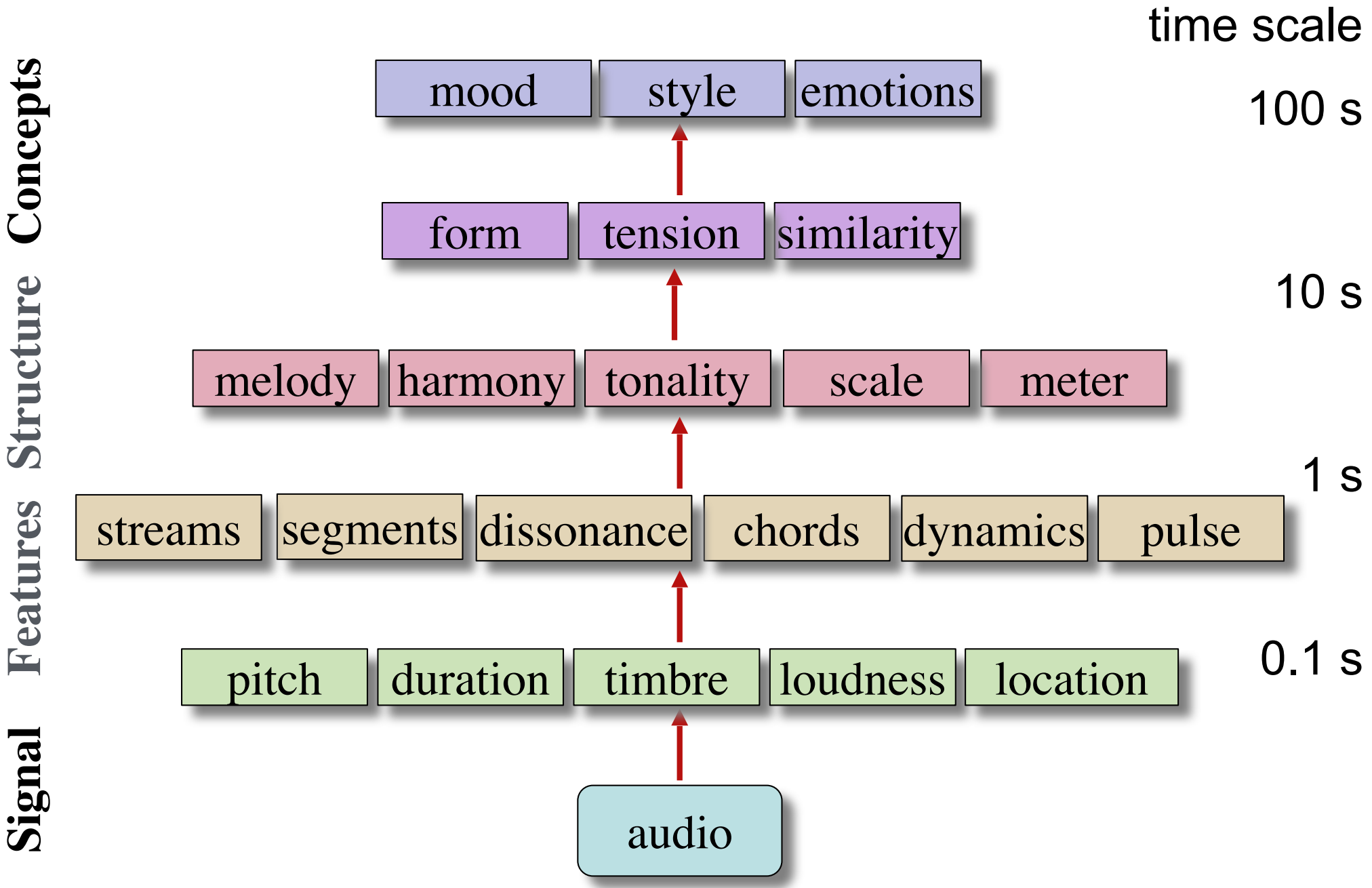
Why is window length important (perceptually)?

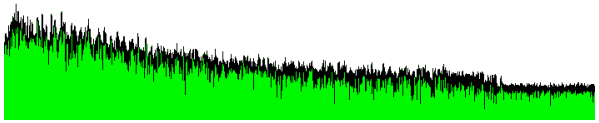
Levels of Music Processing



Musical features: Examples

Low-level / Timbral

brightness 

spectrum 

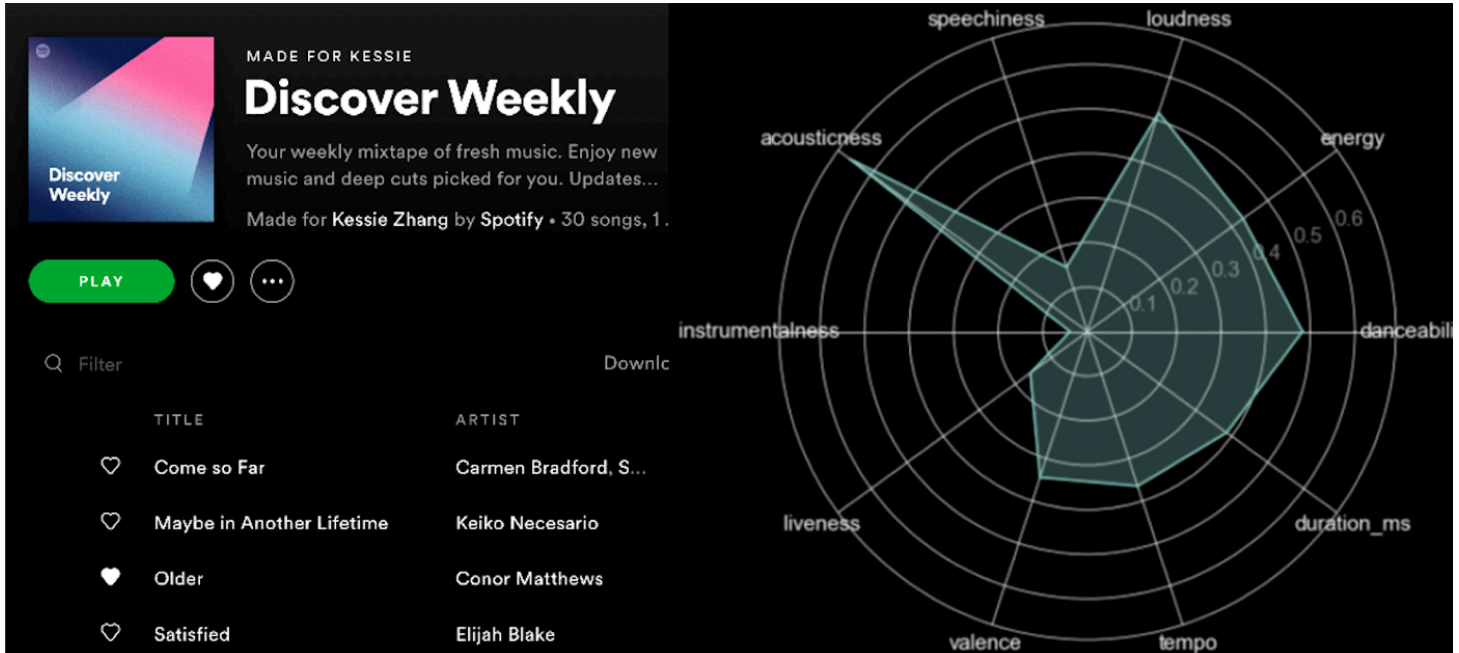
Mid/High-level / Rhythmic

pulse clarity

envelope autocorrelation

Feature Extraction from music

- features in music evolve continuously
- feature extraction relies on summarising this evolution (means, std)

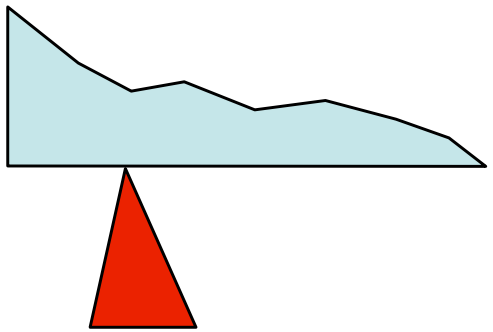


Which sounds brighter? (spectral centroid)

mean = 2167 Hz
std = 751 Hz

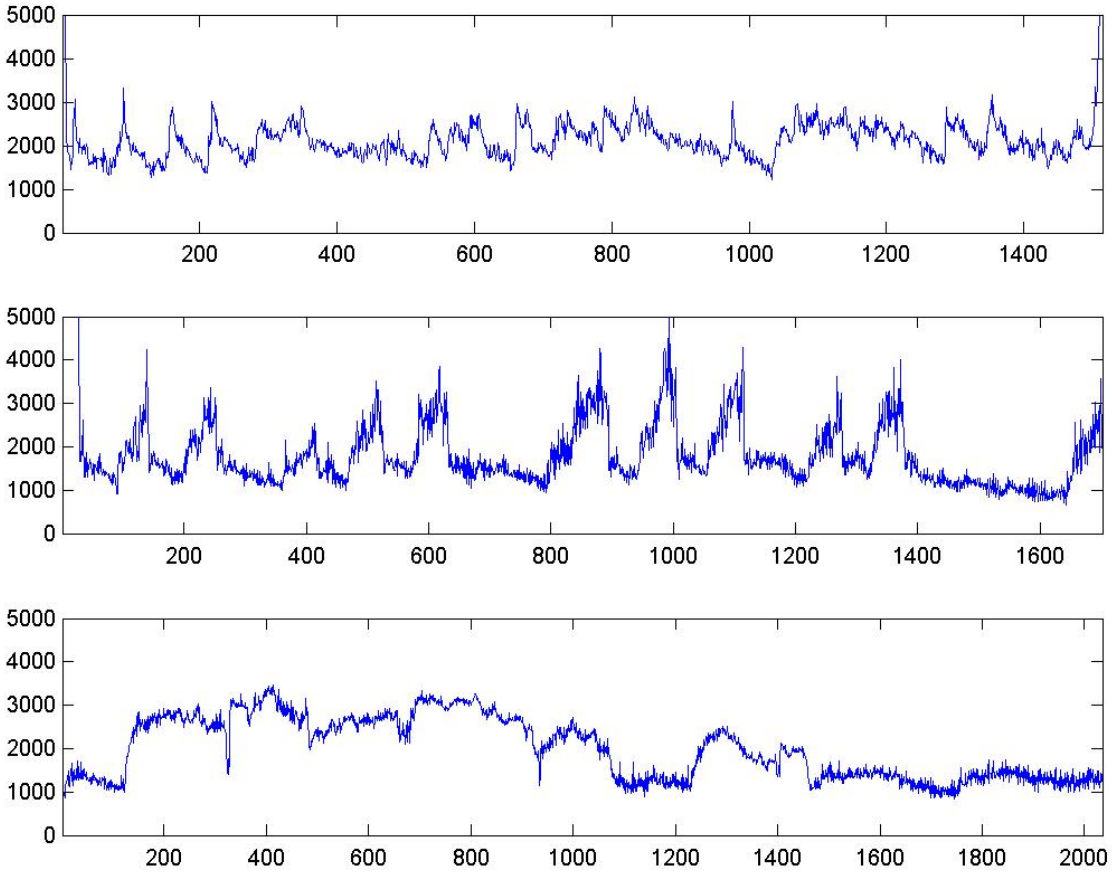
mean = 1953 Hz
std = 1534 Hz

mean = 1993 Hz
std = 706 Hz



Example of frame-based analysis

- spectral centroid of three excerpts



mean = 2167 Hz
std = 751 Hz

mean = 1953 Hz
std = 1534 Hz

mean = 1993 Hz
std = 706 Hz

Features Overview

- Dynamics
- Pitch
- Timbre
- Tempo/rhythm
- Tonality
- Structure

Acoustic features

Temporal

- zero-crossing rate
- low energy

Spectrotemporal

- spectral flux
- sub-band flux

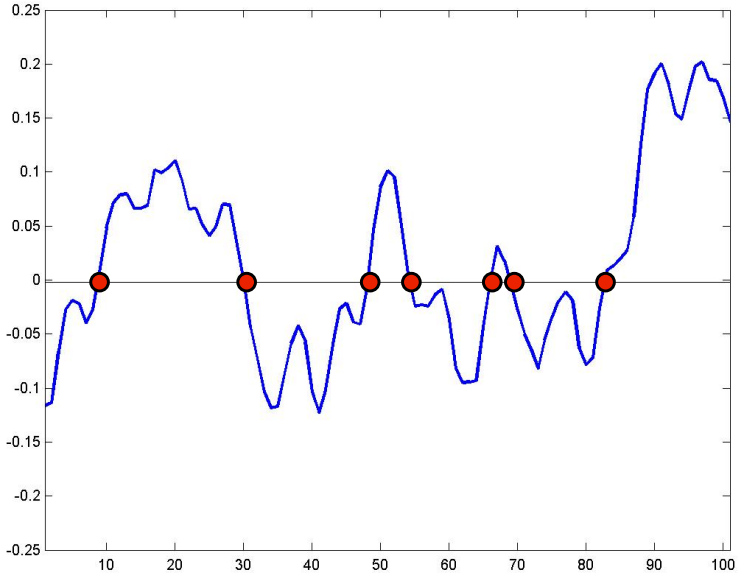
Spectral

- centroid
- high energy-low energy ratio
- entropy
- roll-off 85
- MFCC

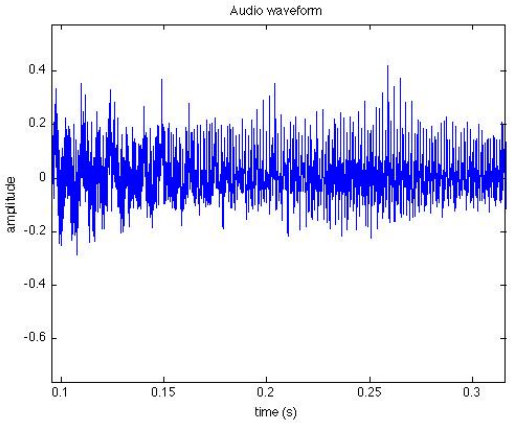
Identify features that might be useful for genre classification based on perceptual relevance

Zero-crossing rate

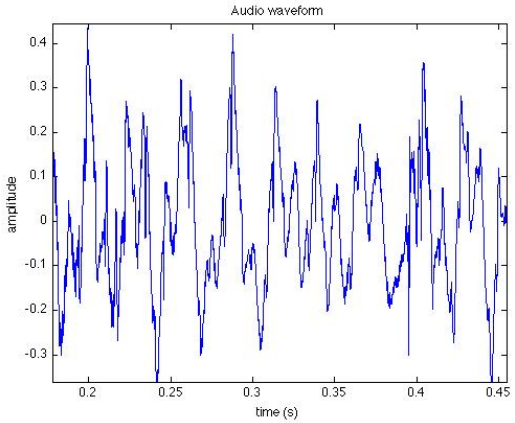
- number of time-domain zero-crossings of the signal per time unit



high:



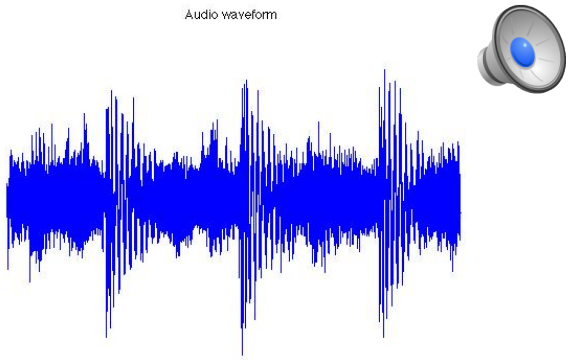
low:



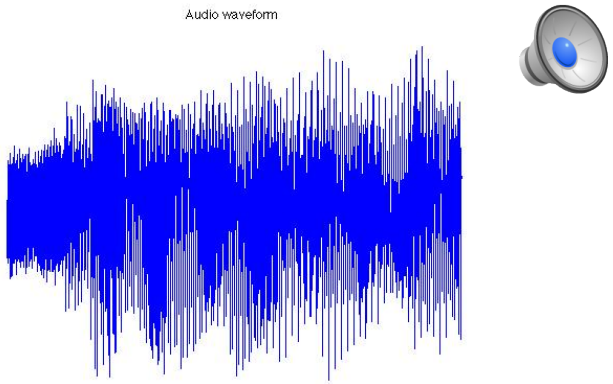
Low Energy

- proportion of signal frames whose energy is below average energy

high:

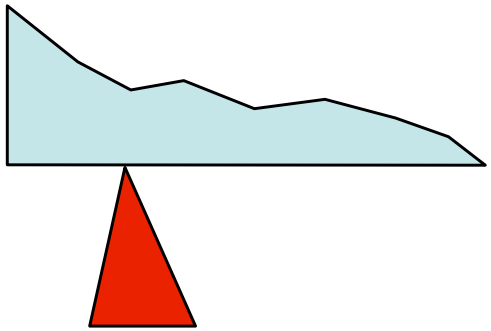


low:



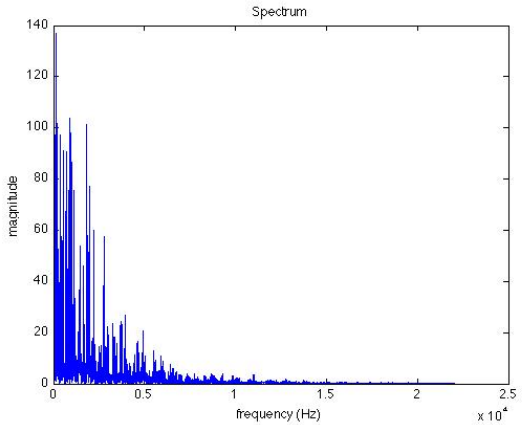
Spectral Centroid

- Center of mass of the spectrum

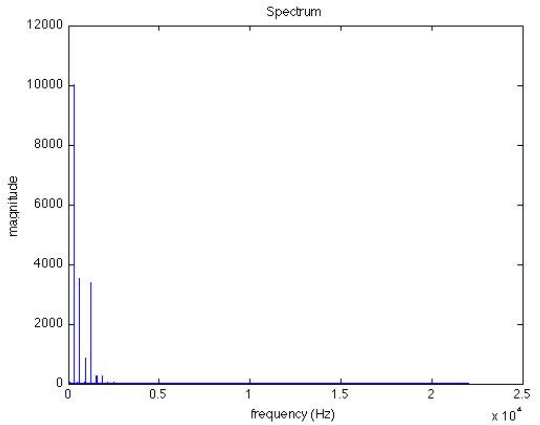


$$sc = \frac{\sum a_i f_i}{\sum a_i}$$

high:



low:

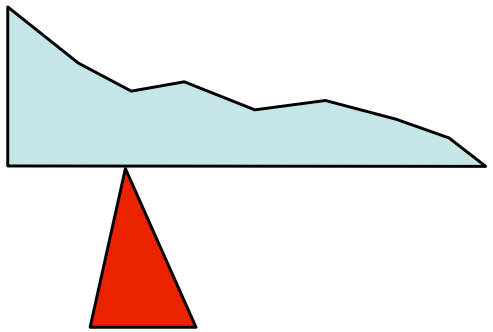


Which sounds brighter?

mean = 2167 Hz
std = 751 Hz

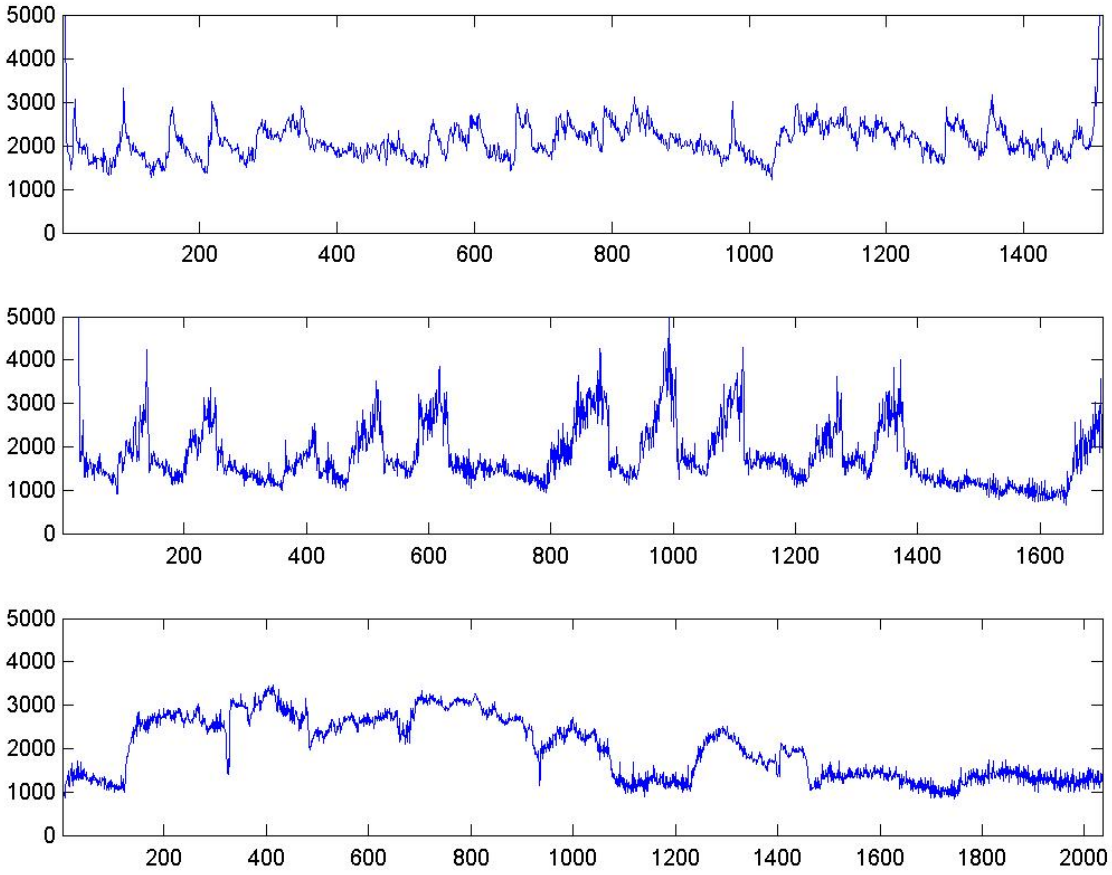
mean = 1953 Hz
std = 1534 Hz

mean = 1993 Hz
std = 706 Hz



Example of frame-based analysis

- spectral centroid of three excerpts



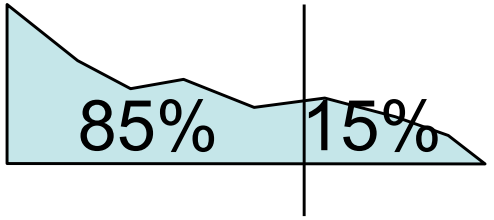
mean = 2167 Hz
std = 751 Hz

mean = 1953 Hz
std = 1534 Hz

mean = 1993 Hz
std = 706 Hz

Spectral Roll-Off

- Frequency, below which a certain fraction (usually 85%) of spectral energy

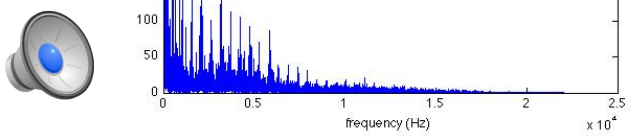


- R such that

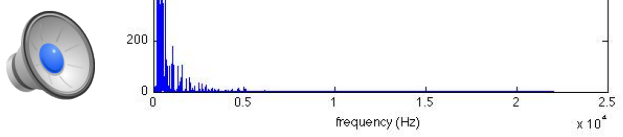
$$\sum_{1}^{R} a_i = 0.85 \sum_{1}^{N} a_i$$

- Measure of spectral shape

high:



low:

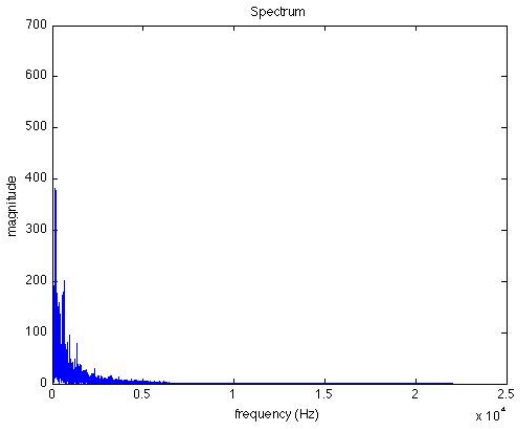


Spectral Irregularity

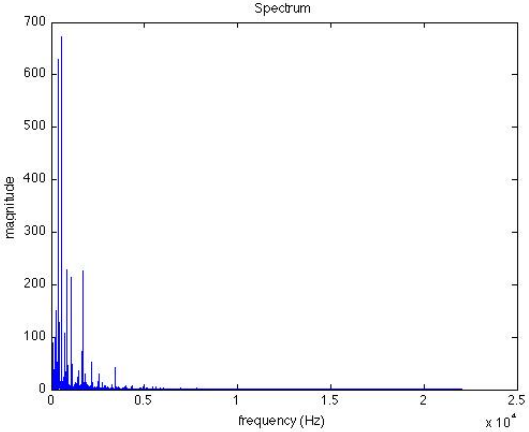
- measure of "jaggedness" of spectrum (Jensen, 1999)

$$irreg = \frac{\sum_{2}^{N} (a_i - a_{i-1})^2}{\sum_{1}^{N} a_i^2}$$

high:



low:

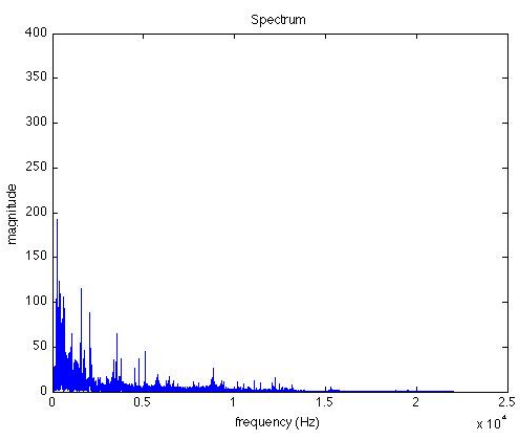


Spectral Entropy

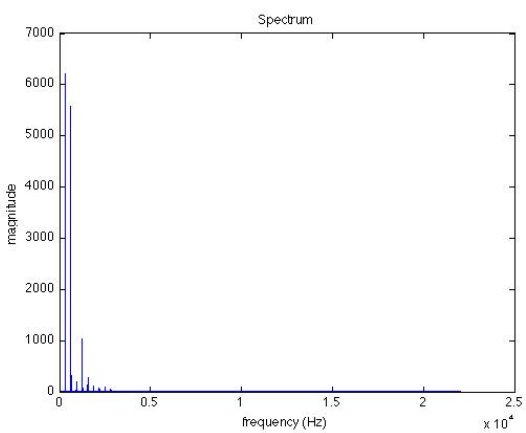
- information-theoretic measure of spectral energy distribution

- high entropy = even distribution of spectral energy (more noise-like?)

high:

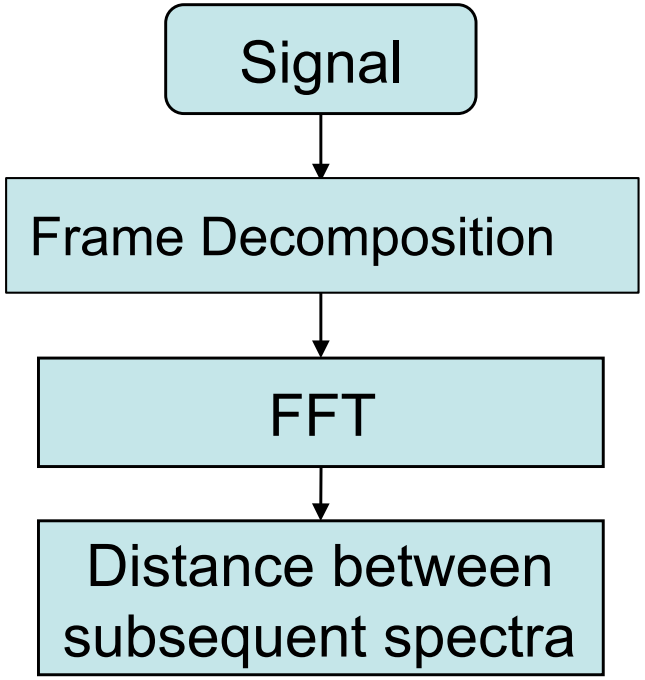


low:



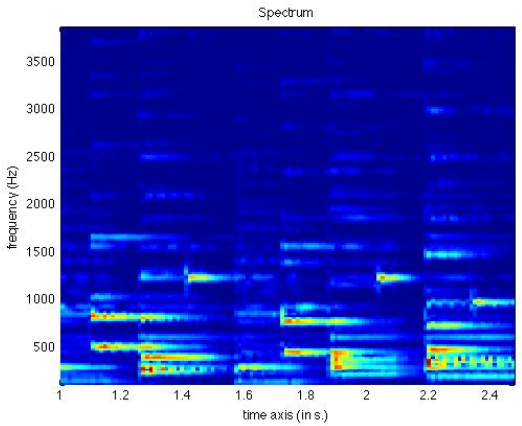
Spectral Flux

- Measure of change over time in spectrum
- Dissimilarity between subsequent spectral frames

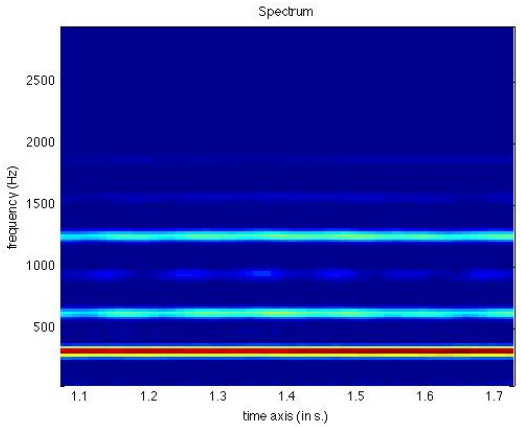


$$flux_i = \sum_{j=1}^{M} (a_{ij} - a_{(i-1)j})^2$$

high:

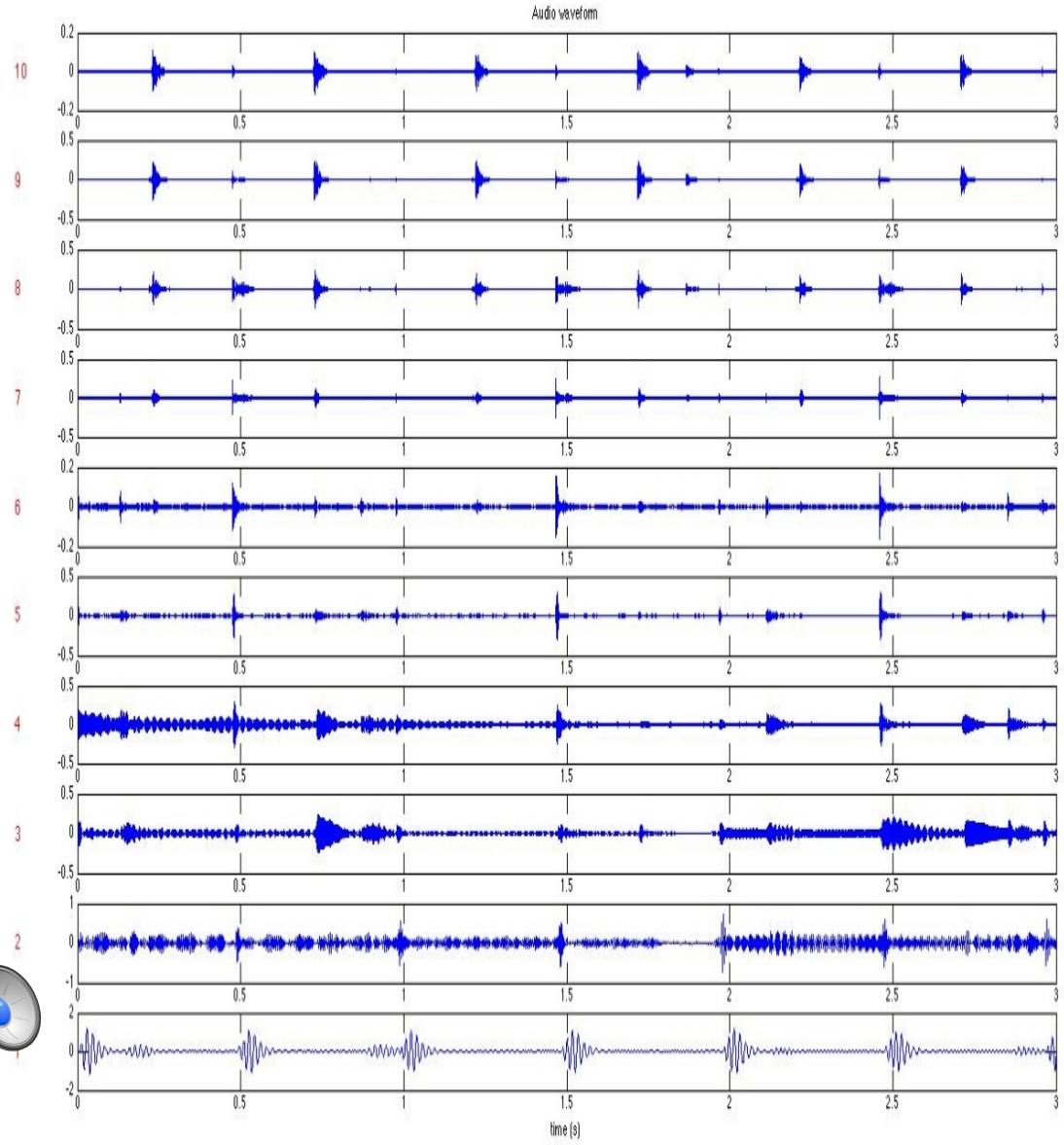


low:



Sub-band Flux

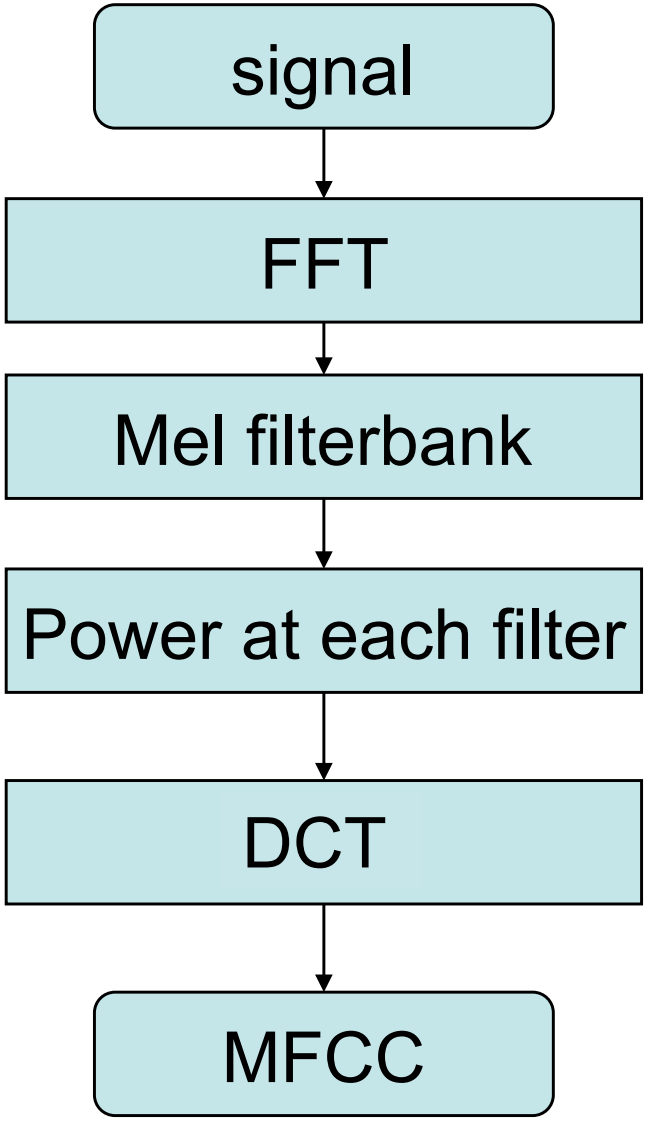
- Octave-scaled spectrum
  - 50 hz
  - 10 bands
- Spectral Flux in each band

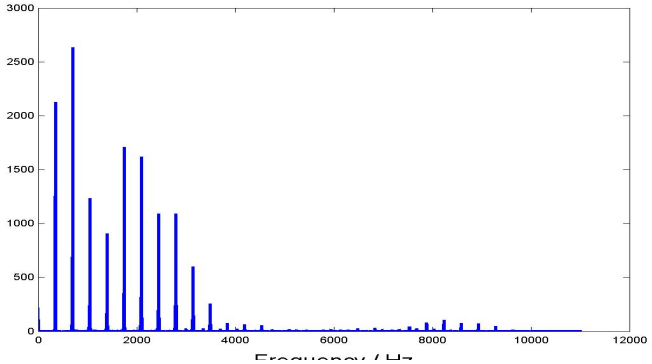


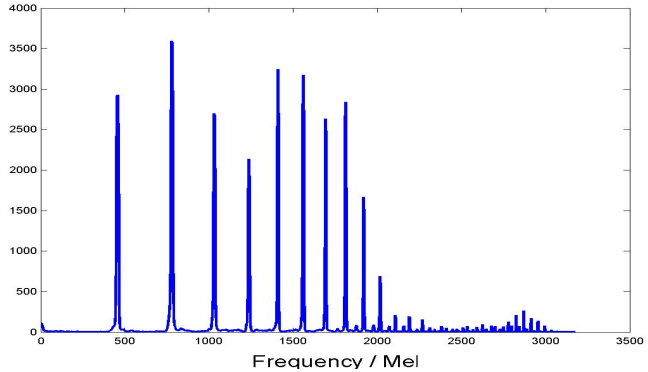
Mel-Frequency Cepstral Coefficients

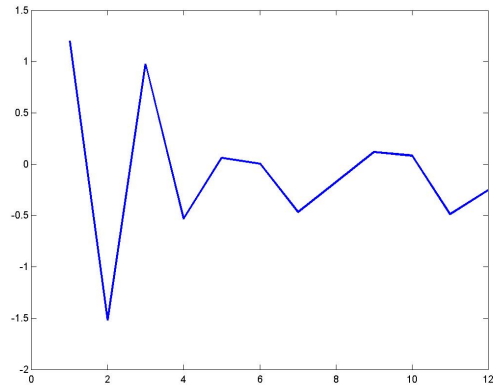
- Descriptor of spectral shape based on perception
- widely used in speech research (e.g. speech recognition)

Mel-frequency Cepstral Coefficients









Significance of MFCC

- provide a representation of the sound spectrum that closely corresponds to perceived distances between timbres (DePoli and Prandoni, 1997; Eronen, 2001; Terasawa et al., 2005)

- similarity in MFCC <-> similarity in perceived timbre

- important in classification of genre, mood, emotion, semantics

Genre Classification

Next class

Identify features that might be useful for genre classification based on perceptual relevance:

Temporal

- zero-crossing rate
- low energy

Spectrotemporal

- roughness
- sub-band flux

Spectral

- centroid
- high energy-low energy ratio
- entropy
- roll-off 85
- MFCC

Rhythm

Key/Tonality

Acoustic features

Rhythm

- tempo
- pulse clarity

Tonality

- chromagram
- mode
- keystrength/keyclarity

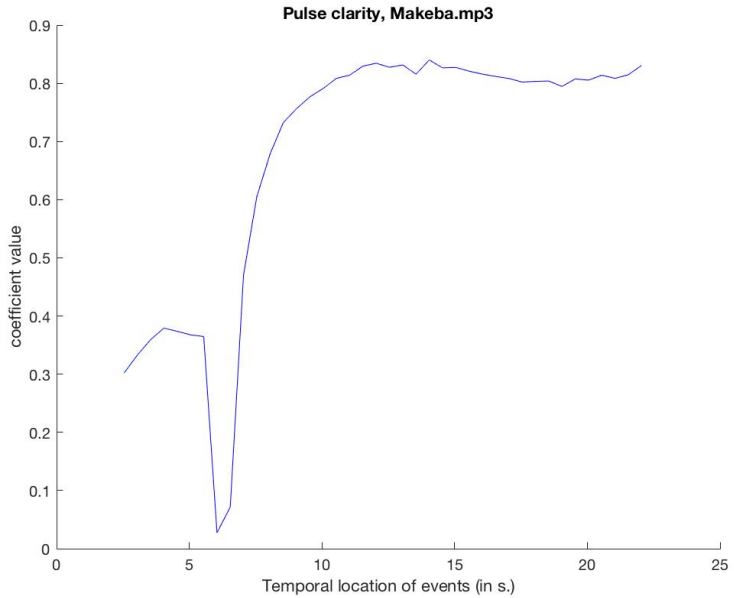
*typically extracted using longer time windows (context-dependent)

Rhythm

Tempo & Pulse Clarity

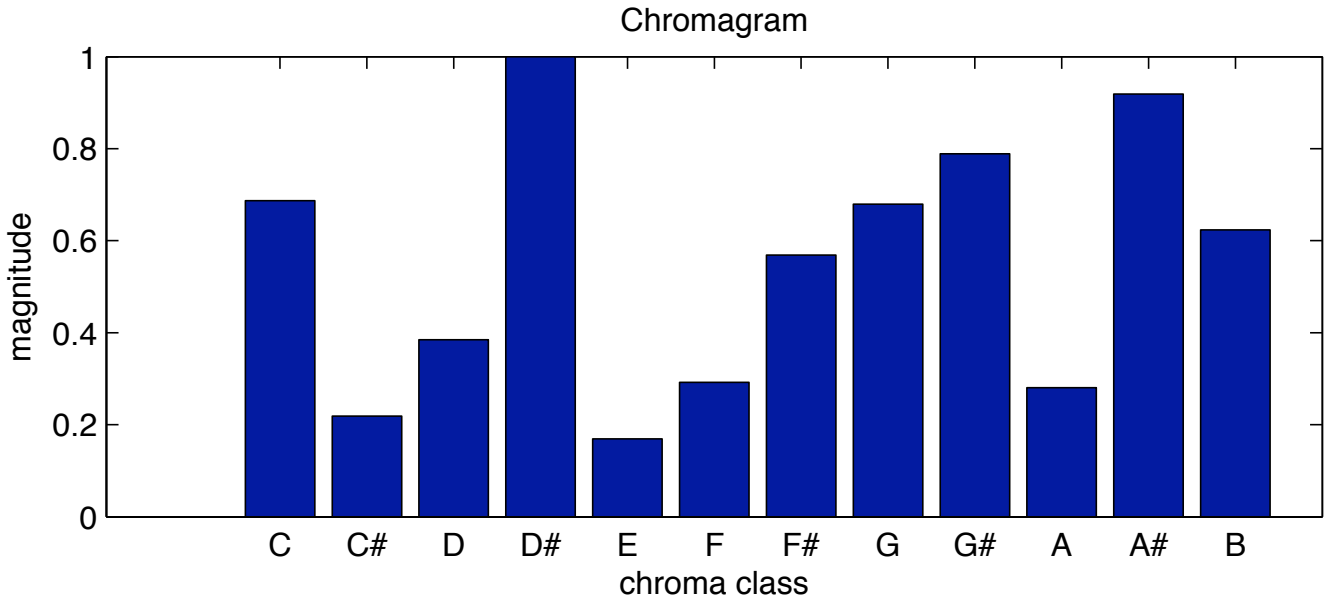
- tempo: estimate of how fast/slow the piece of music is

- pulse clarity/beat salience: how clear the beat is



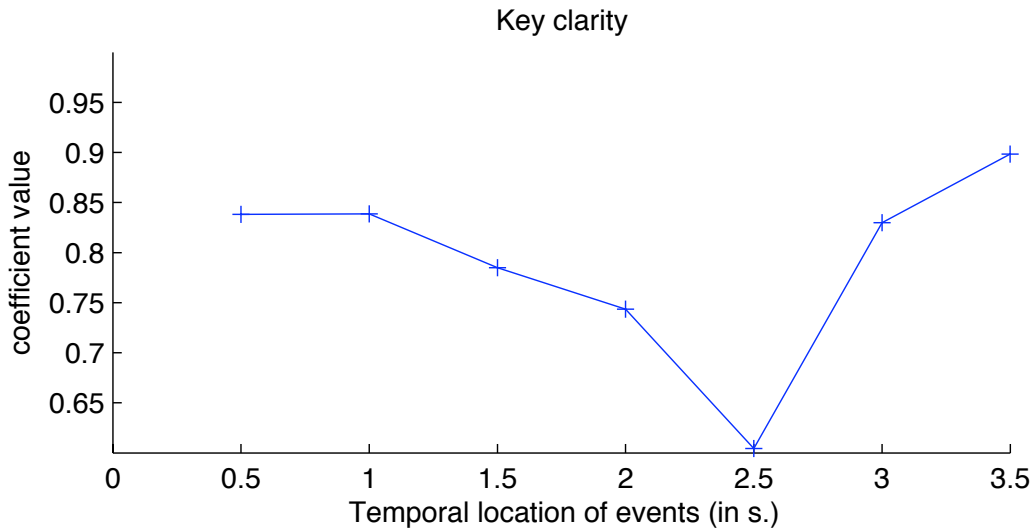
Acoustic features

- chromagram: Harmonic Pitch Class Profile, shows the distribution of energy along the pitches or pitch classes.



Acoustic features

- keystrength: measure of the tonal clarity



- mode: major or minor (roughly depicts “happy” or “sad”)