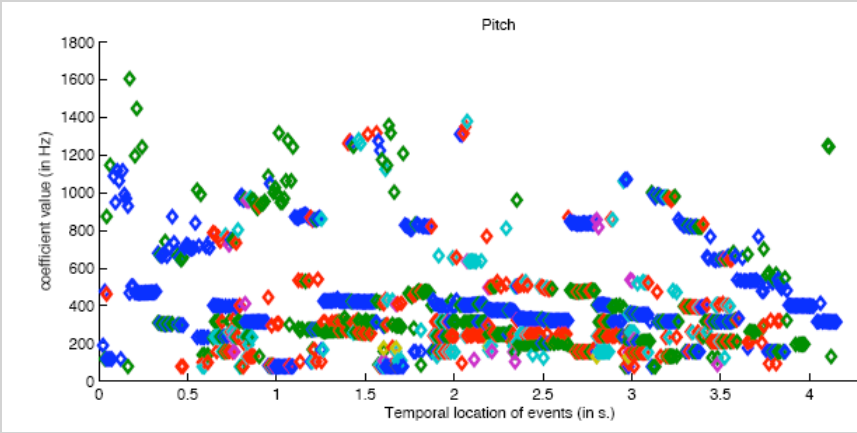


Techniques

❑ Physical features (cont.)

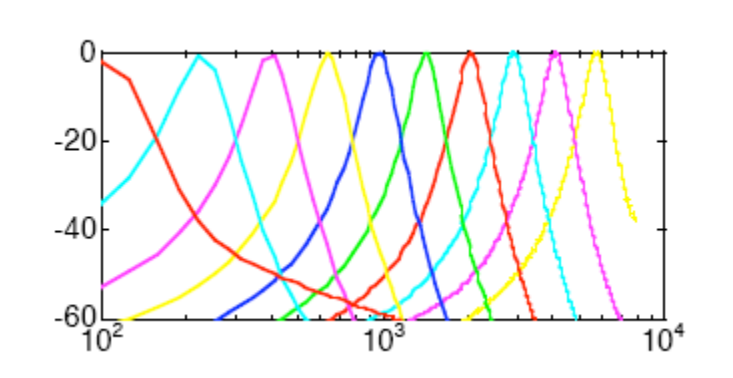
- **Time domain**
  - Auditory model-based F0 detectors, beat detectors (energy-based)

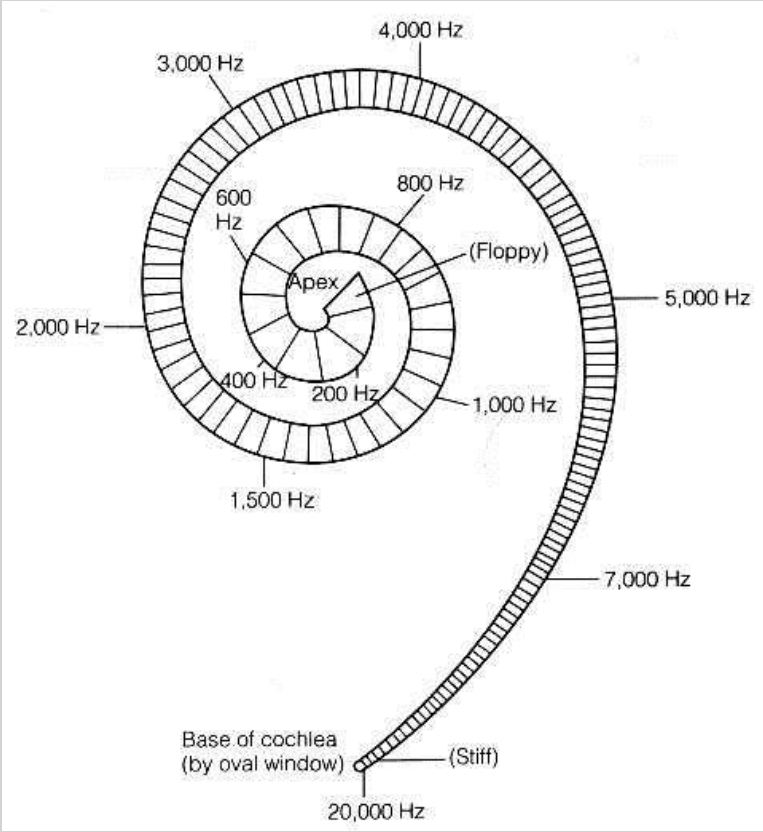




Techniques

- **Perceptual features**
  - **Pitch**
    - **Frequency**
    - Intensity
    - Context
    - Ear physiology (age)

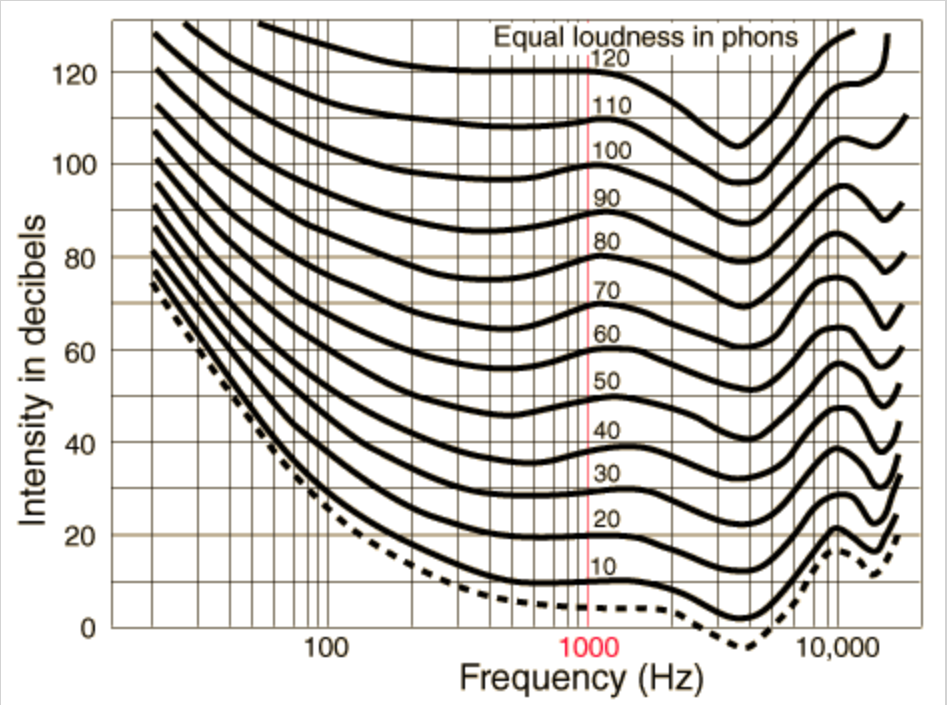




Techniques

- **Perceptual features (cont.)**
  - **Loudness**
    - **Intensity**
    - Frequency
    - Context
    - Ear physiology (age)

Fletcher-Munson equal loudness contours

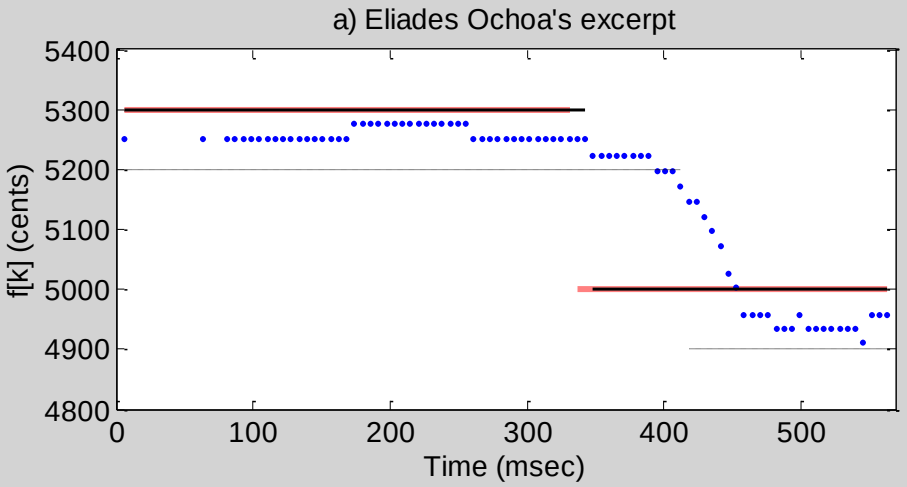


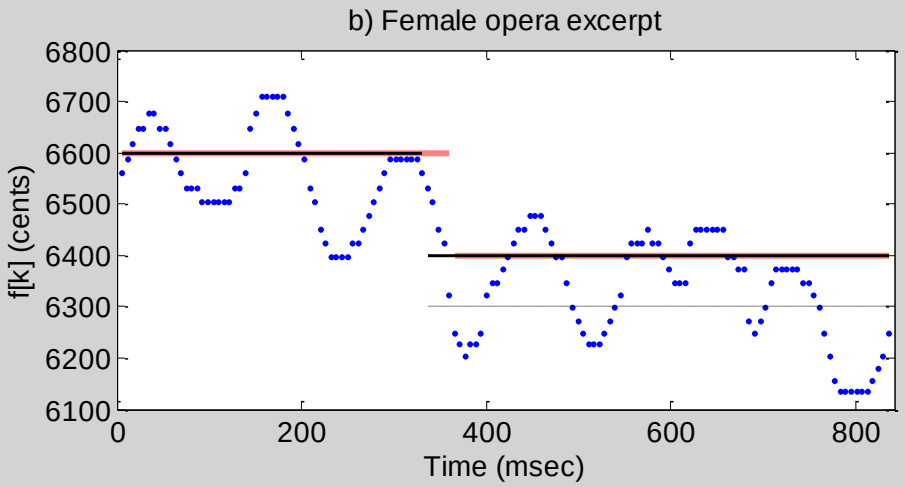
Techniques

- **Perceptual features (cont.)**
  - **Timbre**
    - No physical correlate
    - "what something sounds like":
      - Spectral content at steady-sate
        - Centroid, rolloff, relative amplitudes of harmonic components, inharmonicty…
      - Signal's temporal envelope
        - Attack transient
      - Temporal behavior of the harmonics
  - **Melodic contour**
    - UDUEEUUD
  - **Rhythm contour**
    - LSSLEELS
  - **Beat**

Techniques

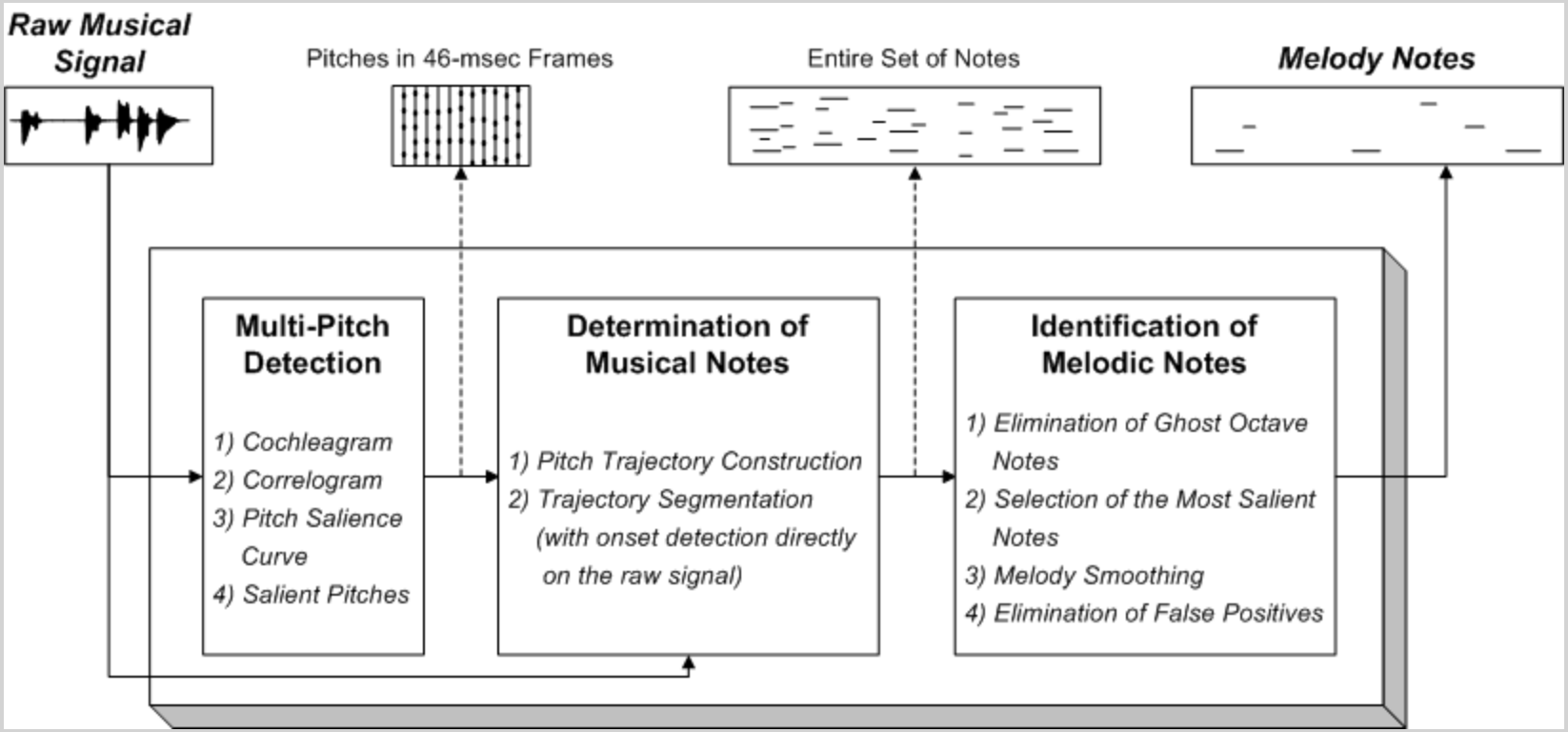
- **Musicological features**
  - **Notes from audio**





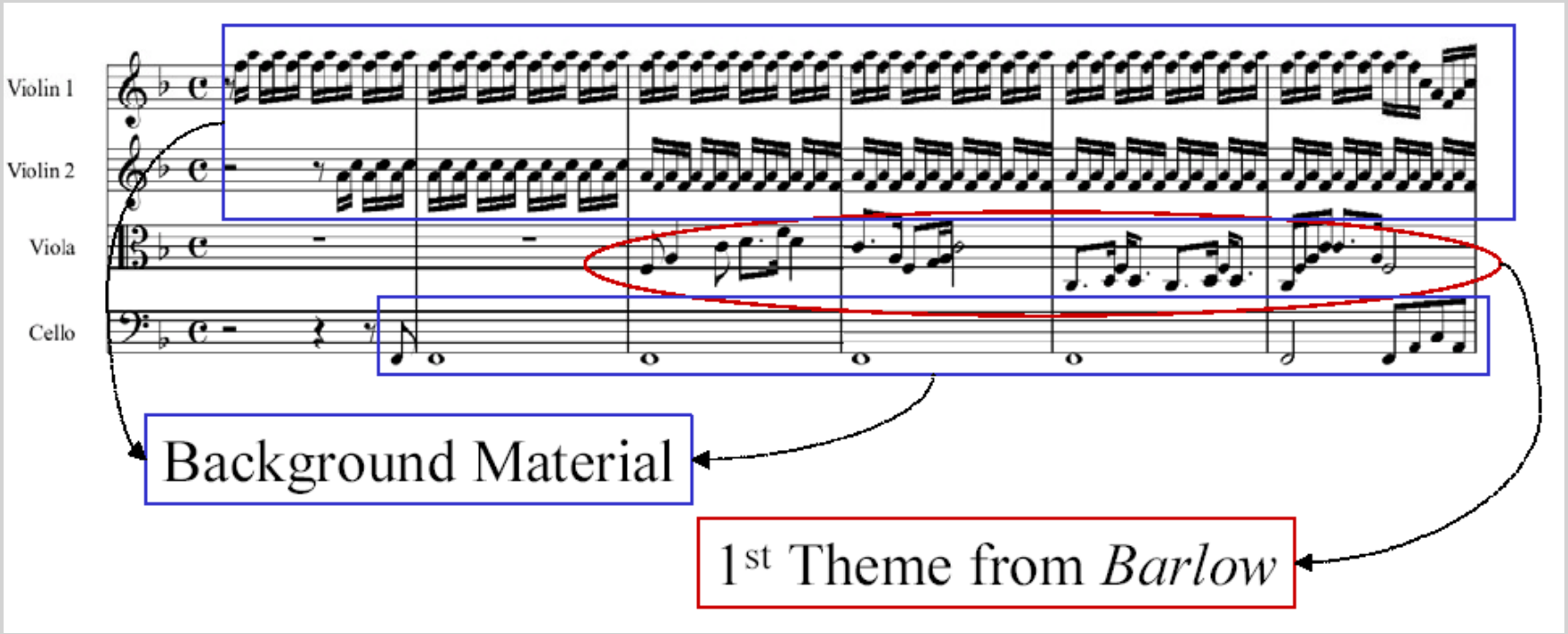
Techniques

- **Musicological features (cont.)**
  - **Melody**



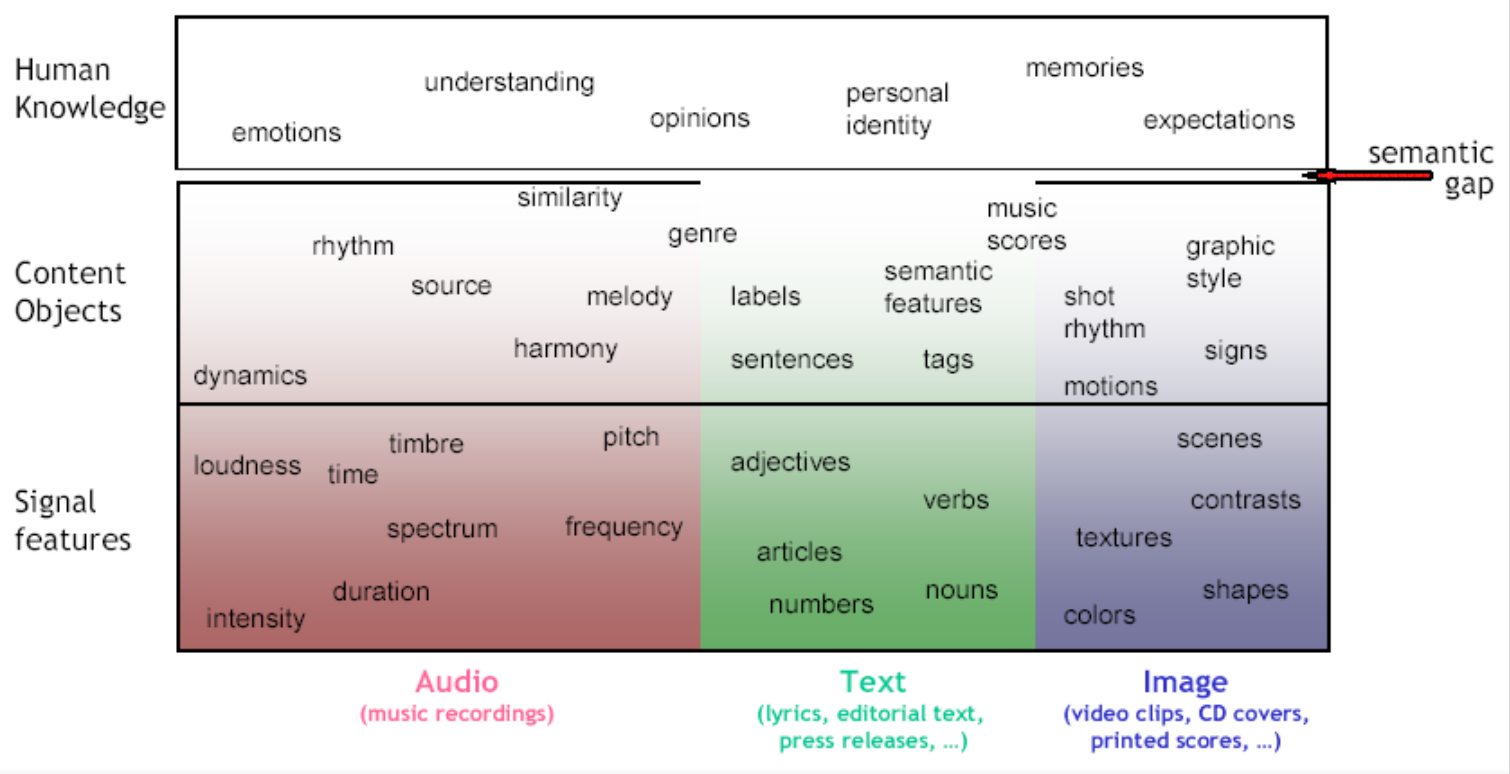
Techniques

- **Musicological features (cont.)**
  - **Themes**



Techniques

- Higher-level features → top-down information flow



© Xavier Serra, 2005

Techniques

- **Higher-level features (cont.)**
  - **Bridge the semantic gap**
  - **Memory, context, expectations**
    - Repetitions, sonic environment, modeling the individual, musicological knowledge
  - **Emotion: *valence* (happy/anxious) and *arousal* (calm/energetic)**
    - Classification approaches resorting to low-level features