Spectral based sound description

Xavier Serra

Music Technology Group

Universitat Pompeu Fabra, Barcelona

http://mtg.upf.edu

Index

- Sinusoidal plus residual model features
- Spectral-based features in Essentia
- Features for sound/music description
- Features for instrument modeling
- Music description

Sinusoidal+residual model features

- Instantaneous frequency and amplitude of partials
- Instantaneous spectrum of residual
- Instantaneous fundamental frequency
- Amplitude and spectral shape of sinusoidal component
- Amplitude and spectral shape of residual

Essentia functionality

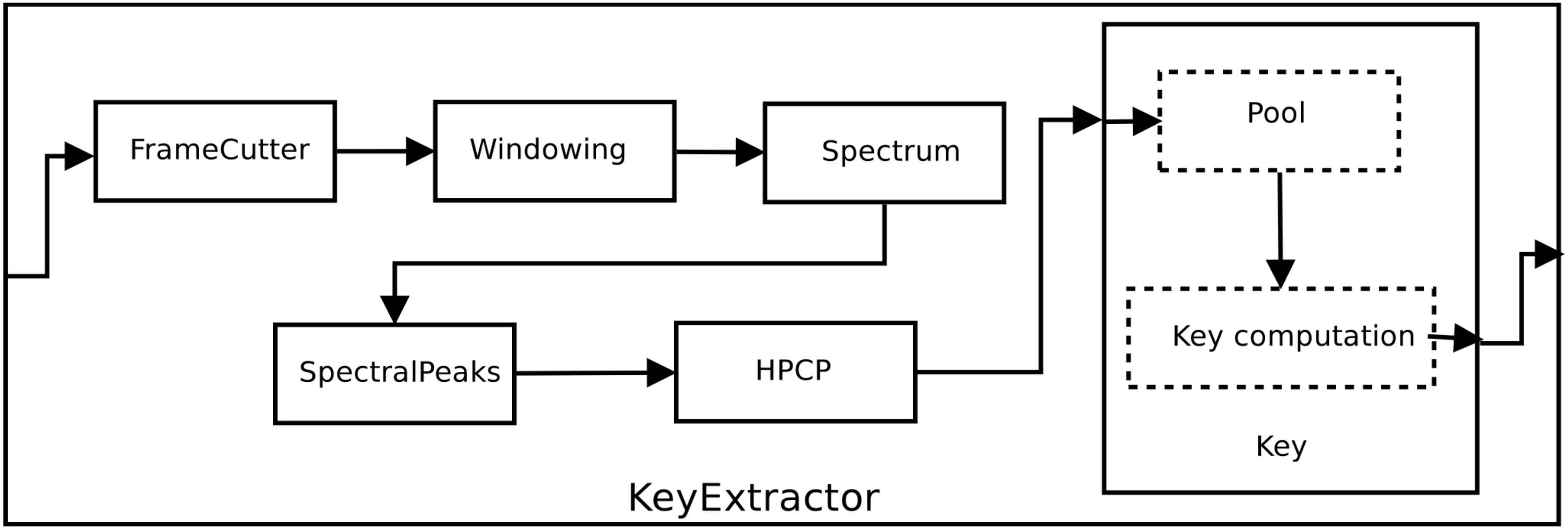
- audio file i/o; standard DSP building blocks; filters
- Descriptors such as
  - spectral: spectral shape, MFCC, Bark/Mel bands
  - time-domain/rhythmic: loudness, dynamics, onsets, beats, beats per minute, danceability
  - tonal: melody, pitch, chroma, key, scale, chords
  - high-level: segmentation, genres, mood (happy/sad), instrumentation (acoustic, electronic, timbre dark/bright, voice male/female)

Spectral features in Essentia

- **BarkBands:** computes the Bark band energies.
- **MelBands:** computes the Mel band energies.
- **ERBBands:** computes the energies in bands spaced on an Equivalent Rectangular Bandwidth scale.
- **MFCC:** computes the Mel-frequency cepstral coefficients of a frame.
- **GFCC:** computes the gammatone feature cepstrum coefficients similar to MFCCs.
- **LPC:** computes the Linear Predictive Coding coefficients of a frame as well as the associated reflection coefficients.
- **HFC:** computes the High-Frequency Content measure.
- **SpectralContrast**: computes spectral contrast of a spectrum.
- **Inharmonicity and Dissonance:** both try to estimate whether an audio frame “sounds” harmonic or not.
- **SpectralWhitening:** whitens the input spectrum.
- **Panning:** computes the panorama distribution of a stereo audio frame.

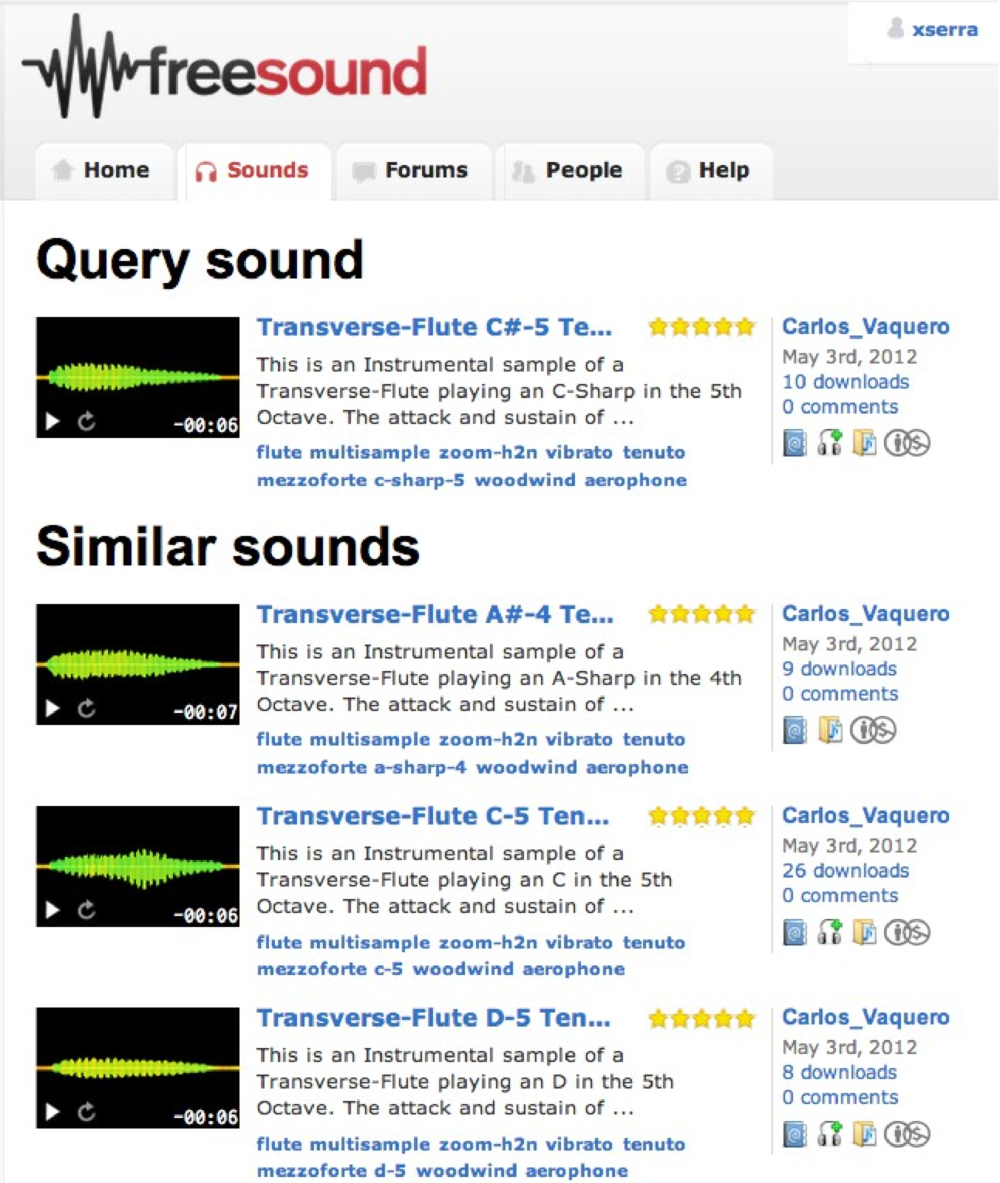
Essentia extractors

Executable extractors built by combining algorithms in a “data-flow” manner



```
import essentia
from essentia.standard import *
from pylab import *

loader = essentia.standard.MonoLoader(filename = 'oboe.wav')
audio = loader()

w = Windowing(type = 'hann')
spectrum = Spectrum()
mfcc = MFCC()

pool = essentia.Pool()

for frame in FrameGenerator(audio, frameSize = 1024, hopSize = 512):
    mfcc_bands, mfcc_coeffs = mfcc(spectrum(w(frame)))
    pool.add('lowlevel.mfcc', mfcc_coeffs)
    pool.add('lowlevel.mfcc_bands', mfcc_bands)

output = YamlOutput(filename = 'mfcc.sig')
output(pool)
```

Sound similarity



Query sound

Transverse-Flute C#-5 Te... ★★★★★

-00:06

This is an Instrumental sample of a Transverse-Flute playing an C-Sharp in the 5th Octave. The attack and sustain of ...

flute multisample zoom-h2n vibrato tenuto mezzoforte c-sharp-5 woodwind aerophone

Carlos_Vaquero
May 3rd, 2012
10 downloads
0 comments

Similar sounds

Transverse-Flute A#-4 Te... ★★★★★

-00:07

This is an Instrumental sample of a Transverse-Flute playing an A-Sharp in the 4th Octave. The attack and sustain of ...

flute multisample zoom-h2n vibrato tenuto mezzoforte a-sharp-4 woodwind aerophone

Carlos_Vaquero
May 3rd, 2012
9 downloads
0 comments

Transverse-Flute C-5 Ten... ★★★★★

-00:06

This is an Instrumental sample of a Transverse-Flute playing an C in the 5th Octave. The attack and sustain of ...

flute multisample zoom-h2n vibrato tenuto mezzoforte c-5 woodwind aerophone

Carlos_Vaquero
May 3rd, 2012
26 downloads
0 comments

Transverse-Flute D-5 Ten... ★★★★★

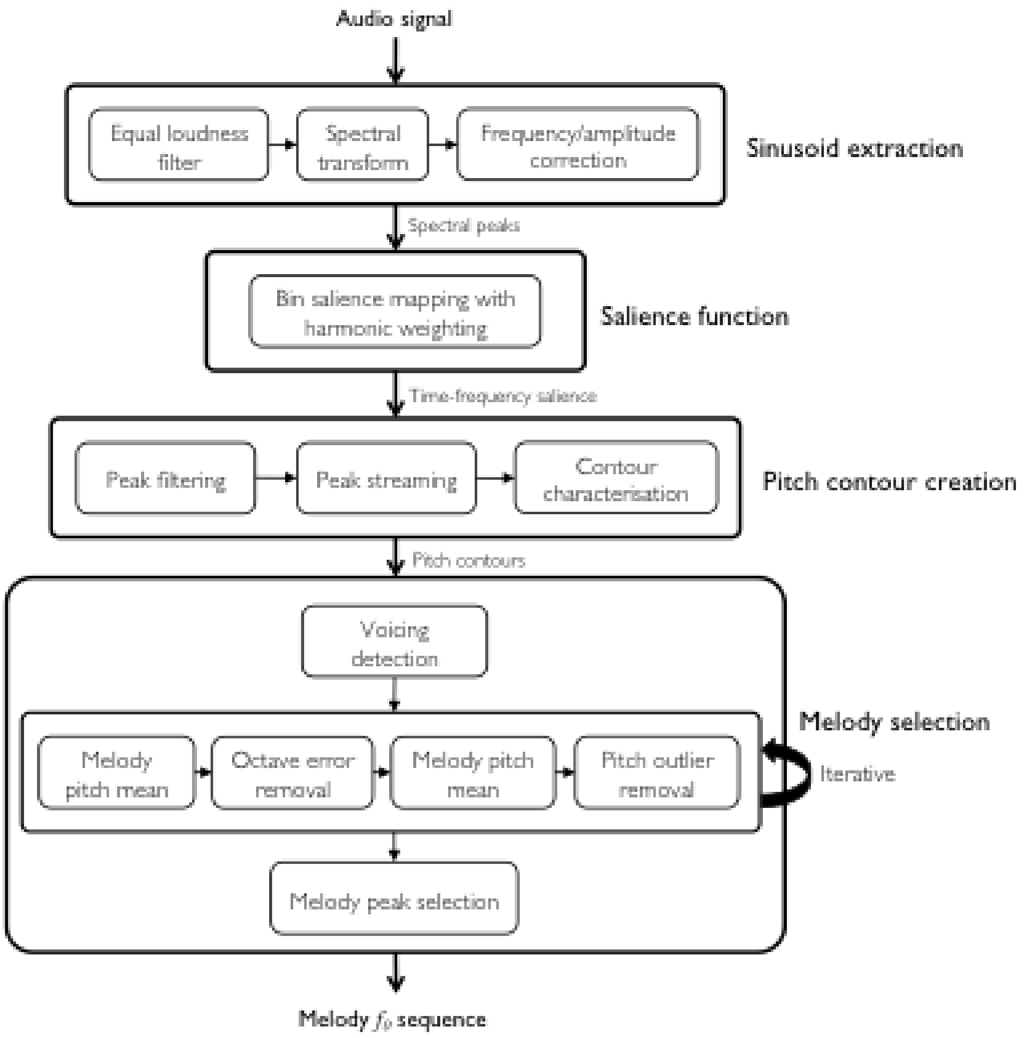
-00:06

This is an Instrumental sample of a Transverse-Flute playing an D in the 5th Octave. The attack and sustain of ...

flute multisample zoom-h2n vibrato tenuto mezzoforte d-5 woodwind aerophone

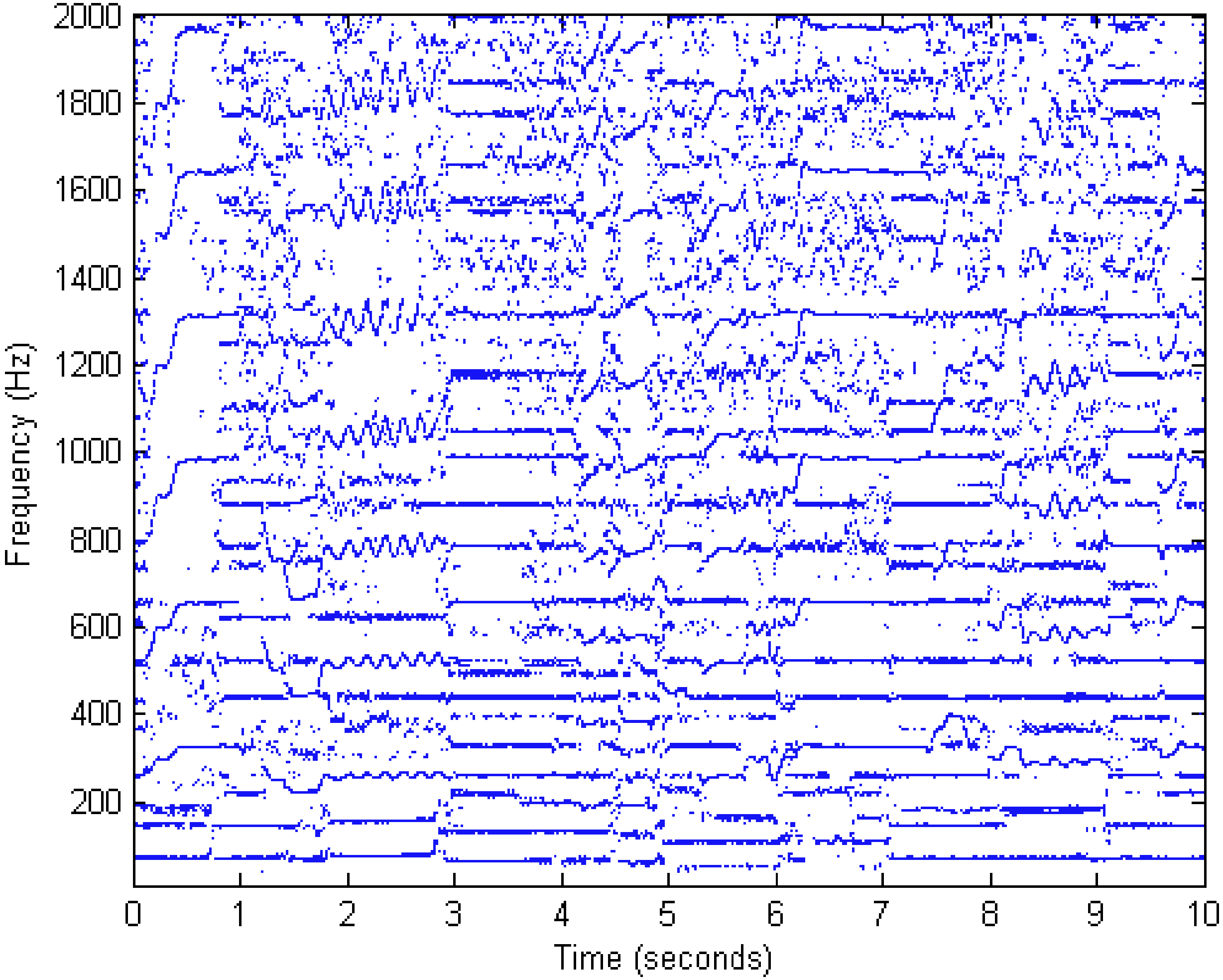
Carlos_Vaquero
May 3rd, 2012
8 downloads
0 comments

Prominent pitch detection

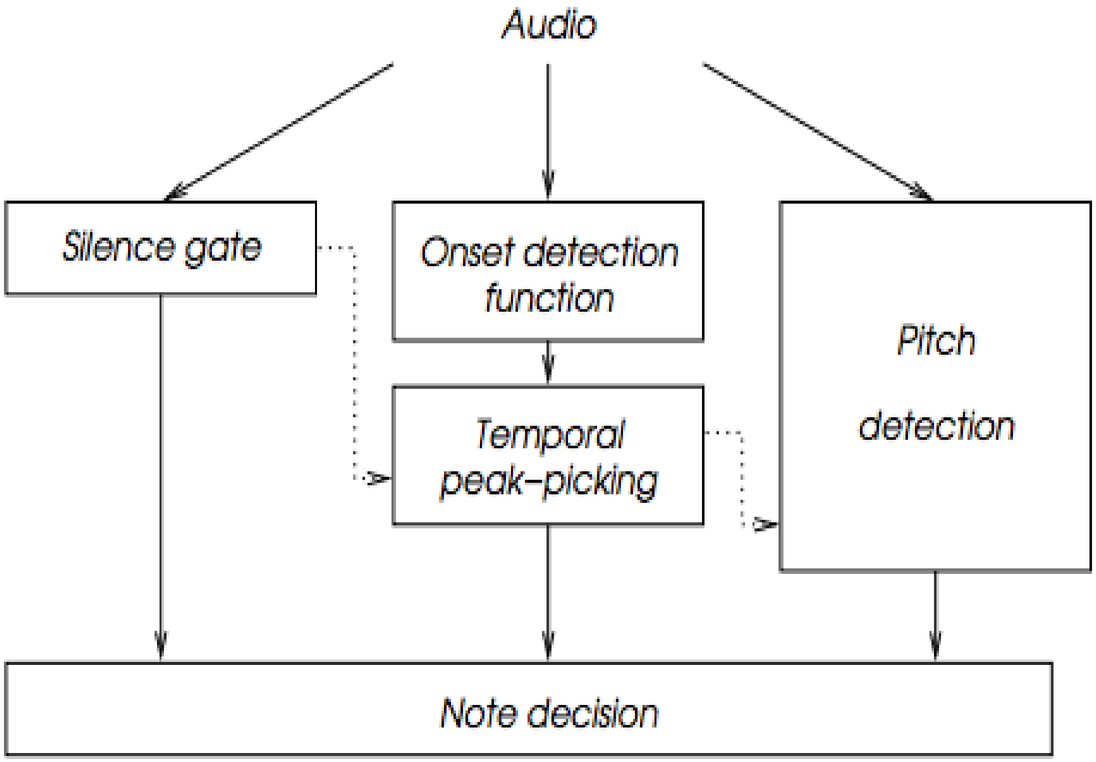


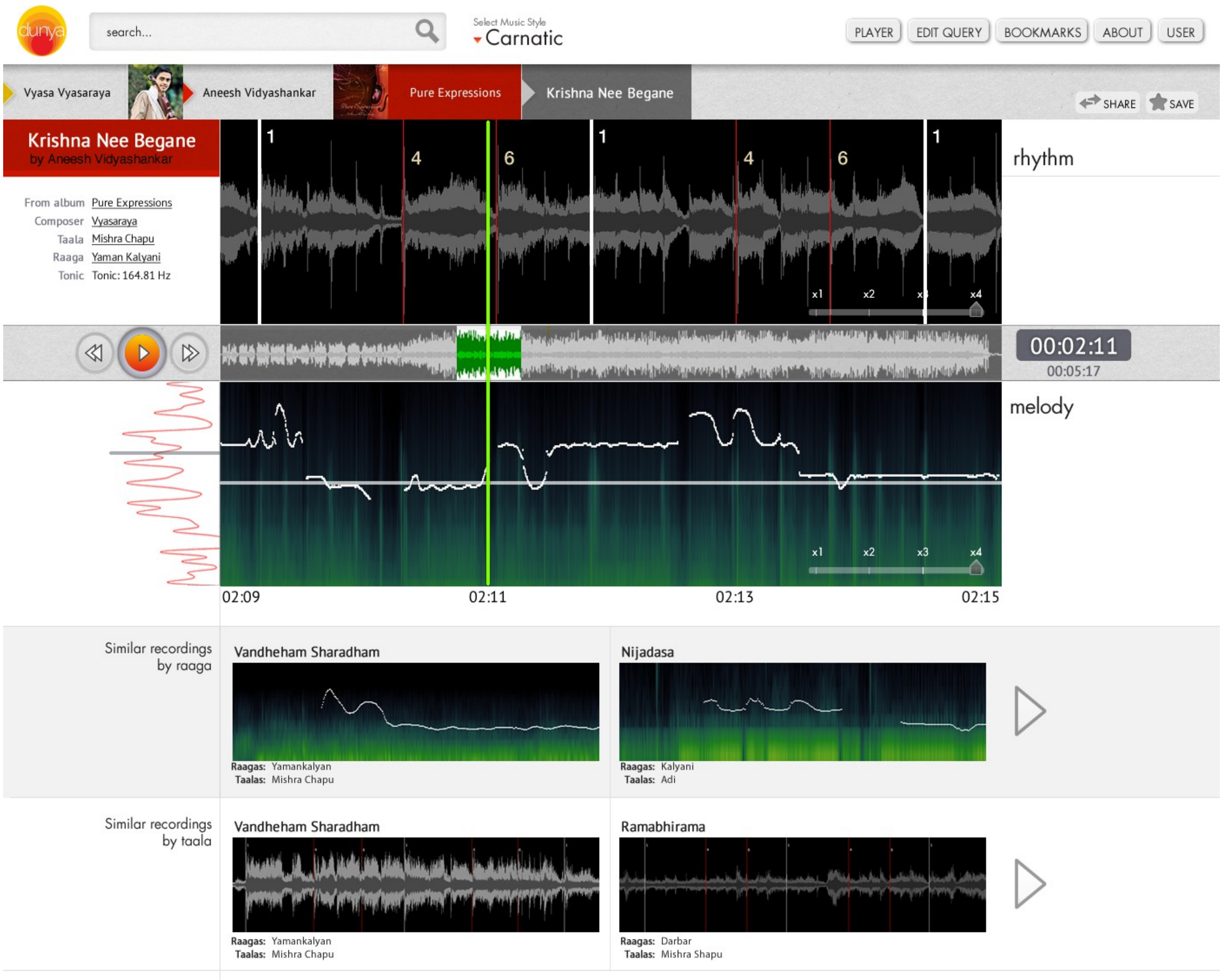
(Salamon, 2013)

Peak tracking in polyphonic signals



Onset detection





Krishna Nee Begane
by Aneesh Vidyashankar

From album Pure Expressions
Composer Vyasaraya
Taala Mishra Chapu
Raaga Yaman Kalyani
Tonic Tonic: 164.81 Hz

00:02:11
00:05:17

Similar recordings by raaga

Vandheham Sharadham
Raagas: Yamankalyan
Taalas: Mishra Chapu

Nijadasa
Raagas: Kalyani
Taalas: Adi

Similar recordings by taala

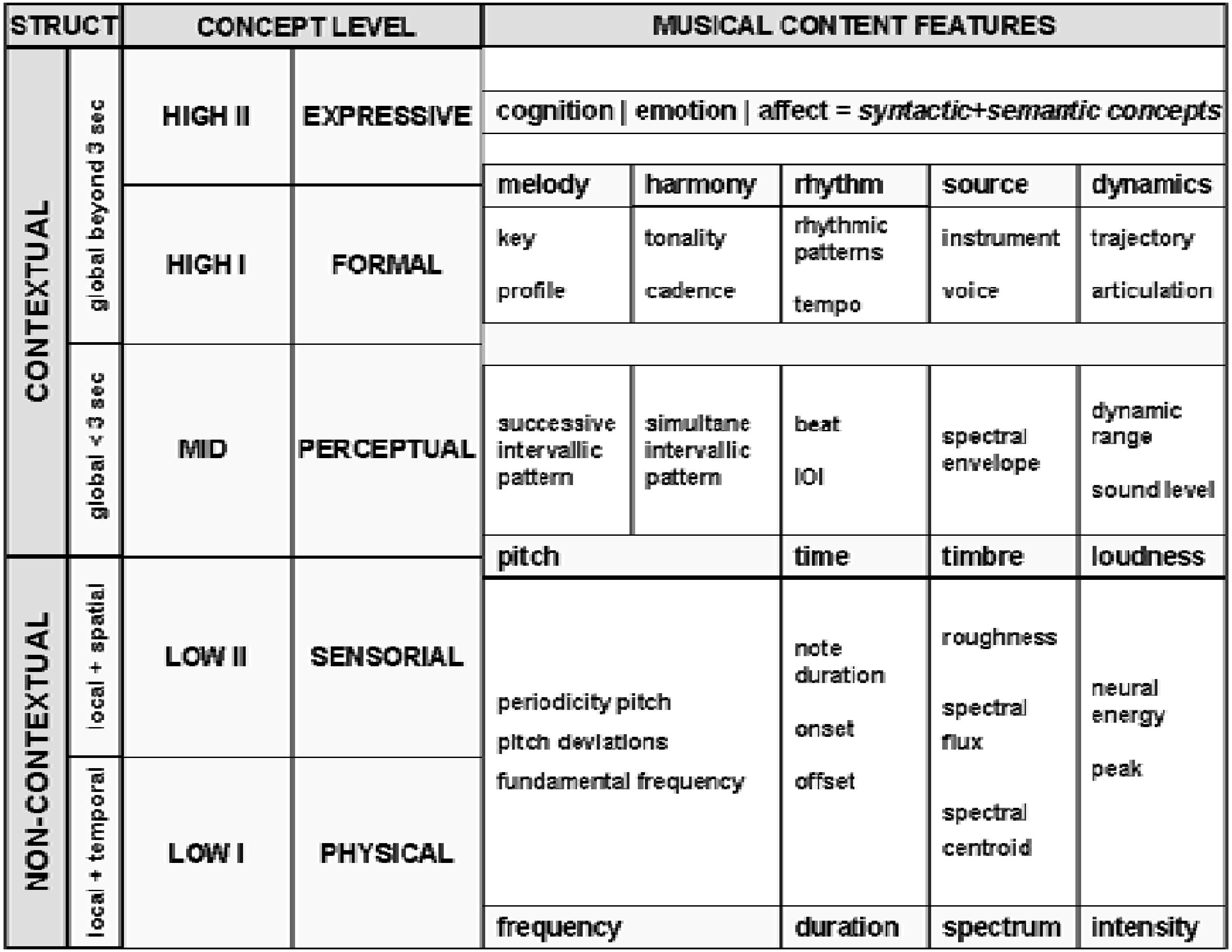
Vandheham Sharadham
Raagas: Yamankalyan
Taalas: Mishra Chapu

Ramabhirama
Raagas: Darbar
Taalas: Mishra Shapu

Instrument model

- Spectral Shape Models (formants, average shape, ...)
- Phase Models (constant, formant-based, ...)
- Frequency Models (harmonic model, piano model)
- Vibrato Models
- Articulation Models (frequency, amplitude functions)
- Residual Models
- Brightness-Loudness model (amplitude versus spectral tilt)

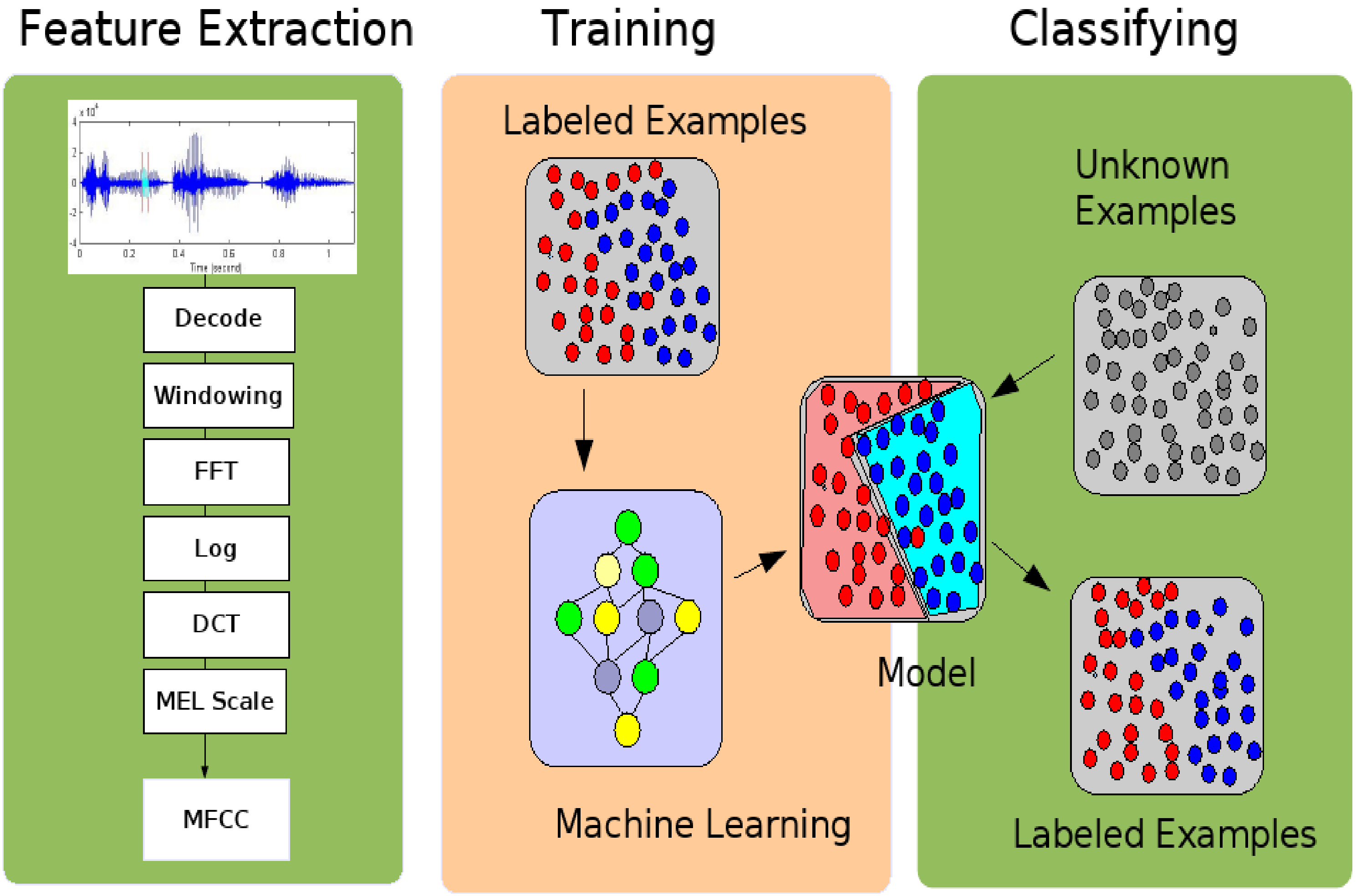
Taxonomy of musical features



STRUCT		CONCEPT LEVEL		MUSICAL CONTENT FEATURES				
CONTEXTUAL	global beyond 3 sec	HIGH II	EXPRESSIVE	cognition | emotion | affect = *syntactic+semantic concepts*				
				melody	harmony	rhythm	source	dynamics
		HIGH I	FORMAL	key profile	tonality cadence	rhythmic patterns tempo	instrument voice	trajectory articulation
	global < 3 sec	MID	PERCEPTUAL	successive intervallic pattern	simultane intervallic pattern	beat IOI	spectral envelope	dynamic range sound level
				pitch		time	timbre	loudness
NON-CONTEXTUAL	local + spatial	LOW II	SENSORIAL	periodicity pitch pitch deviations fundamental frequency		note duration onset offset	roughness spectral flux spectral centroid	neural energy peak
	local + temporal	LOW I	PHYSICAL					
				frequency		duration	spectrum	intensity

Lesaffre et alt.,
2003

Audio content classification



Feature Extraction

Decode

Windowing

FFT

Log

DCT

MEL Scale

MFCC

Training

Labeled Examples

Machine Learning

Model

Classifying

Unknown Examples

Labeled Examples

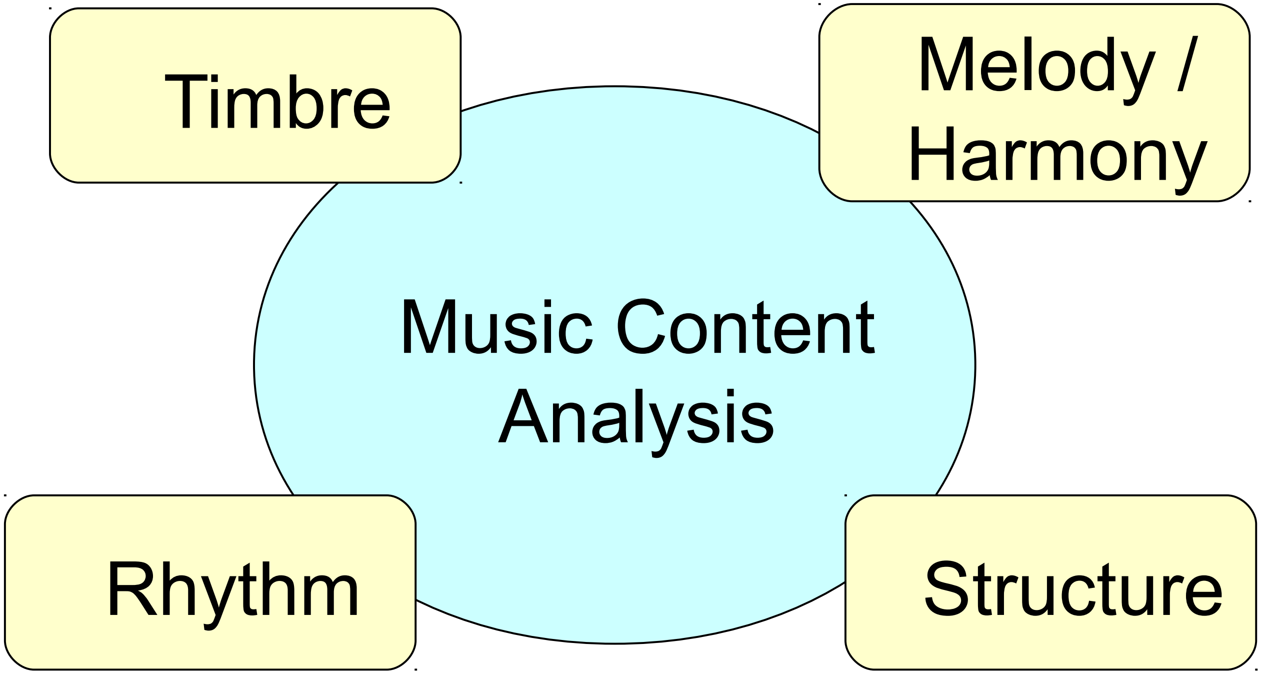
Levels of description

- Low-level (signal-centered) descriptors: computed from the audio signal in a direct or derived (ex: spectral analysis) way: average energy, spectral centroid, MFCCs ….

- Mid-level (object-centered) descriptors: requiring an induction operation or data modeling: key, genre, instrument …

- High-level (user-centered) descriptors: requiring a user model: mood (ex: happy, sad), …

Facets of music content

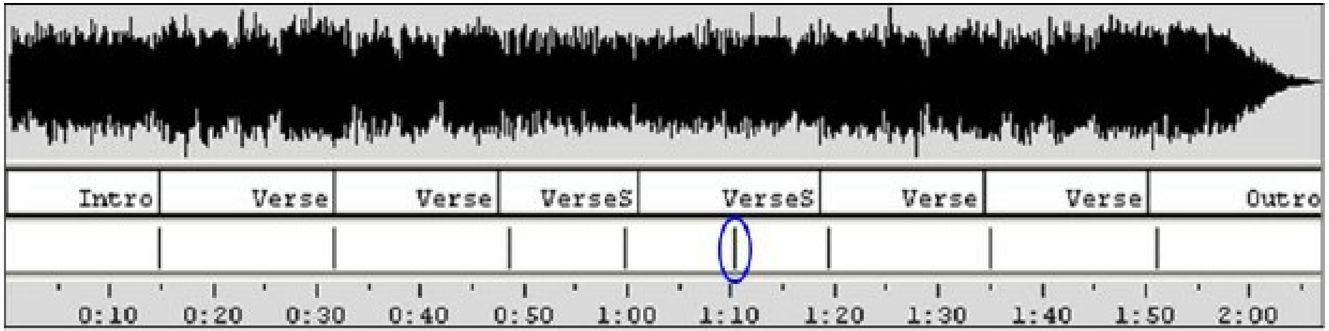


Structure description

- Partitioning the sound stream into *homogeneous* regions

- Detecting special roles for the segmented regions: intro, verse, chorus, bridge,

- Other segments can also be identified: instrumental / singing; solo / ensemble; chords…



(Ong, 2006)

Structure description

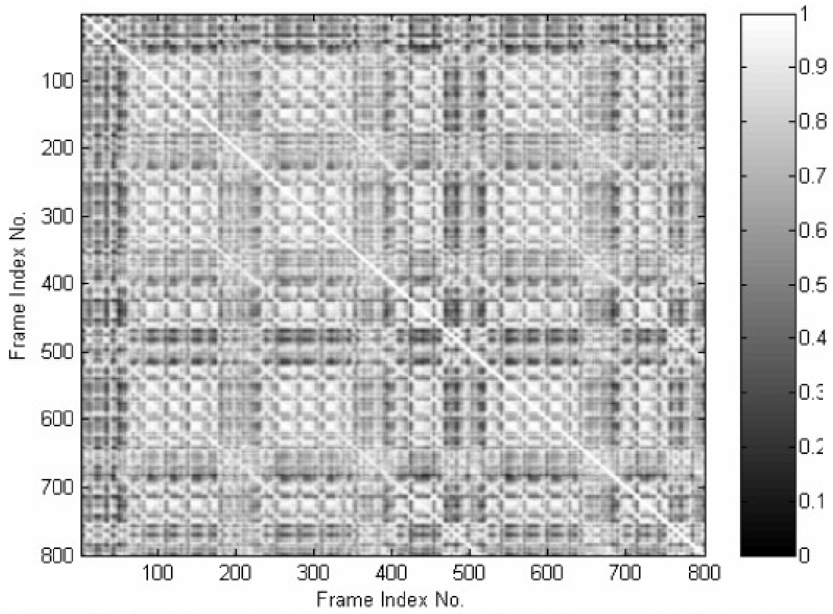
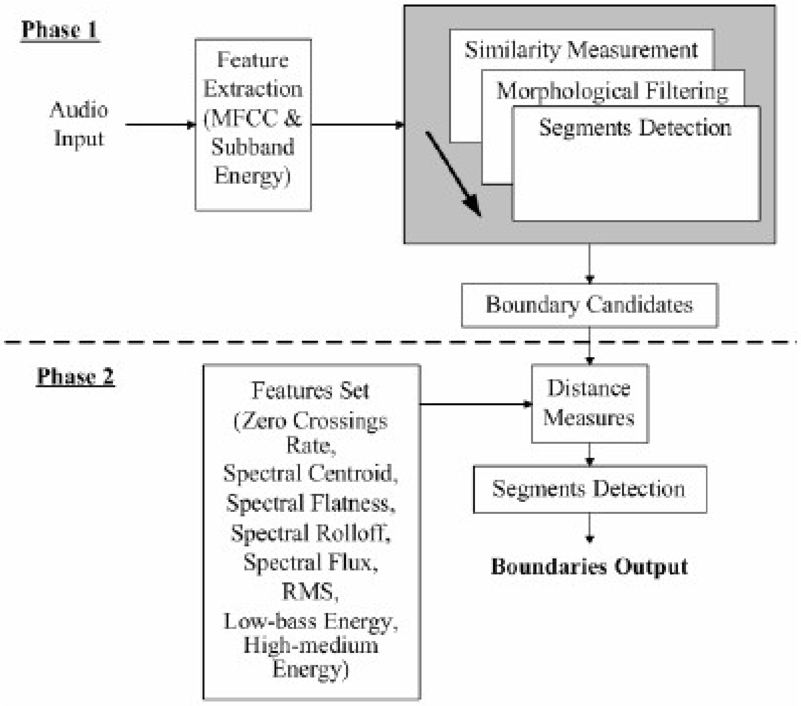


Figure 4.4. Two-dimensional similarity plot of The Beatles' song entitled *I'm a Loser*.

(Ong, 2006)

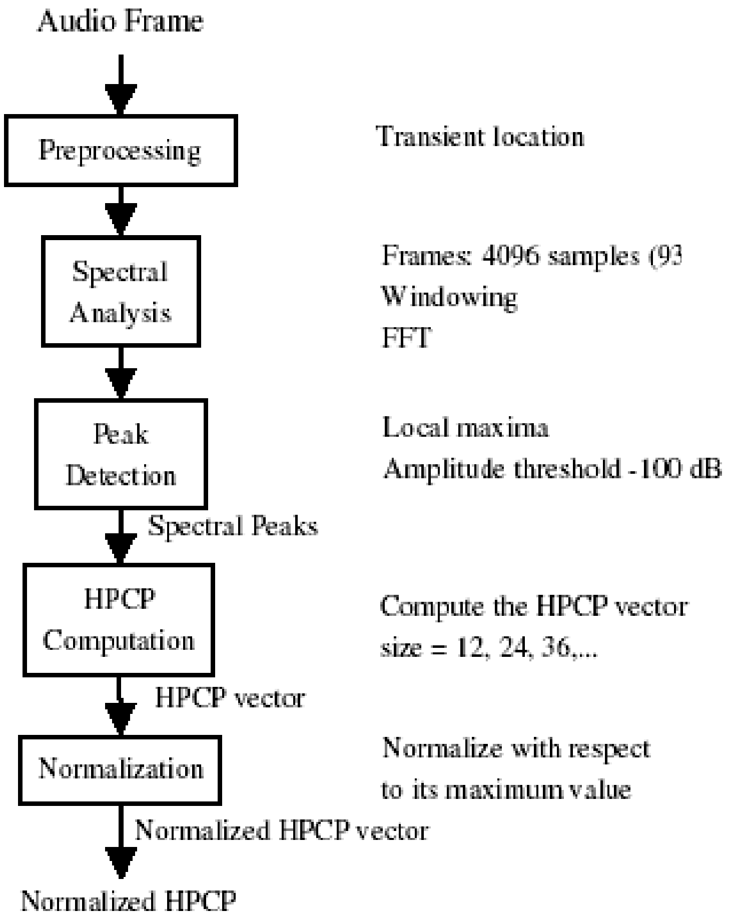


Phase 1
Audio Input
Feature Extraction (MFCC & Subband Energy)
Similarity Measurement
Morphological Filtering
Segments Detection
Boundary Candidates
Phase 2
Features Set (Zero Crossings Rate, Spectral Centroid, Spectral Flatness, Spectral Rolloff, Spectral Flux, RMS, Low-bass Energy, High-medium Energy)
Distance Measures
Segments Detection
Boundaries Output

Tonal description

- Extract:
  - Melody (predominant melody or score)
  - Harmony (chords)
  - Key, modulations
- Much research is related to **automatic transcription** of music *(Klapuri PhD 2004)*
  - Fundamental frequency / Multipitch estimation (*de Cheveigné*)
  - Melody extraction (Predominant pitch, note segmentation)
  - Still unsolved, even for monophonic signals.
- Pitch class distribution of a piece
- Mid and high level features -> apply a tonal model / musical analysis (*Krumhansl, Leman, Temperley, ….*)

Tonal description



Audio Frame

Preprocessing

Transient location

Spectral Analysis

Frames: 4096 samples (93

Windowing

FFT

Peak Detection

Local maxima

Amplitude threshold -100 dB

Spectral Peaks

HPCP Computation

Compute the HPCP vector

size = 12, 24, 36,...

HPCP vector

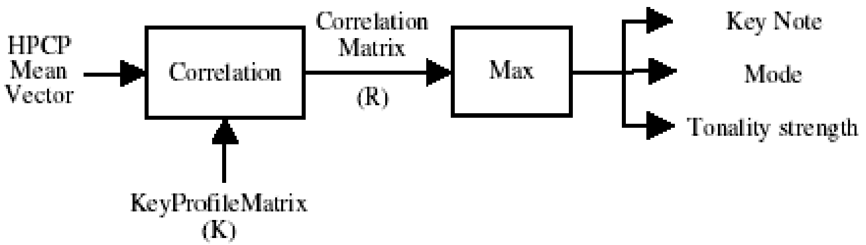
Normalization

Normalize with respect

to its maximum value

Normalized HPCP vector

Normalized HPCP



HPCP Mean Vector

Correlation

KeyProfileMatrix (K)

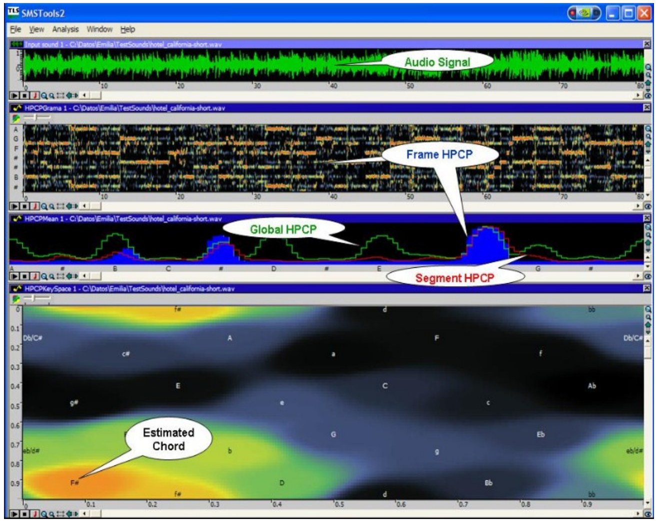
Correlation Matrix (R)

Max

Key Note

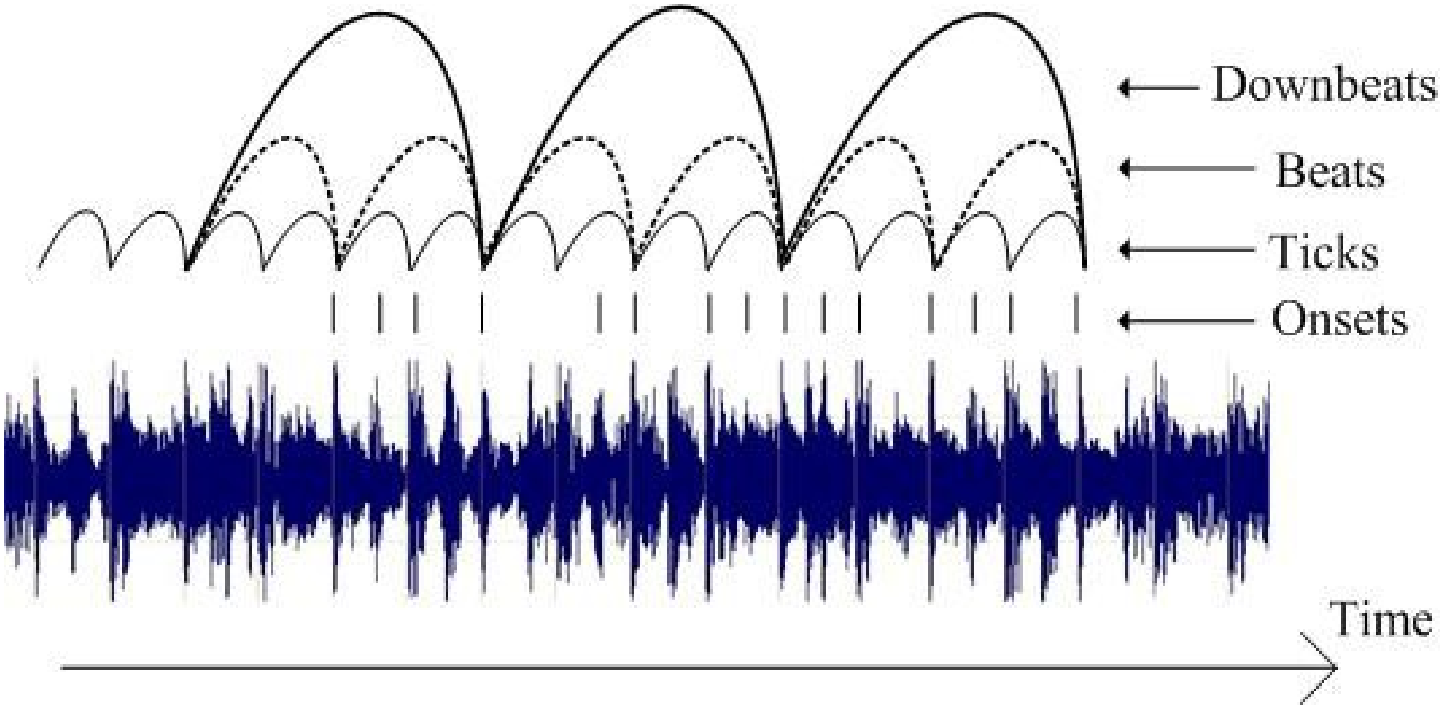
Mode

Tonality strength



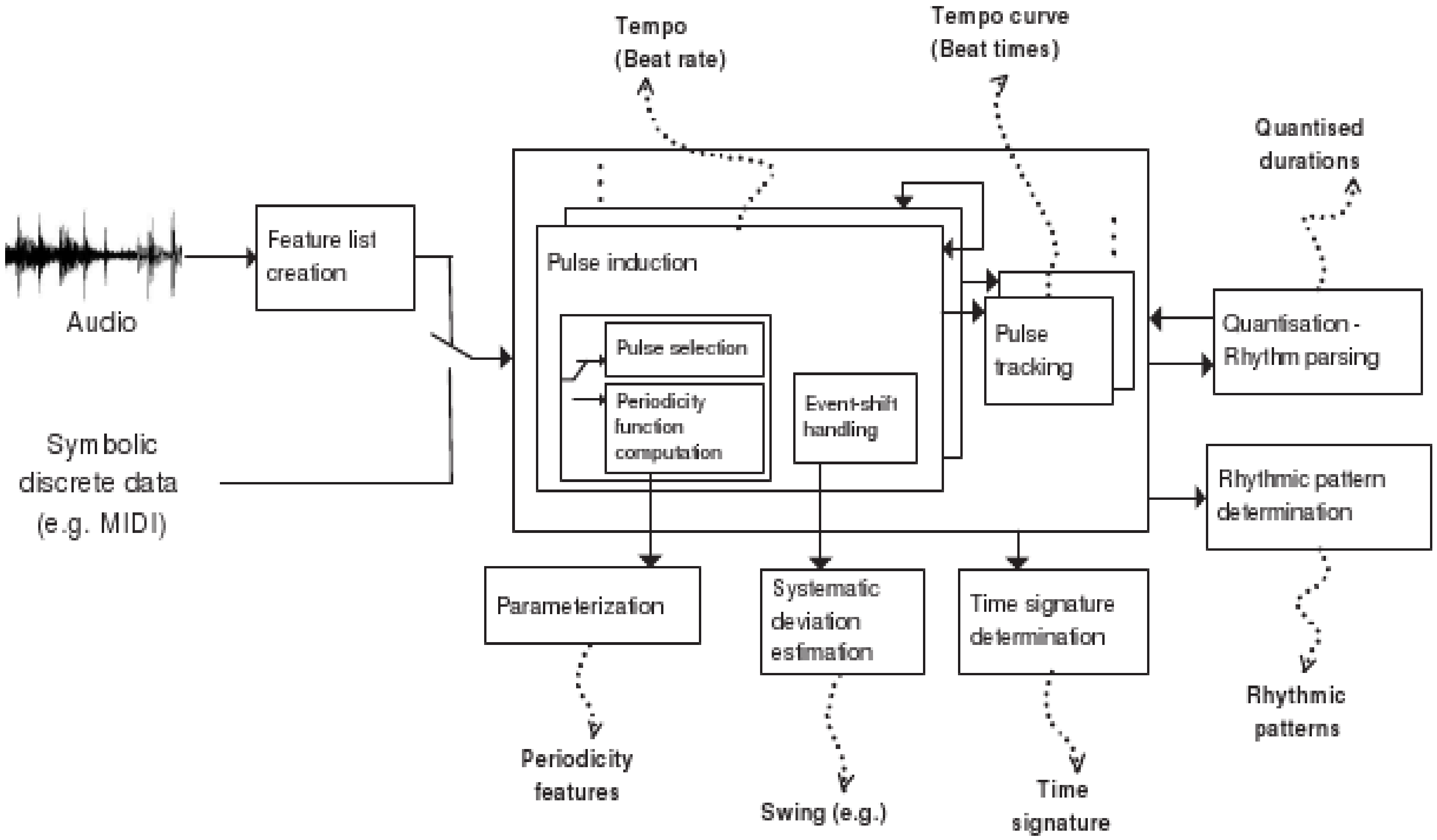
Rhythm description

Extraction of the metrical structure of a piece



(Gouyon, 2005)

Rhythm description



(Gouyon, 2005)

References

- http://essentia.upf.edu
- http://en.wikipedia.org/wiki/Music_information_retrieval
-

Credits