Intelligent Audio Systems: A review of the foundations and applications of semantic audio analysis and music information retrieval

Jay LeBoeuf
Imagine Research
jay{at}imagine-research.com

July 2008

These lecture notes contain hyperlinks to the CCRMA Wiki.

On these pages, you can find additional supplement the lecture material found in the class - providing extra tutorials, support, references for further reading, or demonstration code snippets for those interested in a given topic .

Click on the  symbol on the lower-left corner of a slide to access additional resources.

WIKI REFERENCES...

Review from Day 2

- Name the 4 spectral moments
- What are the 3 major components of a MIR system?
- Why do we have to scale our extracted features?
- Which of these did we really not do at all in Lab 2? And, do you think this was a problem?

- How did the lab go?
- Let's dig into some interesting observations from the lab
- Did you try other audio files – other instrument recognizers?

(demo4tempo.pdf)

Feature extraction

- Feature design and creation uses one's domain knowledge.
- Choosing discriminating features is critical
- Smaller feature space yields smaller, simpler models, faster training, often less training data needed

Developing an innate understanding of the features

Visualizing Features

- 2-D visualization
- Matlab
- Weka's Visualization panes
- SOM, IsoMap, mapping multi-D down to 2D/3D

Listening to Features

- Play examples of Matt Hoffman's work (SoundLab)

Basic system overview

Segmentation

(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction

(Time-based, spectral energy, MFCC, etc)

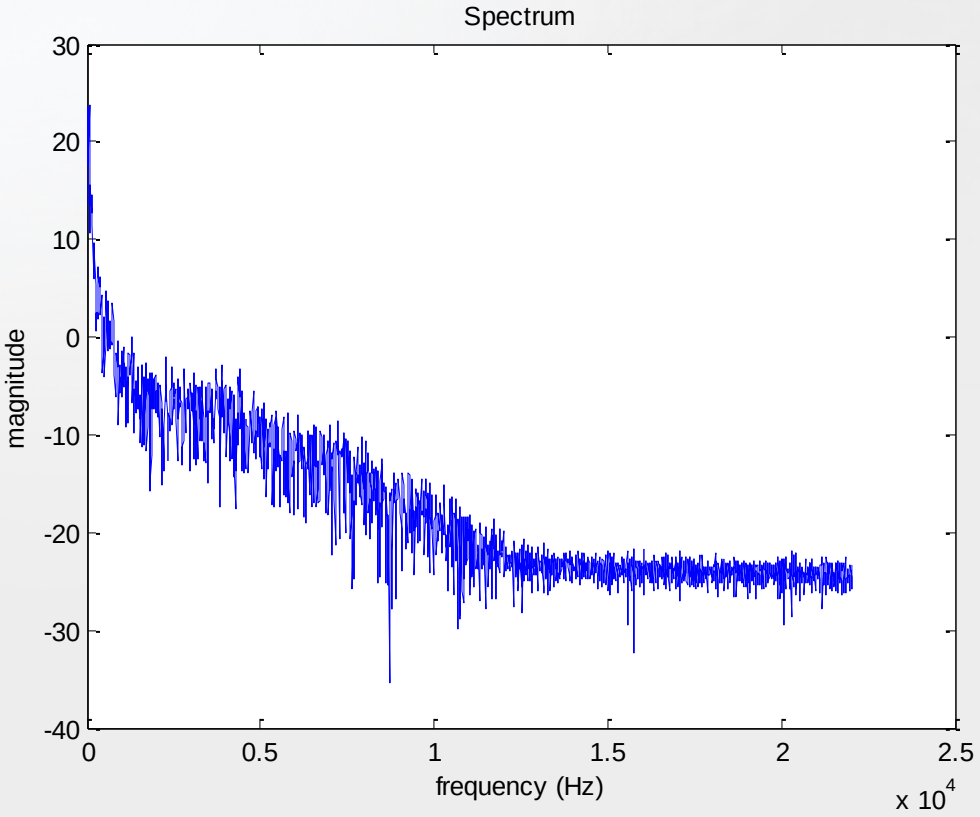
Analysis / Decision Making

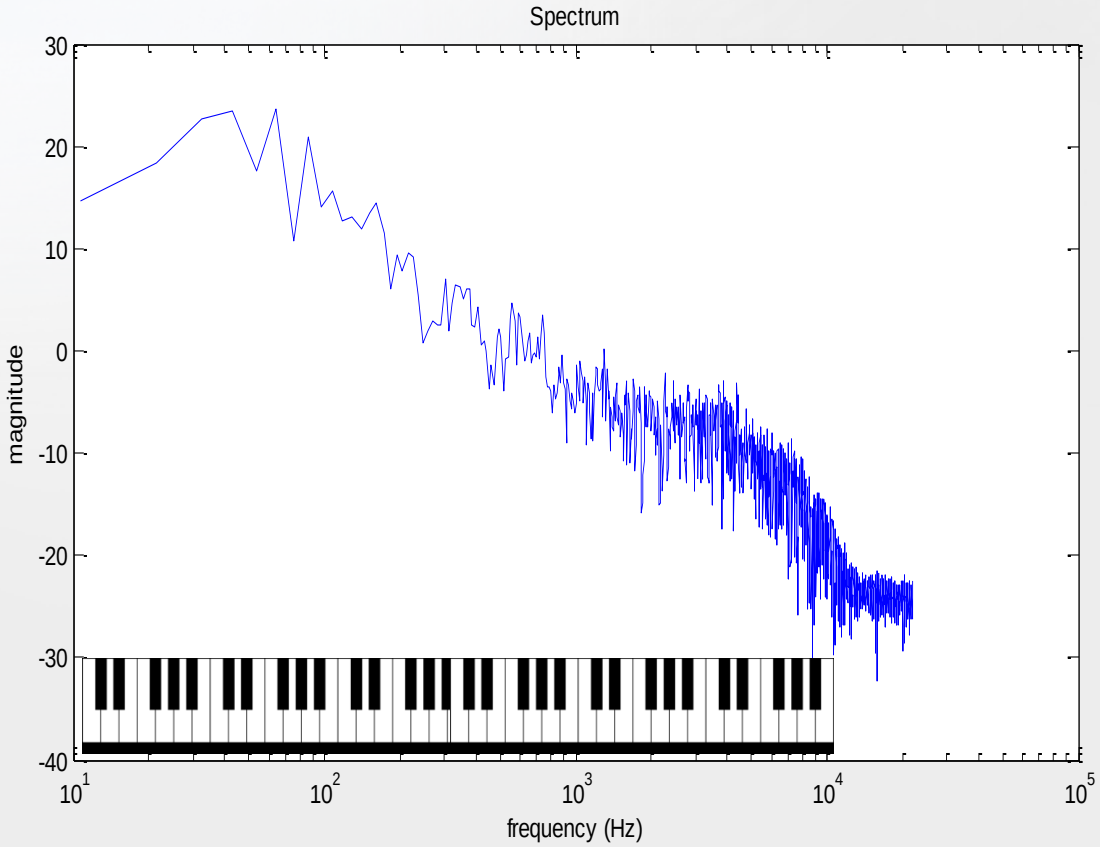
(Classification, Clustering, etc)

Spectral Features

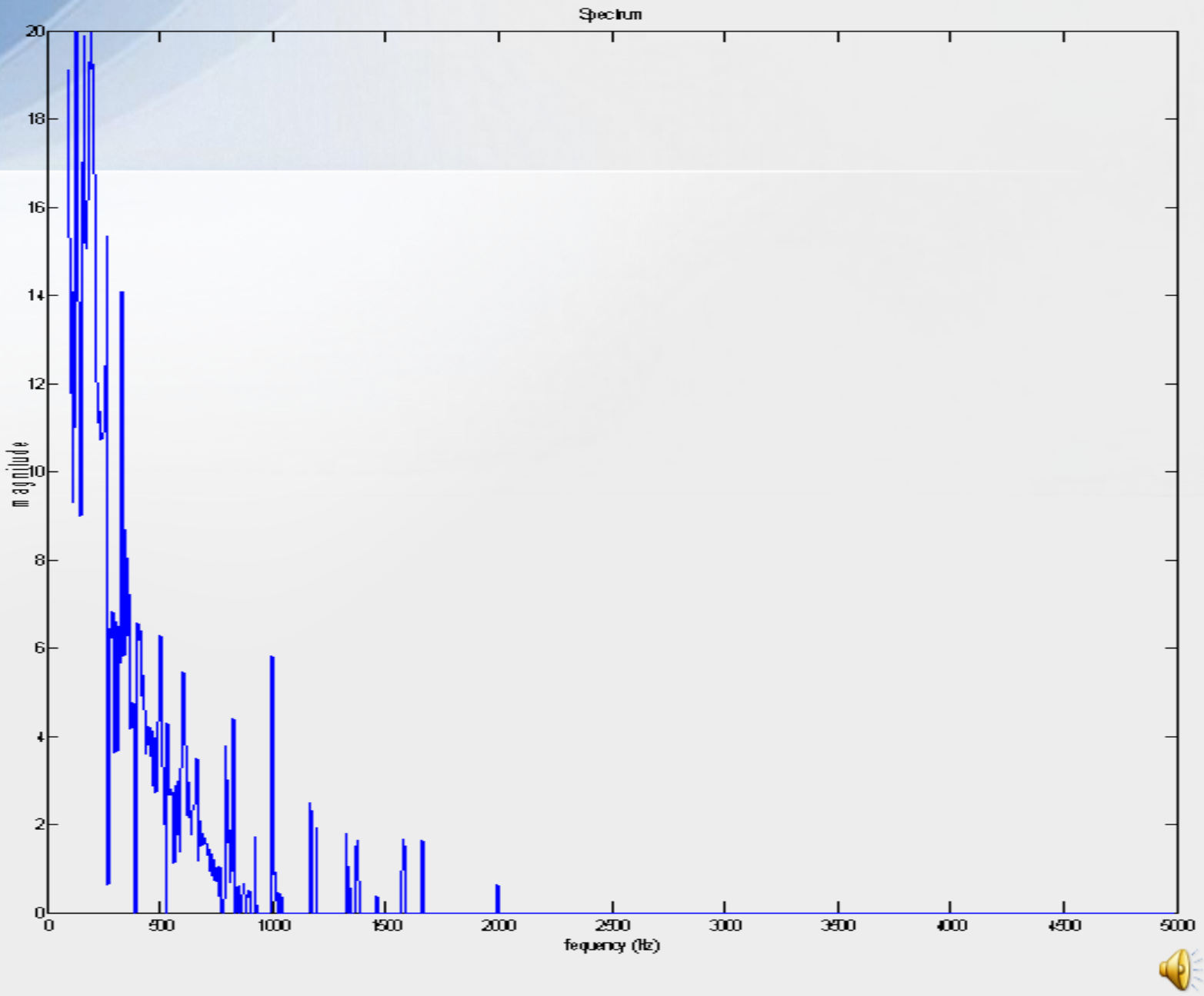
- Spectral Flatness Measure
- Spectral Crest Factor
- Spectral Flux

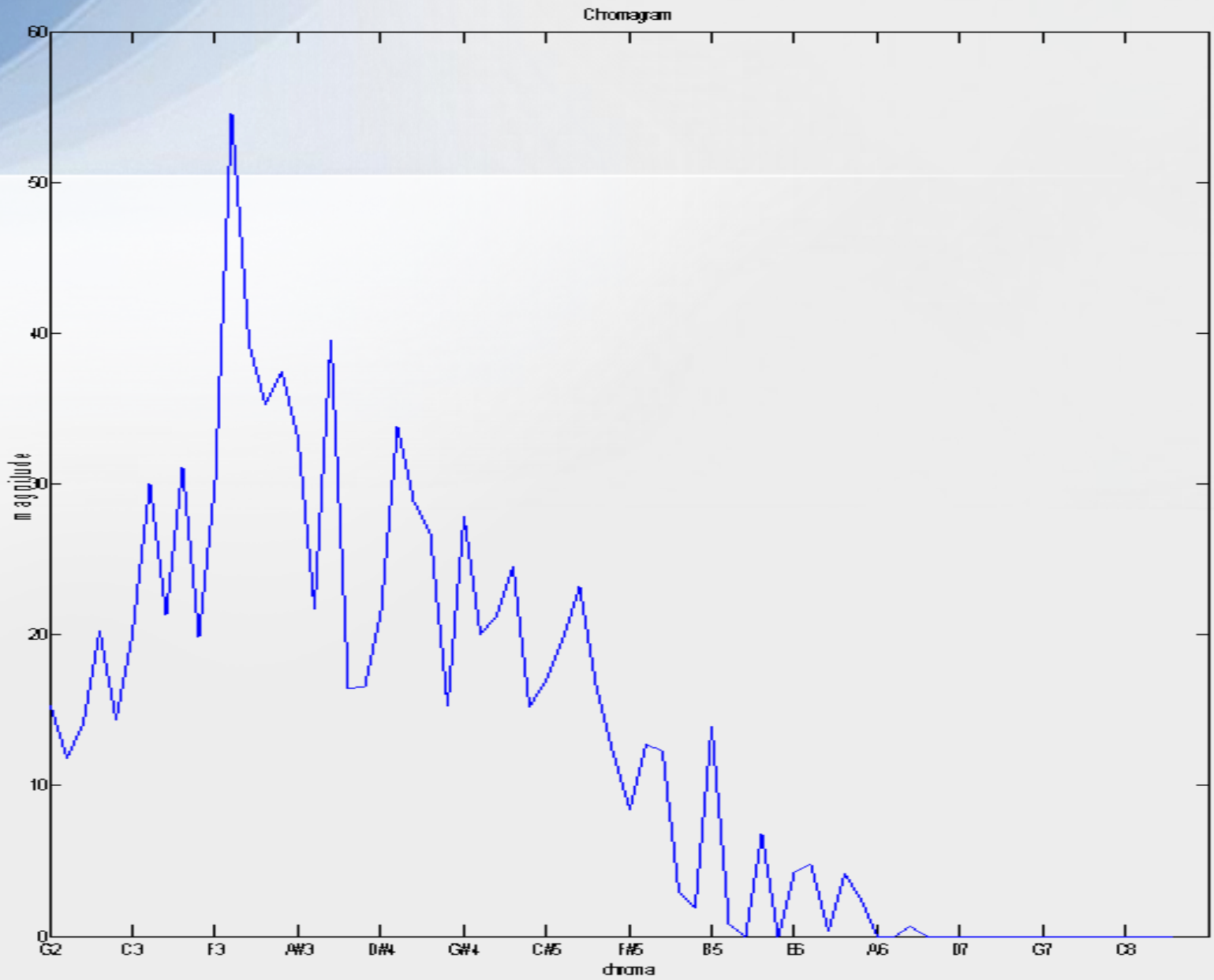
Log Spectrogram

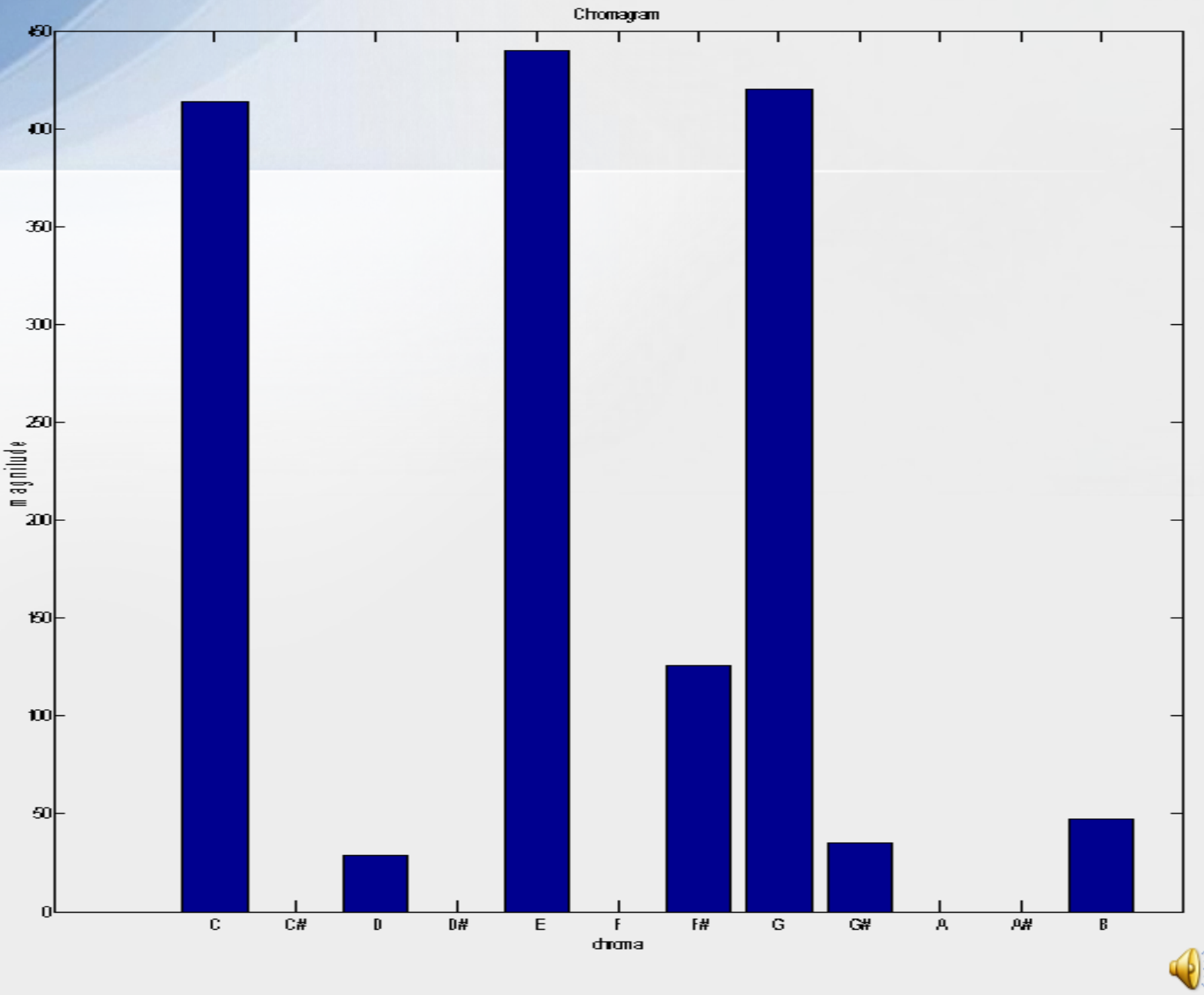


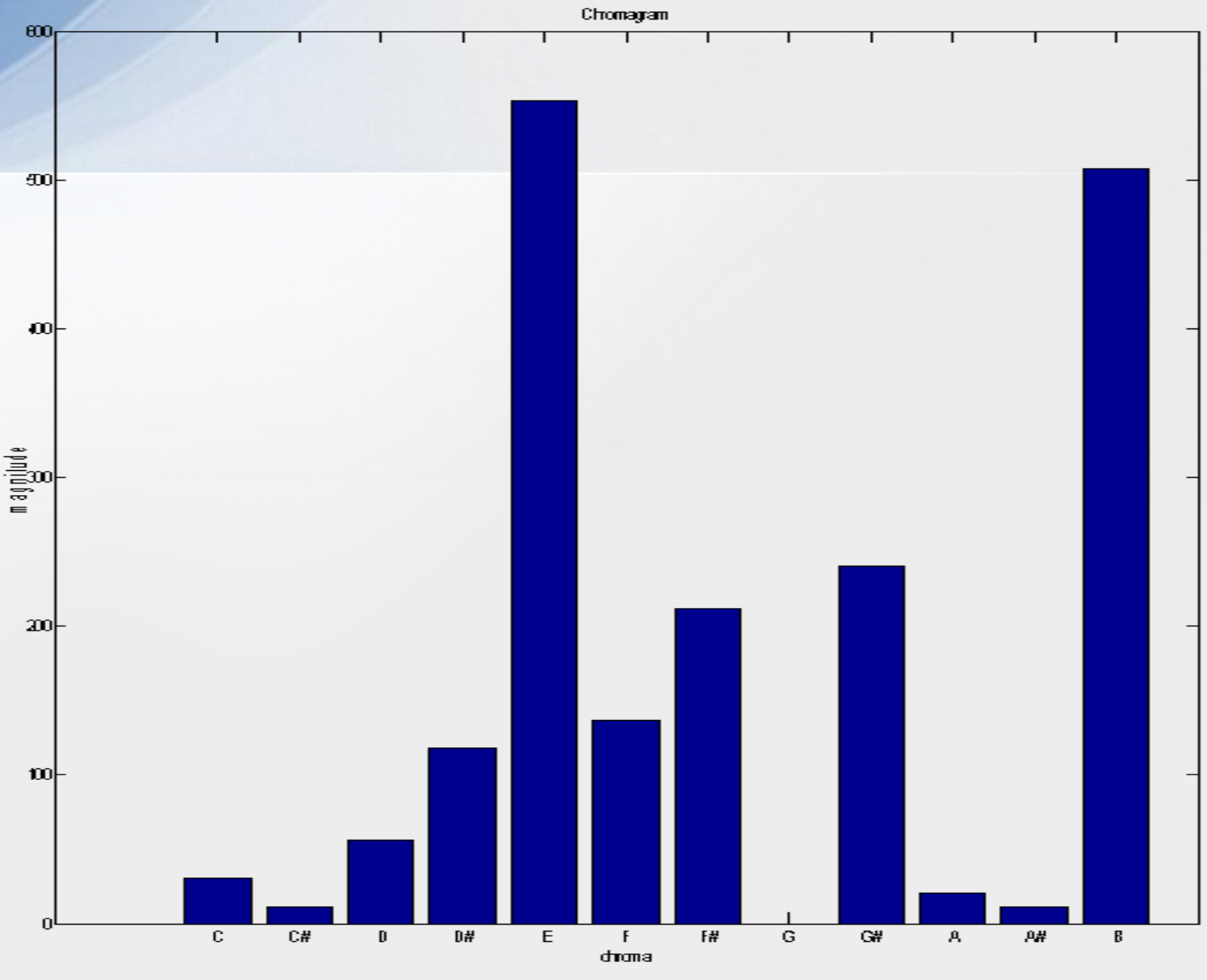

Spectrum
30
20
10
0
-10
-20
-30
-40
magnitude
10^1
10^2
10^3
10^4
10^5
frequency (Hz)

Chroma Bins

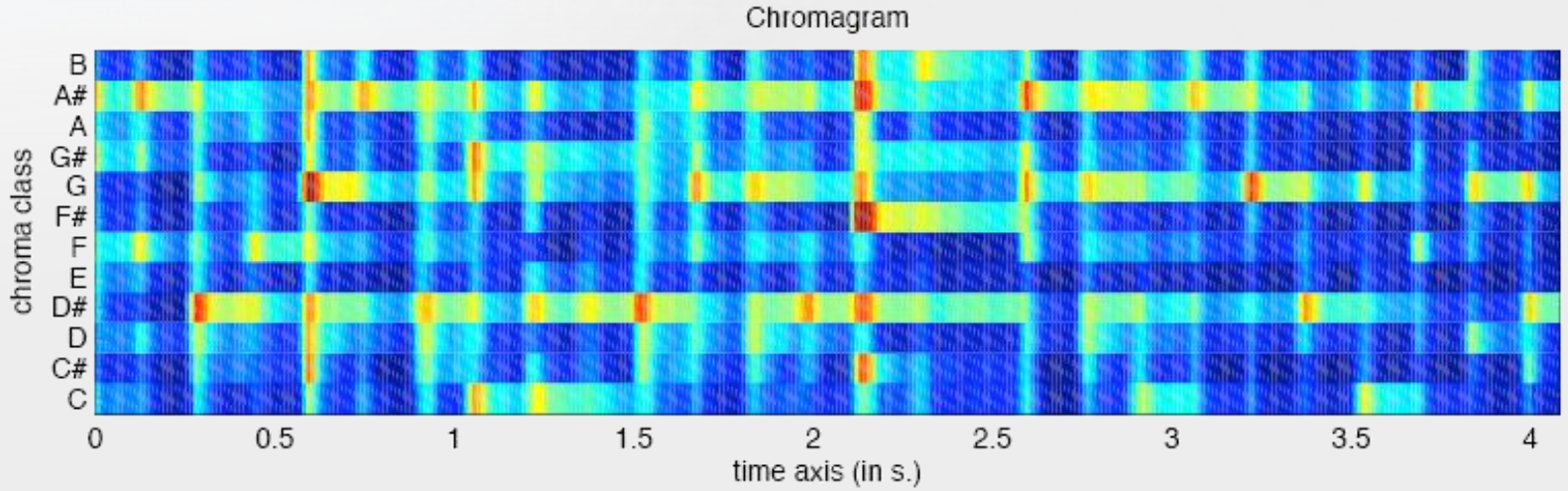

Spectrum
magnitude
fequency (Hz)
0
2
4
6
8
10
12
14
16
18
20
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000


Chromagram
magnitude
chroma
60
50
40
30
20
10
0
G2
C3
F3
A#3
D#4
G#4
C#5
F#5
B5
E6
A6
D7
G7
C8

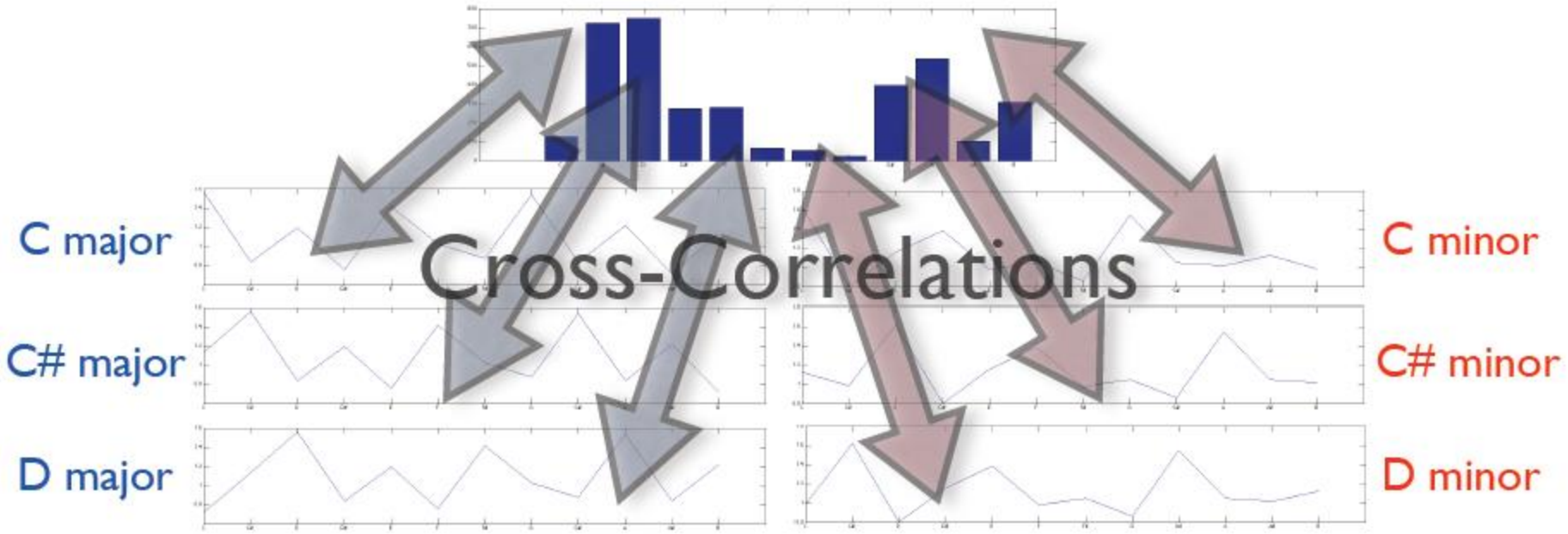

Chromagram
magnitude
chroma
0
50
100
150
200
250
300
350
400
450
C
C#
D
D#
E
F
F#
G
G#
A
A#
B


Chromagram
magnitude
chroma
0
100
200
300
400
500
600
C
C#
D
D#
E
F
F#
G
G#
A
A#
B

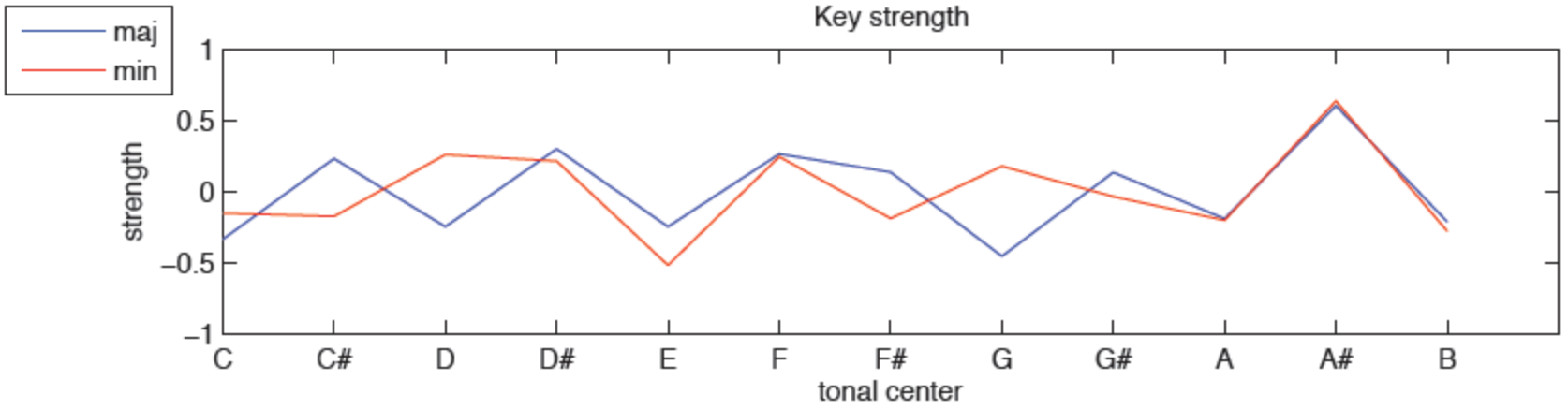
Example



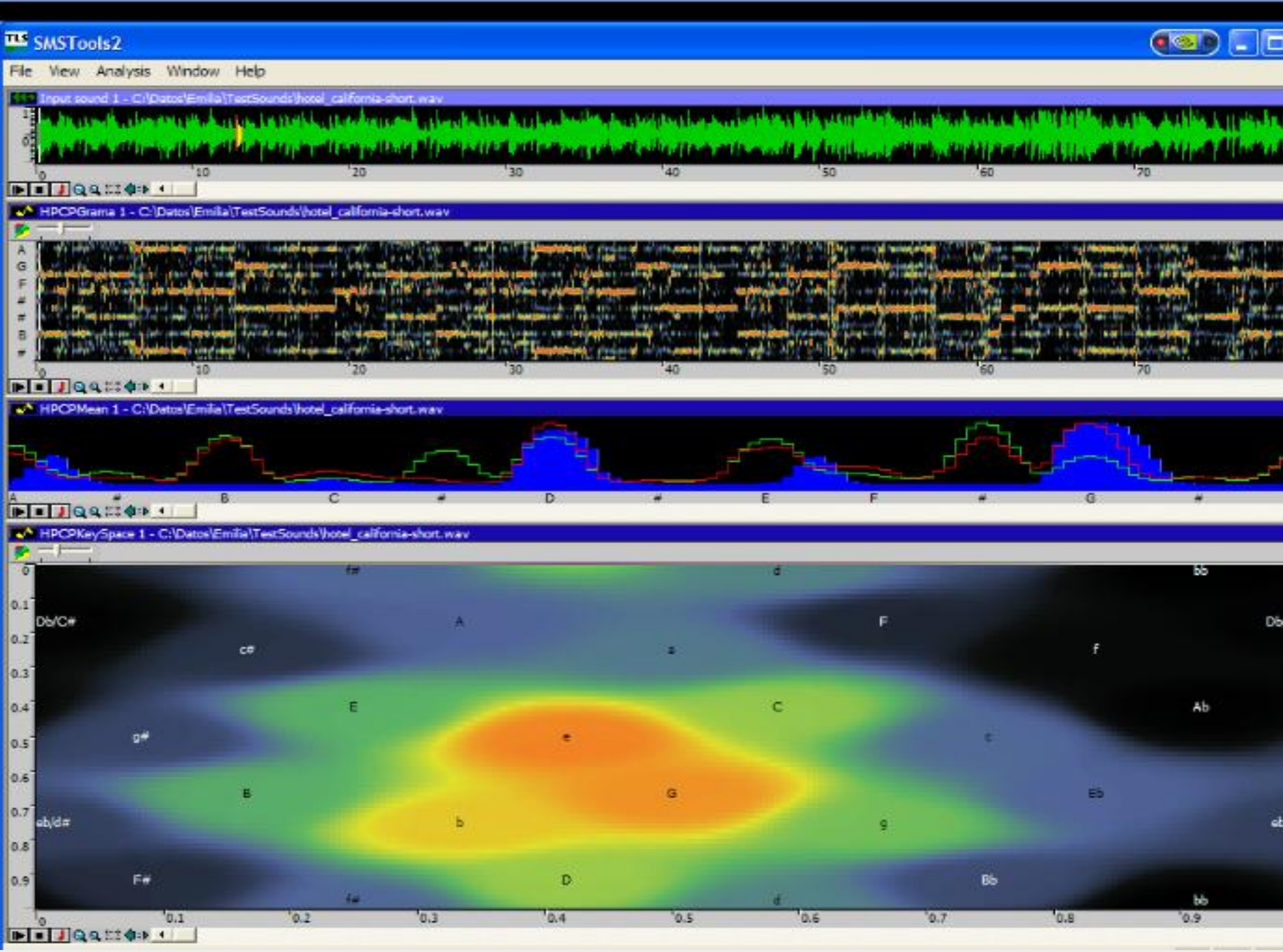
Picture courtesy: Olivier Lartillot

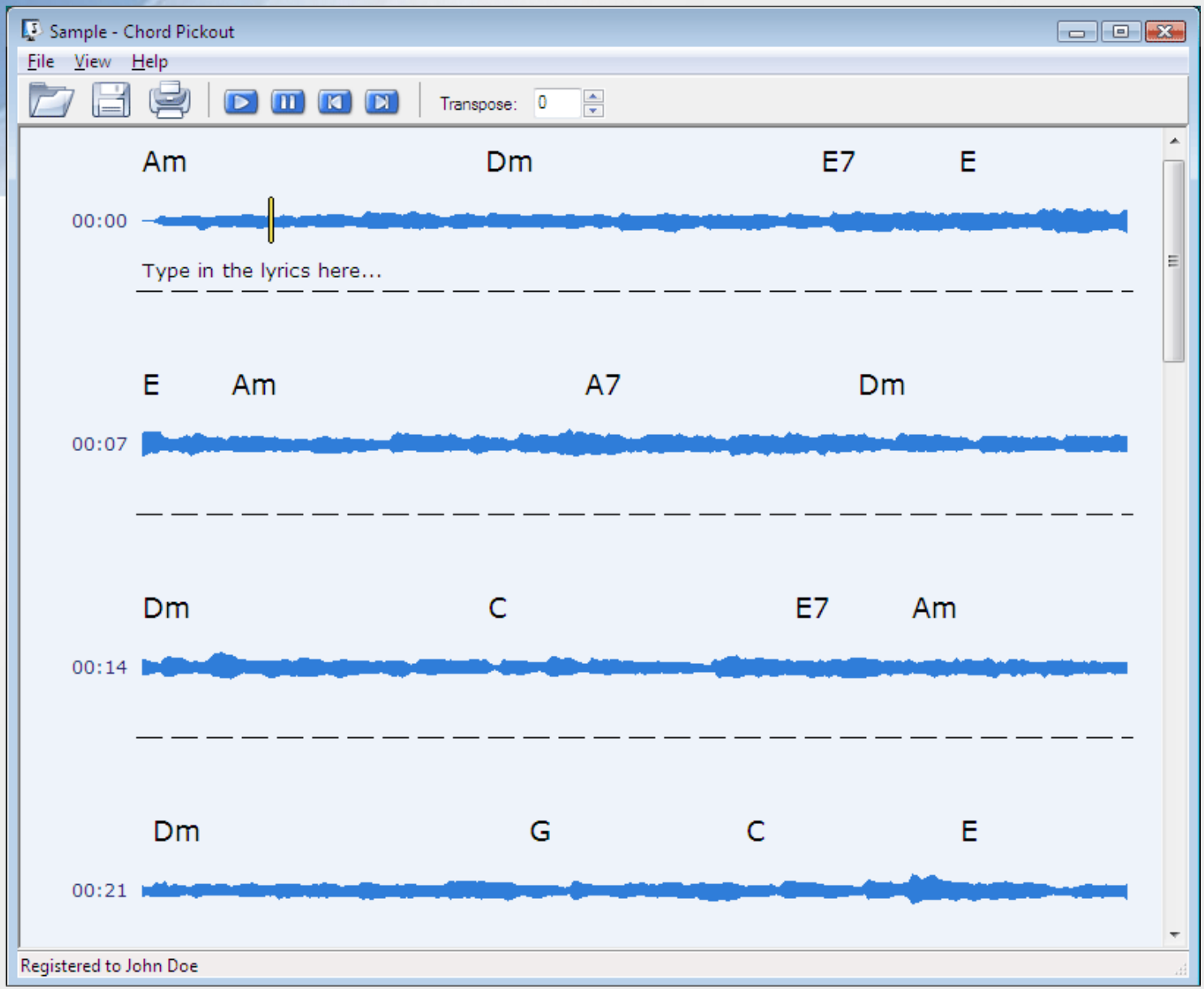


The resulting graph indicate the cross-correlation score for each different tonality candidate.



Picture courtesy: Olivier Lartillot


SMSTools2
File View Analysis Window Help
Input sound 1 - C:\Datos\Emilia\TestSounds\hotel_california-short.wav
HPCPGrama 1 - C:\Datos\Emilia\TestSounds\hotel_california-short.wav
HPCPMean 1 - C:\Datos\Emilia\TestSounds\hotel_california-short.wav
HPCPKeySpace 1 - C:\Datos\Emilia\TestSounds\hotel_california-short.wav
Db/C#
c#
A
E
B
F#
F
f
C
c
G
g
D
d
Ab
Eb
Bb
bb
eb/d#



- http://www.chordpickout.com/index.html

Real-world

- Test on MIDI files first
- For analyzed audio:
  - 91.7% success for 48 Bach preludes
    - 24 major and minor keys
    - Krumhansl-Schmuckler
  - Kyogu Lee: ~84% for 1 of Beatles CD

Final projects

Explain your ideas and implementation
Share of ideas and knowledge.
Define specific, concrete project with measurable goals.

Brainstorm an approach
Research previous approaches
Define your approach (segmentation, features, classifiers, etc)

Provide demonstrations and sound examples.
Define clear goals and evaluation metrics.
Implement

Play with it and evaluate
Make improvements
How can you improve?

Final Projects (2)

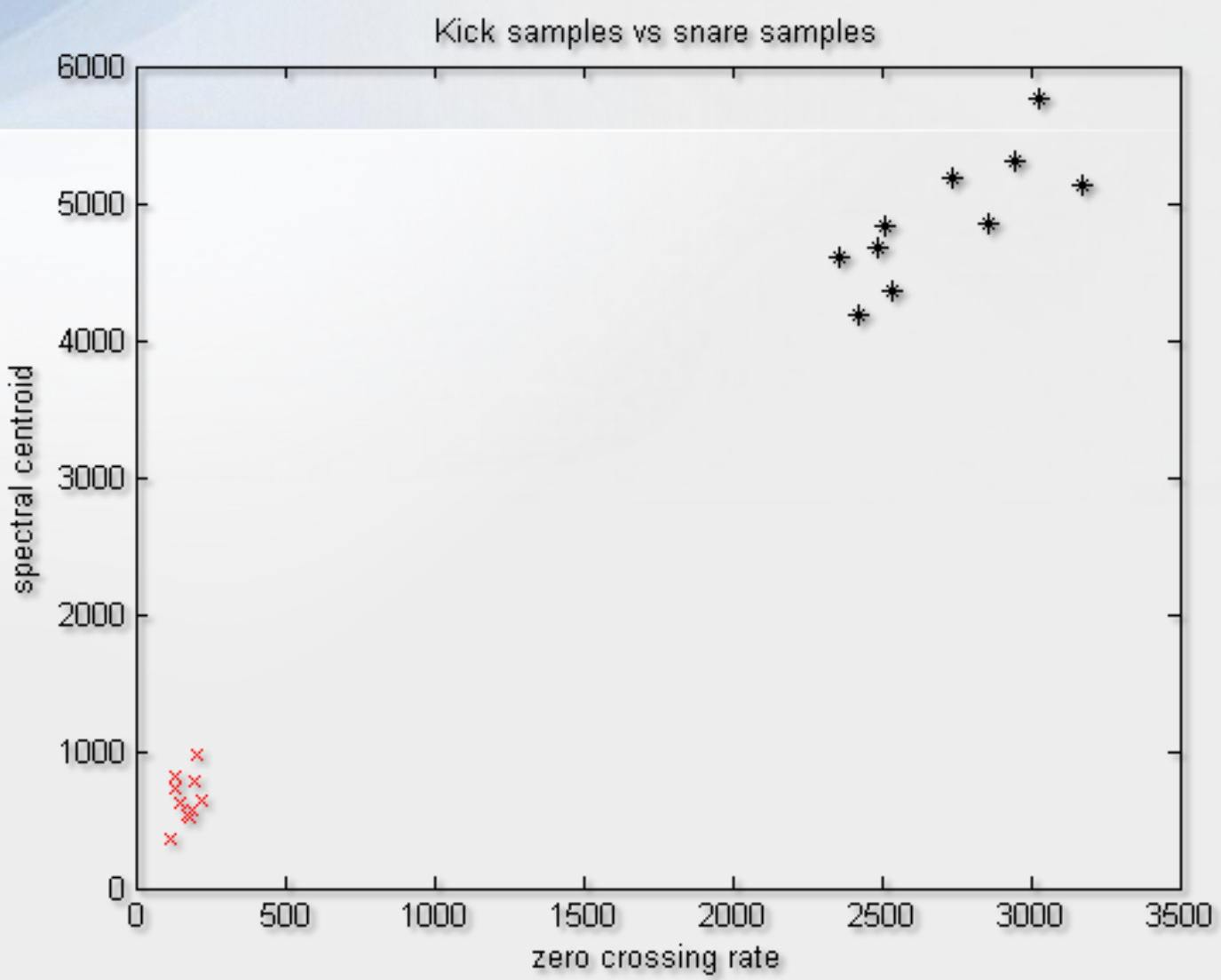
Examples:
https://cm-wiki.stanford.edu/wiki/MIR_workshop_2008#Final_Projects_from_MIR_2008_Workshop

http://en.wikipedia.org/wiki/List_of_Music_Genome_Project_attributes_by_type

Features: Measuring changes

- Δ and Δ Δ
  - Change between frames
  - How quickly the change is occurring

- Spectral flux is the distance between the spectrum of successive frames


Kick samples vs snare samples
spectral centroid
zero crossing rate
0
1000
2000
3000
4000
5000
6000
0
500
1000
1500
2000
2500
3000
3500

X: Cluster (Nom)

Y: spectral_centroidmean (Num)

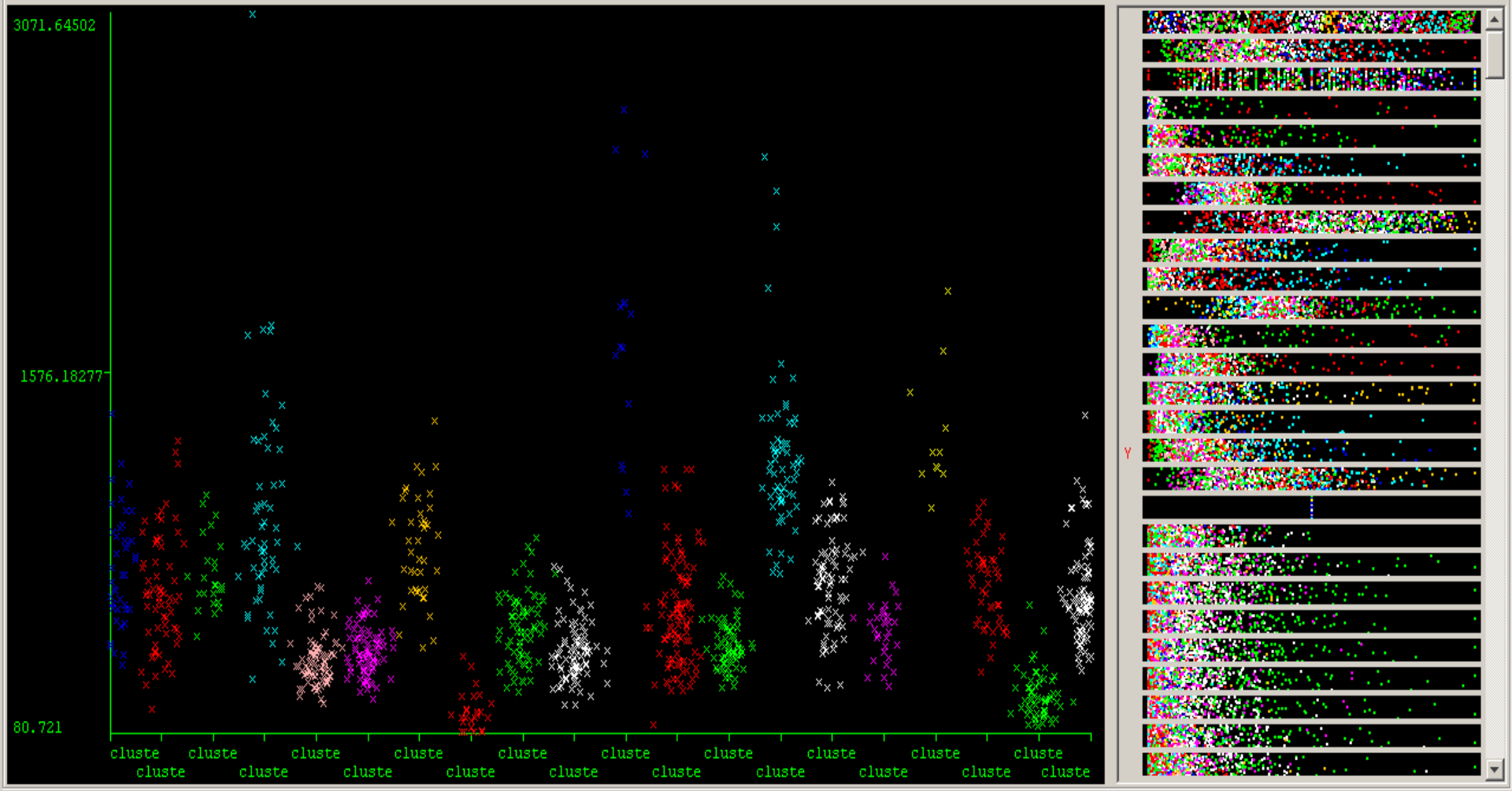
Colour: Cluster (Nom)

Select Instance

Reset | Clear | Save

Jitter

Plot: phonondips_clustered



Class colour

cluster0 cluster1 cluster2 cluster3 cluster4 cluster5 cluster6 cluster7 cluster8 cluster9

cluster10 cluster11 cluster12 cluster13 cluster14 cluster15 cluster16 cluster17 cluster18 cluster19

Day 3 Lab

- Get your Lab 2 working
  - Make sure that training data = 100% accurate
  - Try the test snares and test kicks
    - Write down your accuracy and parameters
    - Change the number of features
    - Add or replace current features with different values
      - (e.g., mirbrightness, mirrolloff)

- Demo - tonality
- Demo – tempo