Multimedia Systems

III. Sound and Music Computing

3.2. Sound and Music Description

António Sá Pinto

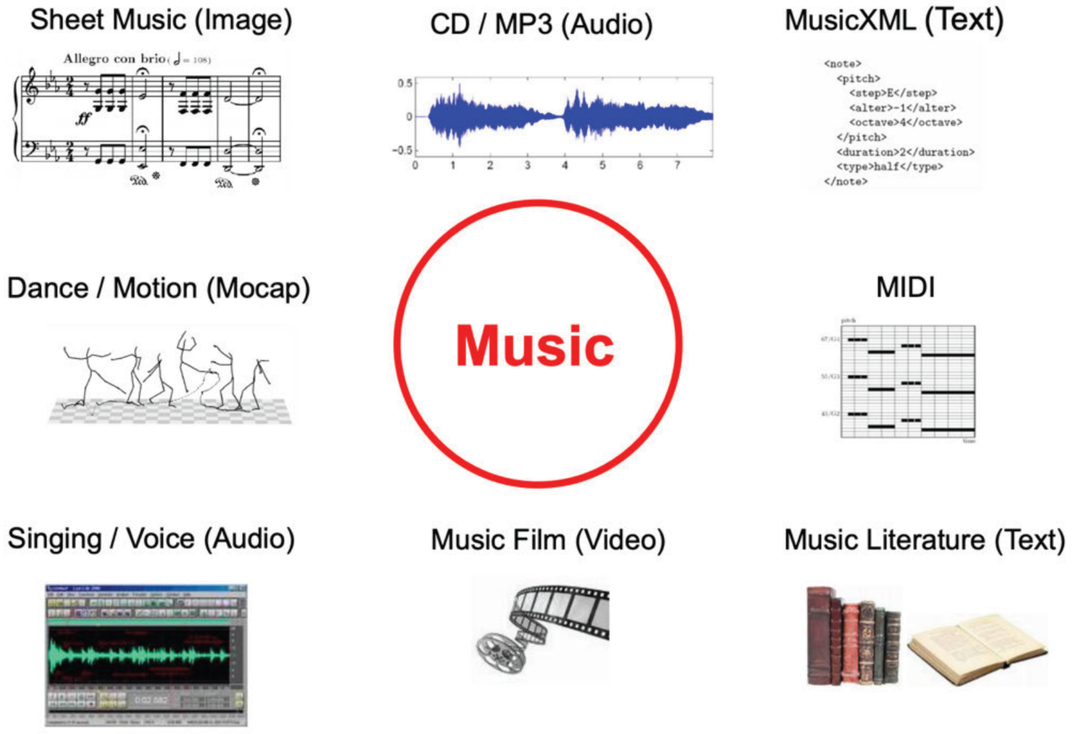
FEUP

Agenda

- The need for Sound and Music Description
- The Semantic Gap
- Standardisation
- Classification of Music Descriptors
- Computation of Low-Level Descriptors

Why?

Way to navigate through ever more numerous sources of information about music



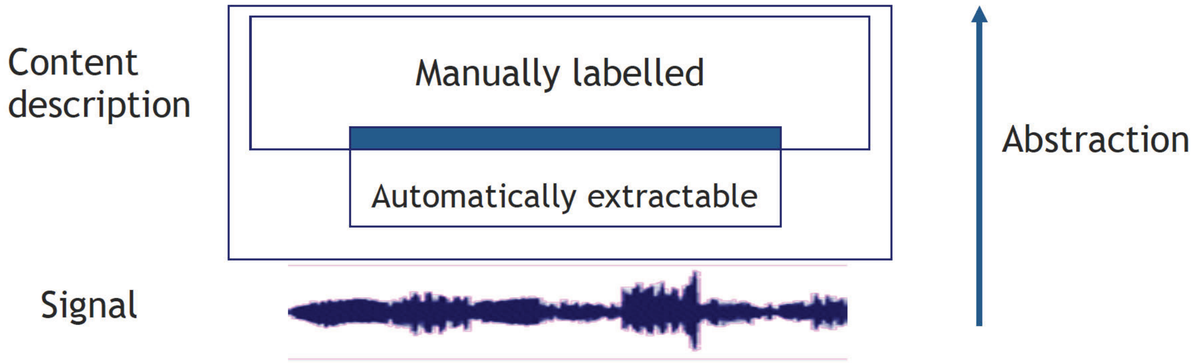
from: (Müller) Fundamentals of Music Processing, *Springer 2015*

Use Cases

Use Case	Specificity	Description
Music Identification	H	Identify a compact disk, provide metadata about an unknown track, mobile music information retrieval: e.g. *shazam.com*
Plagiarism detection	H	Identify mis-attribution of musical performances, mis-appropriation of music intellectual property.
Copyright monitoring	H	Monitor music broadcast for copyright infringement or royalty collection
Versions	H/M	Remixes, live vs. studio recordings, cover songs. Used for database normalization and near-duplicate results elimination
Melody	H/M	Find works containing a melodic fragment
Identical Work / Title	M	Retrieve performances of same opus number or song title
Performer	M	Find music by a specific artist
Sounds like	M	Find music that sounds like a given recording
Performance Alignment	M	Mapping one performance onto another independent of tempo and repetition structure
Composer	M	Find works by one composer
Recommend-ation	M/L	Find music that matches the user's personal profile
Mood	L	Find music using emotional concepts: *Joy, Energetic, Melancholy, Relaxing*
Style / Genre	L	Find music that belongs to a generic category: *Jazz, Funk, Female Vocal*
Instrument(s)	L	Find works with same instrumentation
Music-Speech	L	Radio broadcast segmentation, Music archives cataloguing

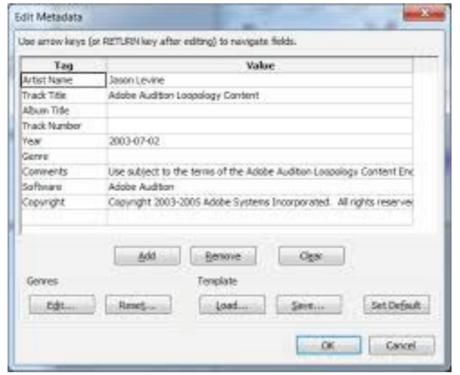
Sound and Music **Content** description

Content: The implicit information that is related to a piece of sound and/or music and that is represented in the piece itself.



Metadata *vs* Content descriptor

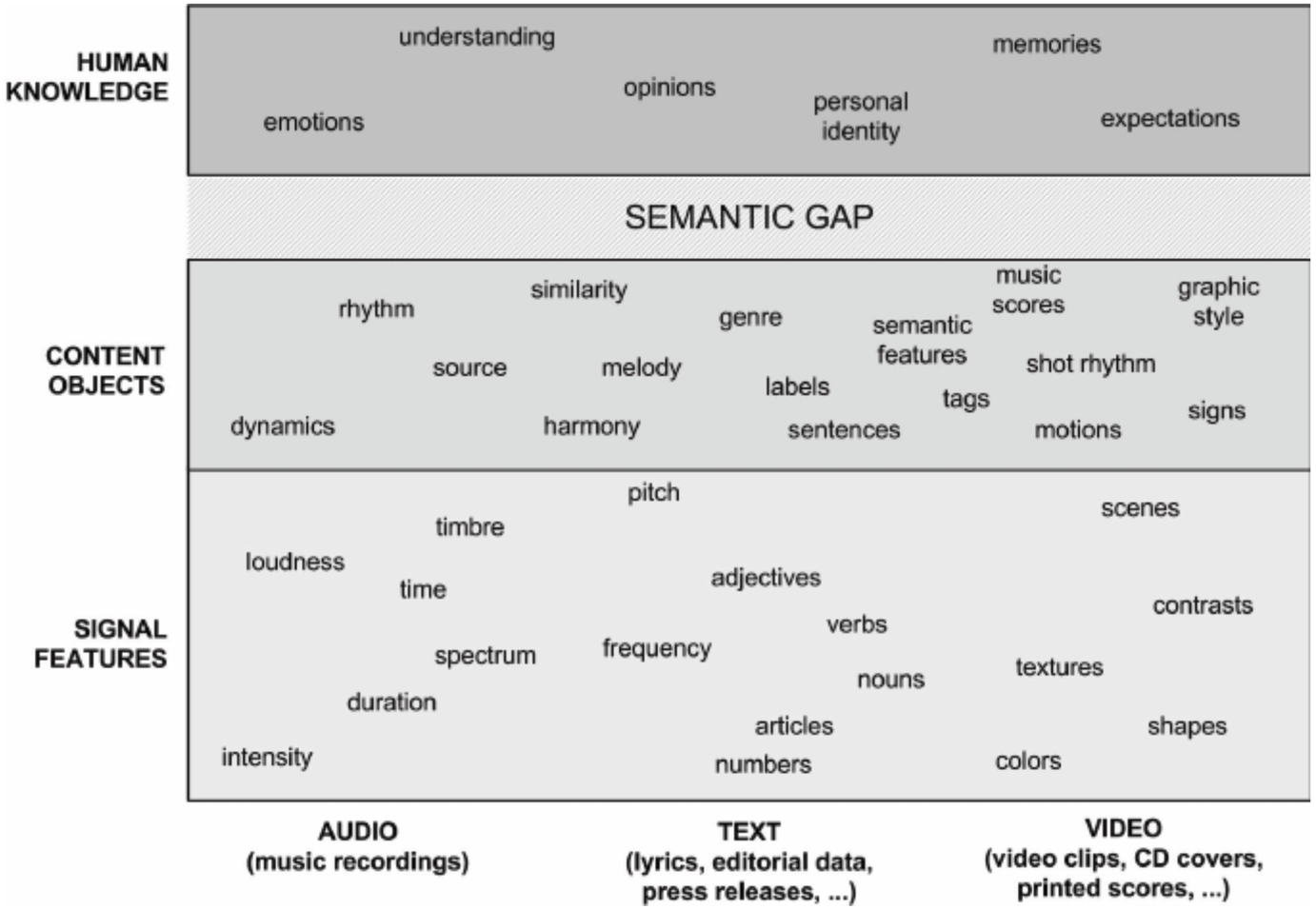
Metadata: any piece of information related to a music piece that can be annotated, extracted, and that is in any meaningful way (i.e. it carries semantic information).



Content descriptor (MPEG-7): a distinctive characteristic of the data which signifies something to somebody.

High-level Description	Data Source	Task Description
Timbre	Audio	Instrument Recognition Percussive, Pitched, Ensemble Recognition
Melody / Bass	Audio / Symbolic	Melody-line extraction Bass-line extraction
Rhythm	Audio	Onset detection Meter identification Meter alignment (bars) Beat (tactus) tracking Tempo tracking Average tempo
Pitch	Audio	Single fundamental freq. Multiple fundamental freq.
Harmony	Audio / Symbolic	Chord label extraction Bass-line extraction
Key	Audio / Symbolic	Modulation tracking Pitch spelling
Structure	Audio / Symbolic	Verse / chorus extraction Repeat extraction
Lyrics	Audio	Singing detection, lyrics-identification, word recognition
Non-Western music	Audio	Micro-tonal tuning systems Non-Western canon of concepts

Semantic Gap



Music Descriptors

Classification:

- Abstraction level;
- Temporal scope;
- Musical facet;

Music Descriptors

I. Abstraction

- **Low-level (signal-centered) descriptors**: computed from the data in direct or derived way. Little sense for most users, easy exploitation by systems.

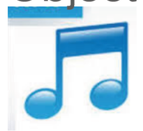
- **Mid-level (object-centered) descriptors**: requiring an induction operation (generalization about the data). More sense for users.

- **High-level (user-centered) descriptors**: requiring an induction operation (user modelling). Bridging the semantic gap.

Signal

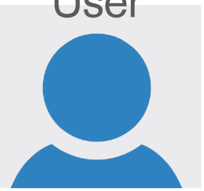
RMS, Energy, Duration, etc.

Object



Key, Melody, Beat, Chords

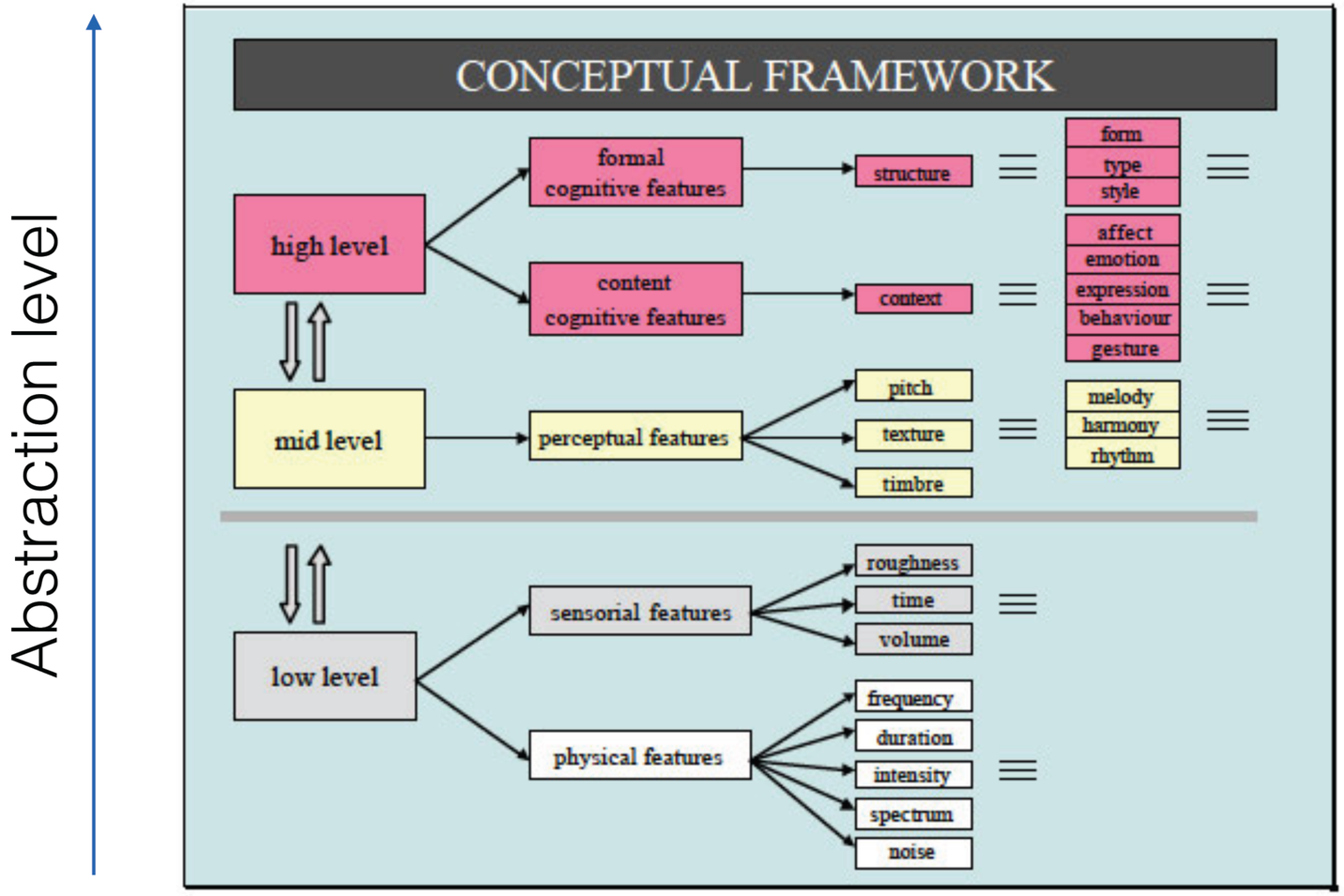
User



Genre, Mood

Music Descriptors

I. Abstraction



Music Descriptors

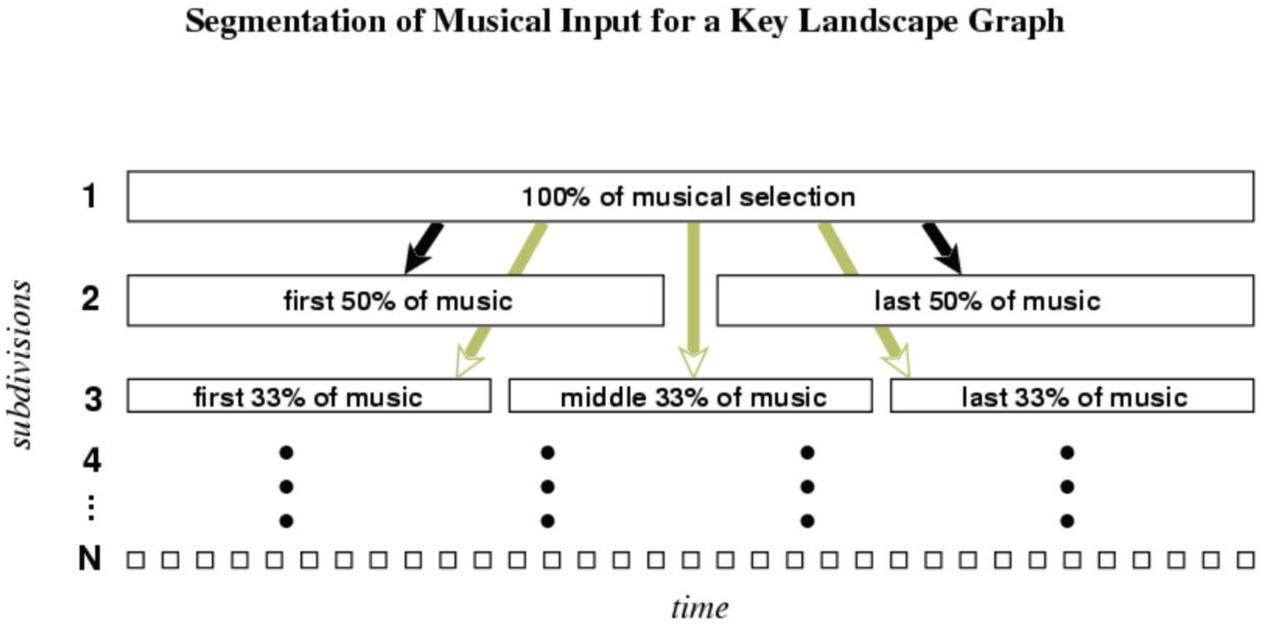
I. Abstraction

STRUCTURE		CONCEPT LEVEL		MUSICAL CONTENT CATEGORIES AND FEATURES				
CONTEXTUAL	GLOBAL DESCRIPTORS	HIGH II	EXPRESSIVE	expression				
				affect experience				
		HIGH I	STRUCTURAL	melody	harmony	rhythm	source	dynamics
				key profile	tonality cadence	patterns tempo	instrument voice	trajectory articulation
		MID	PERCEPTUAL	successive intervallic pattern	simultane intervallic pattern	beat i o i	spectral envelope	dynamic range sound level
NON-CONTEXTUAL	LOCAL DESCRIPTORS	LOW II	SENSORIAL	pitch		time	timbre	loudness
				periodicity pich pitch deviations fundamental frequency		note- duration onset offset	roughness spectral flux spectral- centroid	peak neural- energy
		LOW I	ACOUSTICAL	frequency		duration	spectrum	intensity

Music Descriptors

II. Temporal Scope

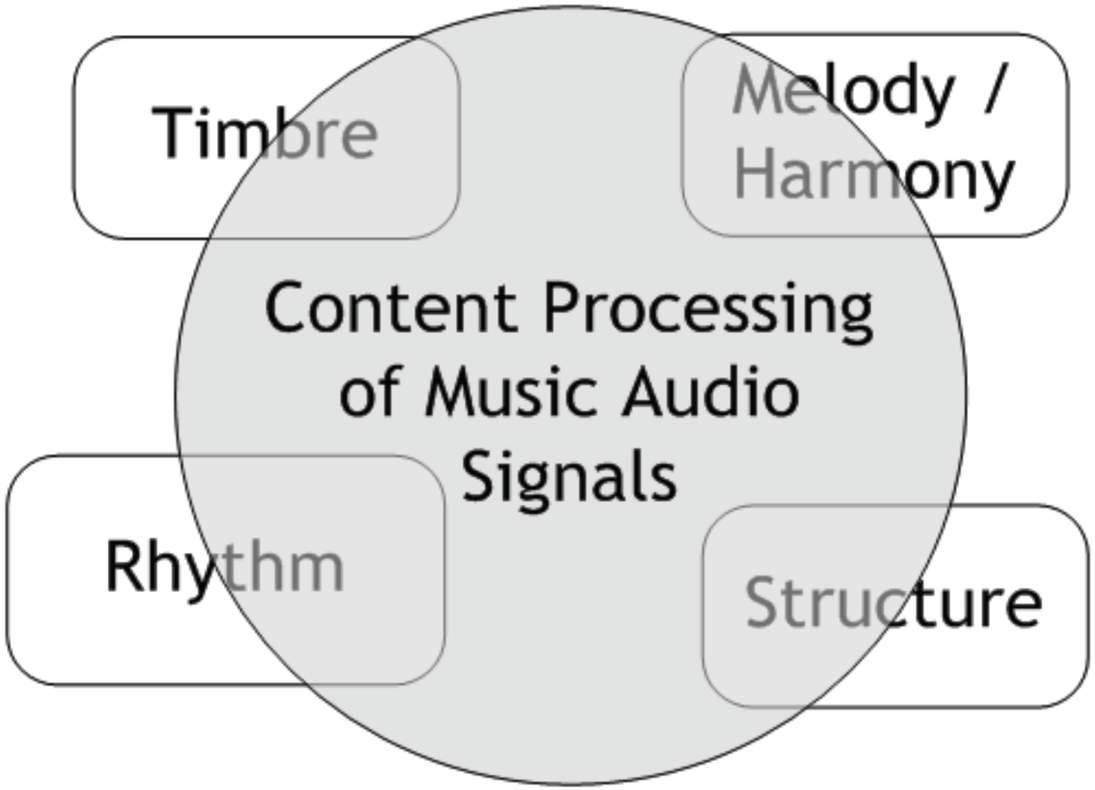
- **Instantaneous descriptors**: time instant (frame-based).
- **Local/segment descriptors**: time window.
- **Global descriptors**: musical piece.



Music Descriptors

III. Musical Facet

- Melody;
- Rhythm;
- Harmony/tonality;
- Timbre/instrumentation;
- Dynamics;
- Structure/Segmentation;



Standards

No standardization initiatives has been widely adopted within MIR tools and descriptors

- Dublin Core (DC, Dublin Core Metadata Initiative): 16 metadata elements (Title, Creator, Subject, Description, Publisher, Date, …);

- MPEG-7 (ISO/IEC Moving Picture Expert Group). Description of multimedia content. Descriptors, Description Schemes.

Specific Content Descriptors (e.g. MPEG-7) are generally used in MIR, but not in a standardized way.

Low-level descriptors

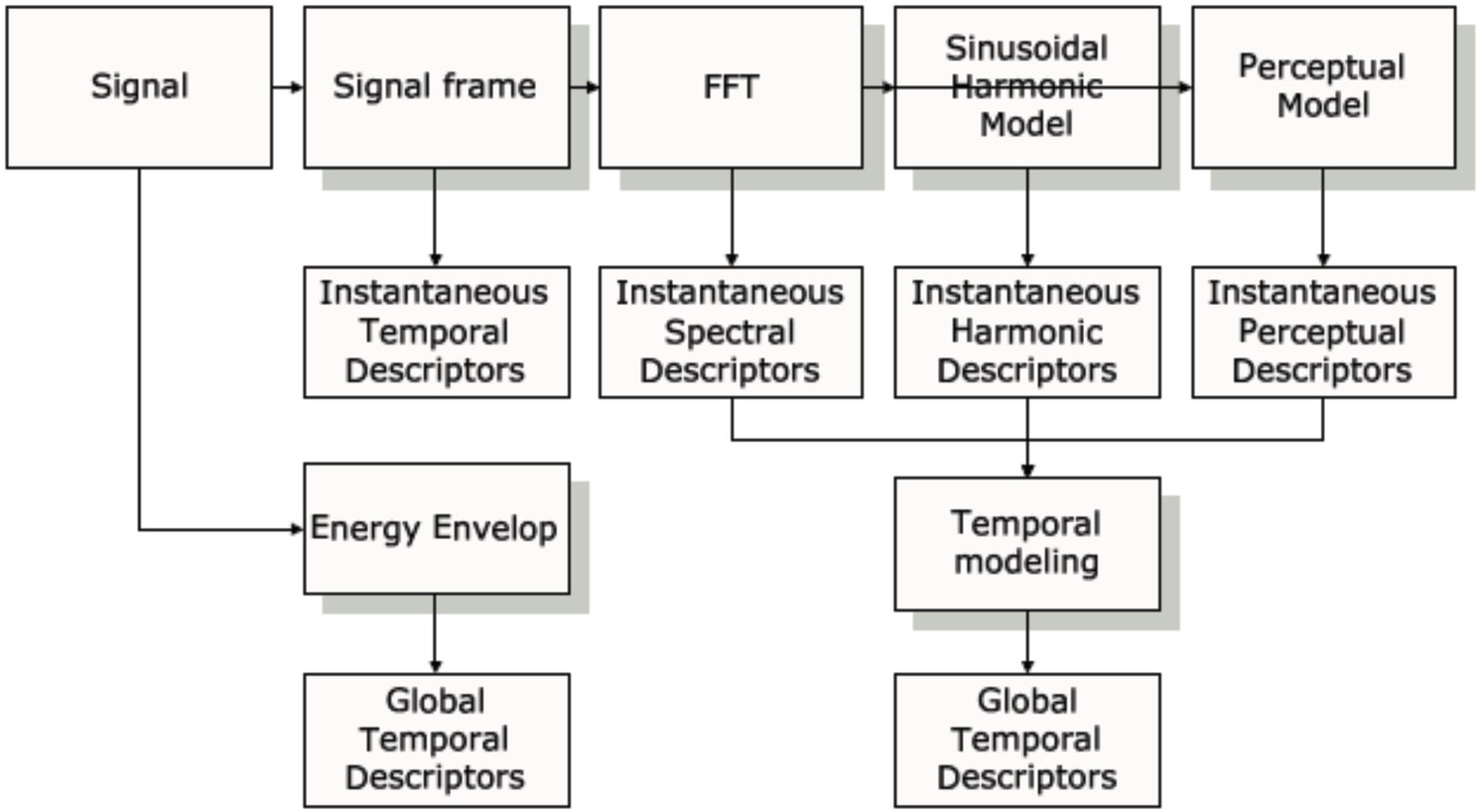
- Computed from the audio signal in a direct or derived way.
- Little meaning to users, easily exploited by computer systems.

Low-level descriptors

Important points

- Basis of further description
- They should represent correctly the sound in the way we want to describe
- They should be:
  a) Deterministic;
  b) Computable for any signal (e.g. white noise, sinusoid, silence, …);
  c) Robust (application dependent, e.g. for mp3, noise addition, resampling, stereo vs mono, etc.)

Low-level descriptors



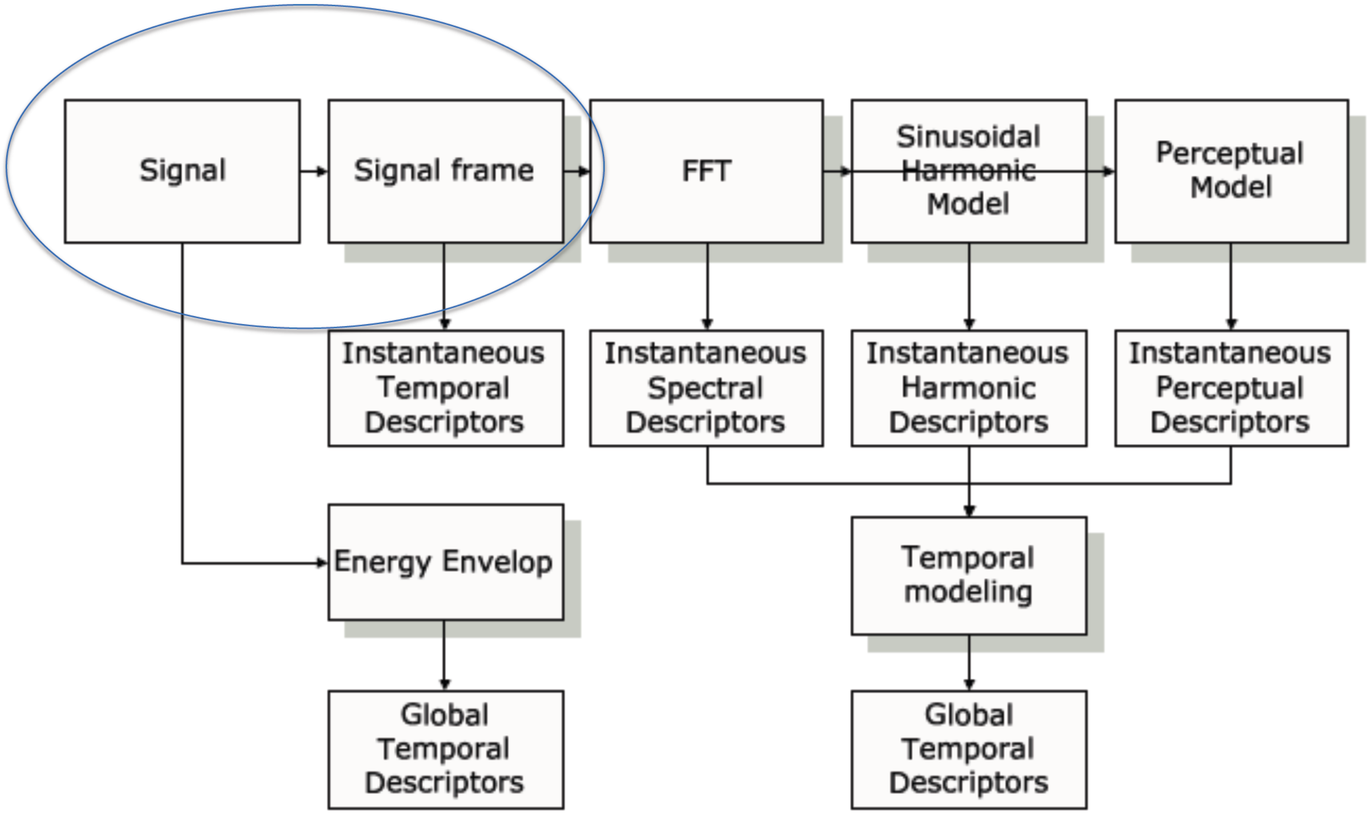
from: (Peeters) A large set of Audio features for sound description (similarity and classification) in the CUIDADO project. *2004*

Low-level descriptors

1. Windowing (frame)

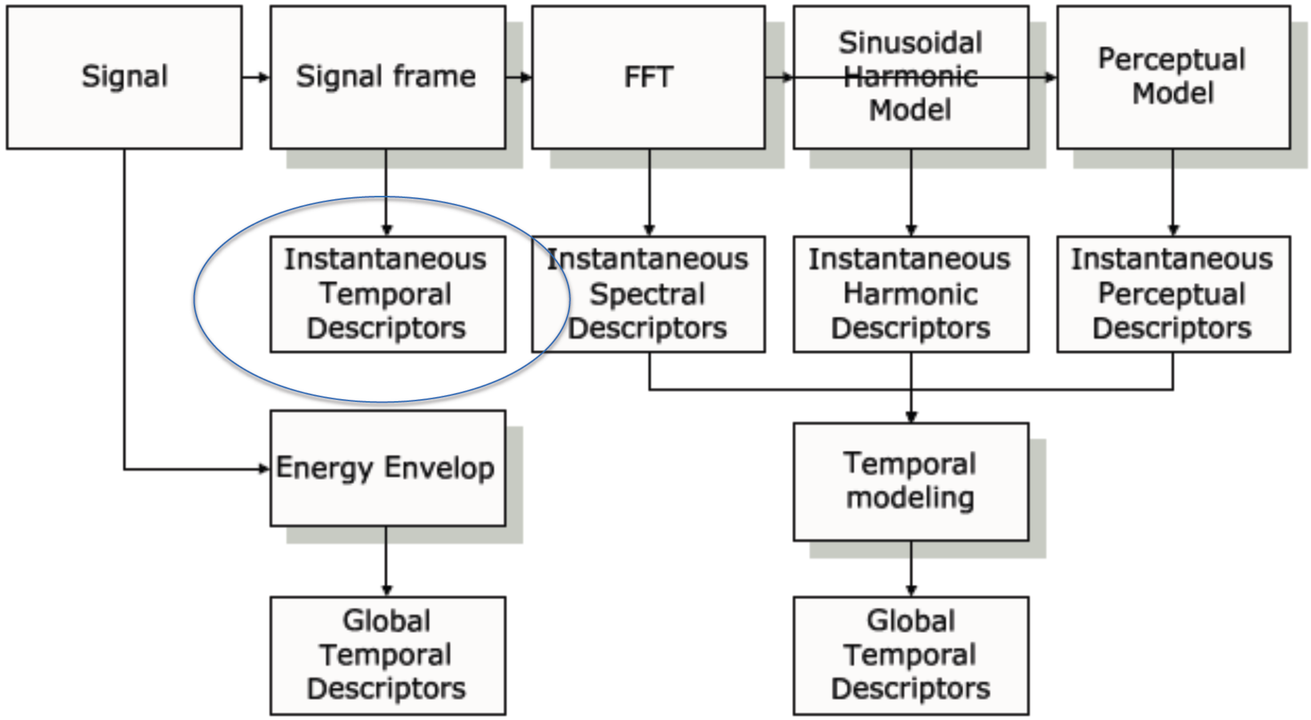
- Size of window,
- Overlap factor (hop size);
- Shape of window.

Influence in the analysis!!!



Low-level descriptors

2. Temporal feature computation



- Log-attack time
- Temporal Centroid
- Zero-Crossing Rate
- Energy

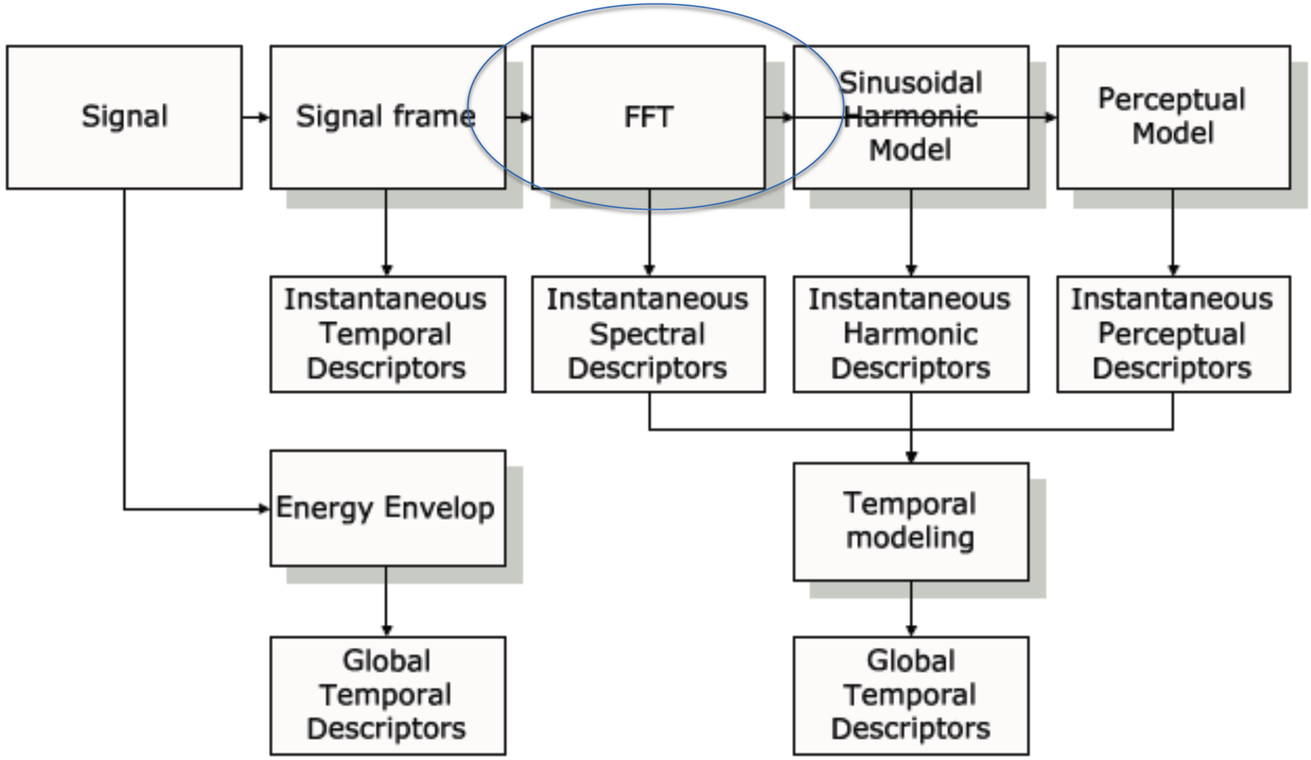
Low-level descriptors

3. Spectral Analysis:

DFT(FFT) → STFT

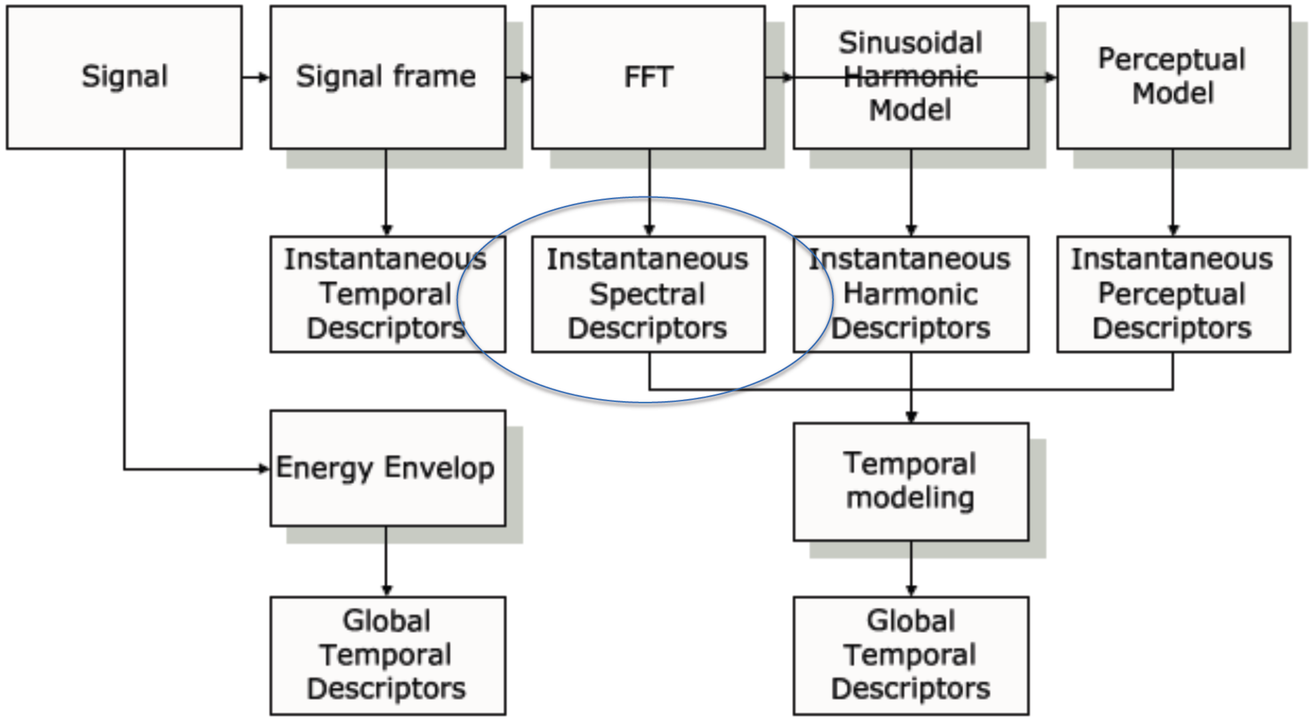
FFT size?

Influence in the analysis!!!



Low-level descriptors

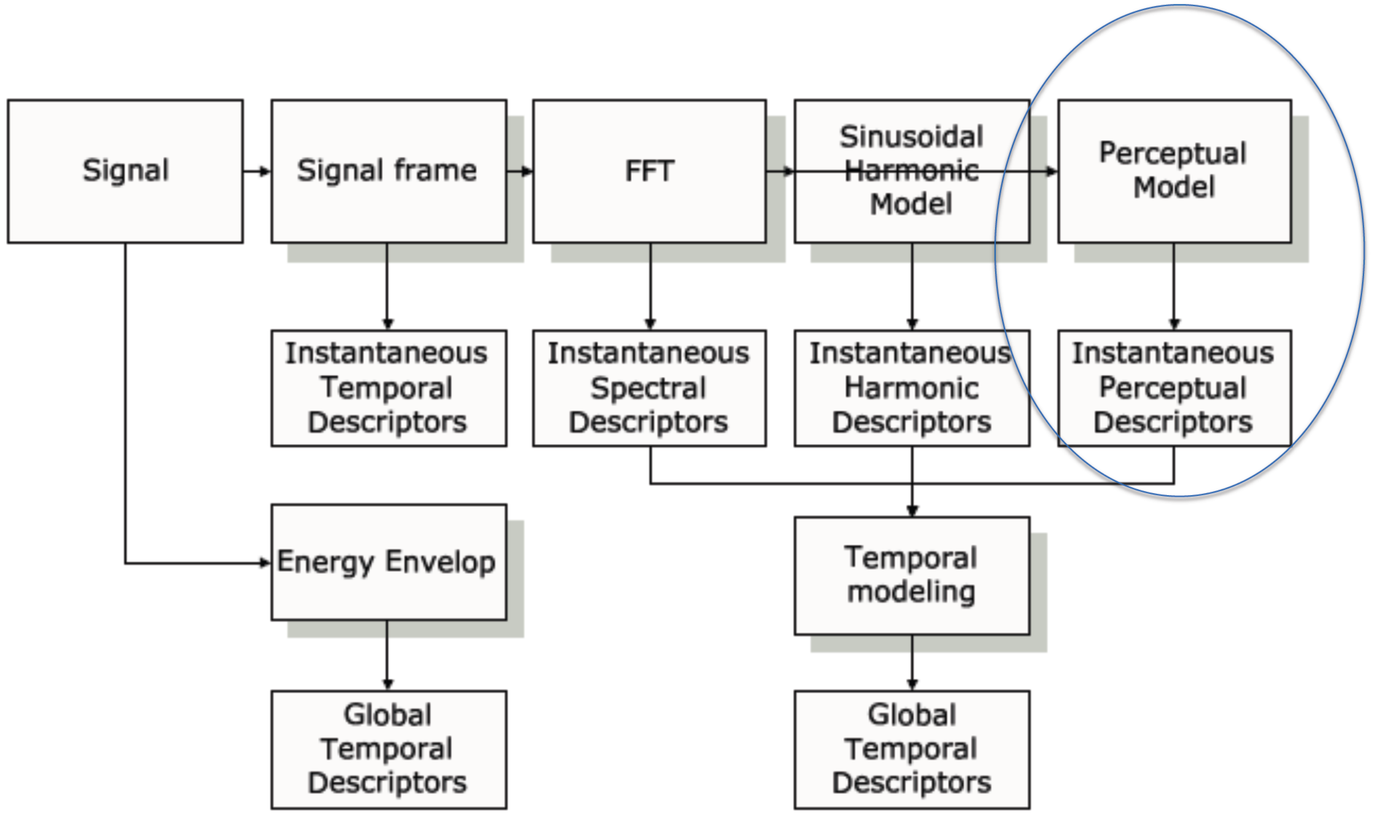
4. Spectral Feature computation



- **Spectral centroid, spread**
- **Spectral skewness, kurtosis**
- **MFCC's**

Low-level descriptors

4. Perceptual features
computation



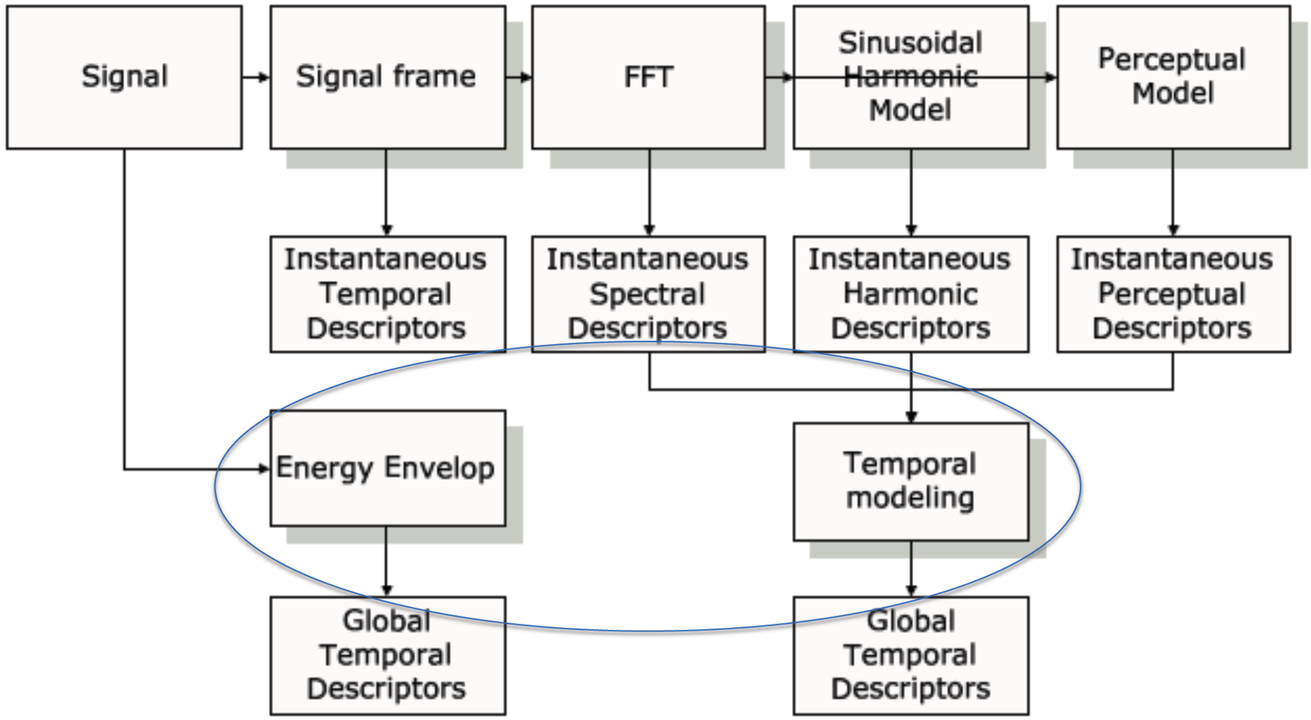
- **Loudness (bark bands)**
- **Sharpness (spectral centroid using bark-band loudness)**

Low-level descriptors

5. Temporal evolution of features

- Derivative;
- Differential normalized with its magnitude (better emulation with human audition)

- **Deltas**



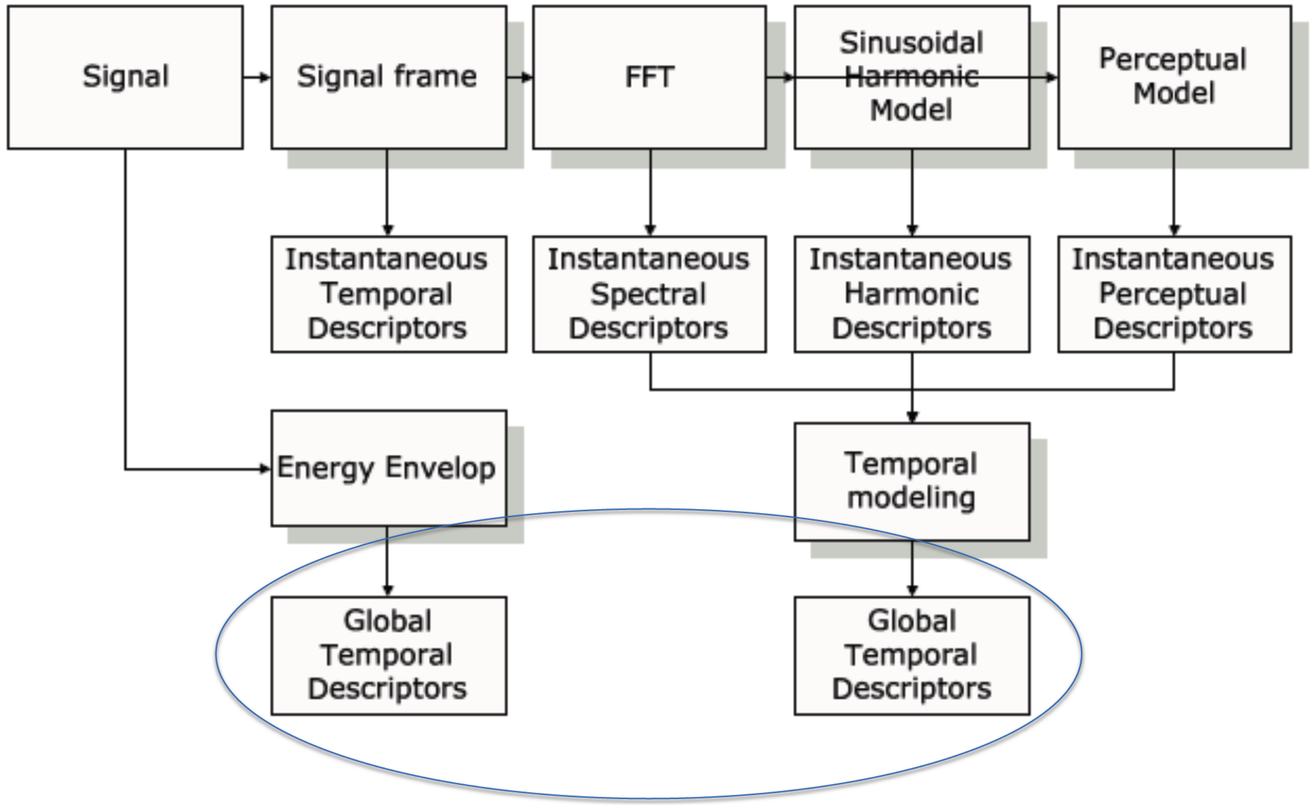
Low-level descriptors

5. Global features

- Statistic of instantaneous descriptors:
  - Moments (mean, variance, etc.)
  - Modulations

- Analysis of temporal evolution:
  - structural description
  - Segmentation
  - repetitions

- **Variance of Log-Attack Time**
- **Mean of spectral centroid**



Resources

General

- A large set of audio features for sound description (similarity and classification) in the CUIDADO project, G. Peeters, (2004)
- Instrument sound description in the context of MPEG-7, G. Peeters, S. McAdams, P. Herrera, ICMC (2000)

Software

Python

- Librosa
- Essentia

Matlab

- The Timbre Toolbox

3.2. Sound and Music Description

.