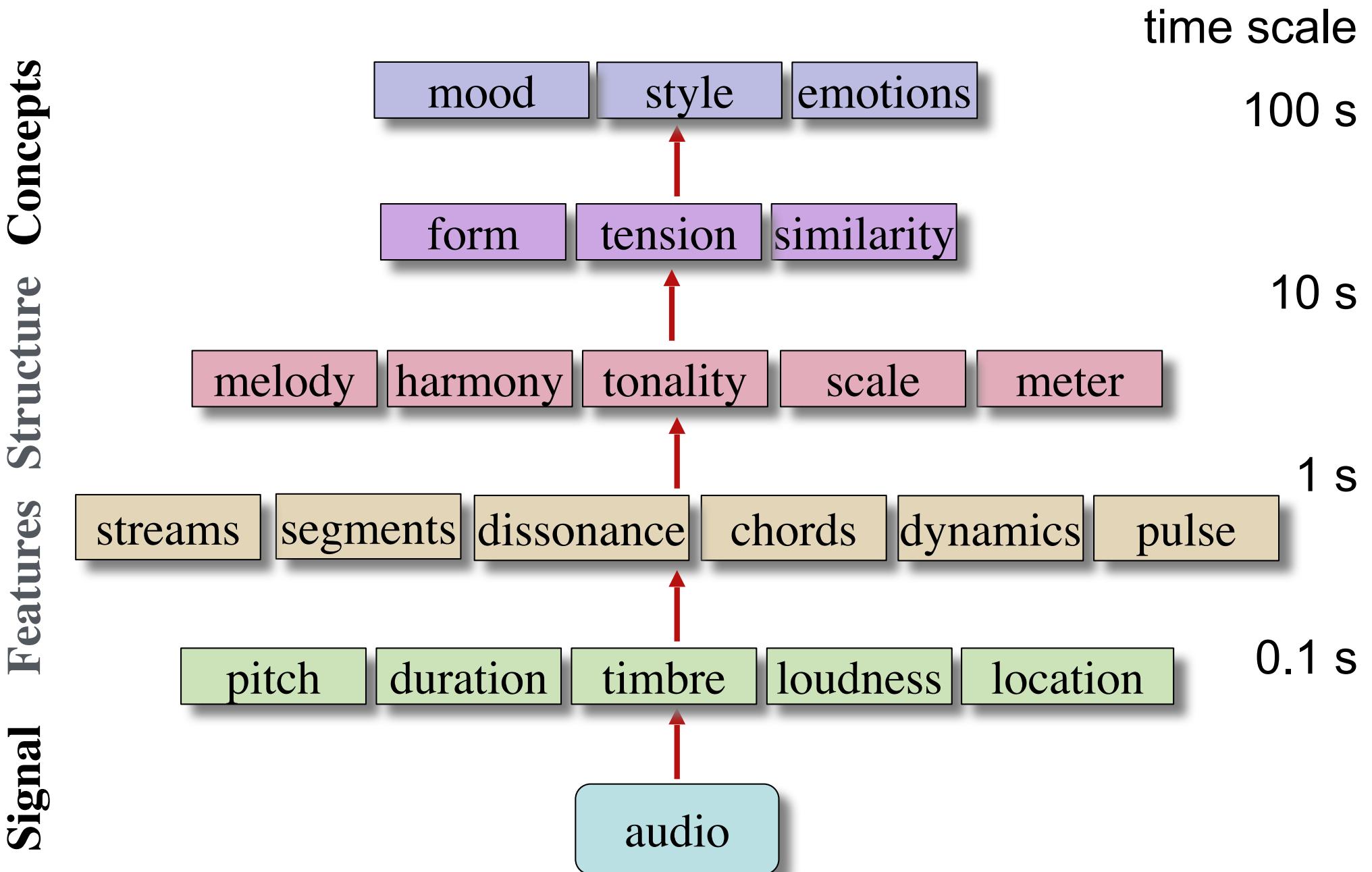


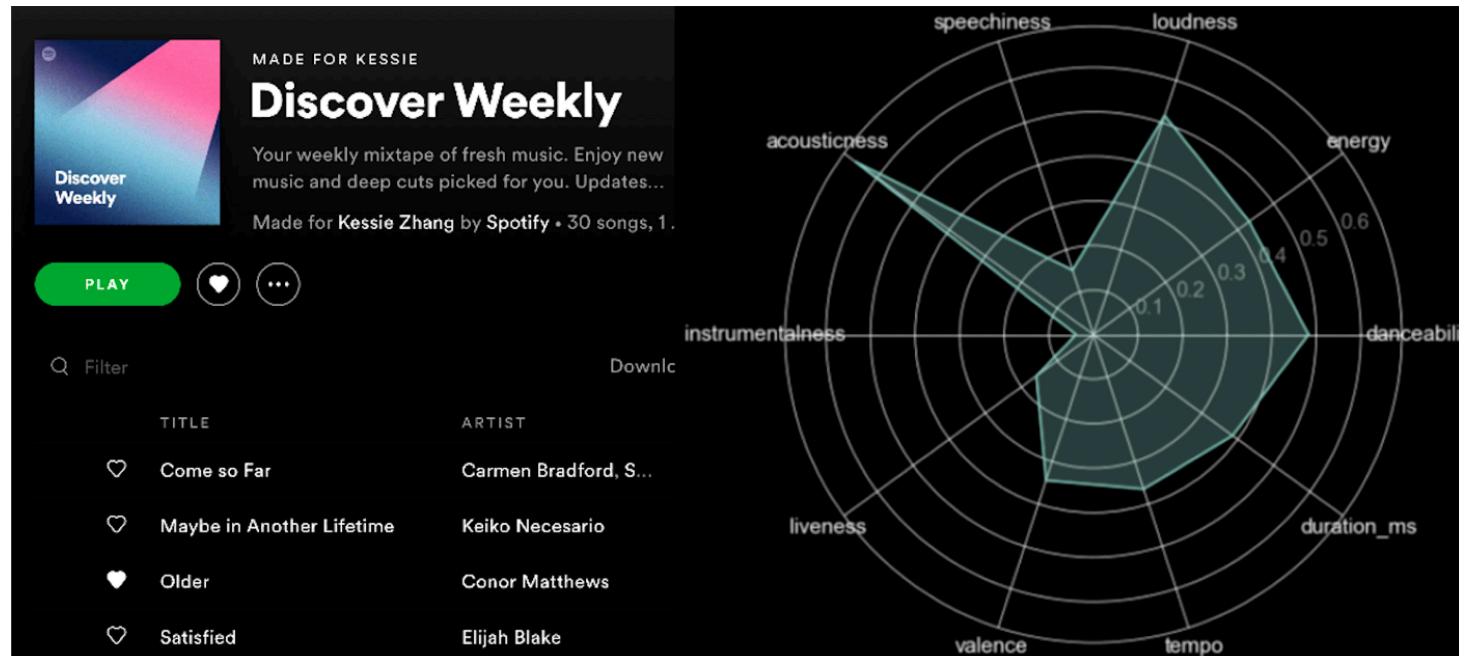
Why is window length
important (perceptually)?

Levels of Music Processing



Feature Extraction from music

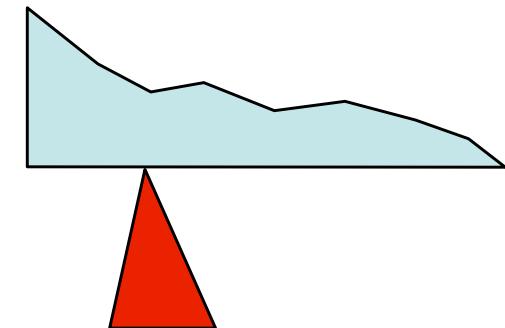
- features in music evolve continuously
- feature extraction relies on summarising this evolution (means, std)



Spectral Centroid

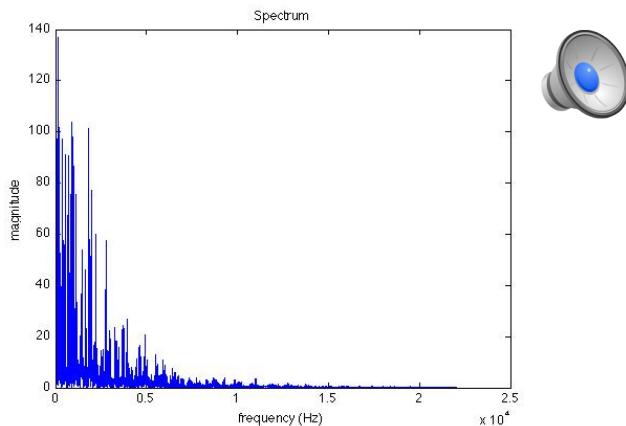
- Center of mass of the spectrum

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

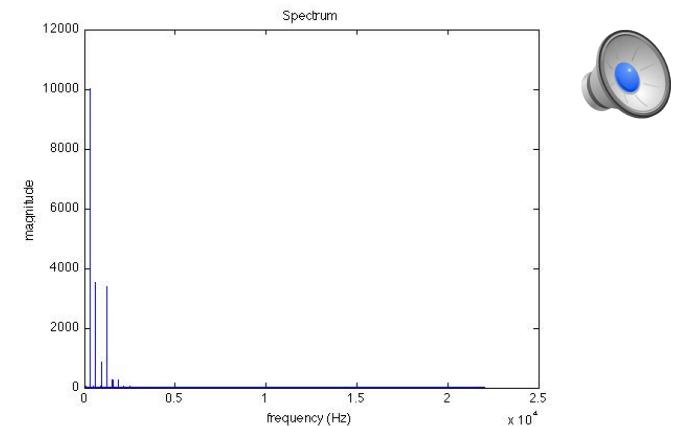


Which sounds brighter?

high:

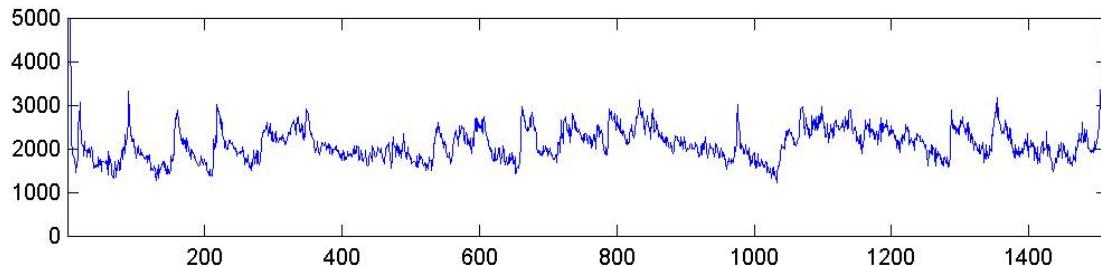


low:

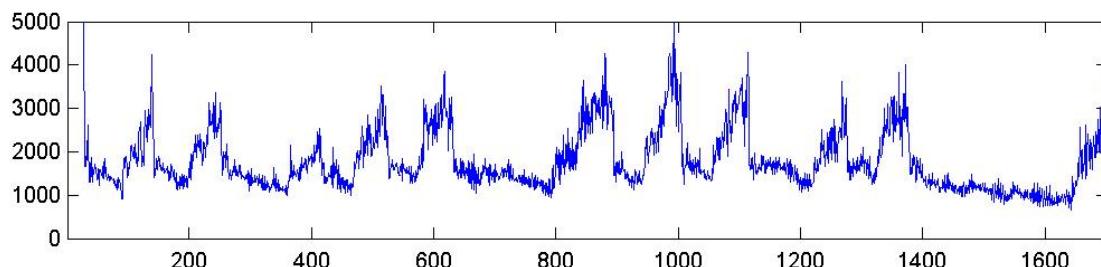


Which sounds brighter? (spectral centroid)

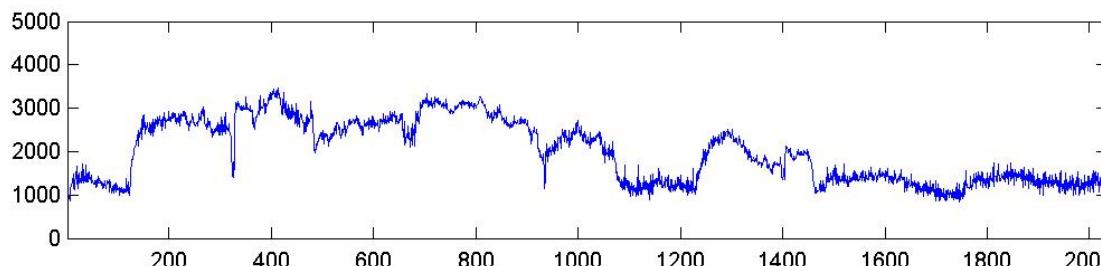
- spectral centroid of three excerpts



mean = 2167 Hz
std = 751 Hz



mean = 1953 Hz
std = 1534 Hz

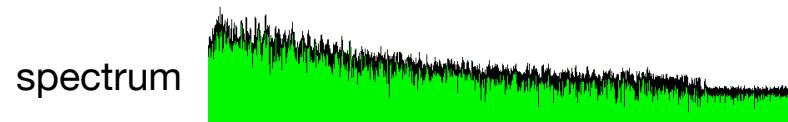


mean = 1993 Hz
std = 706 Hz

Musical features: Examples

Low-level / Timbral

brightness 



Mid/High-level / Rhythmic

pulse
clarity

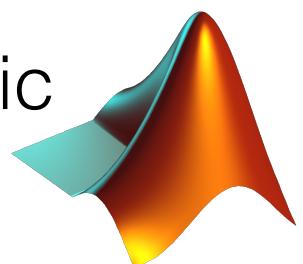
envelope
autocorrelation



MIR Toolbox



- Matlab toolbox for musical feature extraction from audio
- main developer Olivier Lartillot
 - currently @ RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion (University of Oslo)
- started within project Tuning the brain for music (EU FP7, 2006-9)



Features Overview

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

Acoustic features

Temporal

- zero-crossing rate
- low energy

Spectrotemporal

- spectral flux
- sub-band flux

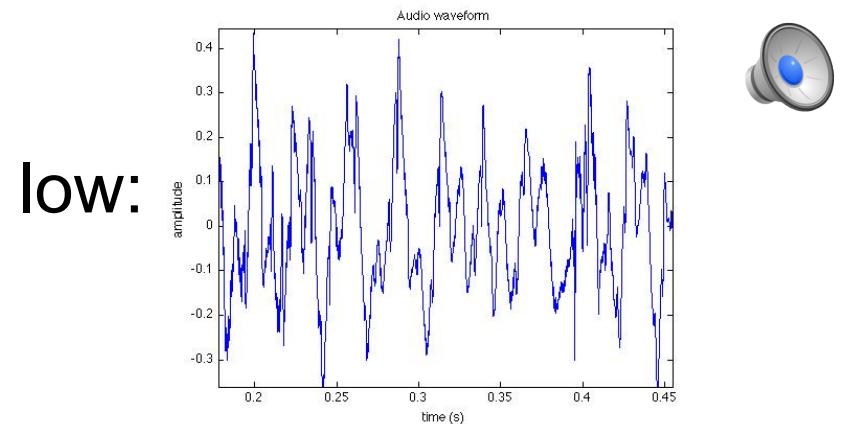
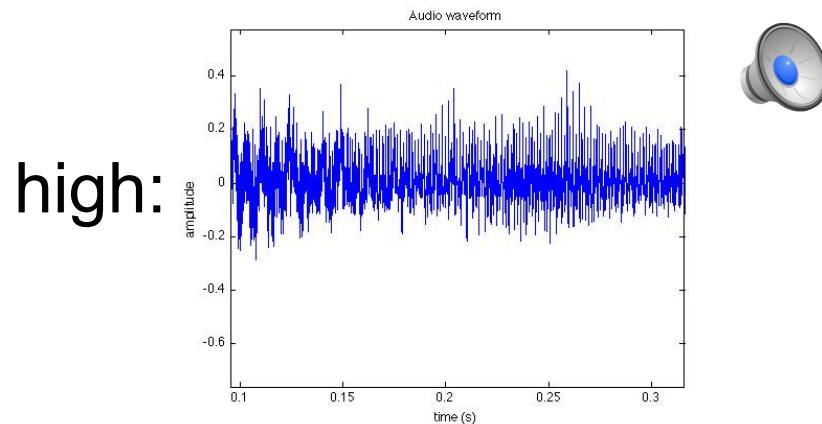
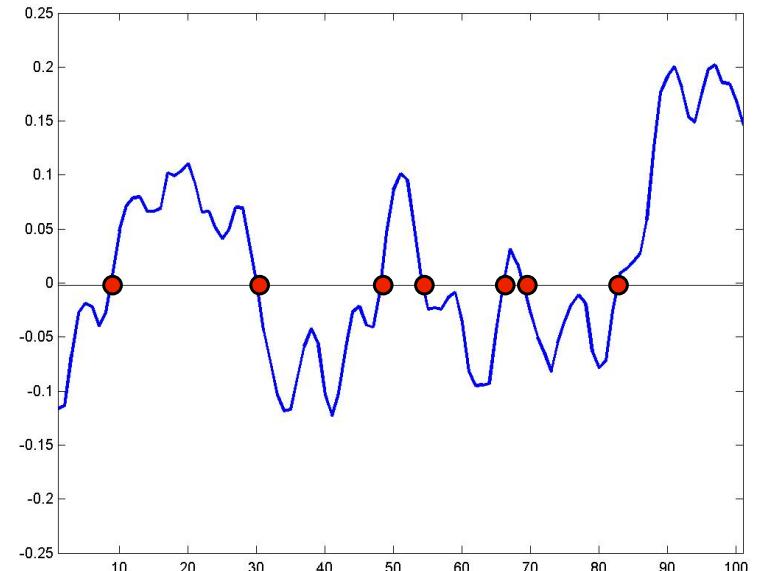
Spectral

- 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

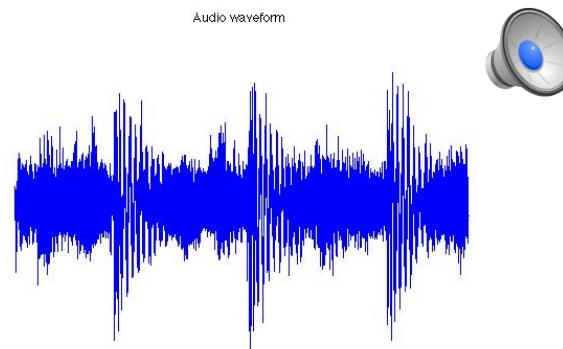
- number of time-domain zero-crossings of the signal per time unit



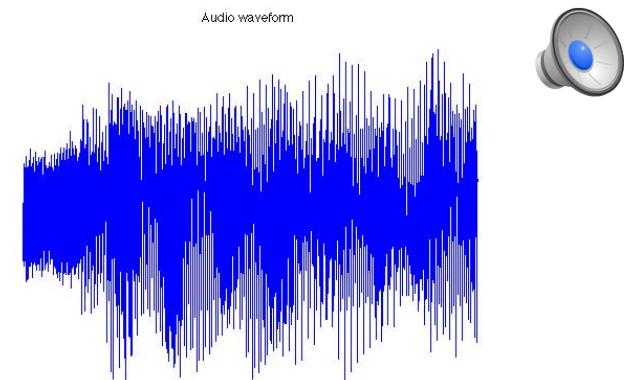
Low Energy

- proportion of signal frames whose energy is below average energy

high:



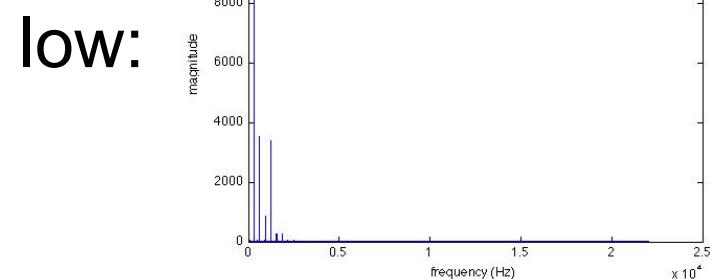
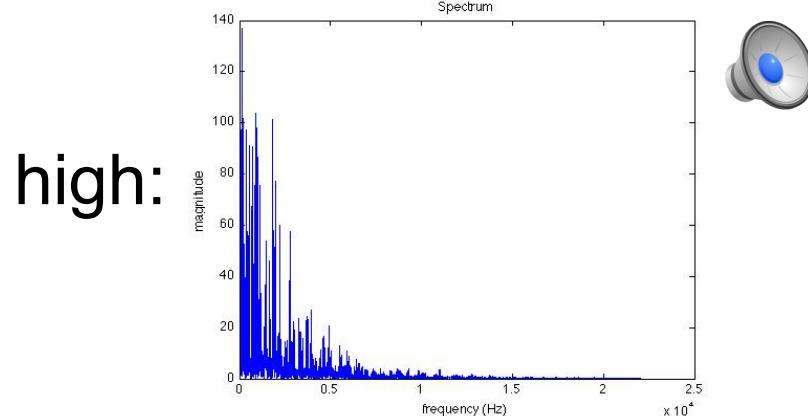
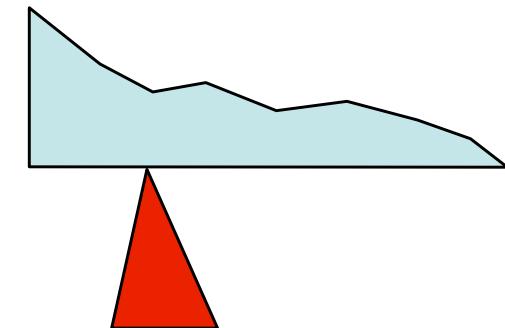
low:



Spectral Centroid

- Center of mass of the spectrum

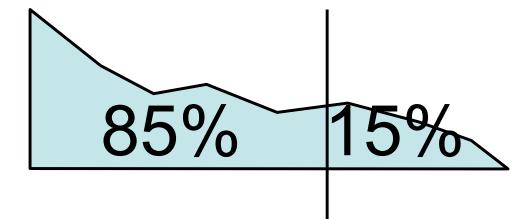
$$sc = \frac{\sum a_i f_i}{\sum a_i}$$



Spectral Roll-Off

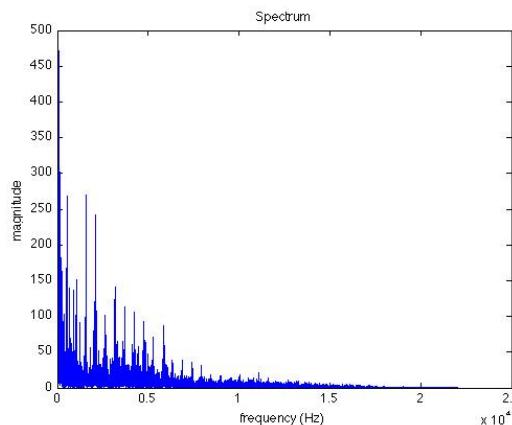
- Frequency, below which a certain fraction (usually 85%) of spectral energy
- R such that

$$\sum_1^R a_i = 0.85 \sum_1^N a_i$$

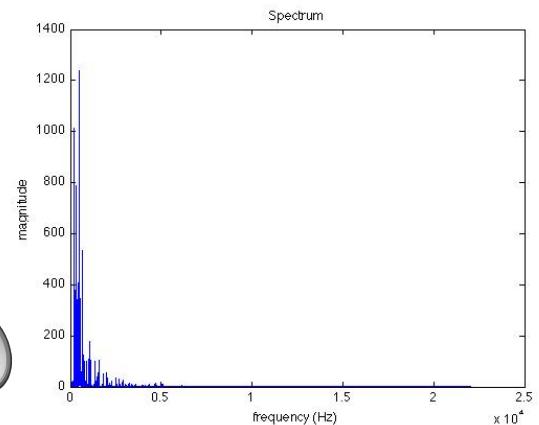


- Measure of spectral shape

high:



low:

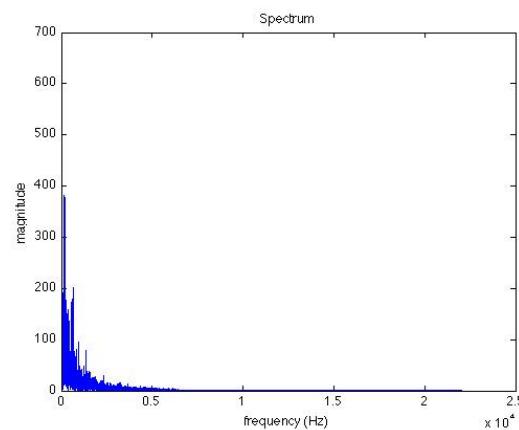


Spectral Irregularity

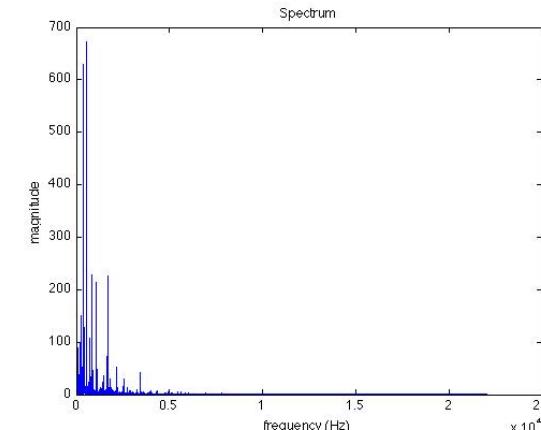
- measure of "jaggedness" of spectrum (Jensen, 1999)

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

high:



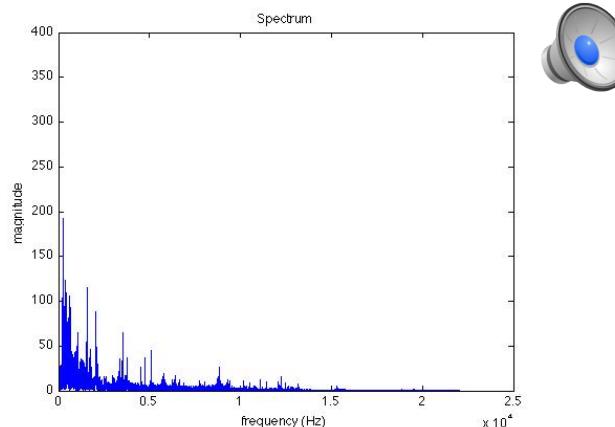
low:



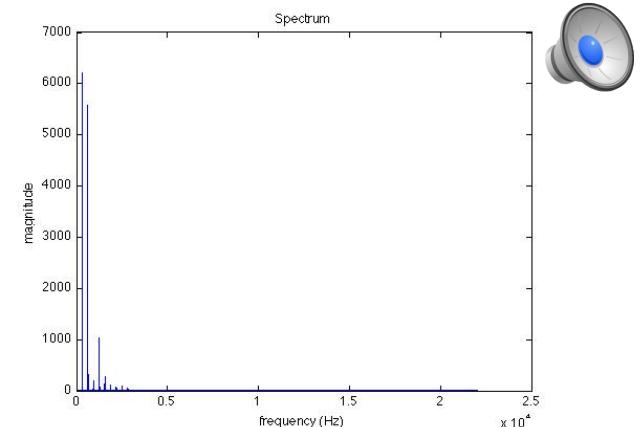
Spectral Entropy

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

high:



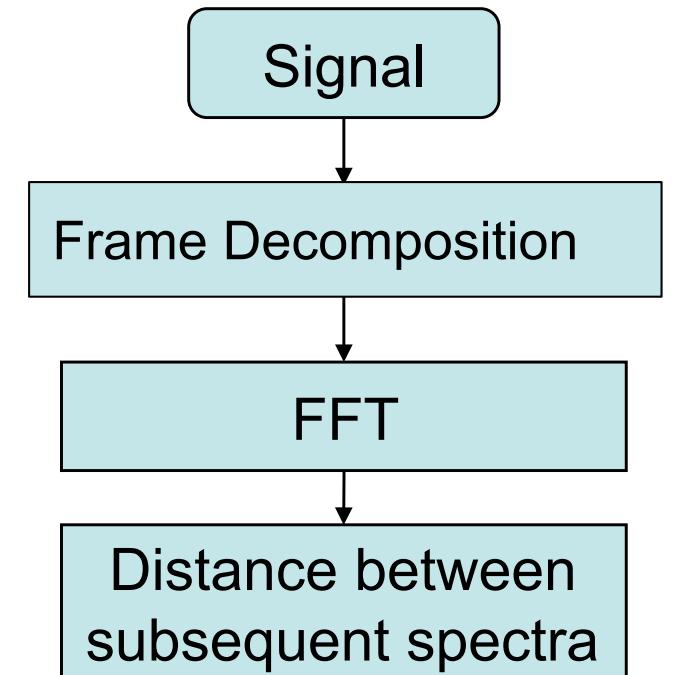
low:



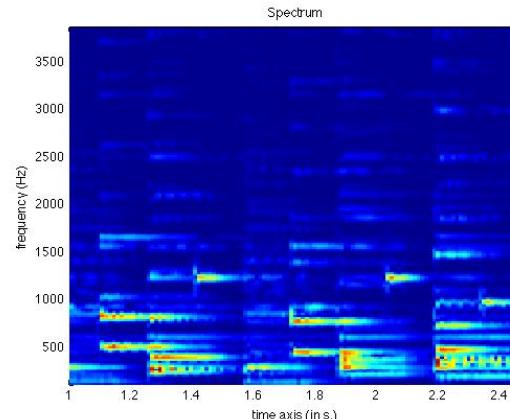
Spectral Flux

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

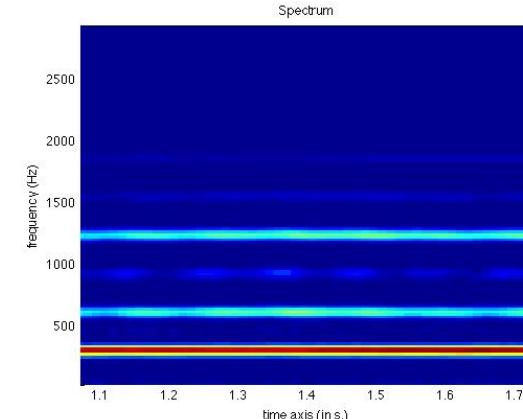
$$flux_i = \sum_{j=1}^M (a_{ij} - a_{(i-1)j})^2$$



high:

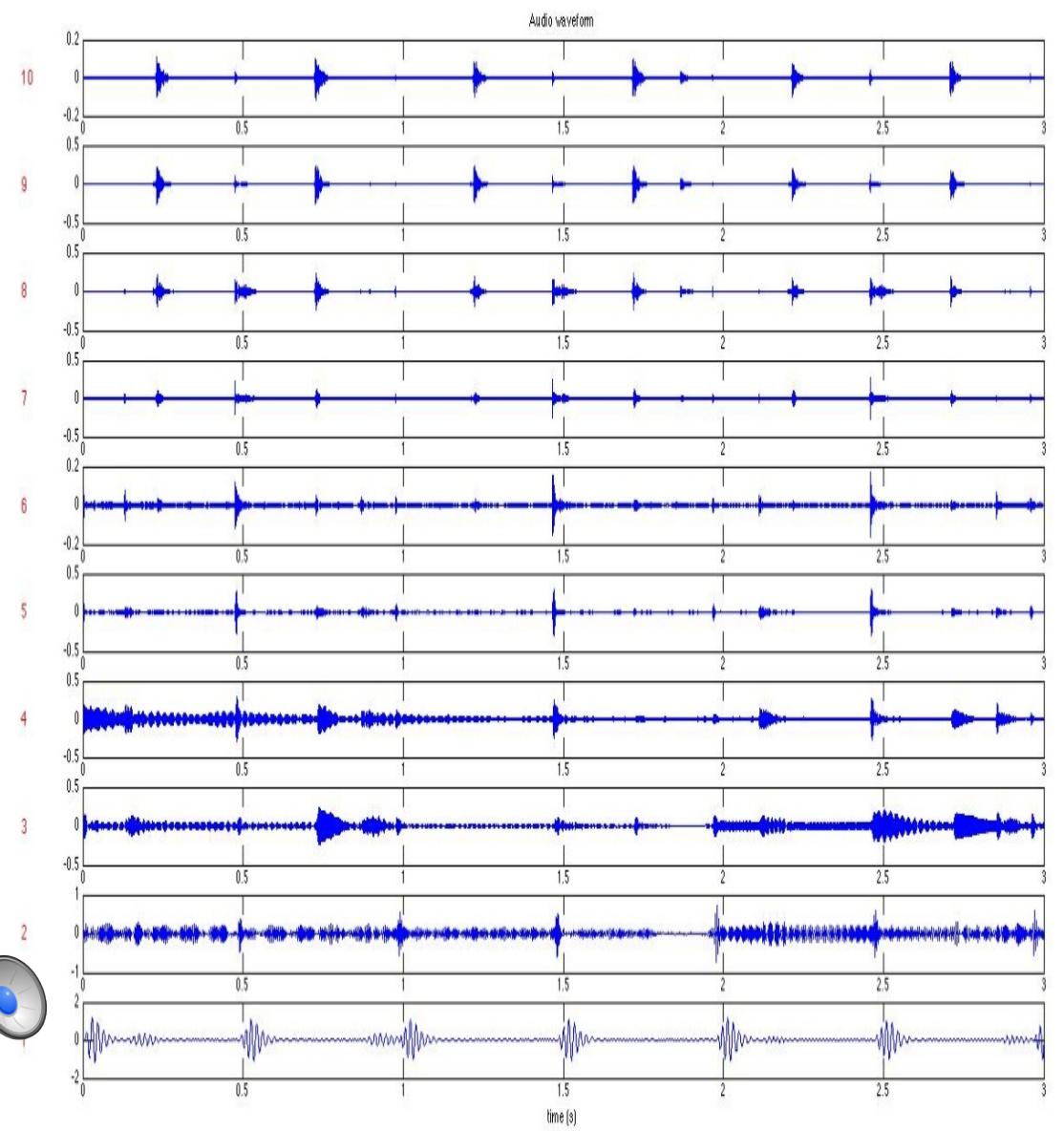


low:



Sub-band Flux

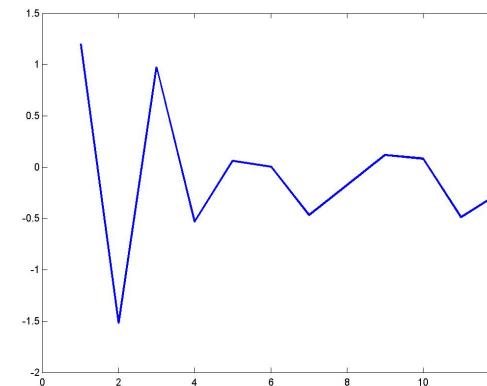
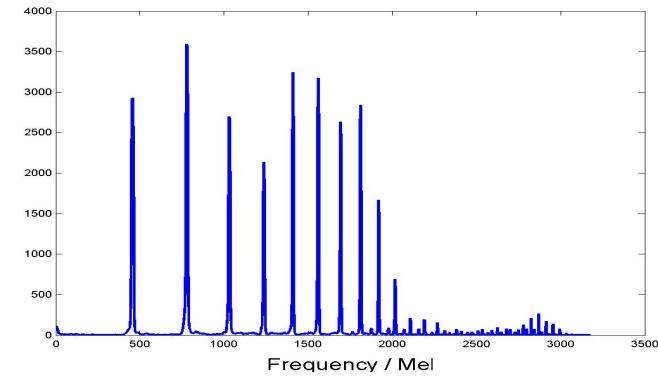
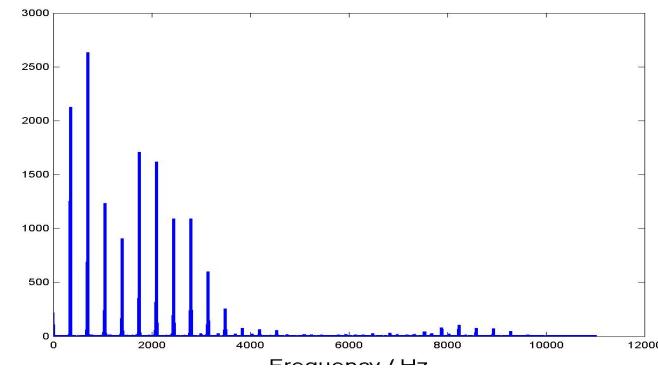
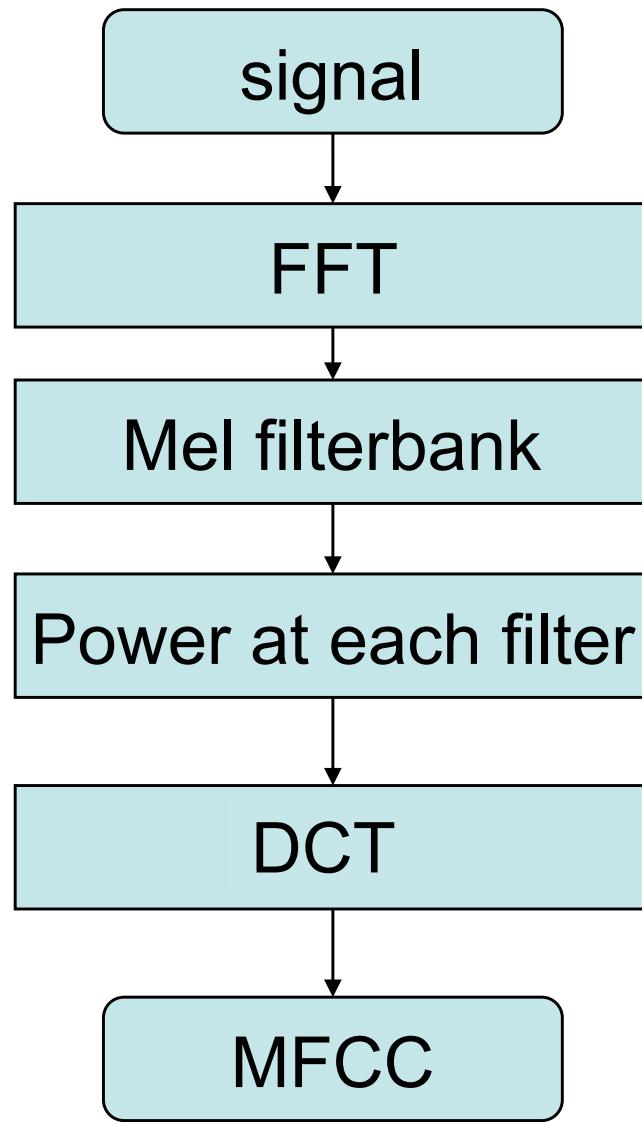
- Octave-scaled spectrum
 - 50 hz
 - 10 bands
- Spectral Flux in each band



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

Acoustic features

Rhythm

- tempo
- pulse clarity

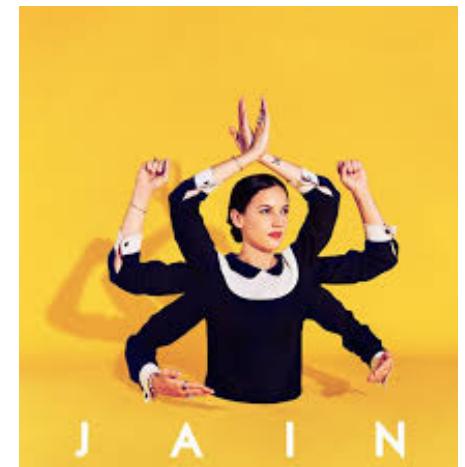
Tonality

- chromagram
- mode
- keystrength/keyclarity

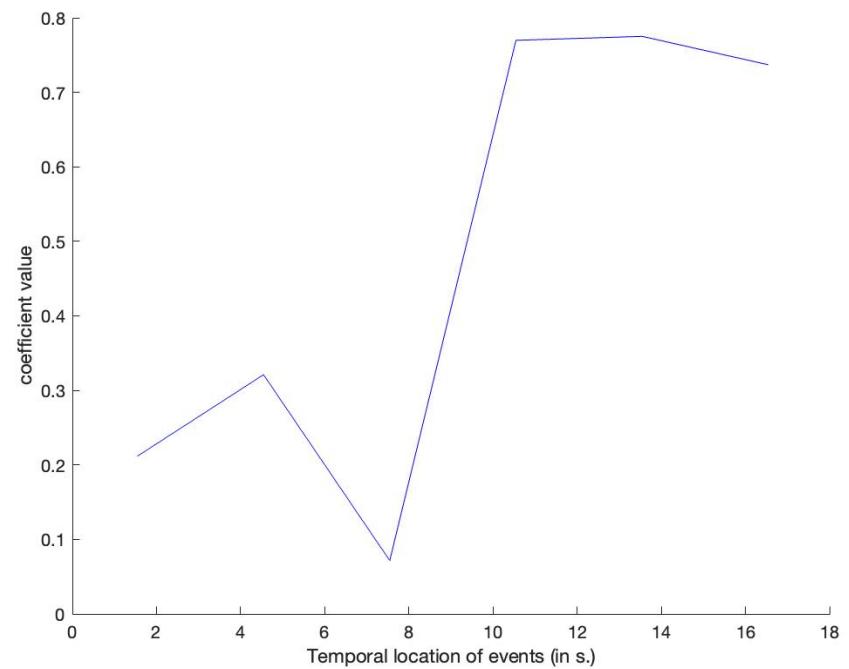
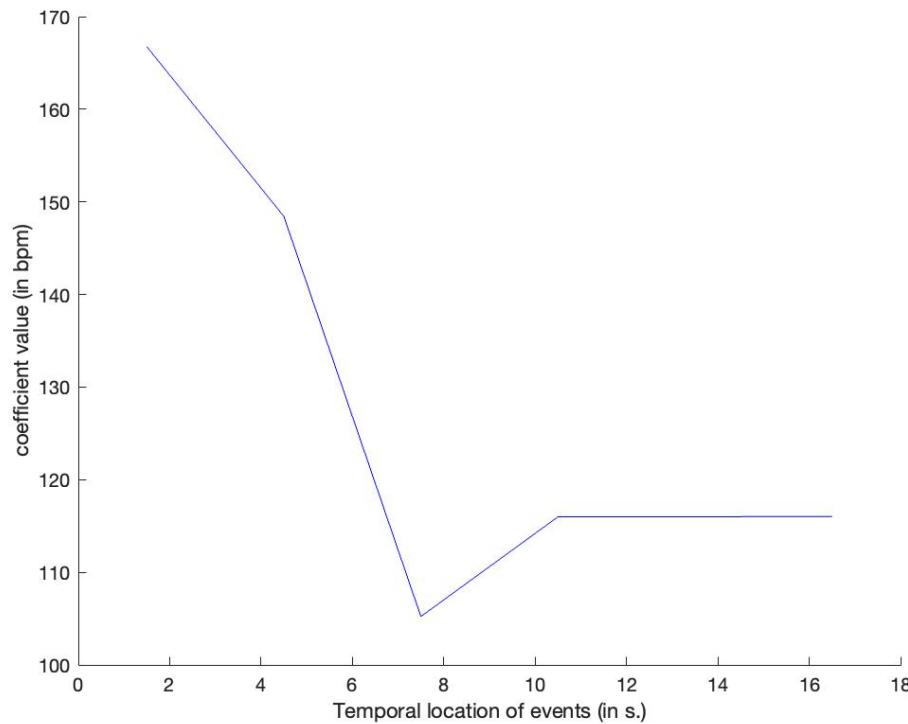
*typically extracted using longer time windows (context-dependent)

Tempo & Pulse Clarity

- tempo: estimate of how fast/slow the piece of music is
- pulse clarity/beat salience: how clear the beat is

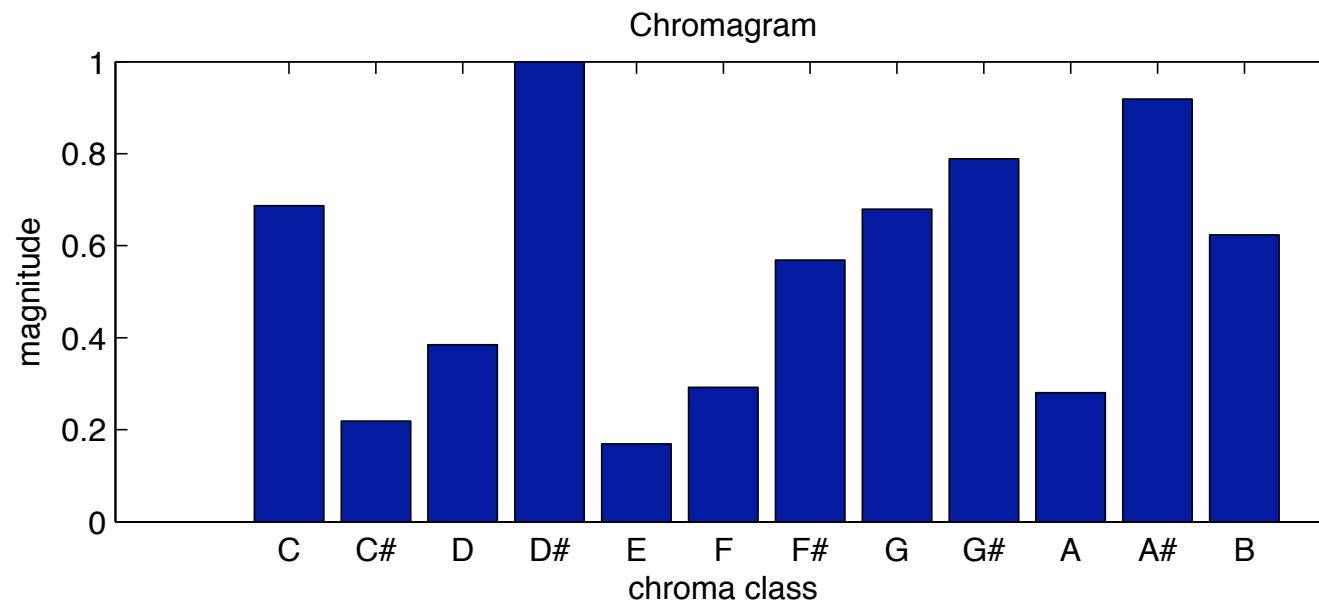


Tempo & Pulse Clarity



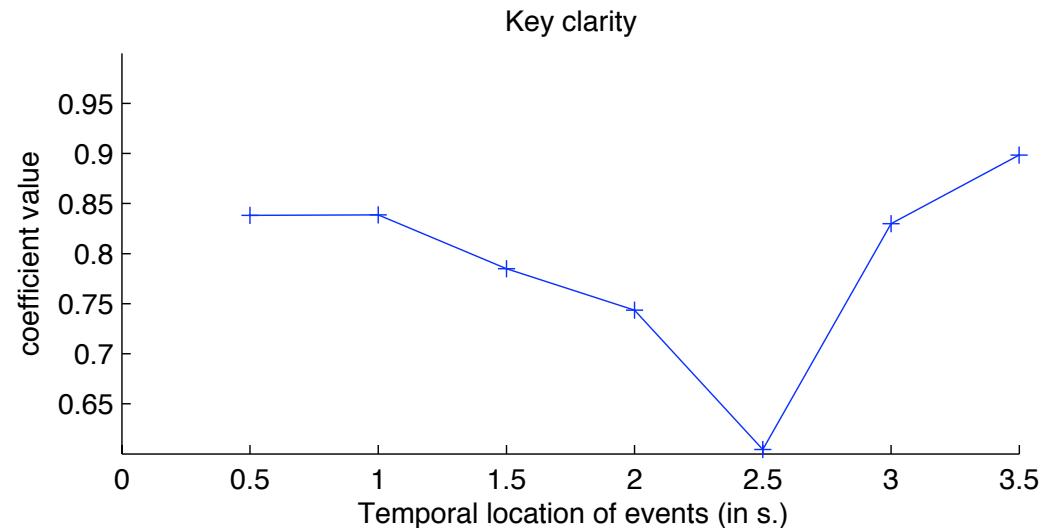
Acoustic features

- 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”)



MIR Toolbox

MIR Toolbox



- Matlab toolbox for musical feature extraction from audio
- main developer Olivier Lartillot; Petri Toivainen
- started within project Tuning the brain for music (EU FP7, 2006-9)

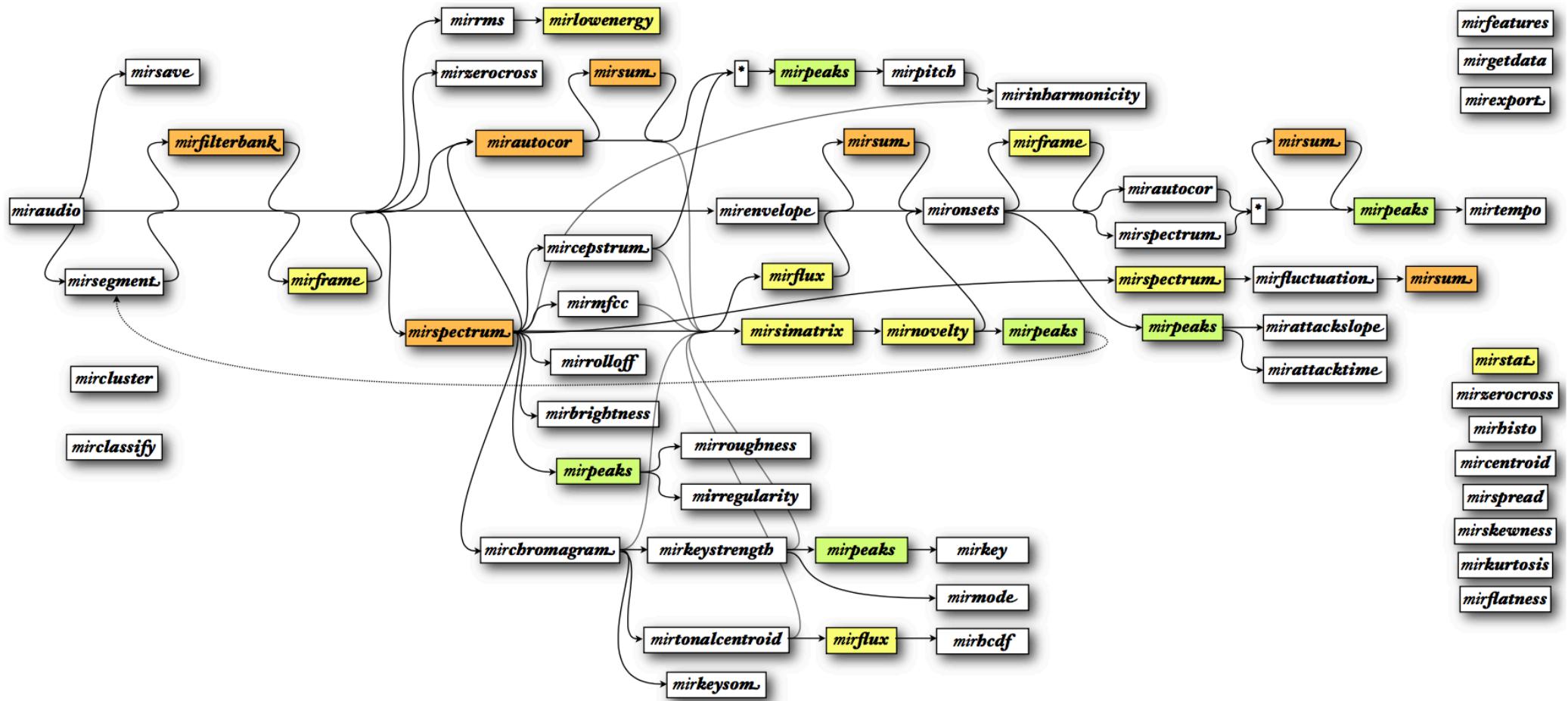
Overview

1. General Principles
2. Syntax & Basic Operations
3. Signal Processing Operators
4. Audio and Musical Descriptors
5. Statistical Analysis

General Principles

- **Modular framework**
 - Building blocks can be reused, reordered, ...
- **Simple and adaptive syntax**
 - User can focus on the general design.
 - MIRtoolbox takes care of the technical details.
- **Free software, open source**
 - Capitalized expertise of the research community, for everybody

General Principles



MIRtoolbox Reliances

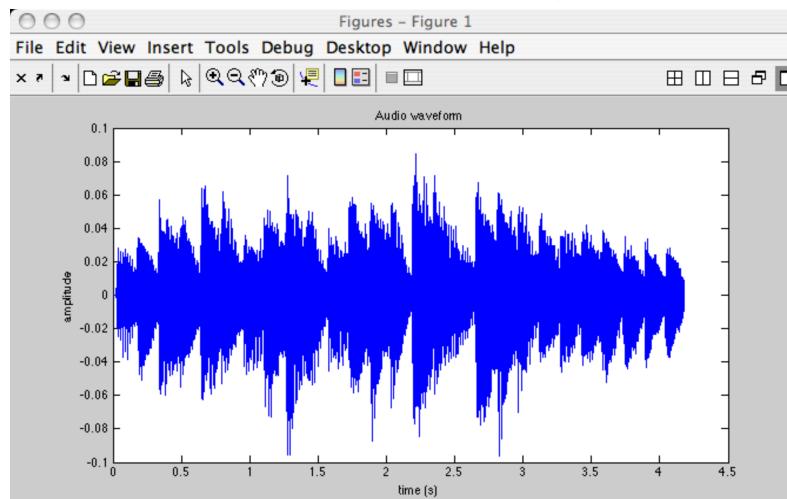
- Requires:
 - Matlab, ≥ version 7
 - Signal Processing Toolbox (MathWorks)
- Distribution includes free softwares:
 - Auditory Toolbox, by Malcolm Slaney
 - Netlab toolbox, by Ian Nabney
 - SOM Toolbox, by Esa Alhoniemi et al.
- Code integrated from GPL project:
 - Music Analysis Toolbox by Elias Pampalk

Basic Operations

- *miraudio('mysong.wav')*



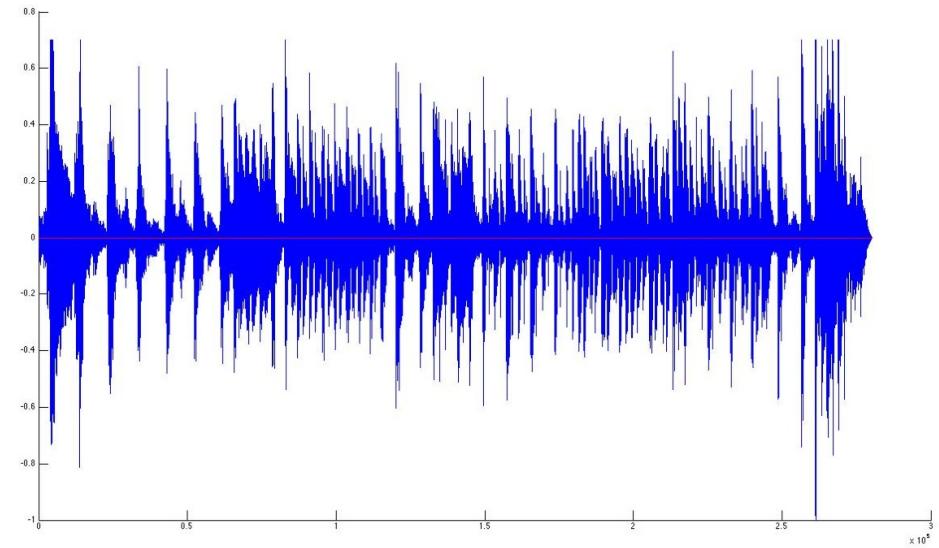
miraudio('mysong.mp3')



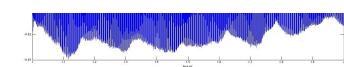
- *miraudio('Folder')*
 - 'Folder' = all files in Current Directory

Basic Operations

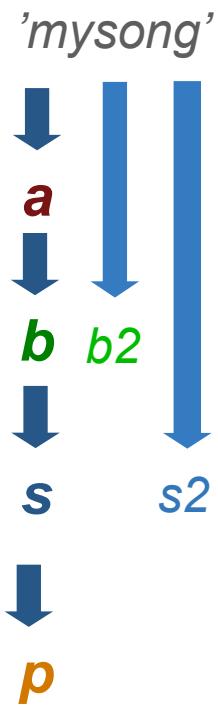
- *miraudio(... , ‘Center’)*
(... , ‘Sampling’,)
(... , ‘Trim’)
(... , ‘TrimStart’)
(... , ‘TrimEnd’)
(... , ‘Extract’, t1,



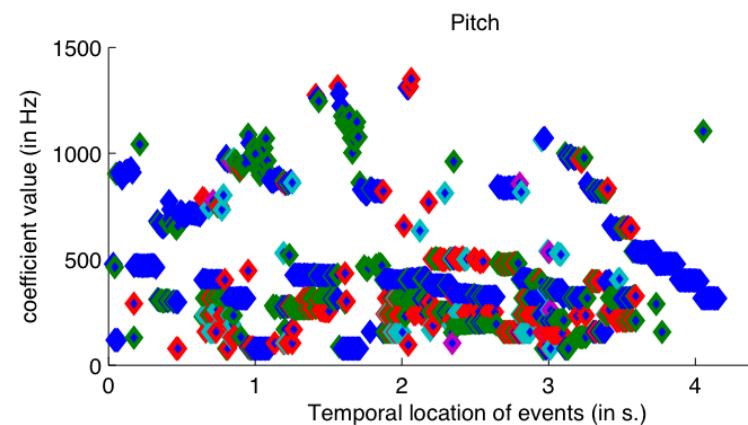
- *Audio Summation*
 - $a1 = \text{miraudio}(\text{‘melody.wav’})$
 - $a2 = \text{miraudio}(\text{‘accompaniment.wav’})$
 - $a = a1 + a2$



Flow of operators



- **a** = *miraudio*('mysong')
- **b** = *miraudio*(**a**, 'Extract', 0, 4)
 - **b2** = *miraudio*('mysong', 'Extract', 0, 4)
 - *mirplay*(**b**)
 - *mirsave*(**b**, 'excerpt.wav')
- **s** = *mirspectrum*(**b**)
 - **s2** = *mirspectrum*('mysong')
- **p** = *mirpitch*(**s**, 'Frame')
 - *mirgetdata*(**p**)



mirspectrum

Fourier Transform

- *mirspectrum*(...)

, ‘**Min**’, 0 (in Hz)

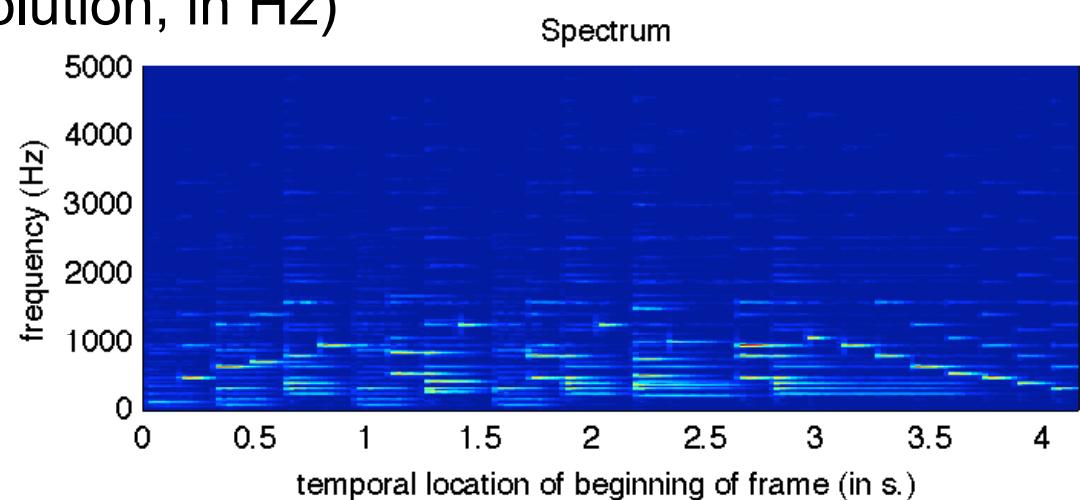
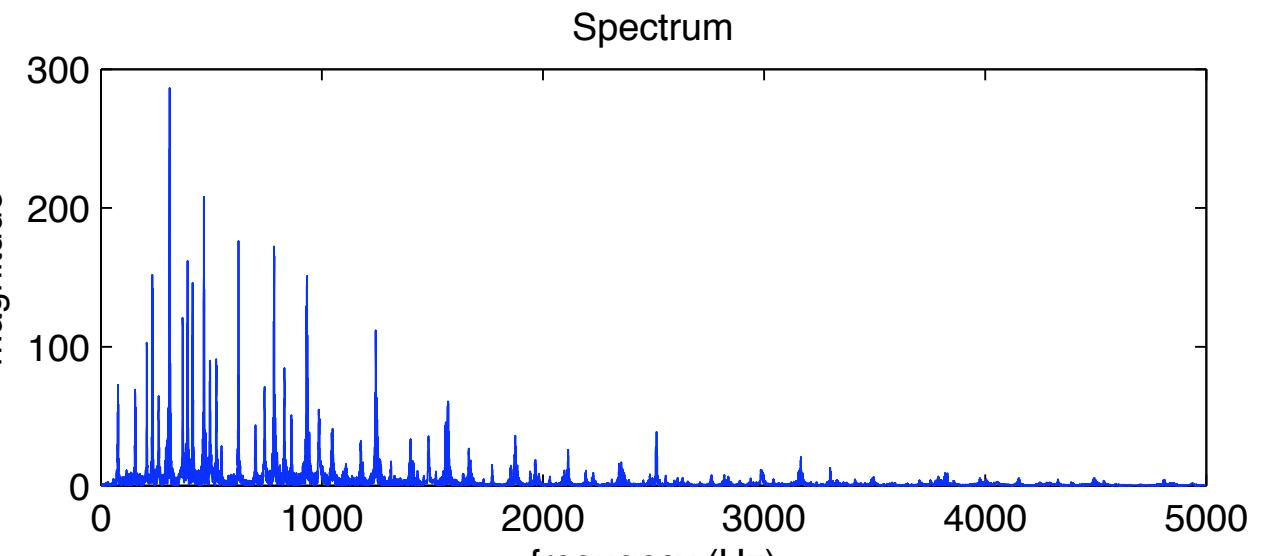
, ‘**Max**’, 5000 (in Hz)

, ‘**Window**’, ‘hamming’

, ‘**MinRes**’, *r*

(minimum frequency resolution, in Hz)

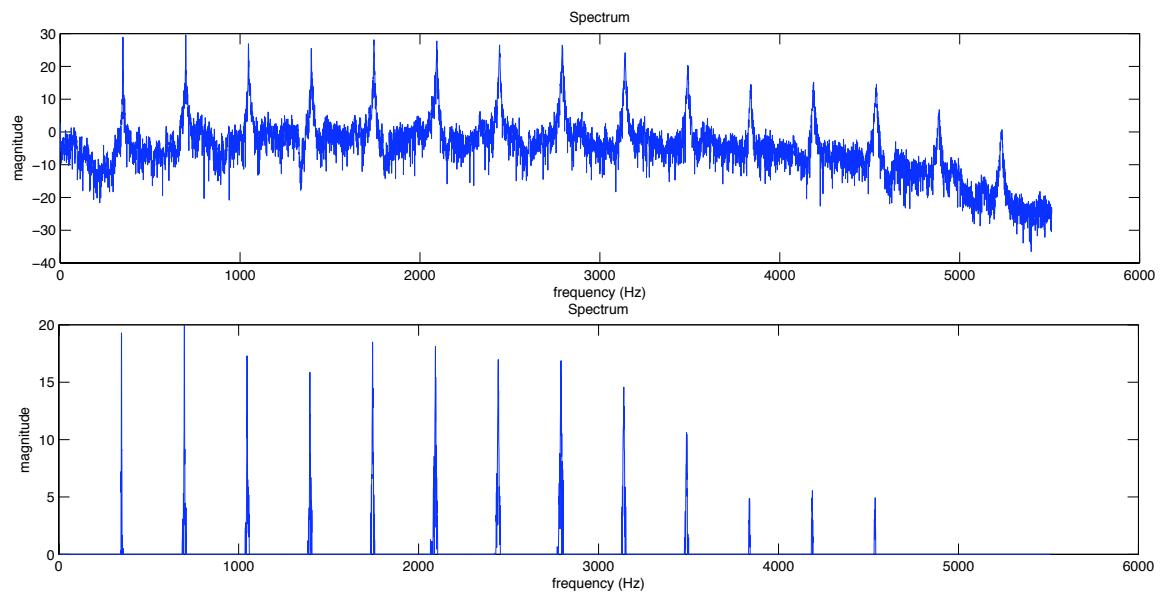
, ‘**Frame**’, *s*)



mirspectrum

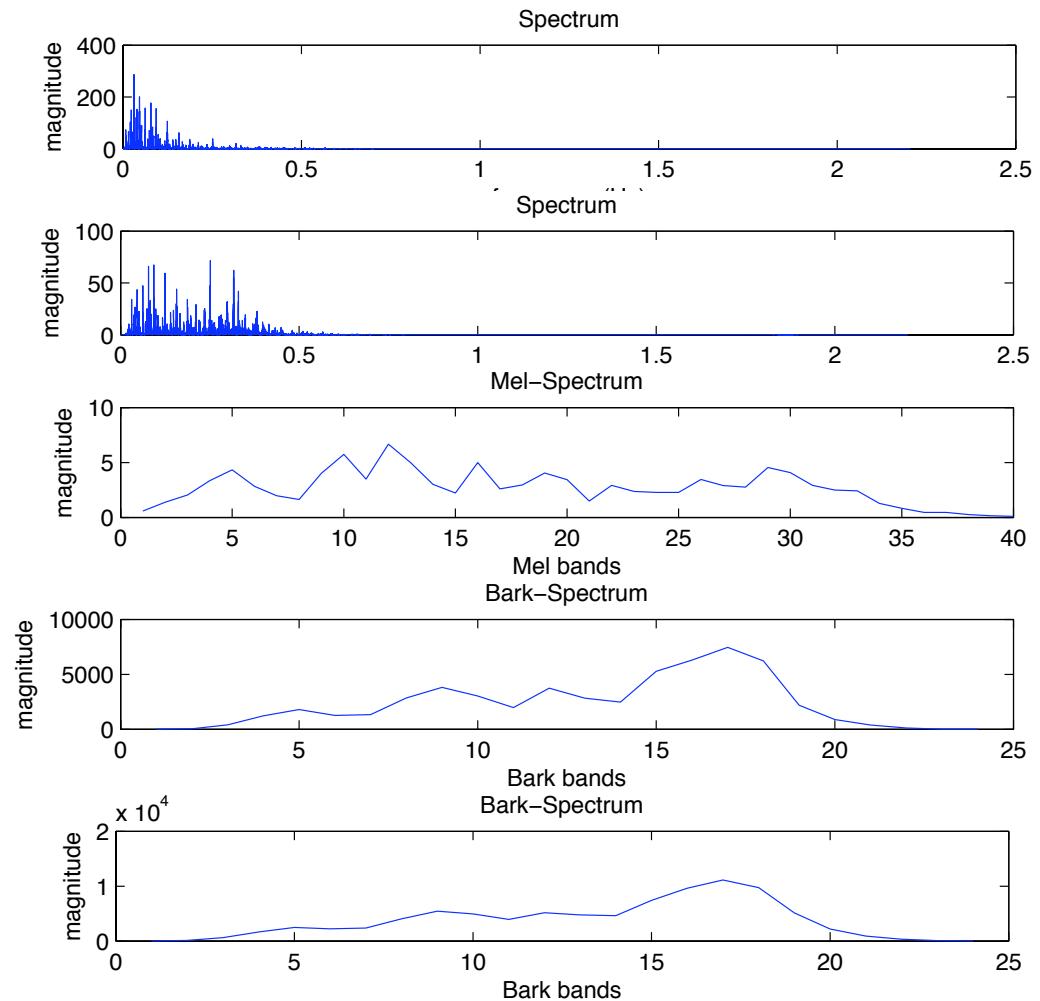
post-processing options

- *mirspectrum(..., 'Normal')*
normalizes w.r.t. energy.
- *mirspectrum(..., 'Power')*
squares the energy.
- *mirspectrum(..., 'dB')*
in dB scale
- *mirspectrum(..., 'dB', th)*
only highest energy, range
of *th* dB



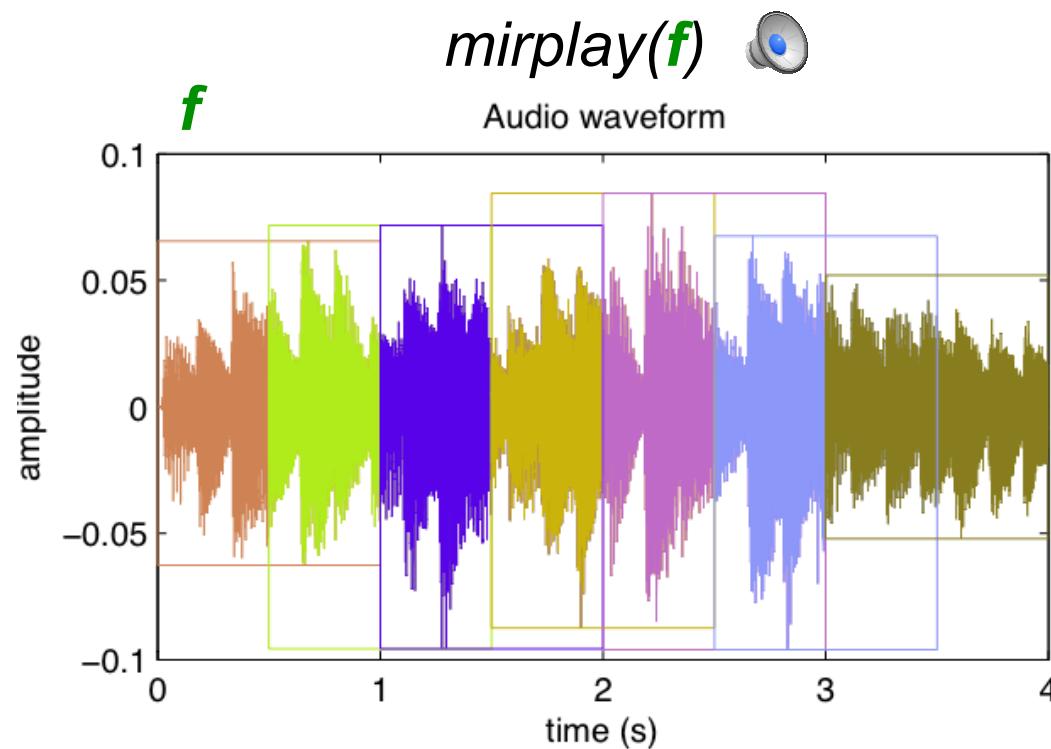
mirspectrum auditory models

- *mirspectrum(..., 'Terhardt')*:
Outer ear modeling
- *mirspectrum(..., 'Mel')*:
Mel-band spectrum
- *mirspectrum(..., 'Bark')*:
Bark-band spectrum
- *mirspectrum(..., 'Mask')*:
Masking effects along bands



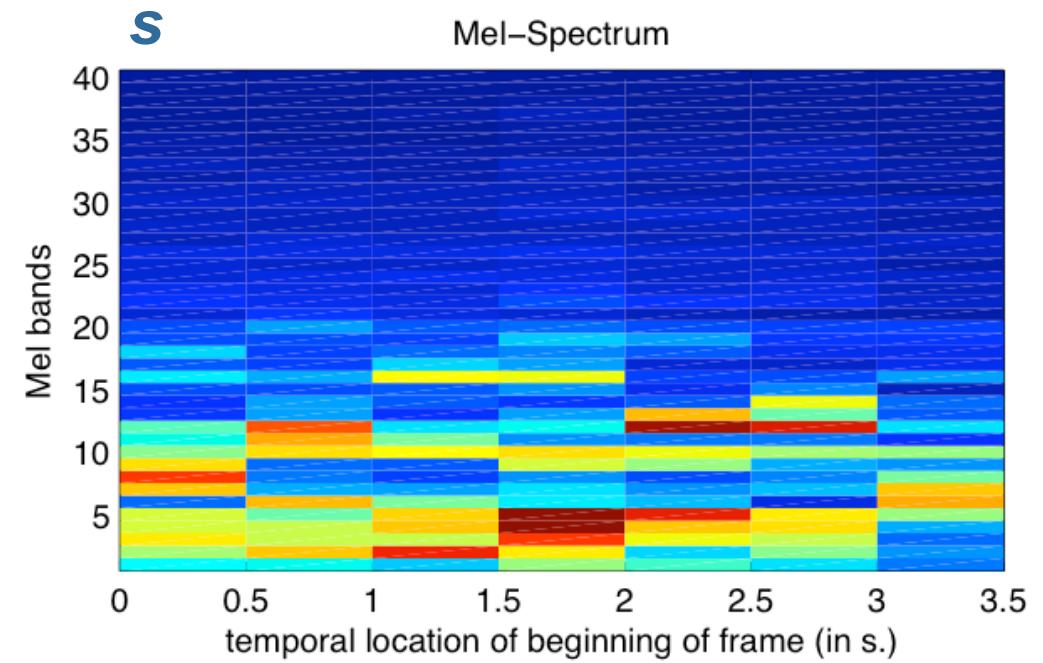
Frame decomposition

$f = \text{mirframe}('mysong', \dots$
'Length', .1, 's', **'Hop'**, 20, '%')



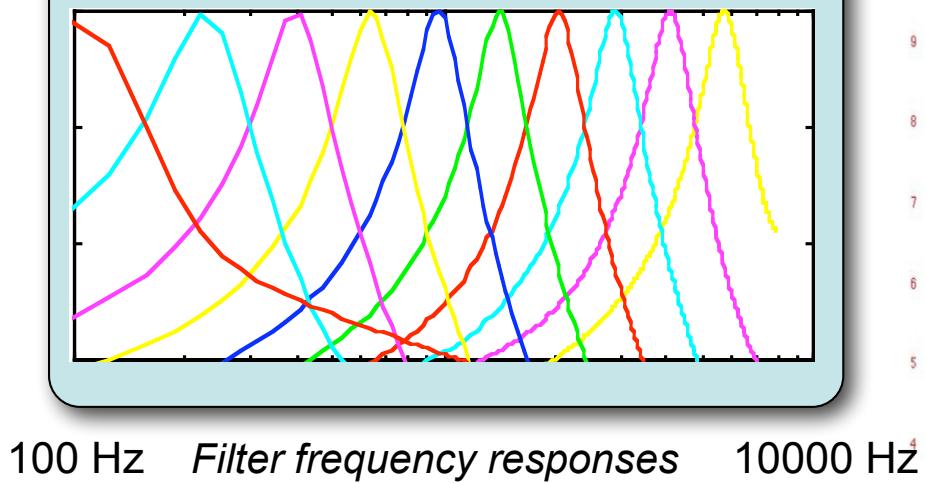
$s = \text{mirspectrum}(f, 'Mel')$

$s = \text{mirspectrum}('mysong', \dots$
'Frame', .1, .2, 'Mel')



Filterbank decomposition

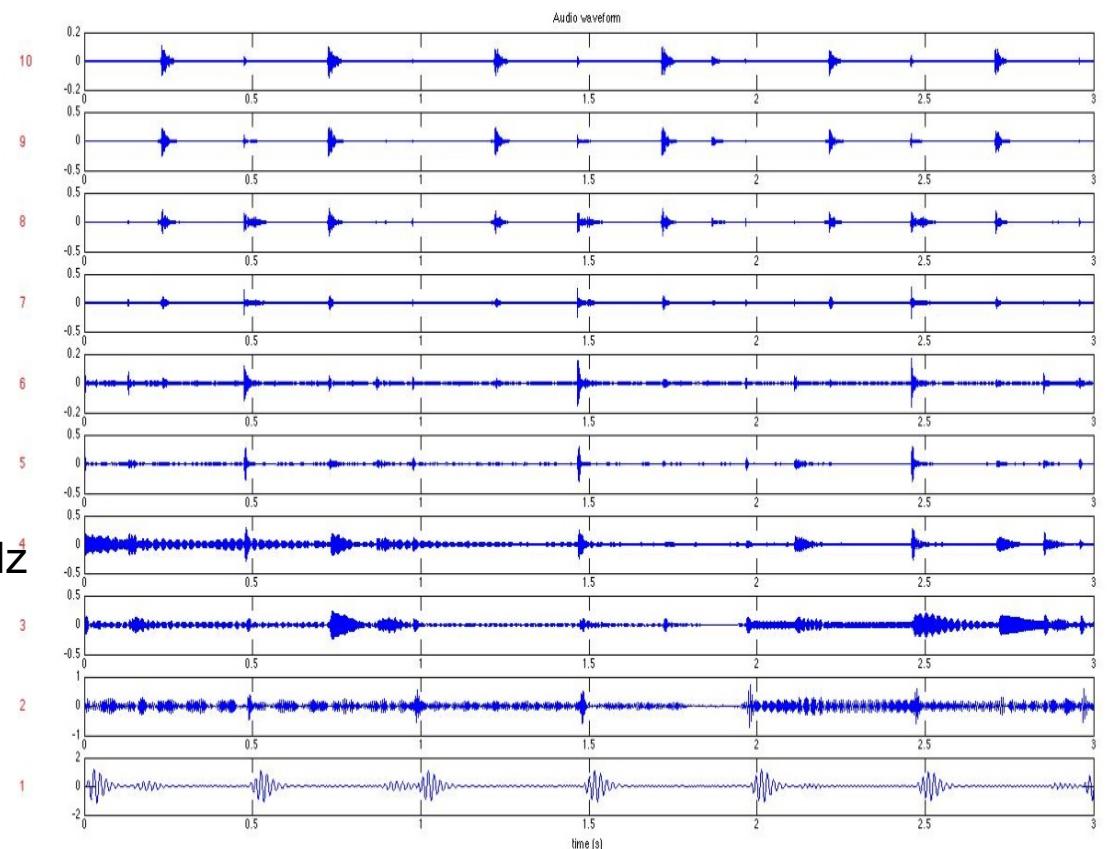
using Gammatone filterbank



100 Hz *Filter frequency responses* 10000 Hz

f = *mirfilterbank*('mysong', ...
 'NbChannels', 10)

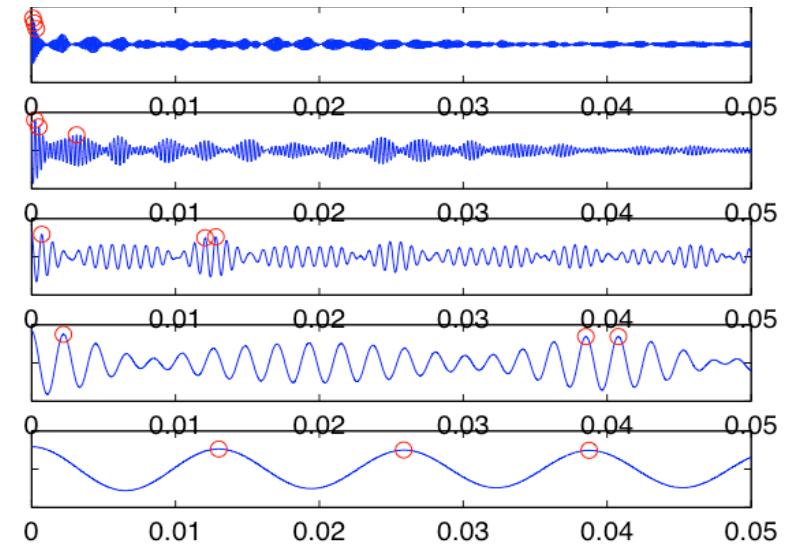
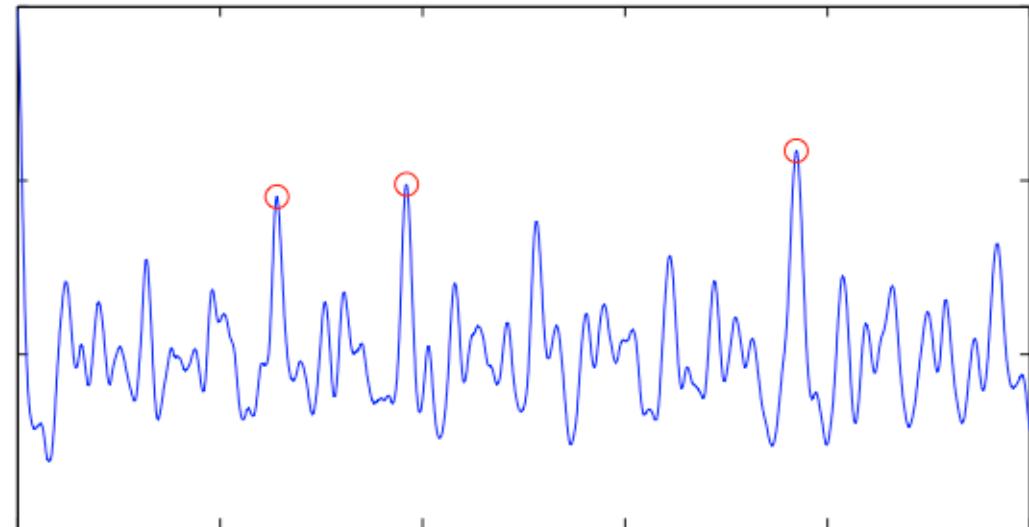
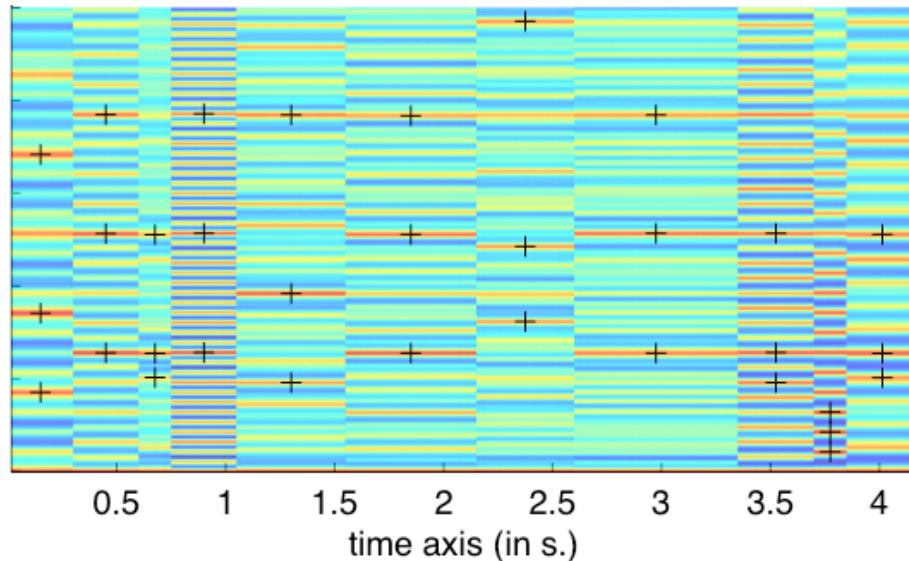
mirplay(**f**)



Peak picking (for different kinds of data)

*mirpeaks(...,
'Total', 3, 'NoBegin')*

- Adapts to various input types:
 - Multi-frame, multi-segment:





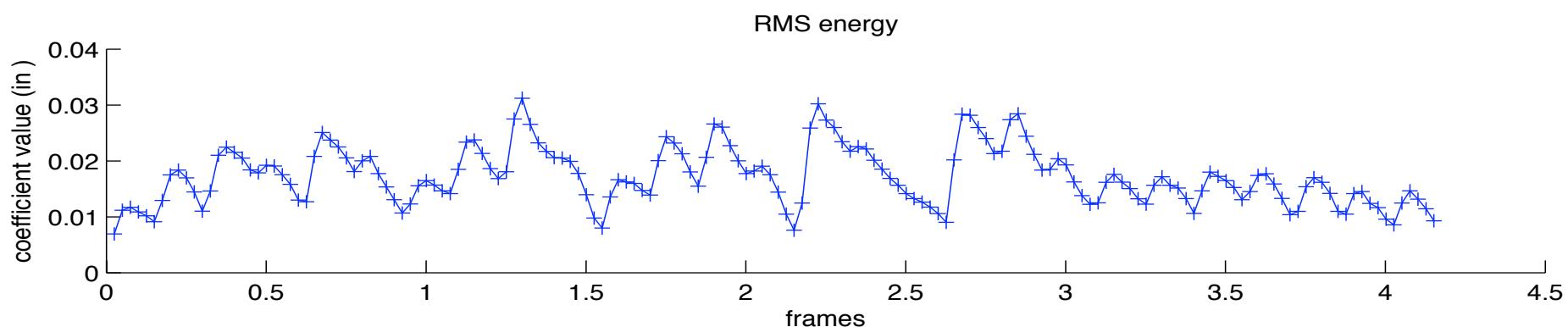
MIR Toolbox

Features Overview

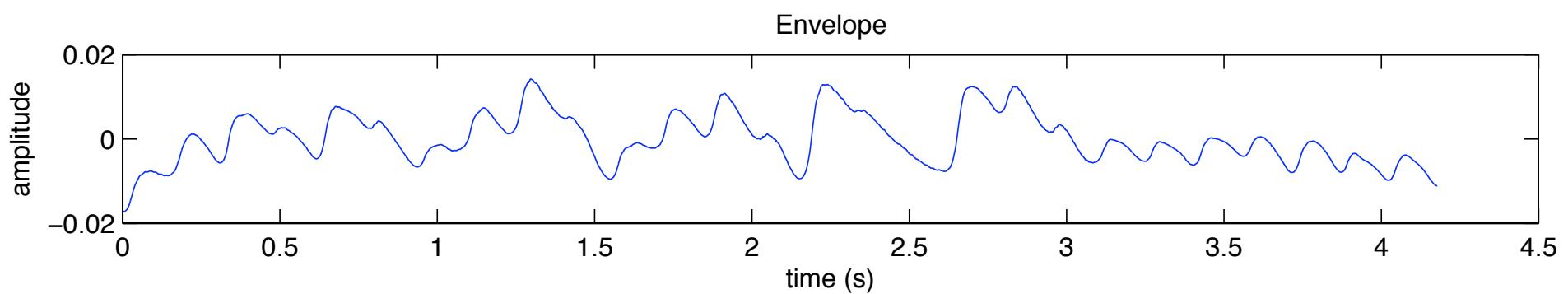
- Dynamics
- Tempo /rhythm
- Pitch
- Timbre
- Tonality
- Structure

(DYNAMICS)Energy / Envelope

- `mirrms('mysong', 'Frame')`

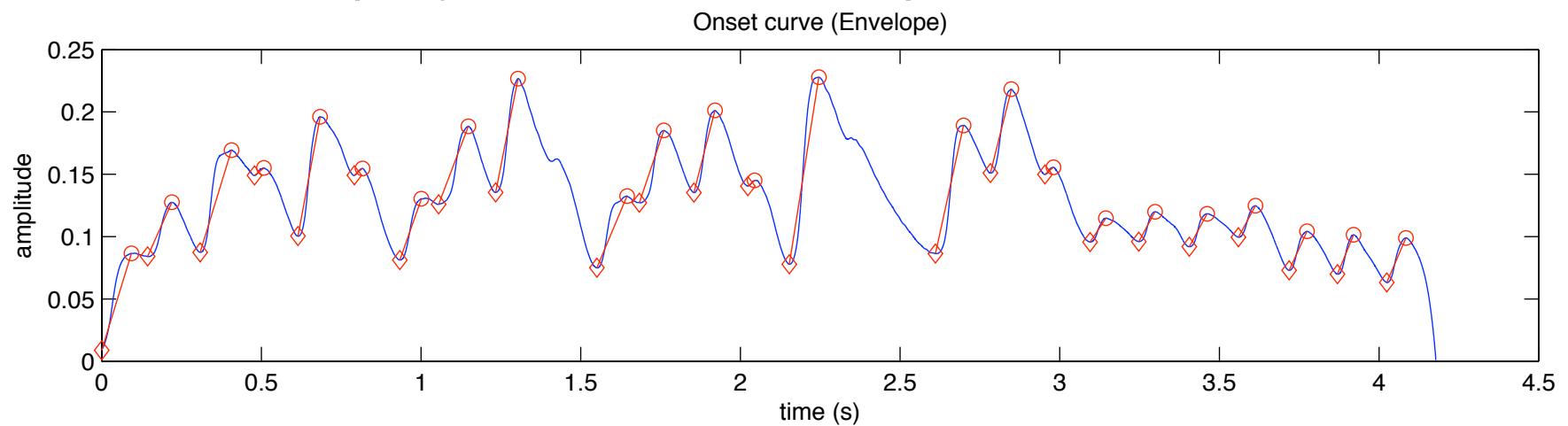


- `mirenvelope('mysong')`

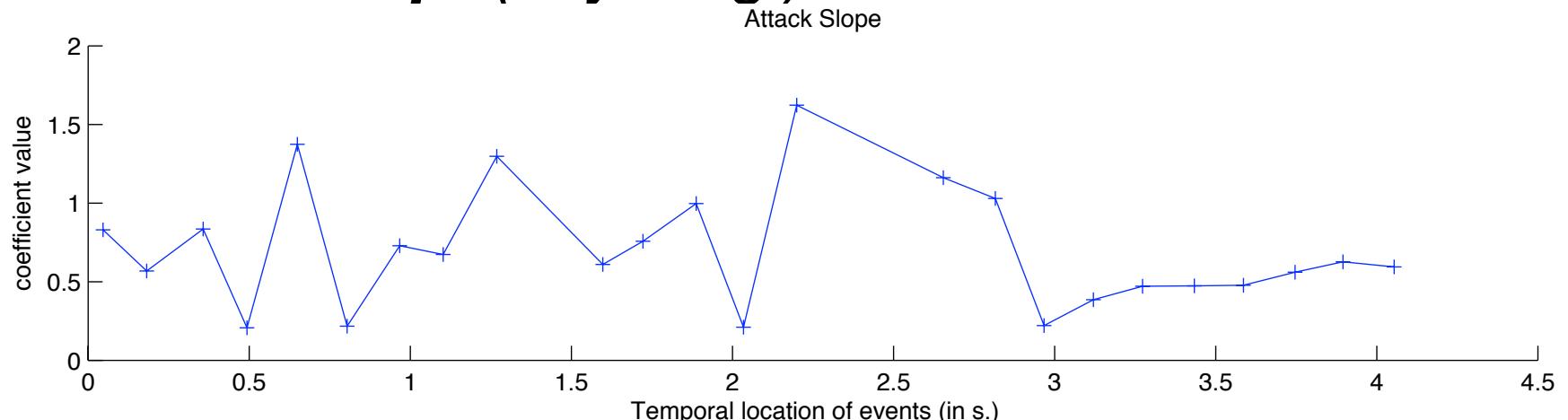


Onsets, attacks, etc.

- ***mironsets*(‘mysong’, ‘Attack’)**

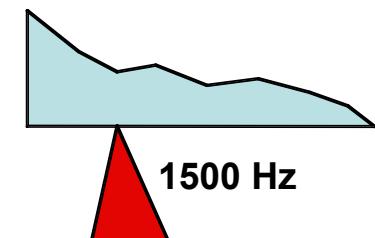
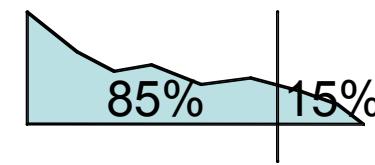
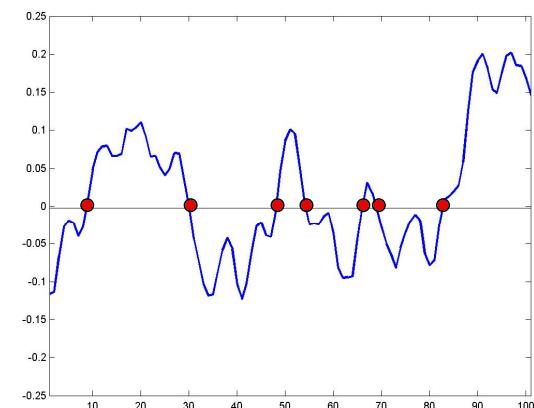


- ***mirattackslope*(‘mysong’)**



Timbre

- ***mirzerocross***
 - Zero-crossing rate
- ***mirrolloff***
 - 85% or 95%
- ***mirbrightness***
- ***mirmfcc***
 - Mel-Frequency Cepstral Coefficients
- ***mirroughness***
 - Sensory Dissonance
- ***mirregularity***



Tempo

- Roughly:

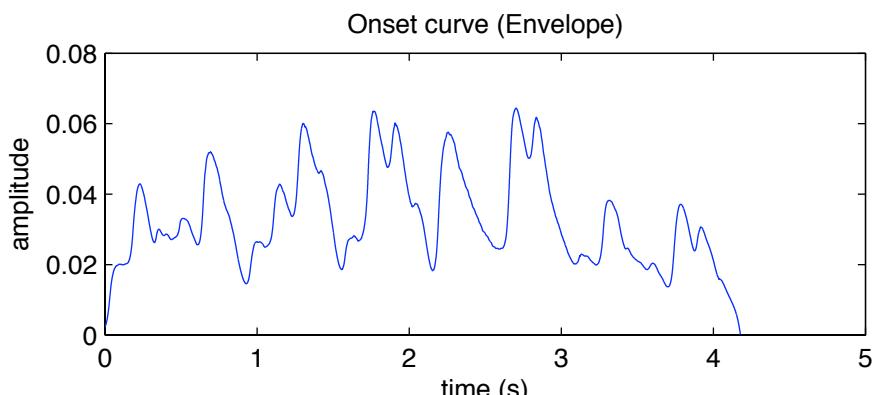
- ***o*** = *mironsets*('mysong',
'Detect', 'No')
- ***do*** = *mironsets*(***o***, 'Diff')
- ***ac*** = *mirautocor*(***do***)
- ***pa*** = *mirpeaks*(***ac***, 'Total', 1)

- In short:

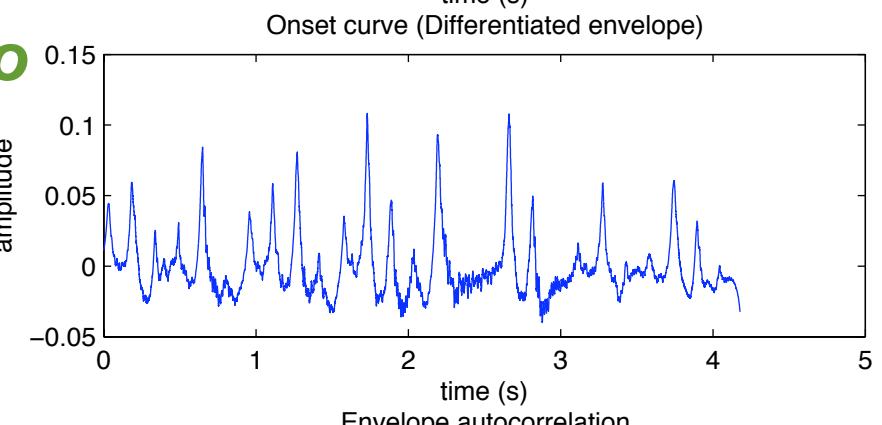
- $[t, \text{pa}] = \text{mirtempo}('mysong')$

$$t = 129.6333 \text{ bpm}$$

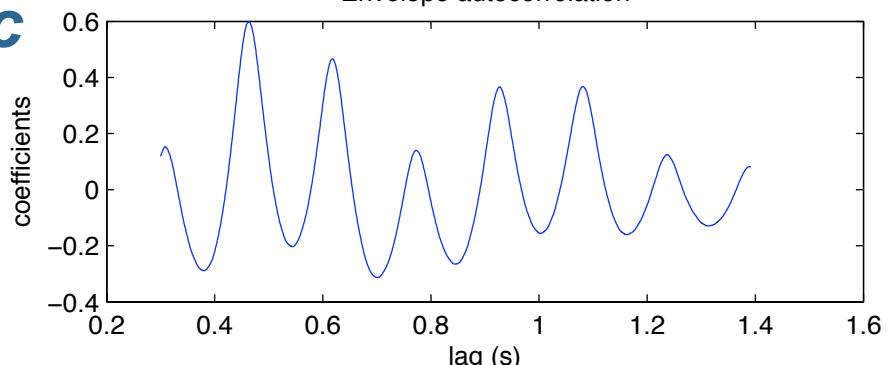
- ***o***



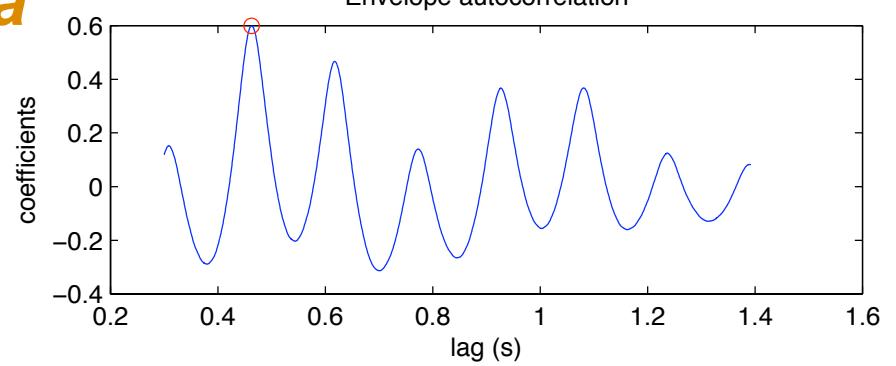
- ***do***



- ***ac***



- ***pa***



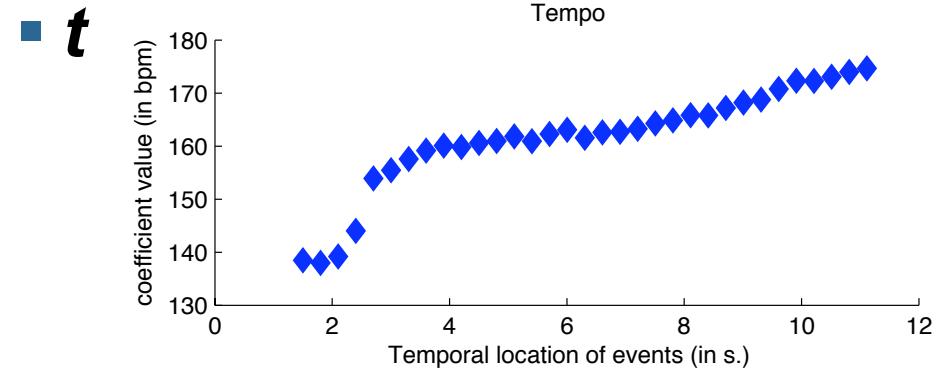
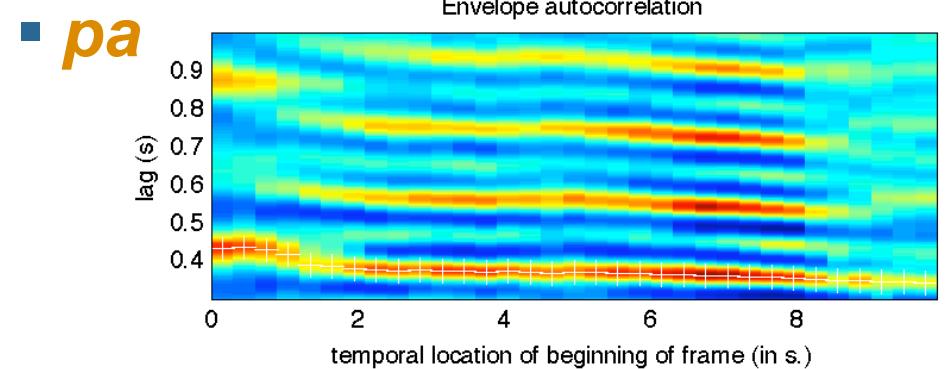
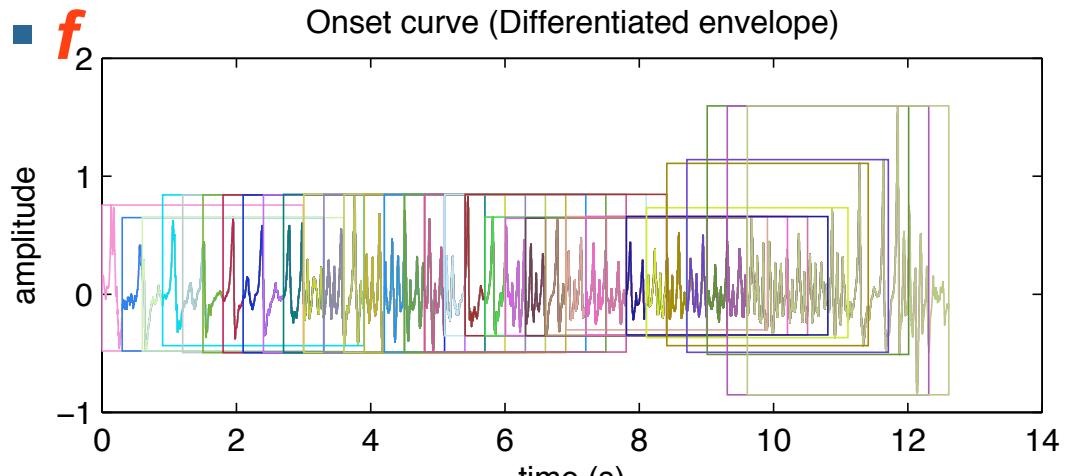
Tempo (temporal evolution)

- Roughly:

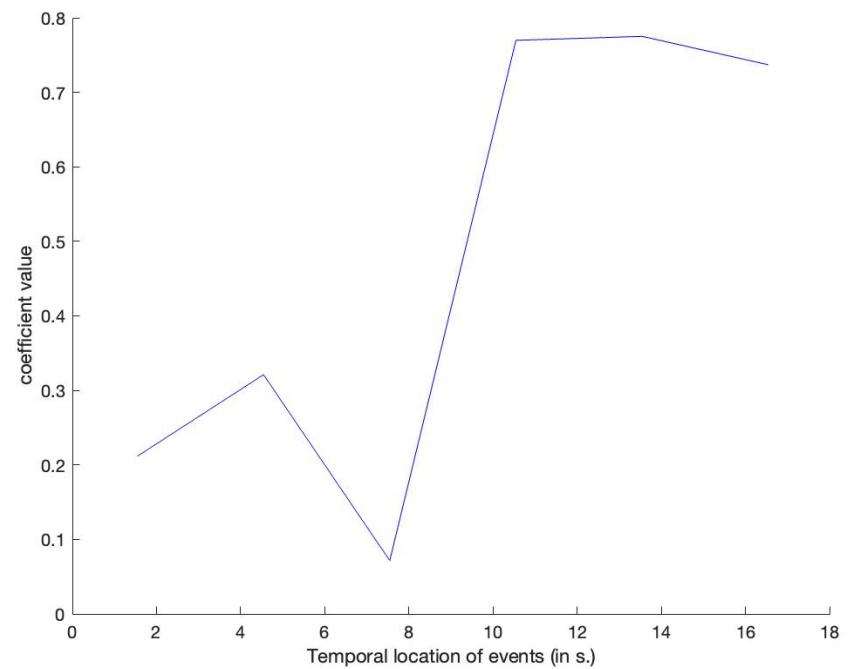
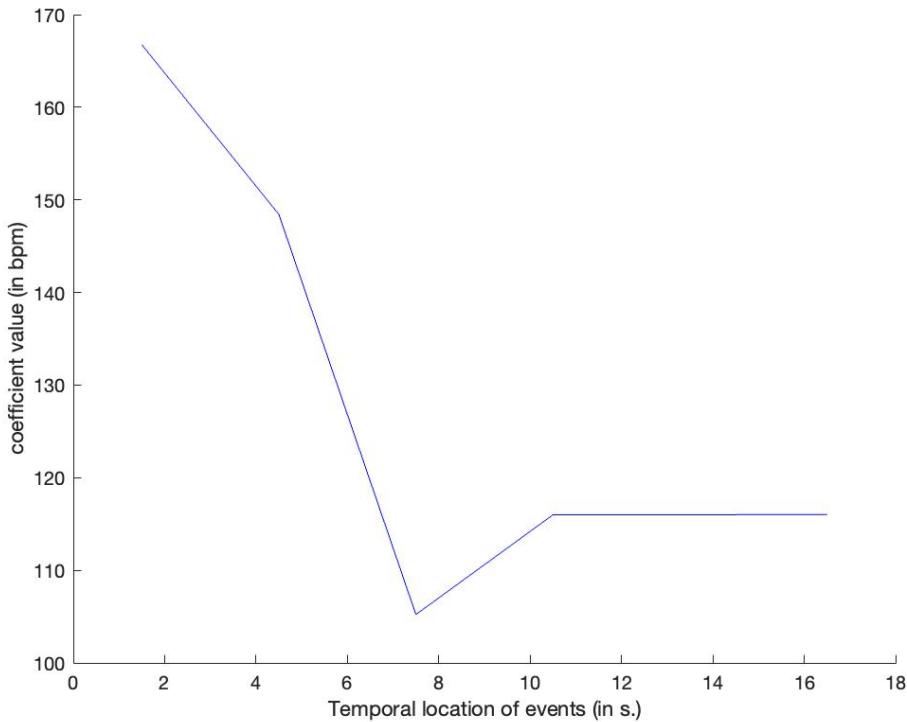
- $\text{o} = \text{mironsets}(\text{'mysong'},$
 $\text{'Detect'}, \text{'No'})$
- $\text{do} = \text{mironsets}(\text{o}, \text{'Diff'})$
- $\text{f} = \text{mirframe}(\text{do})$
- $\text{ac} = \text{mirautocor}(\text{f})$
- $\text{pa} = \text{mirpeaks}(\text{ac}, \text{'Total'}, 1)$

- In short:

- $[\text{t}, \text{pa}] = \text{mirtempo}(\text{'mysong'},$
 $\text{'Frame'})$



tempo vs pulse clarity

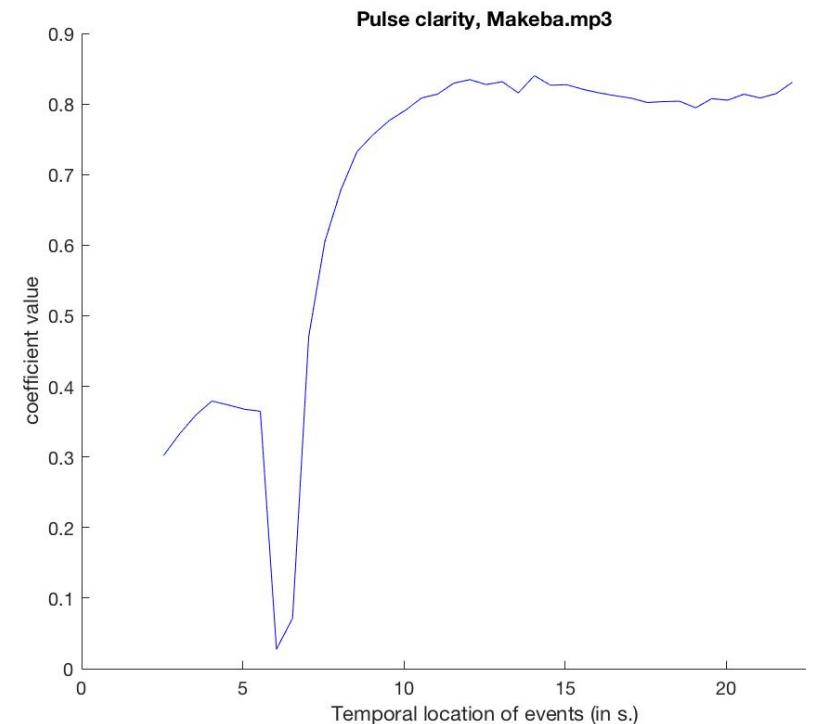


Pulse Clarity



- **b=miraudio('Makeba.mp3','Extract',5,25);**
- **mirpulseclarity(b,'Frame')**

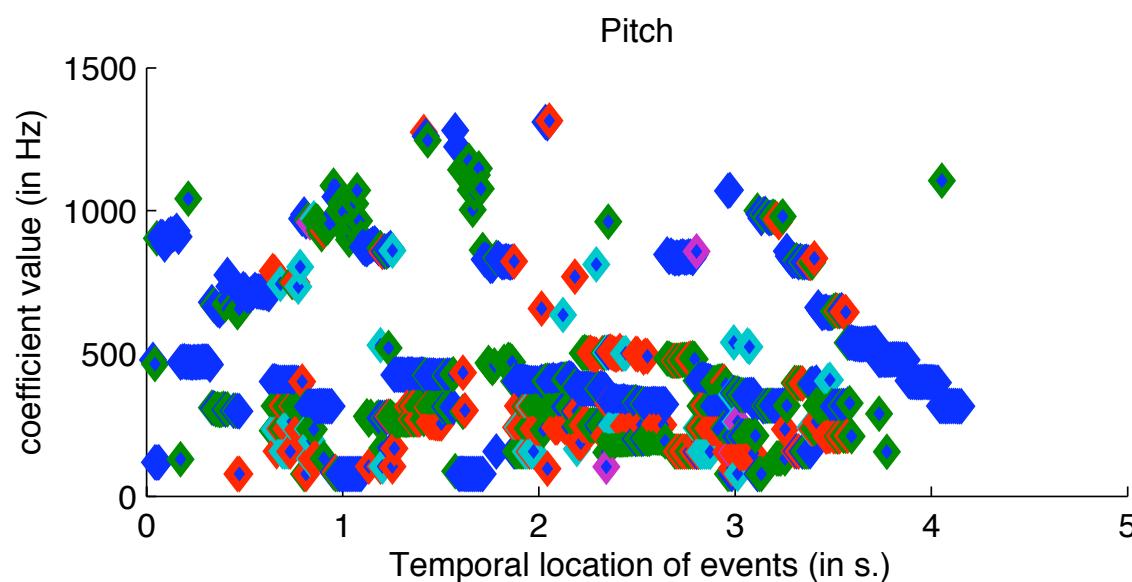
- **mirpulseclarity(miraudio('Makeba.mp3','Extract',5,25),'Frame')**



Pitch

- $[p, pa] = \text{mirpitch}('mysong', 'Frame')$
- $\text{mirplay}(p)$

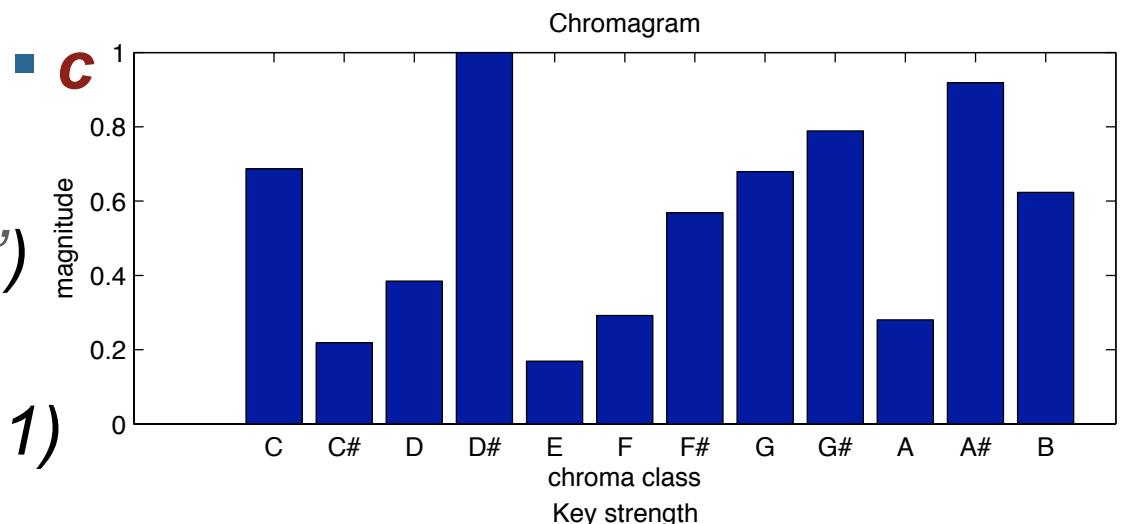
- p



Tonal Analysis

- Roughly:

- $\mathbf{c} = \text{mirchromagram}(\text{'mysong'})$
- $\mathbf{ks} = \text{mirkeystrength}(\mathbf{c})$
- $\mathbf{pk} = \text{mirpeaks}(\mathbf{ks}, \text{'Total'}, 1)$

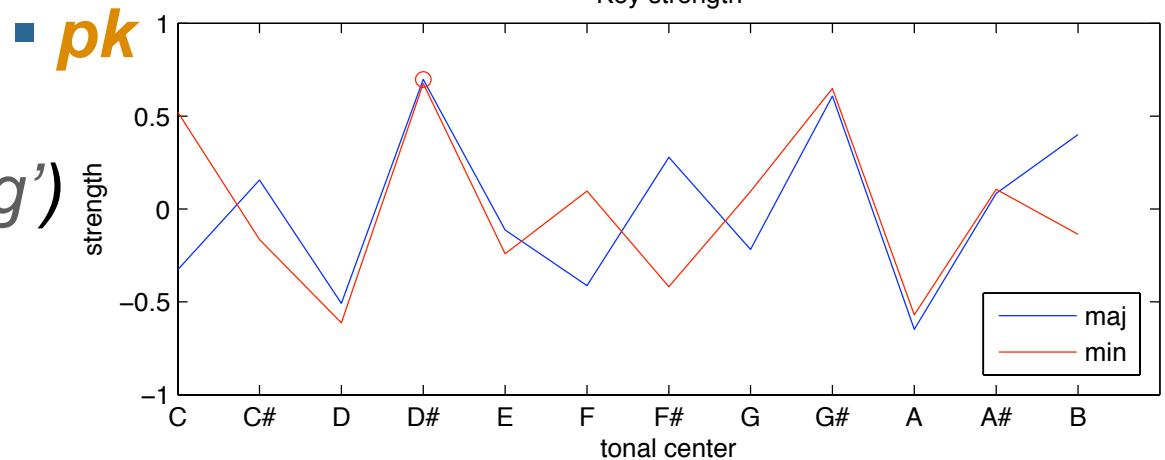


- In short:

- $[\mathbf{k}, \mathbf{kc}] = \text{mirkey}(\text{'mysong'})$

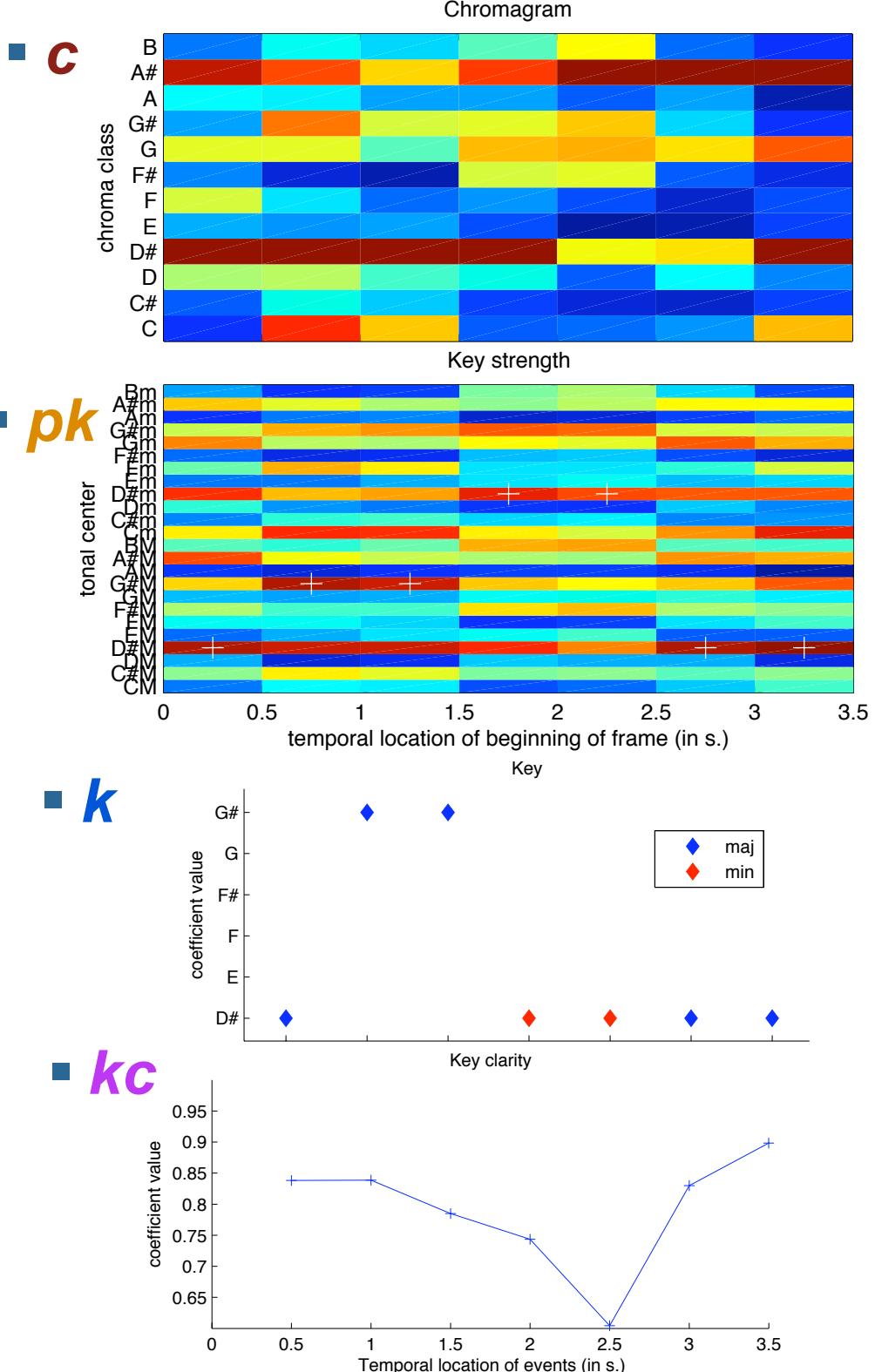
$\mathbf{k} = \text{D\# maj}$

$\mathbf{kc} = 0.69797$



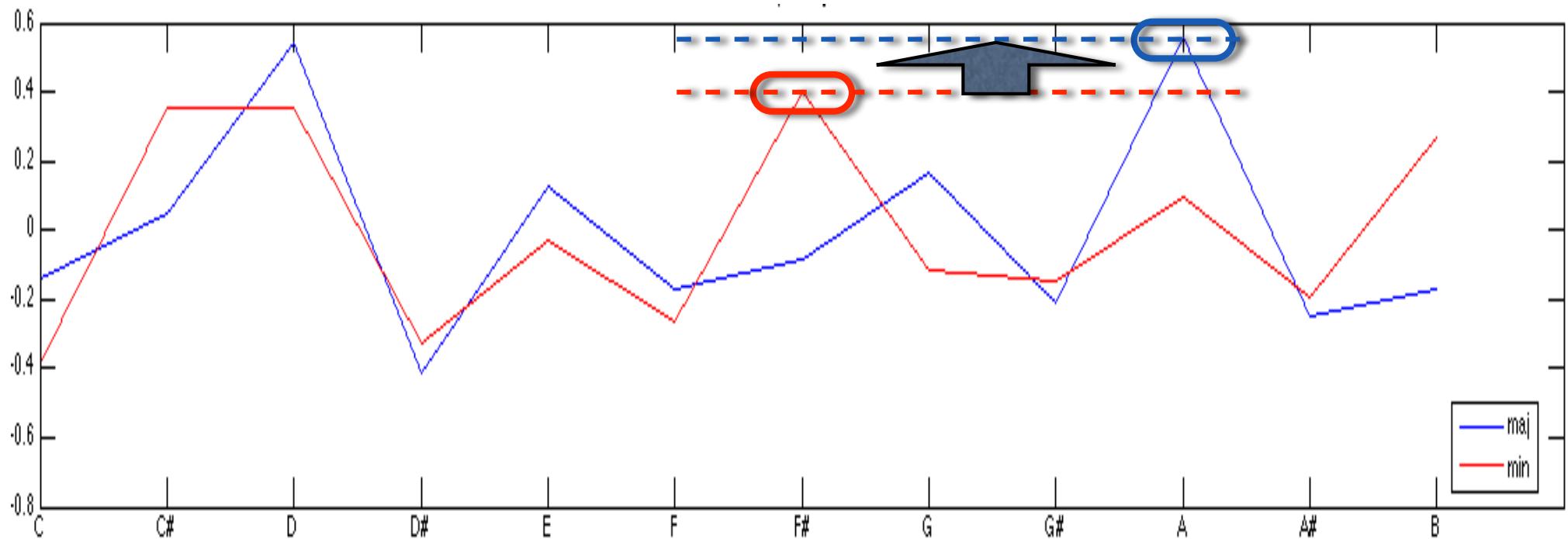
Tonal Analysis (temporal evolution)

- Roughly:
 - \mathbf{c} = *mirchromagram('mysong', 'Frame')*
 - \mathbf{ks} = *mirkeystrength(c)*
 - \mathbf{pk} = *mirpeaks(ks, 'Total', 1)*
- In short:
 - $[\mathbf{k}, \mathbf{kc}] = \text{mirkey}('mysong')$



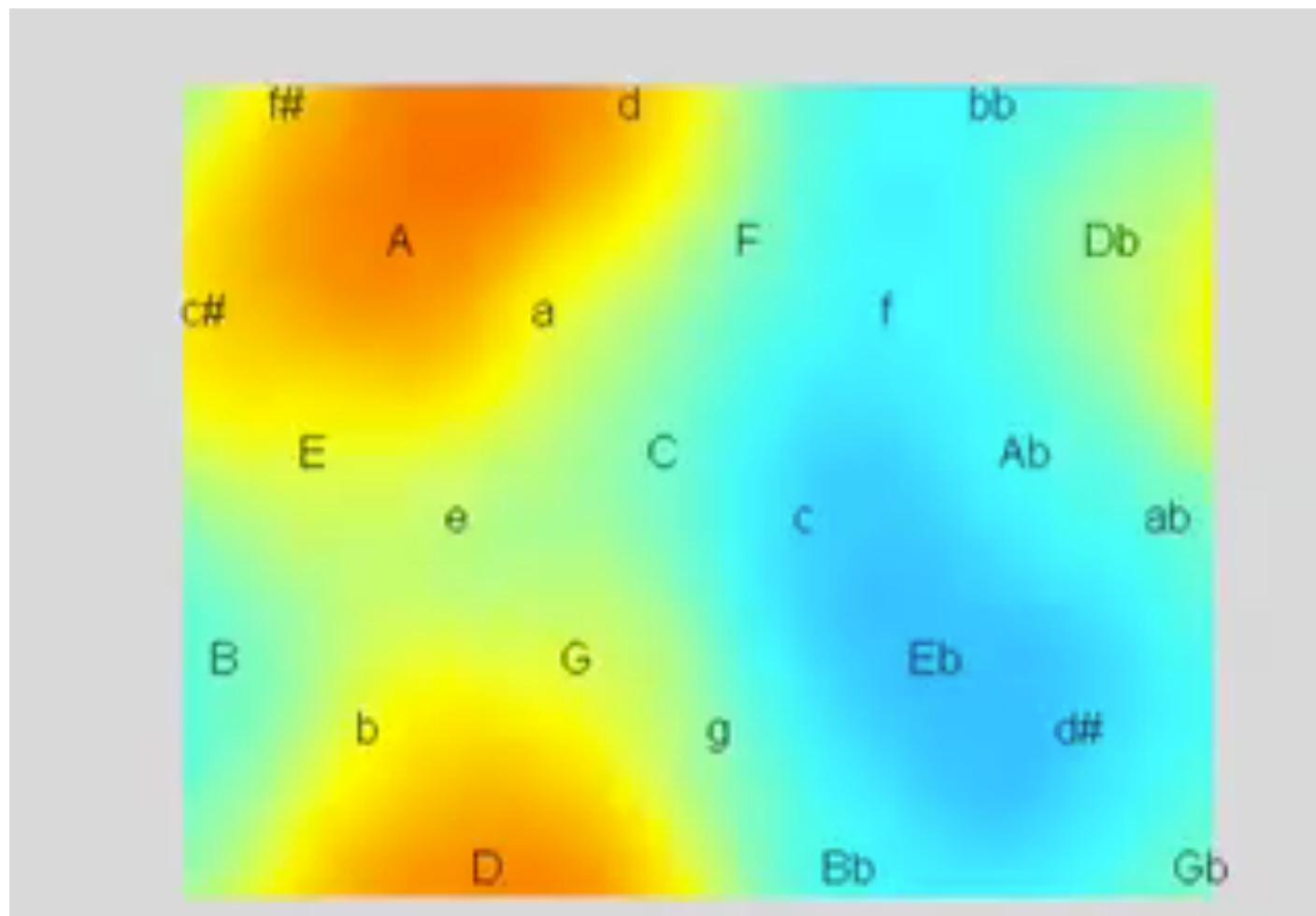
Modal Analysis

- *mirmode('mysong')*



Tonal Analysis

- *mirkeysom* ('mysong', 'Frame')



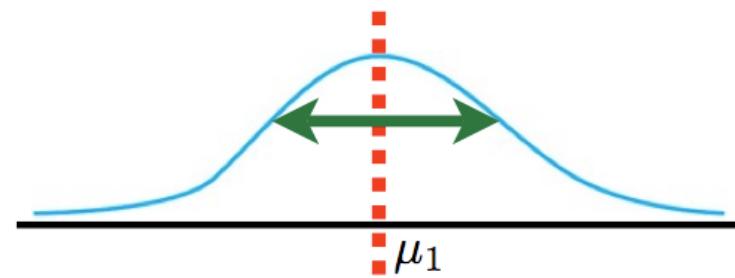
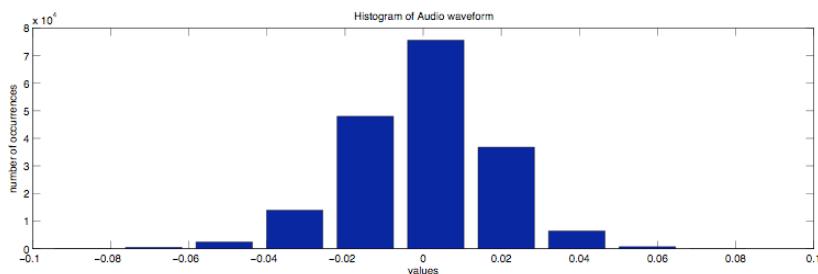
Statistics

- ***mirstat***

- mean
- standard deviation
- slope
- periodicity

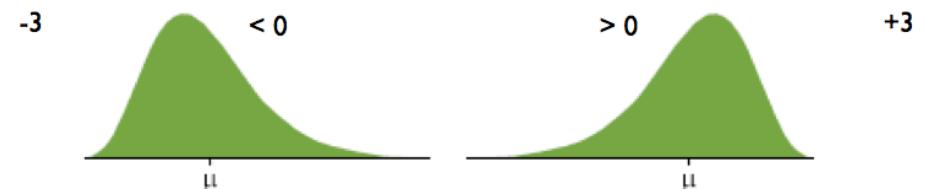
- ***mirhisto***

- distribution histograms



- ***moments***

- ***mircentroid***
- ***mirspread***
- ***mirskeowness***
- ***mirkurtosis***



mirfeatures

- *mirzerocross*
- *mircentroid*
- *mirbrightness*
- *mirspread*
- *mirkewness*
- *mirkurtosis*
- *mirrolloff*
- *mirentropy*
- *mirflatness*
- *mirroughness*
- *mirregularity*
- *mirinharmonicity*
- *mirmfcc*
- *mirfluctuation*
- *mirattacktime*
- *mirattackslope*
- *mirlowenergy*
- *mirflux*
- *mirpitch*
- *mirchromagram*
- *mirkeystrength*
- *mirkey*
- *mirmode*
- *mirhcdf*
- *mirtempo*
- *mirpulseclarity*

Saving & Exporting

a = *miraudio*('mysong', 'Extract', 0, 1)

- *mirsave(a, 'mysample')*

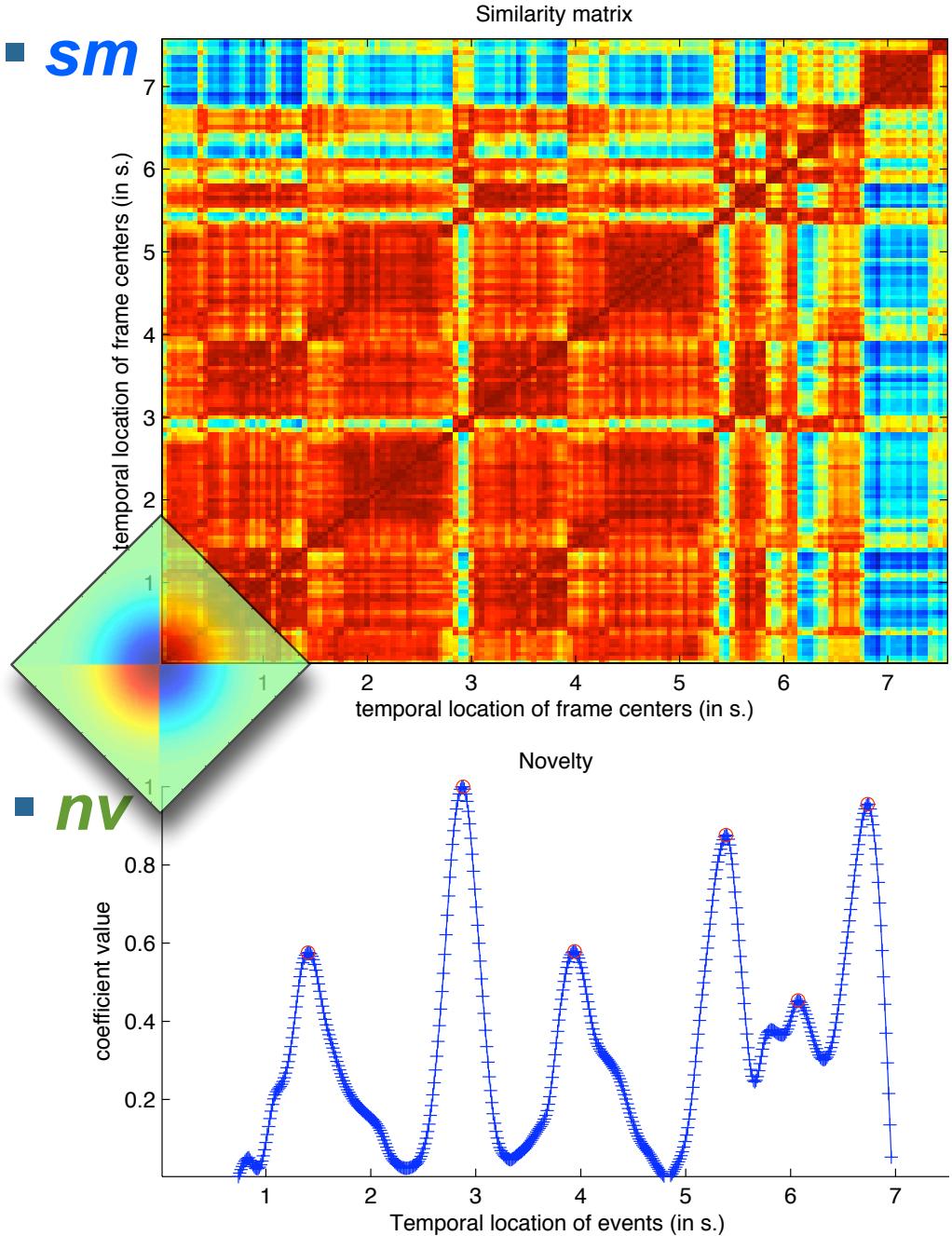
p = *mirwhatever...*

- *mirexport('res.txt', p, q, r, ...)*
 - Excel, Word, etc..
- *mirexport('res.arff', p, q, r, ...)*
 - Weka, etc..

Musical Feature(s)
to
Musical Structure

Structural analysis

- $\text{x} = \text{mirspectrum}(\text{'mysong'}, \text{'Frame'})$
- $= \text{mirmfcc}(\text{'mysong'}, \text{'Frame'})$
- whatever...
- $\text{sm} = \text{mirsimmatrix}(\text{x})$
- $\text{nv} = \text{mirnovelty}(\text{sm})$



Structural analysis

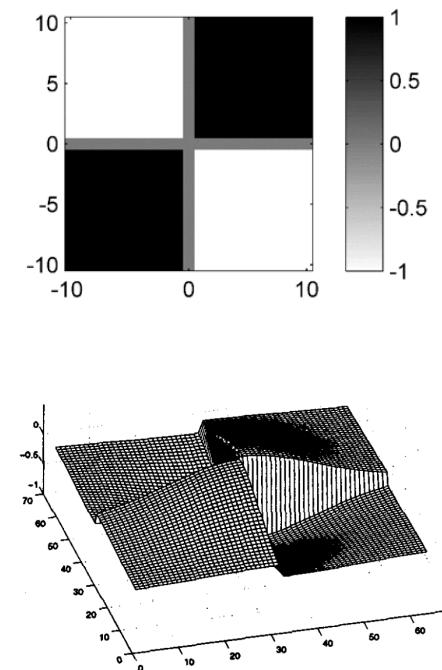
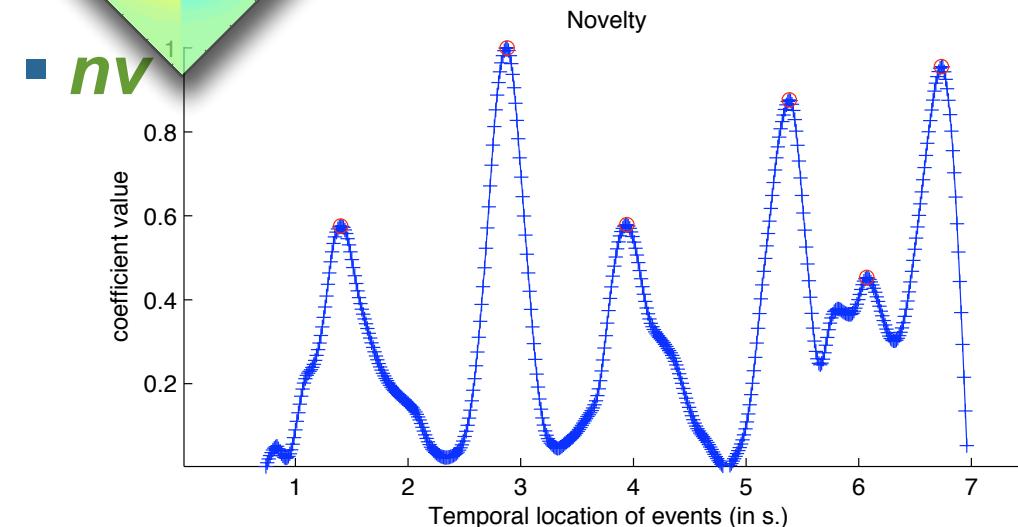
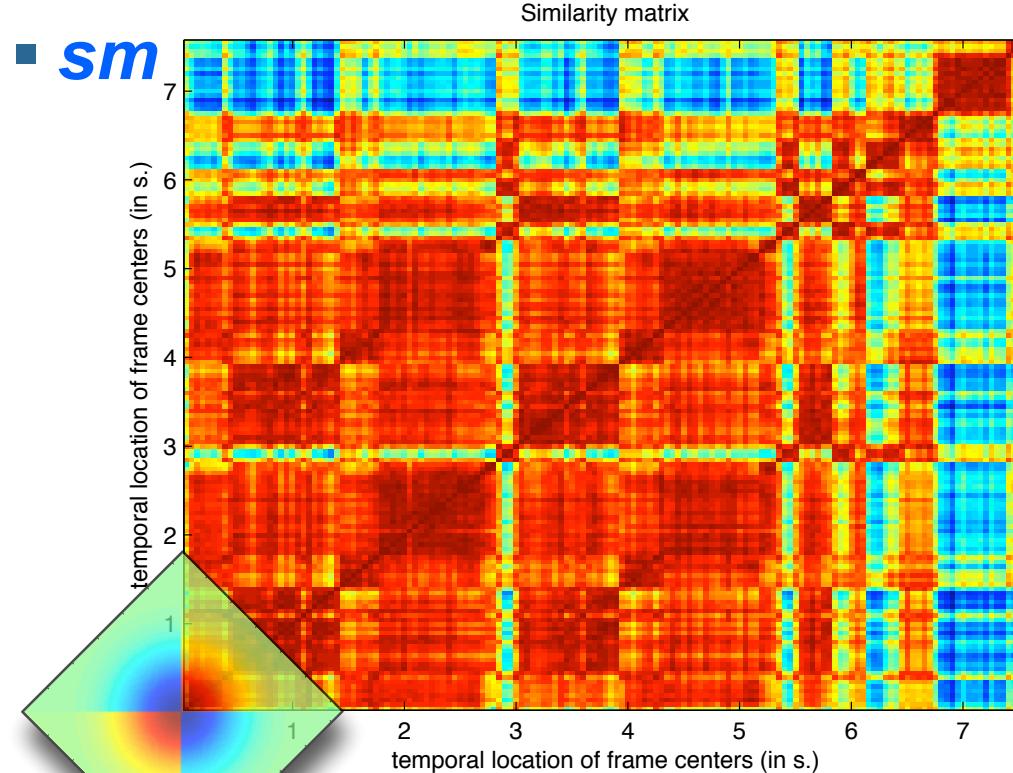


Figure 3. 64×64 checkerboard kernel with Gaussian taper

Tonal Analysis

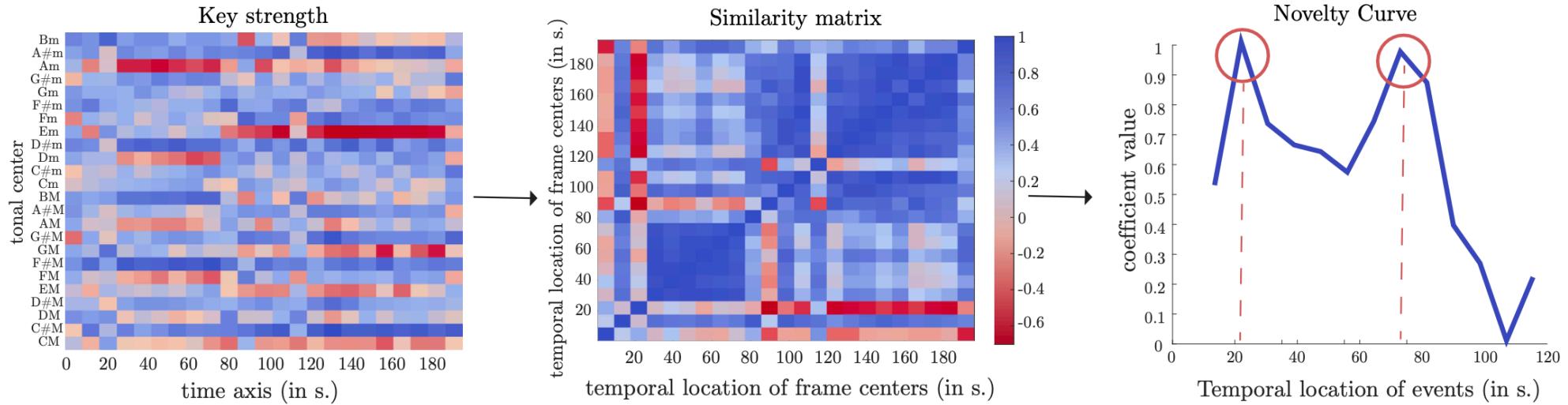


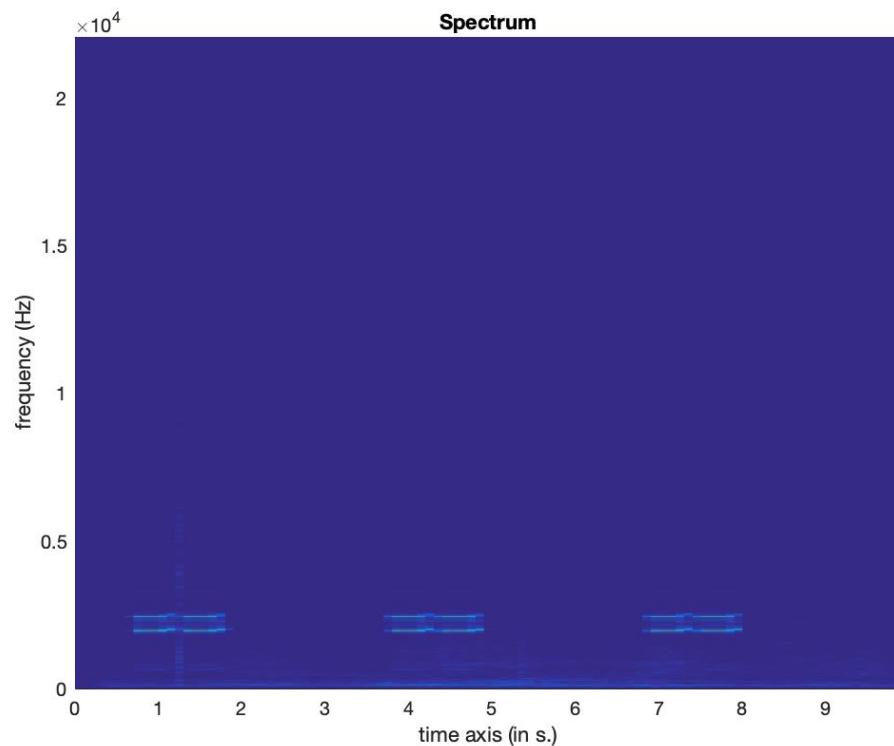
Figure 3. Music segmentation pipeline. We segment all tracks from the soundtrack to ensure a cohesive listening experience. We extract keystrength [21] that captures tonal properties of a soundtrack (left), compute the self-similarity matrix (center), and use that to calculate the novelty curve [24]. The peaks of this curve are used to segment the track.

SONUS TEXERE! AUTOMATED DENSE SOUNDTRACK CONSTRUCTION FOR BOOKS USING MOVIE ADAPTATIONS

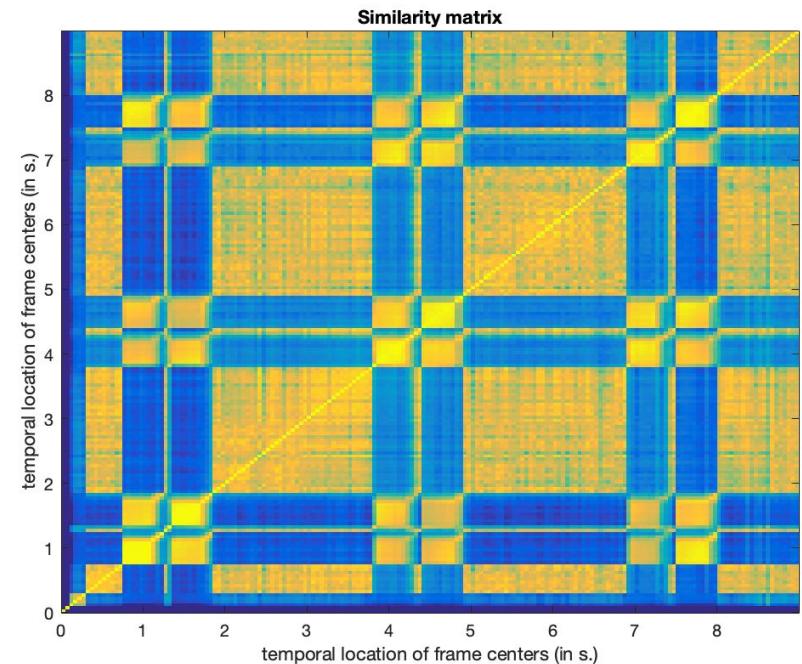
Jaidev Shriram Makarand Tapaswi Vinoj Alluri

International Institute of Information Technology, Hyderabad

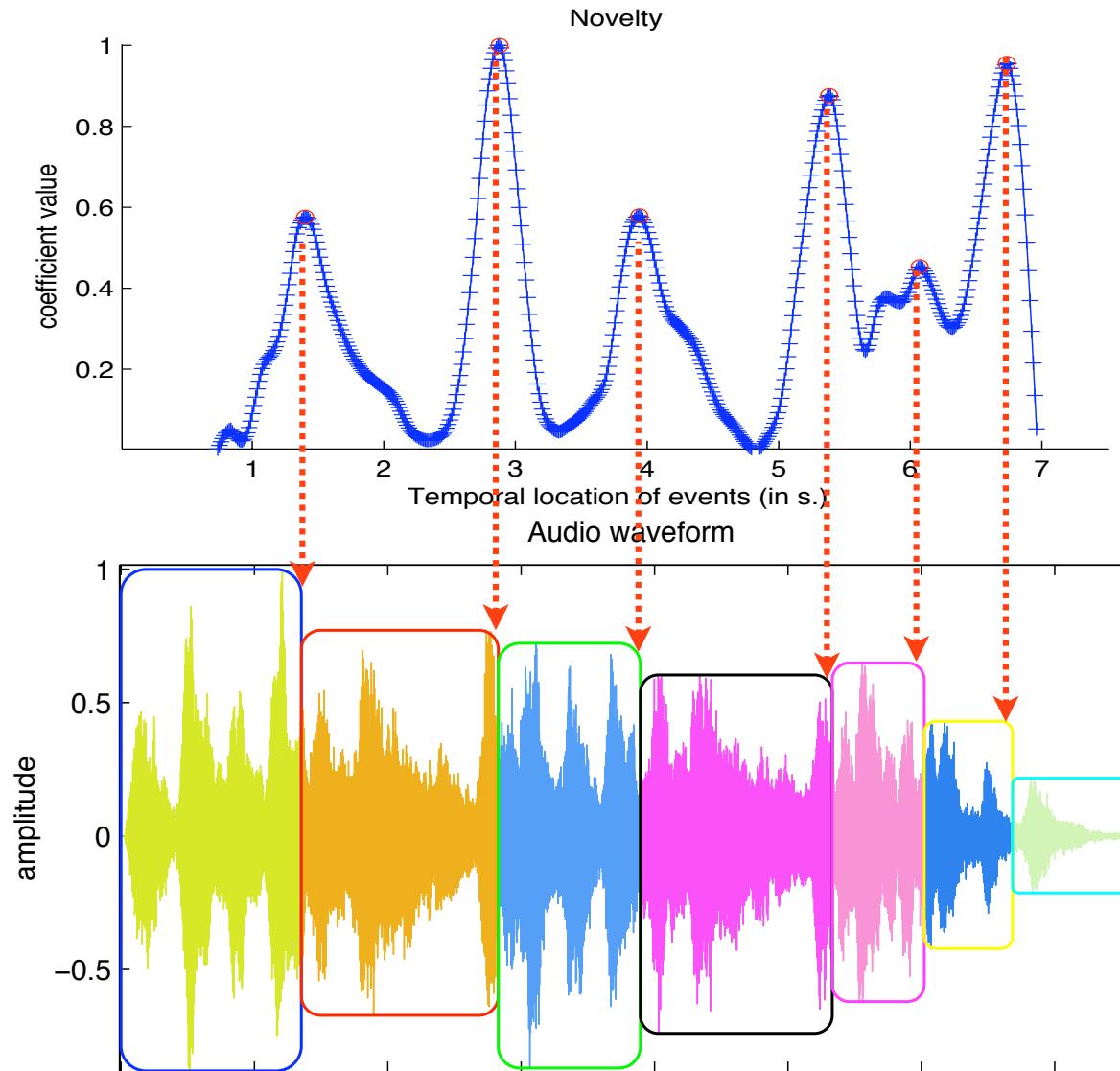
<https://auto-book-soundtrack.github.io/>



How does the structure of this look like?

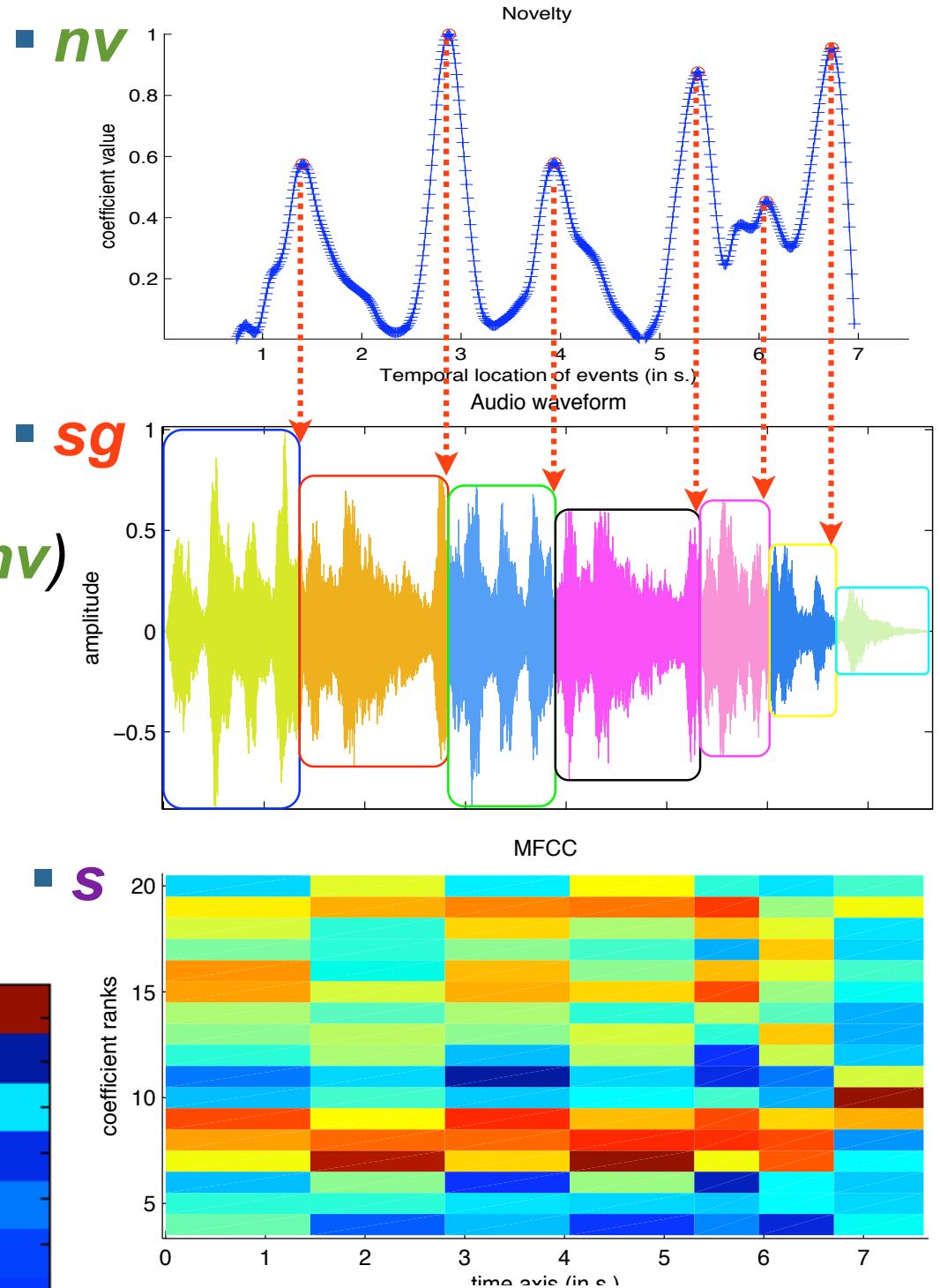
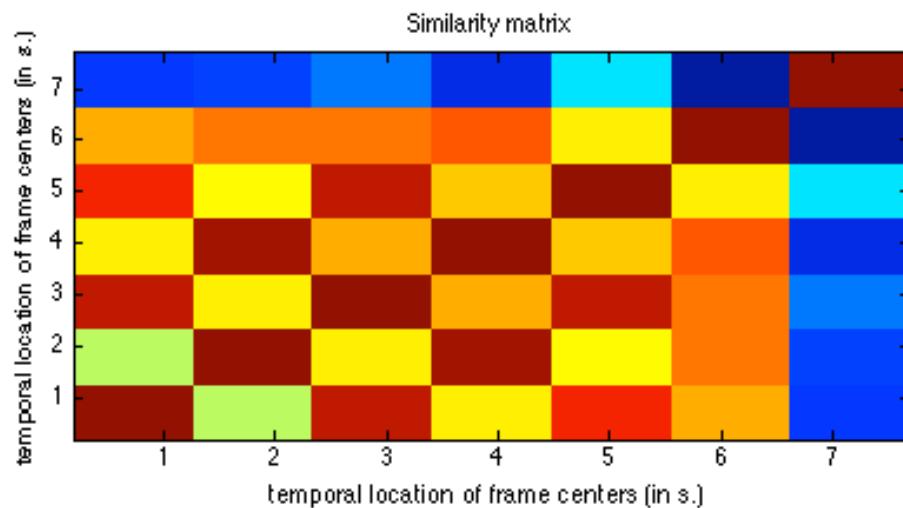


Segmentation

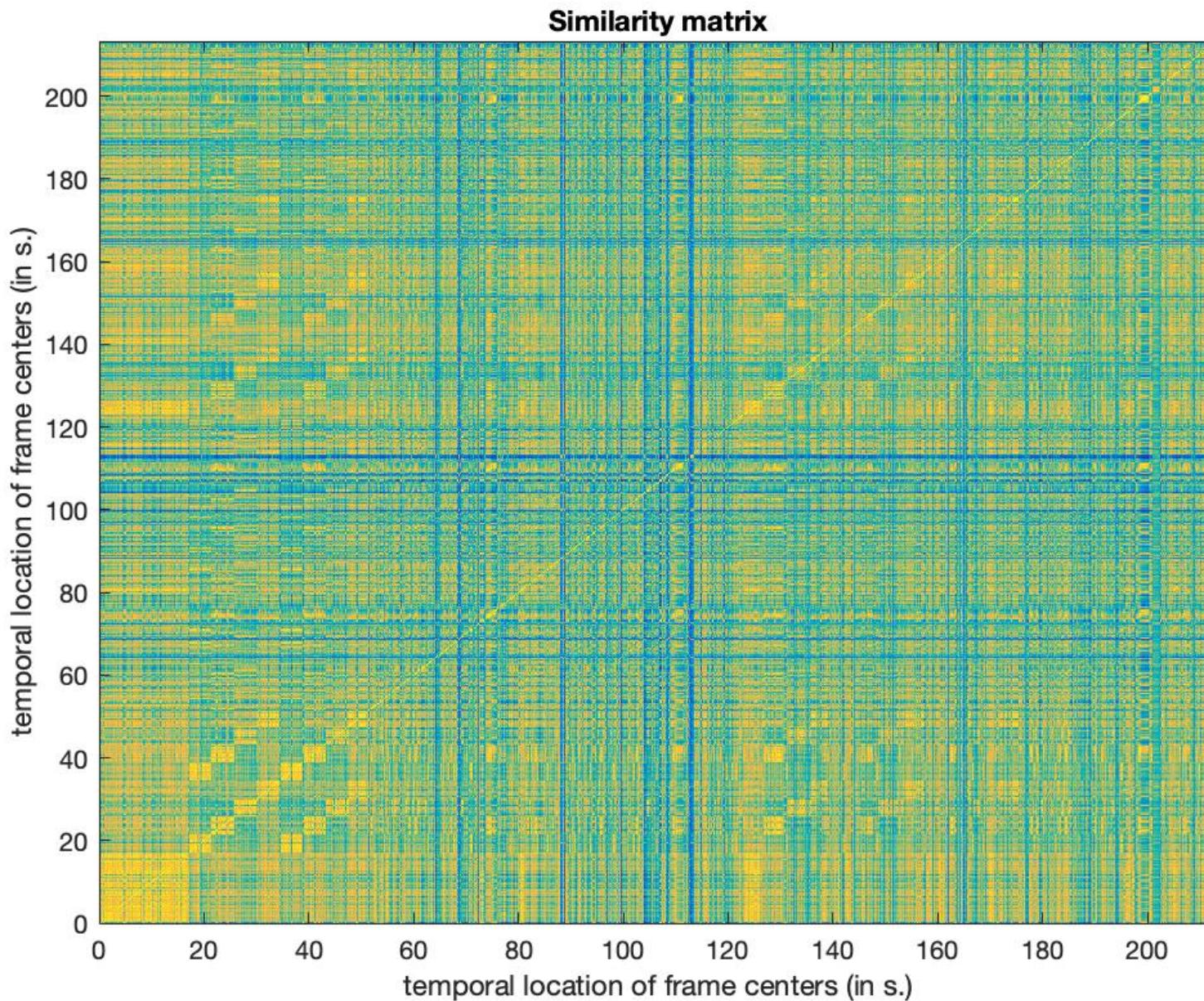


Segmentation

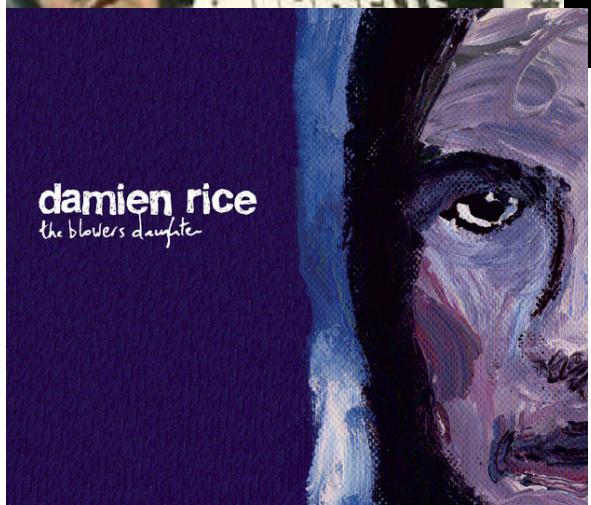
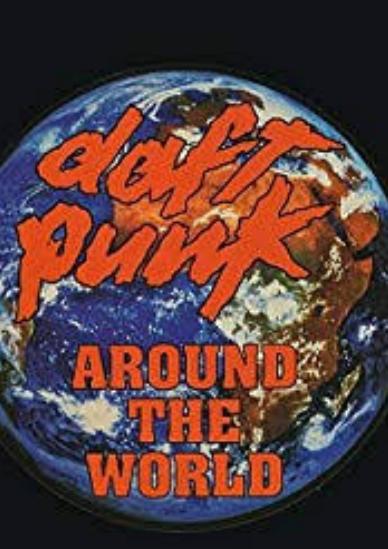
- $nv = \text{mirnovelty}(sm)$
- $sg = \text{mirsegment}(\text{'mysong'}, nv)$
- $sg = \text{mirsegment}(\text{'mysong'})$
 - $\text{mirplay}(sg)$
- $s = \text{mirmfcc}(sg)$
- $sm = \text{mirsimatrix}(s)$



What can you say about the structure of this piece?

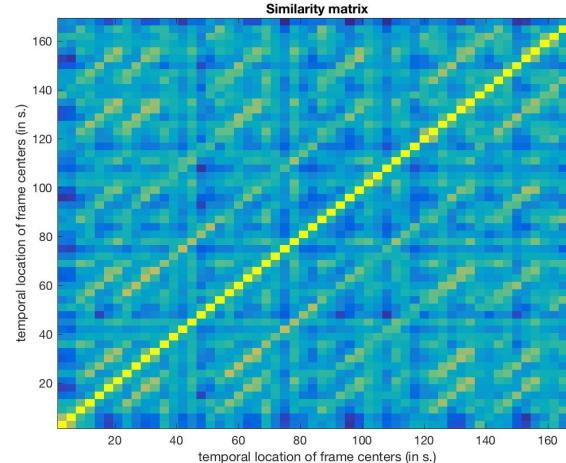
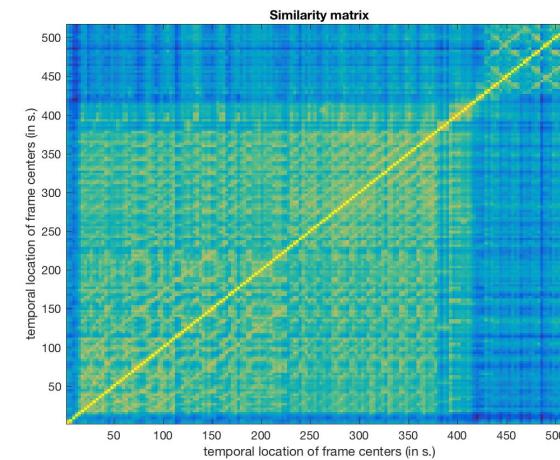
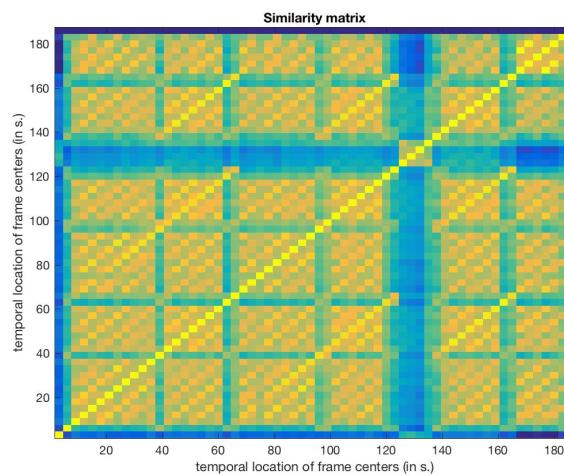
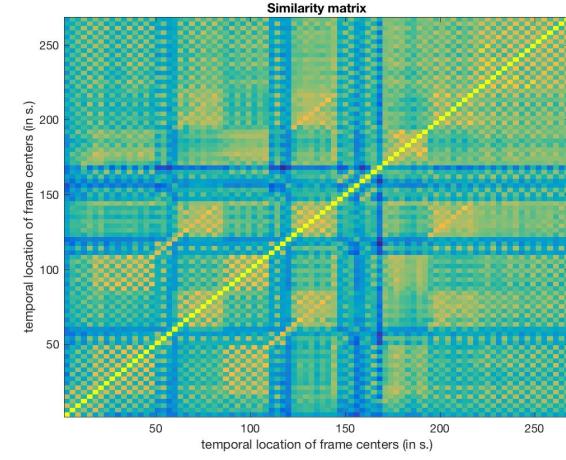
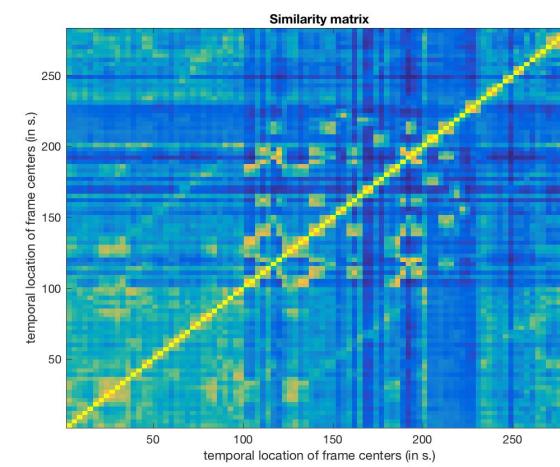
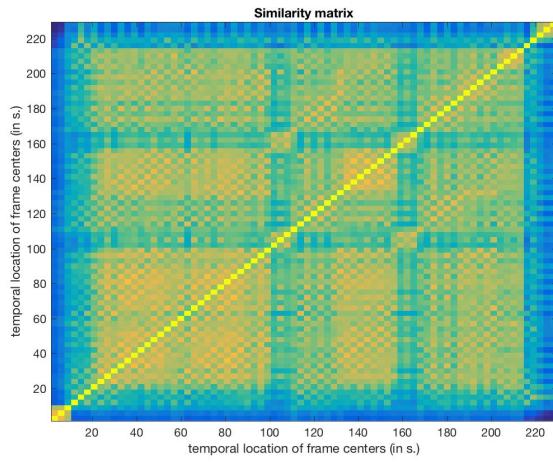
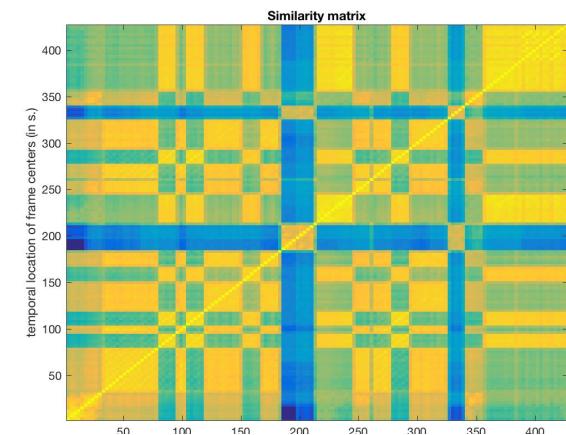
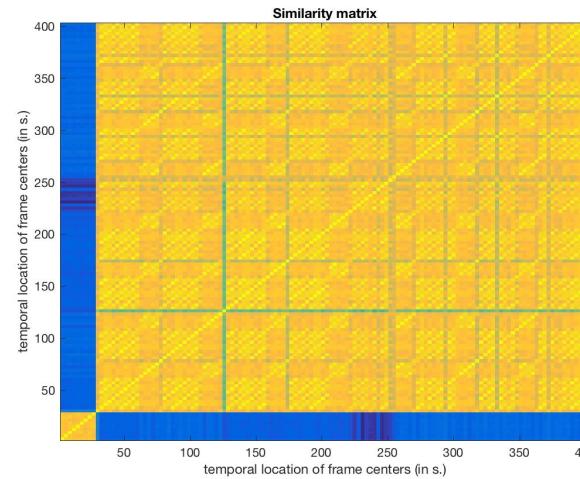
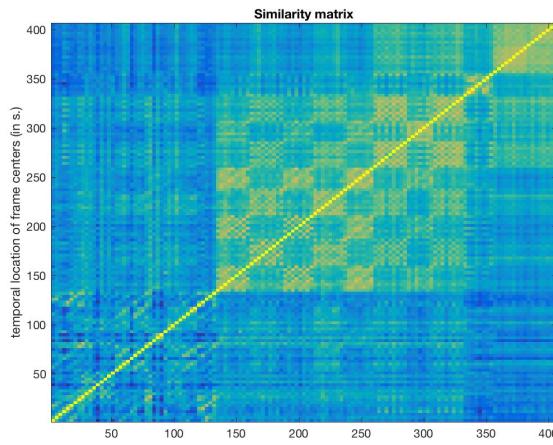


<https://1drv.ms/u/s!AtoLT6JDyxo-iWSxEYJILzdglJ3w?e=nbu6Fm>



?



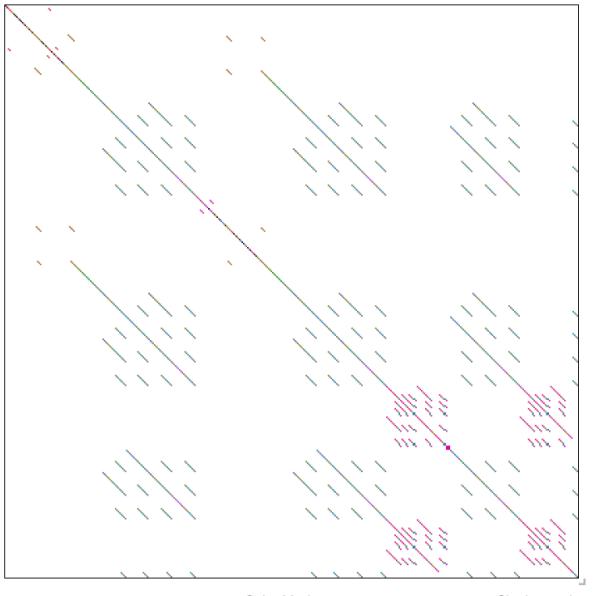


<https://shorturl.at/bjz28>

Visualizing Music(al Structure)

Lyrics

Vox



Color Mode
 black & white colorful color title

Single-word matches

Mobile mode



Carly Rae Jepsen - Call Me Maybe

I threw a wish in the well,
Don't ask me, I'll never tell
I looked to you as it fell,
And now you're in my way
I'd trade my soul for a wish,
Promise you're mine for a kiss
I wasn't looking for this,
But now you're in my way
Your stare was holdin',
Ripped jeans, skin was showin'
Hot night, wind was blowin'
Where you think you're going,
Baby?
Hey, I just met you,
And this is crazy,
But here's my number,
So call me, maybe!
It's hard to look right
At you baby,
But here's my number,
So call me, maybe!
And all the other boys,
Try to chase me,
But here's my number,
So call me, maybe!
I beg, and borrow and steal
At first sight and it's real
I didn't know I would feel it,
But it's in my way
Your stare was holdin',
Ripped jeans, skin was showin'
Hot night, wind was blowin'
Where you think you're going,
Baby?
Hey, I just met you,
And this is crazy,
Hey, I just met you,
And this is crazy,

I missed you so, so bad (bad,
bad)

It's hard to look right
At you baby,

But here's my number,
So call me, maybe!

Hey, I just met you,
And this is crazy,

But here's my number,
So call me, maybe!

And all the other boys,
Try to chase me,

But here's my number,
So call me, maybe!

Before you came into my life
I missed you so bad

I missed you so bad

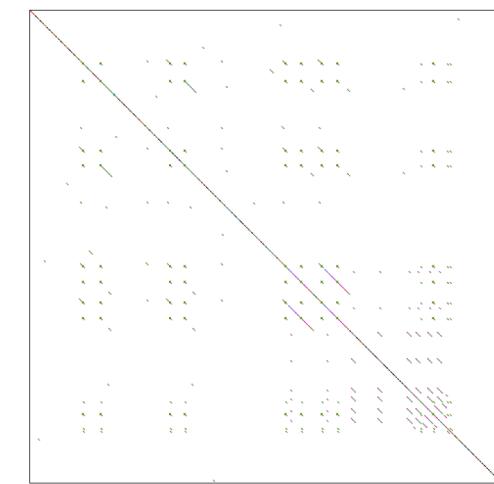
I missed you so, so bad

Before you came into my life
I missed you so bad

I missed you so bad

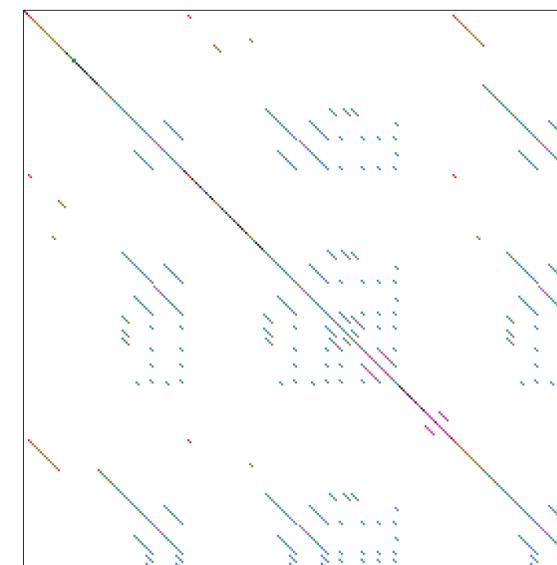
And you should know that

So call me, maybe!



Color Mode
 black & white colorful color title
[Permalink](#)

Single-word matches



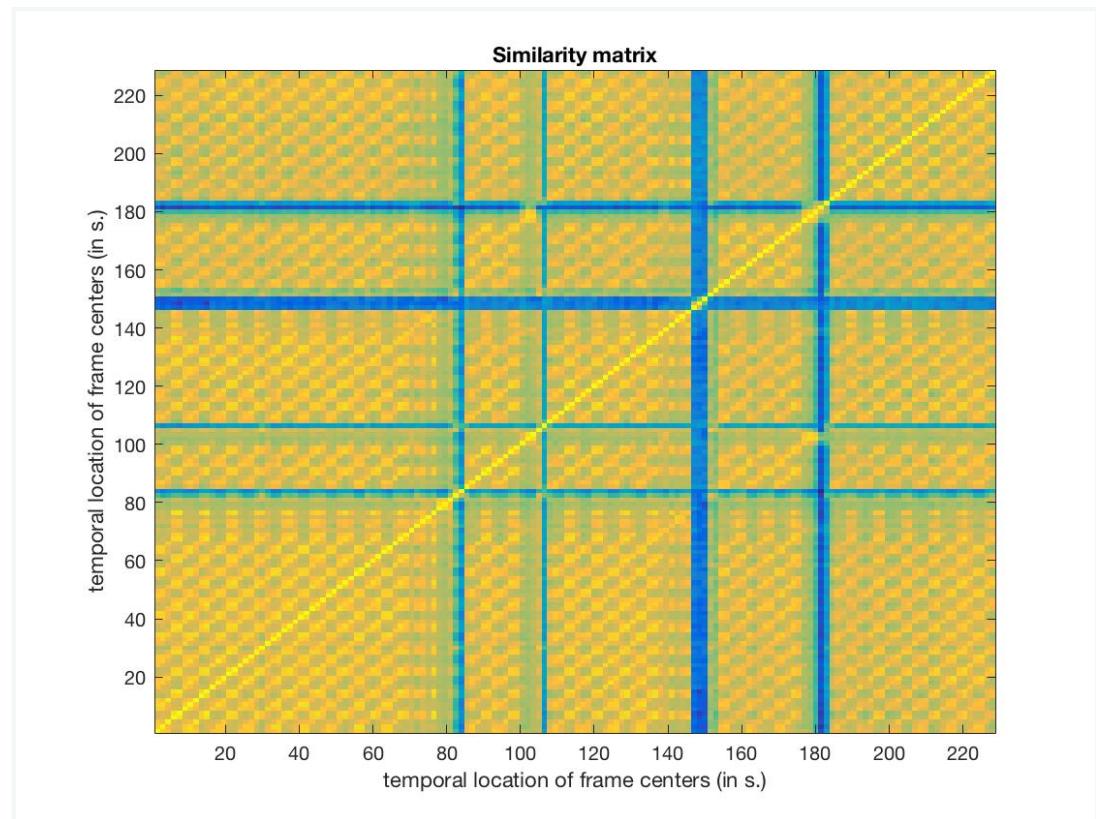
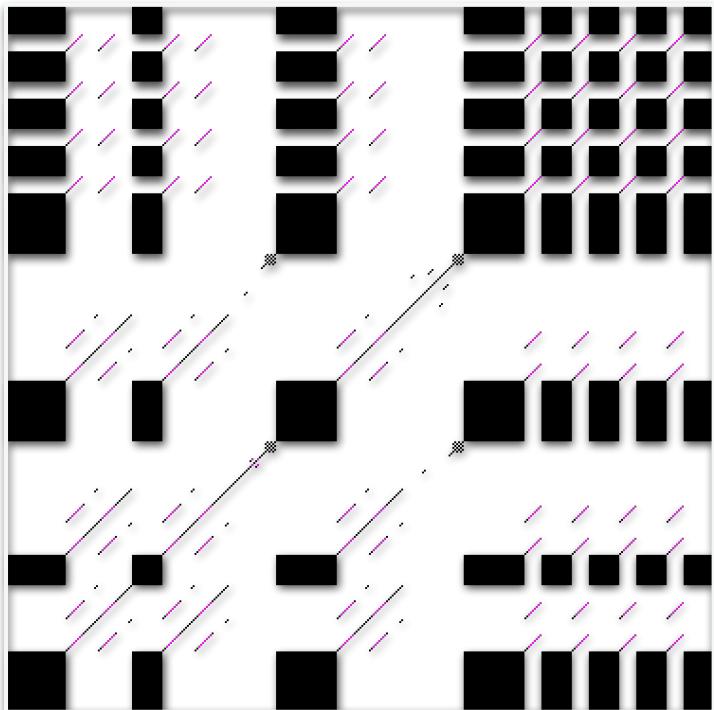
Color Mode

Single-word matches

Mobile mode

Custom			
Aao Lavan Maiboori Nu Aan Jaan Di Pasoon Nu Zehar Bane Haan Teri Pee Jaavon Main Poori Nu Aana Si Oh Nahi Aaya Dil Bhang Bang Mera Taikraya Kega Bol Ke Dosa Jaave Paawan Gheyo Dee Choori Nu Ranwan Ch Banwan Ch On Nu Lukavan Koiy Mainu Na Roke Mere Dhol Judaiyan Di Tainu Khabar Kivein Hove Aa Jaave Dil Tera Poora Vi Na Hove Haan Baniya Banayai Di Galbaat Kivein Hove Aa Jaave Dil Tera	Poora Vi Na Hove Bhool Gayi Maiboori Nu Duniya Di Dastoor Nu Saath Tera Hai Bathera Poora Kar Zaroon Nu Aana Si Oh Nai Aaya Raasta Na Dikhaya Sardaan Na Hove Dil Humara De Sahare Khwhishhat Adhoori Nu On Nu Lukavan Waaari Main Jaavan Man Tainu Bulawan Gall Saari Taan Hove Dildaran Di Sab Yaaran Di Aazzani Na Hove Dildaran Di Sab Yaaran Di Aazzani Na Hove Aa Chale Leke Tuje Hai Jahan Silile Tu Hai Wahn Hai Teri Kam Banade Saja De Panah De Poora Vi Na Hove Oh Haan Baniya Baneijan Di Galbaat Kivein Hove	Aa Jaave Dil Tera Poora Vi Na Hove Mere Dhol Judaiyan Di Sardaan Na Hove Mere Dhol Judaiyan Di Poora Vi Na Hove Mere Dhol Judaiyan Di Sardaan Na Hove Dildaran Di Sab Yaaran Di Aazzani Na Hove Dildaran Di Sab Yaaran Di Aazzani Na Hove Mero Dhol Judaiyan Di Tainu Khabar Kivein Hove Aa Jaave Dil Tera Poora Vi Na Hove Bana De Saja De Panah De Humein Bana De Saja De Panah De Humein	Aag Lawan Maiboori Nu Aan Jaan Di Pasoon Nu Zehar Bane Haan Teri Pee Jaavon Main Poori Nu Ranwan Ch Banwan Ch Oh Nu Lukavan Koy Mainu Na Roke Mere Dhol Judaiyan Di Tainu Khabar Kivein Hove Aa Jaave Dil Tera Poora Vi Na Hove Haan Baniya Banayai Di Galbaat Kivein Hove Aa Jaave Dil Tera

Repetition in Music



Genre Classification

Where does one begin?



What does “genre” (musically) mean?

How do you describe “genre”?

Genre Classification

What assumptions do we make?

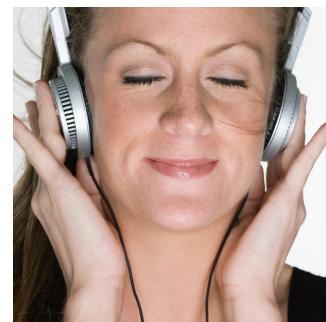
Classes are **acoustically** and **perceptually** separable

How do we go about selecting relevant acoustic features and parameters thereof?

Genre Classification

Assumption: **Classes are separable**

Perceptual



Genre 1

Genre 2

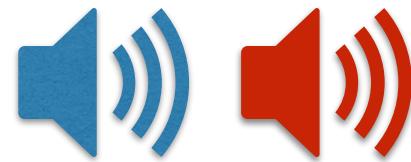


what are the perceptual features that distinguish these genres?

Genre Classification

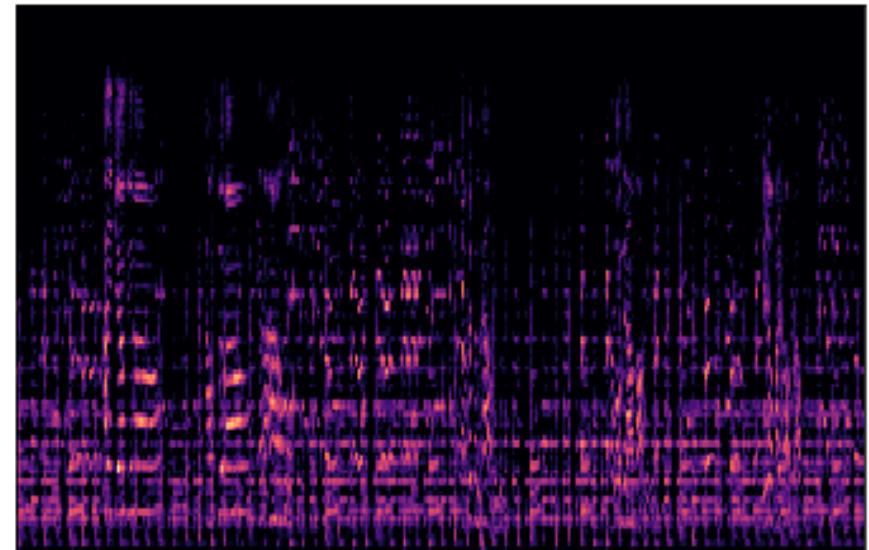
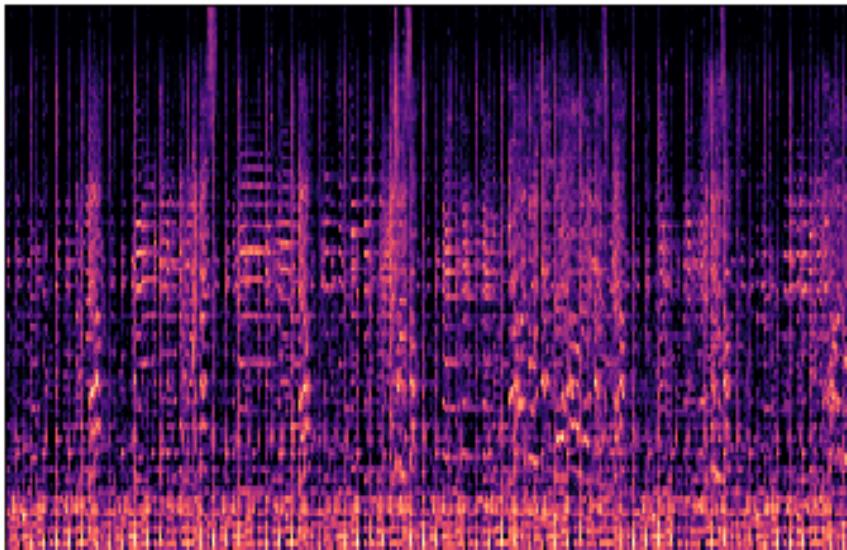
Assumption: **Classes are separable**

Genre 1



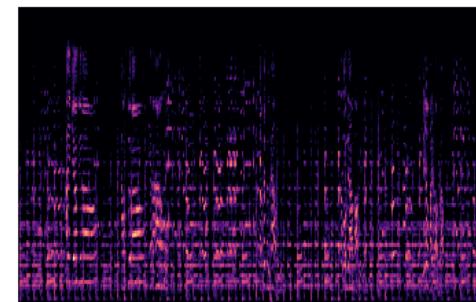
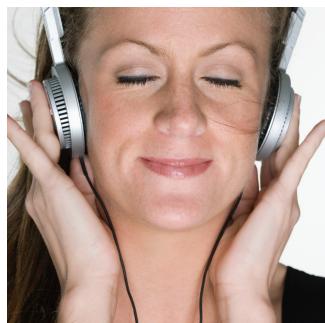
Genre 2

Acoustically

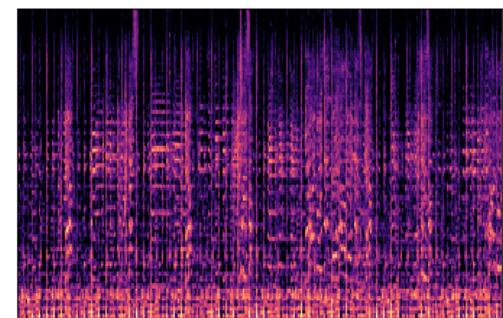


Genre Classification

based on perceptual features can you identify the acoustic features ?



Genre 1



Genre 2

Genre Classification

So what sort of features do i choose?

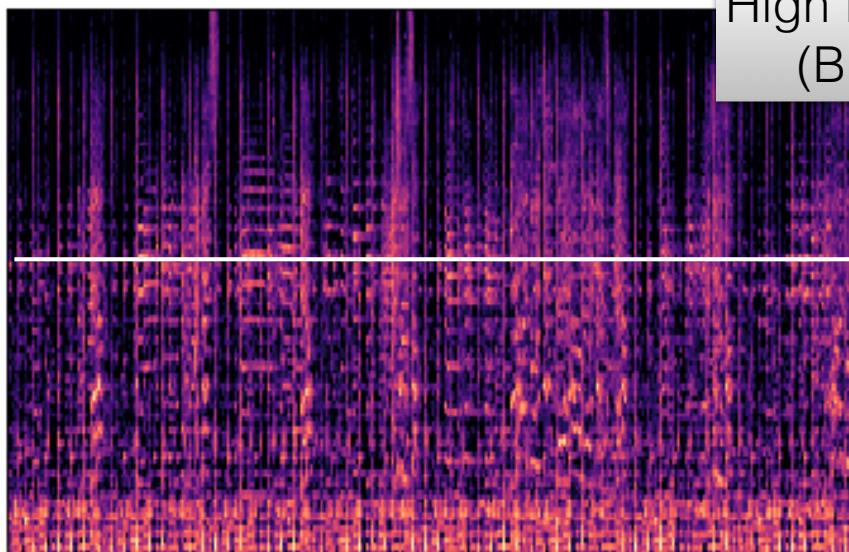
Timbre

Rhythm

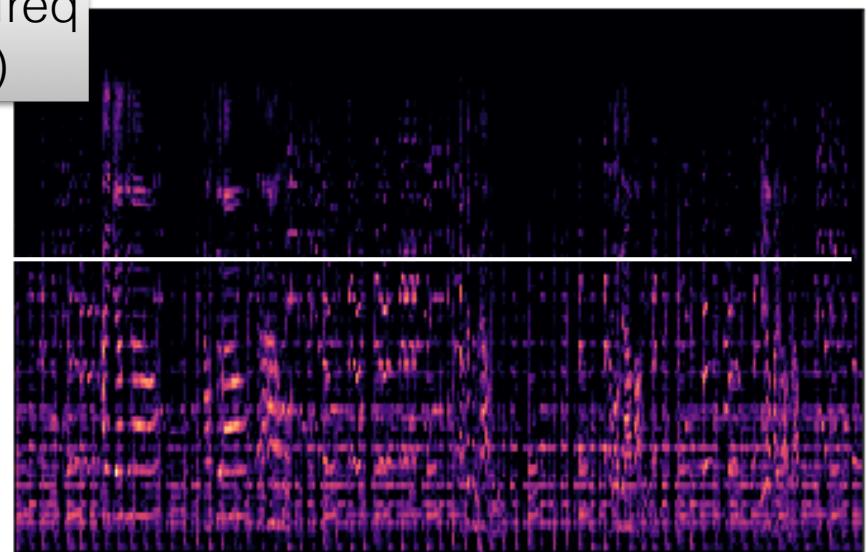
Key/Tonality

Genre 2

Genre 1



High freq/Low freq
(Brightness)



Spectral Centroid

Genre Classification

So what sort of features do i choose?

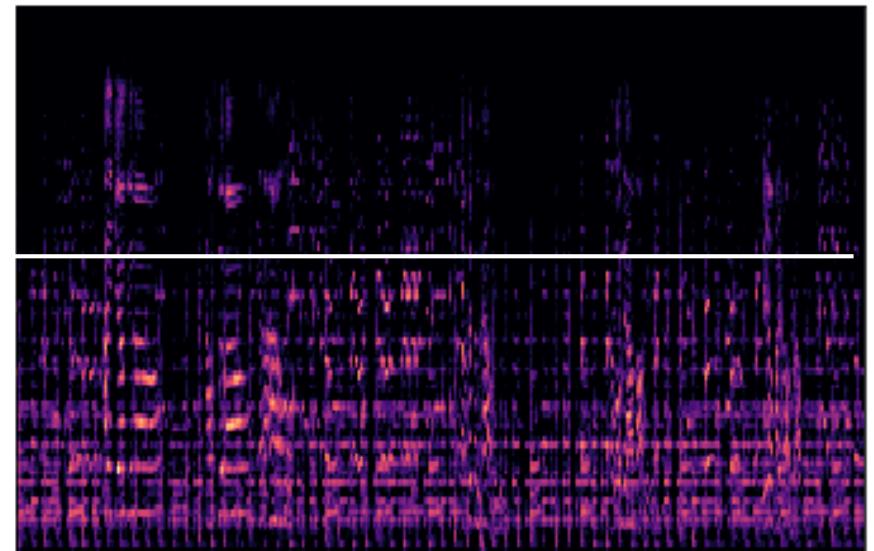
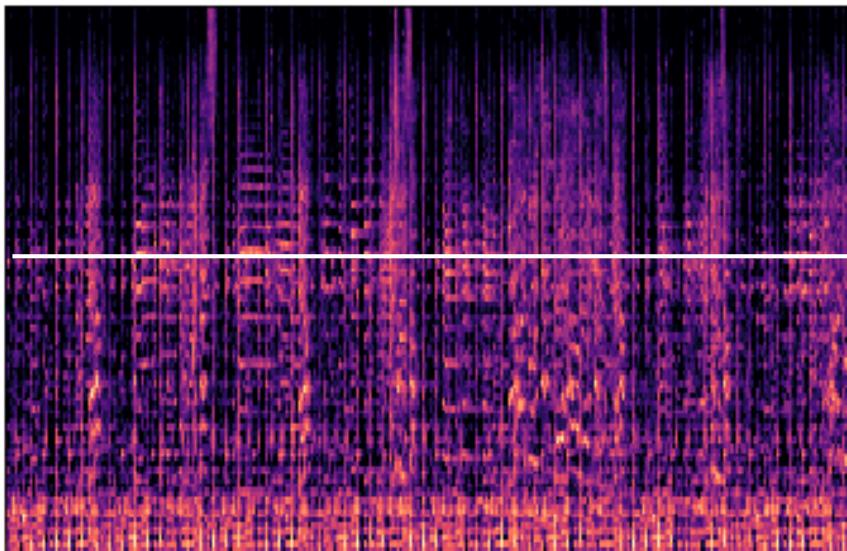
Timbre

Rhythm

Key/Tonality

Genre 2

Genre 1

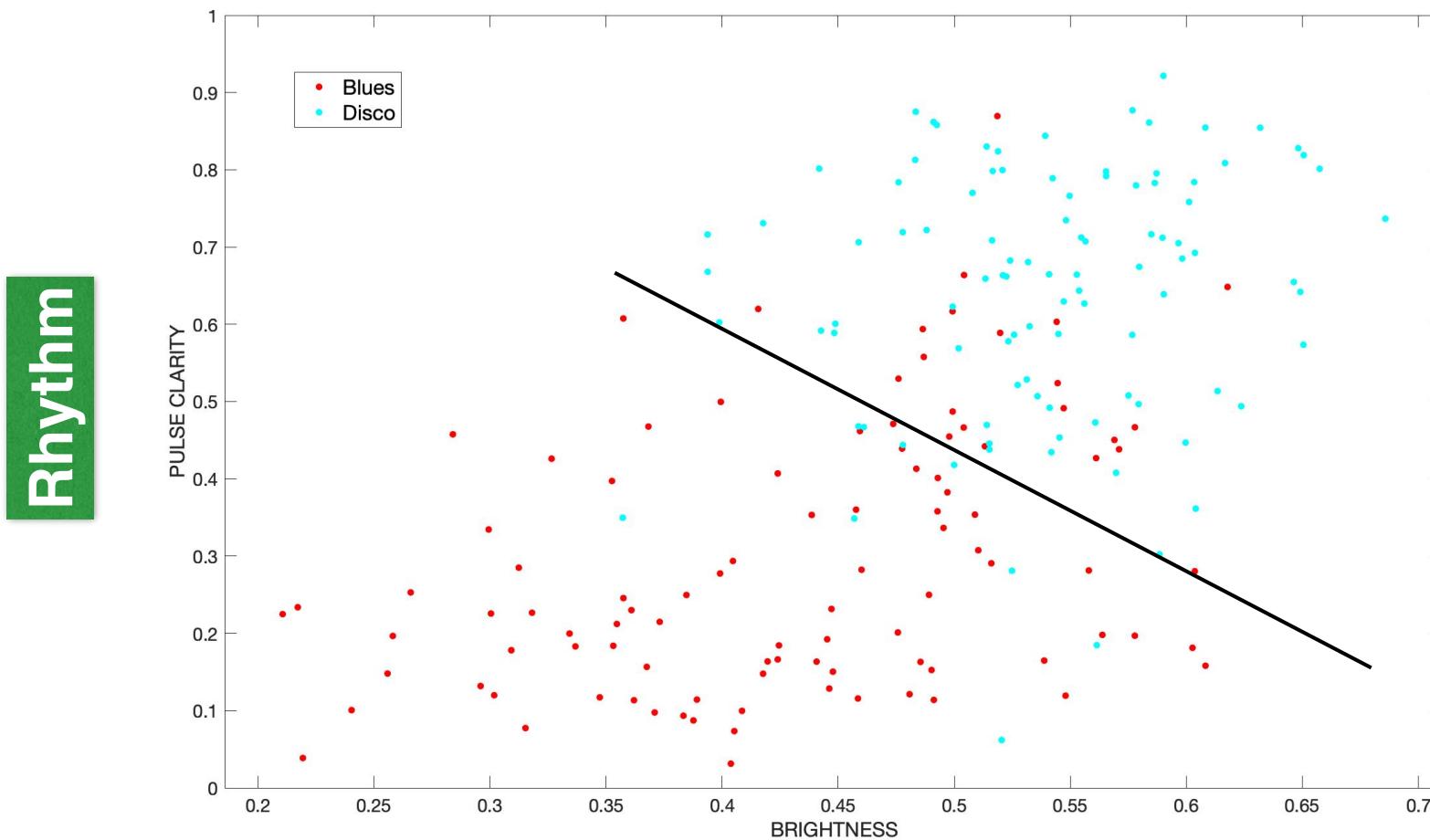


Tempo?

Pulse Clarity?

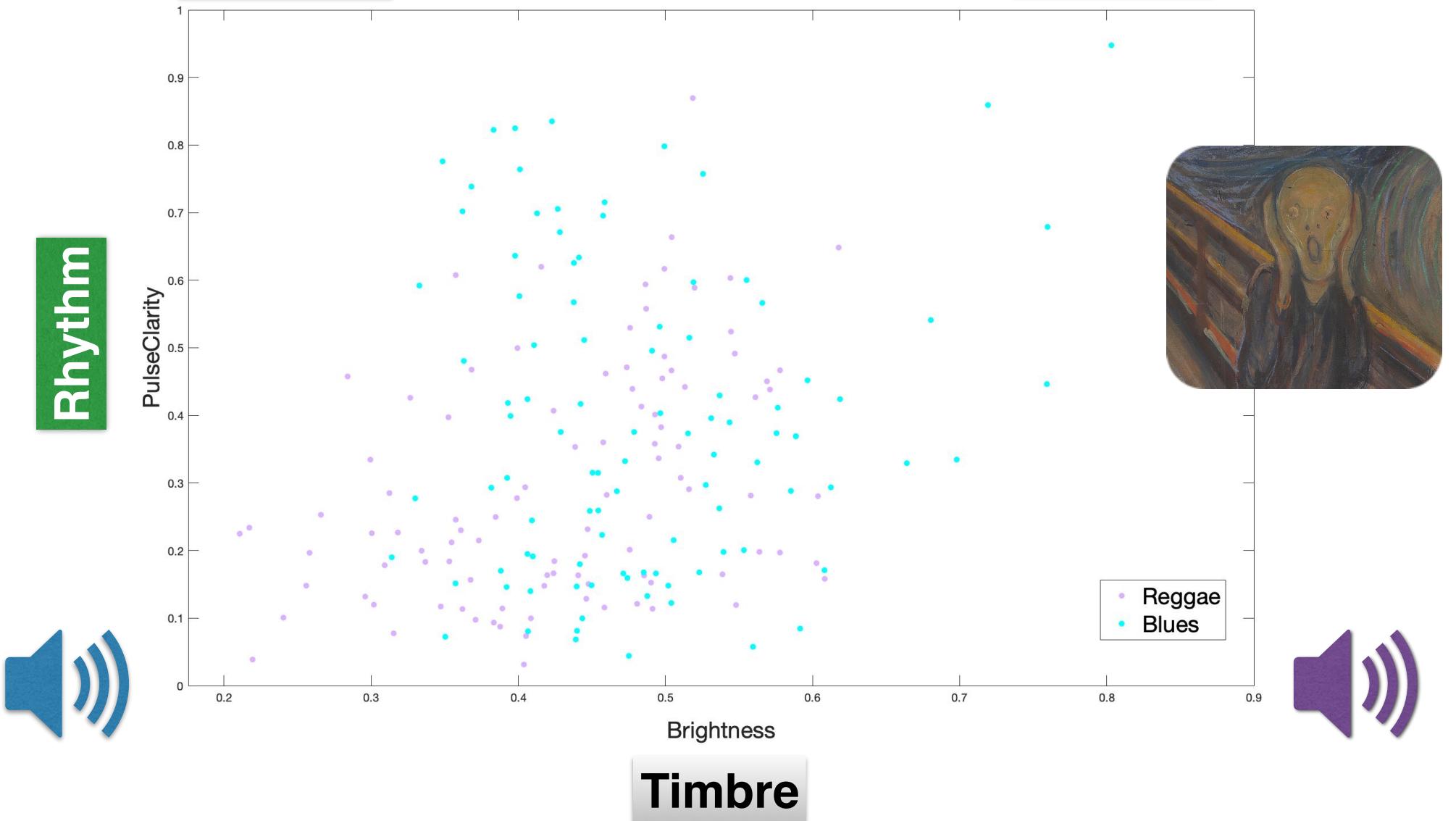
Genre Classification

Acoustic feature selection



Genre Classification

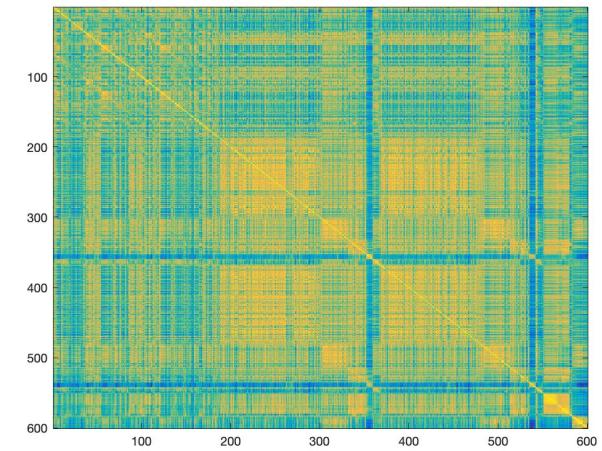
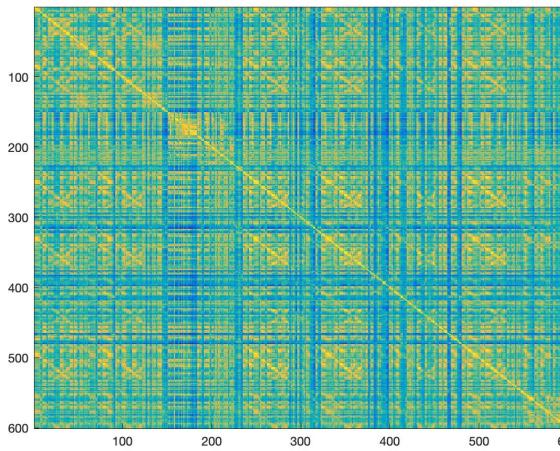
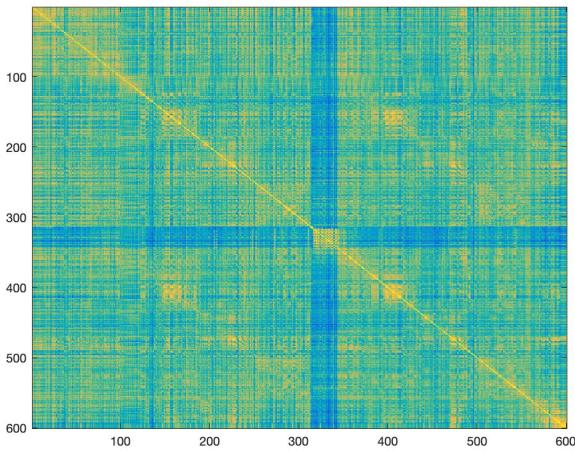
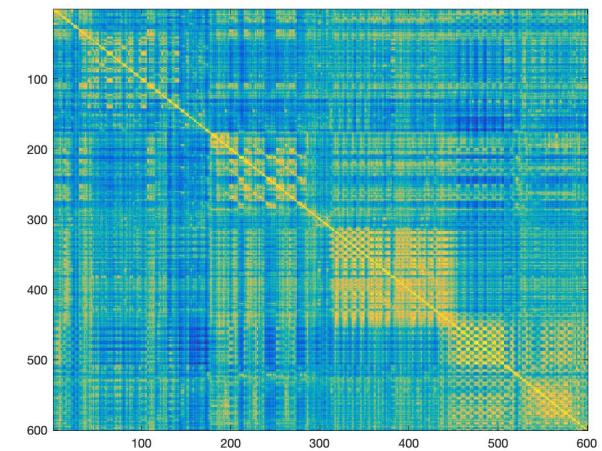
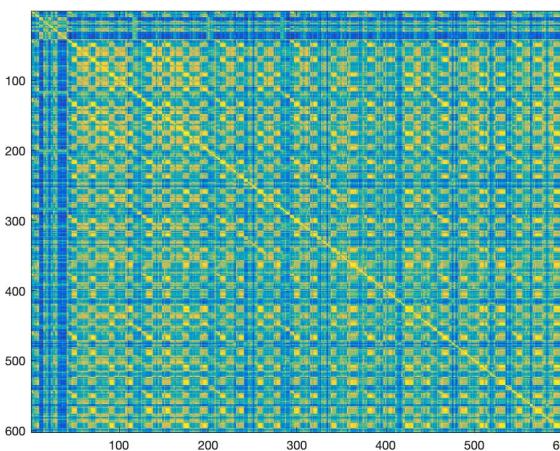
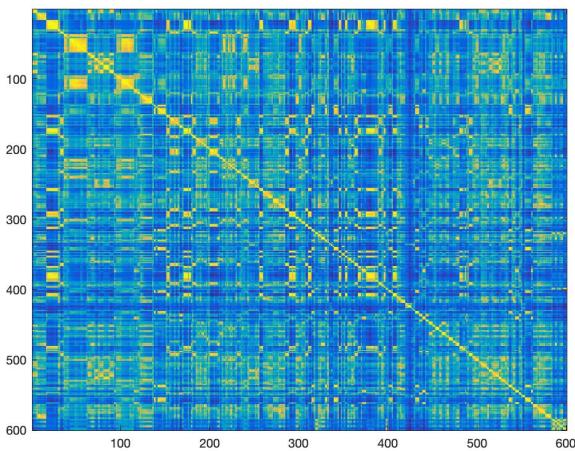
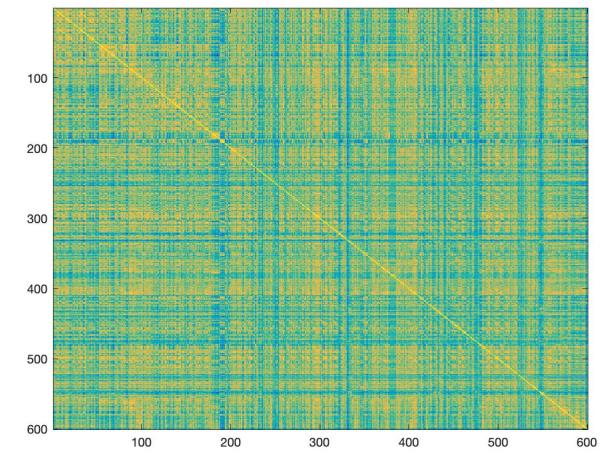
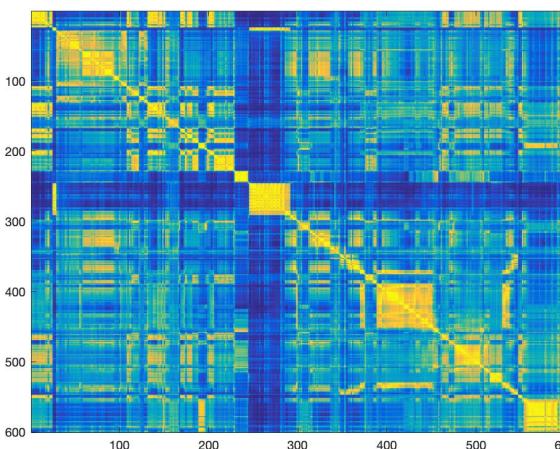
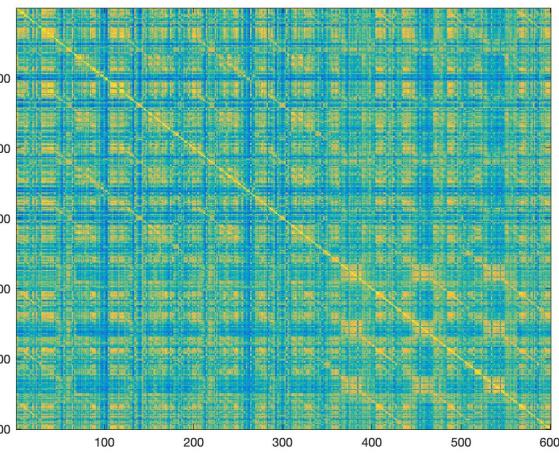
Genre 1 Acoustic feature selection **Genre 2**





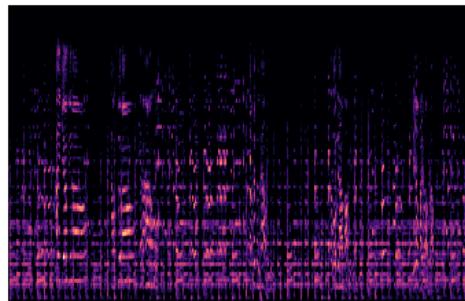
Visualizing Genres (Structure) guess the genre?



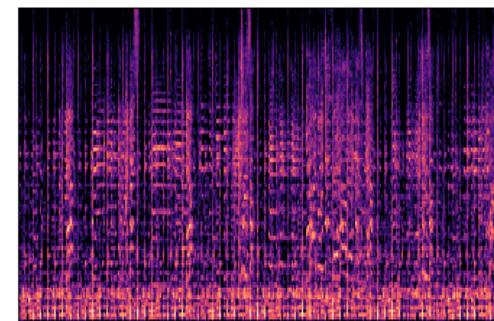


Genre Classification

can lyrics-based similarity matrices be used to identify genres ?



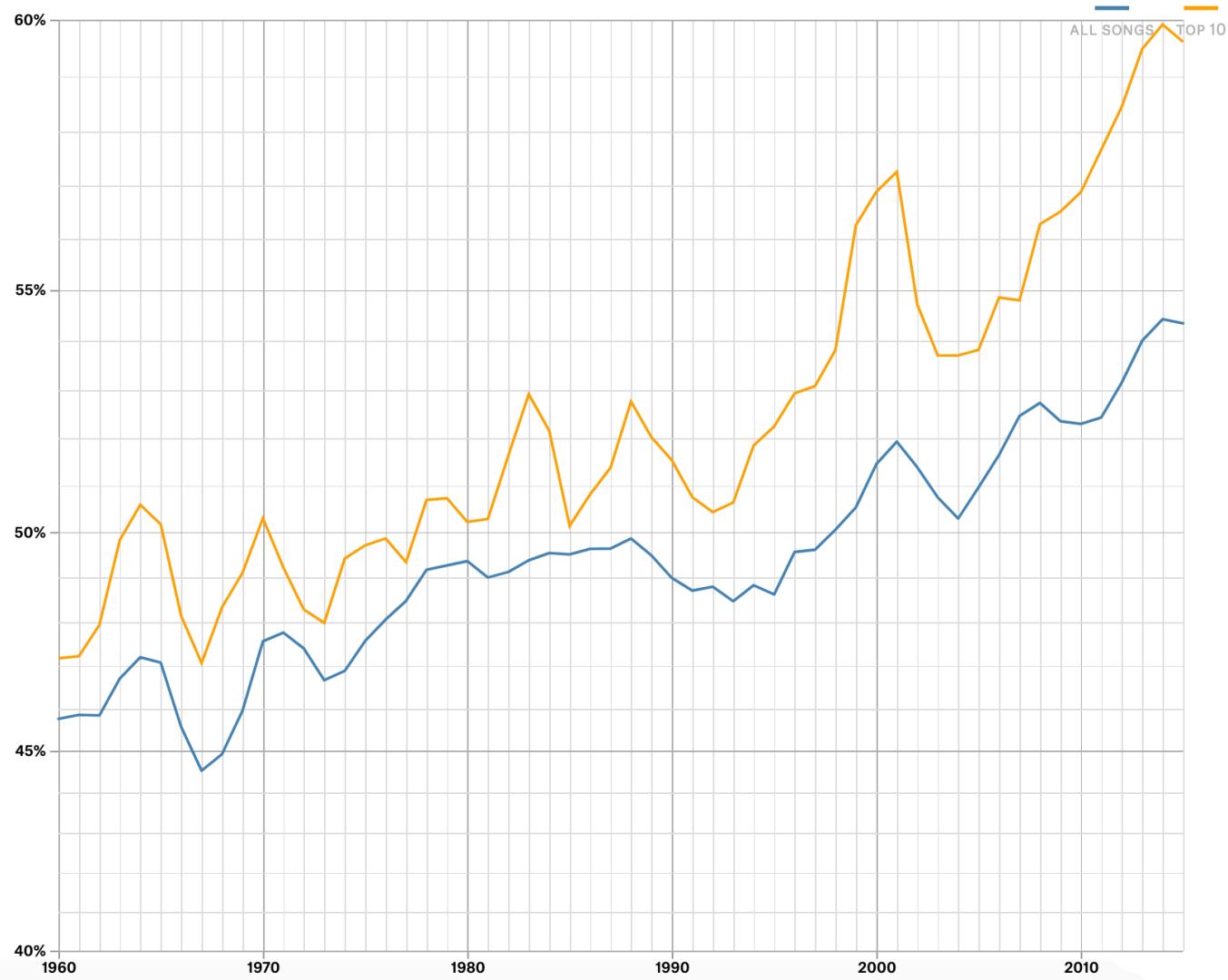
Genre 1



Genre 2

Repetition in Music

Repetition of Popular Music, by Year



Compressibility in Music

