DAY 1

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

Kyogu Lee
Gracenote
Kglee{at}ccrma.stanford.edu

June 2009

These lecture notes contain hyperlinks to the CCRMA Wiki.

On these pages, you can find supplemental material for lectures - 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…

Administration

- https://cm-wiki.stanford.edu/wiki/MIR_workshop_2009
- Daily schedule
- Introductions
  - Our background
  - A little about yourself
  - List of your region of interest, and any specific items of interest that you'd like to see covered.

Example Seed…

Why MIR?

- Find specific item
- Find something vague
- Find something interesting or new

Commercial Applications

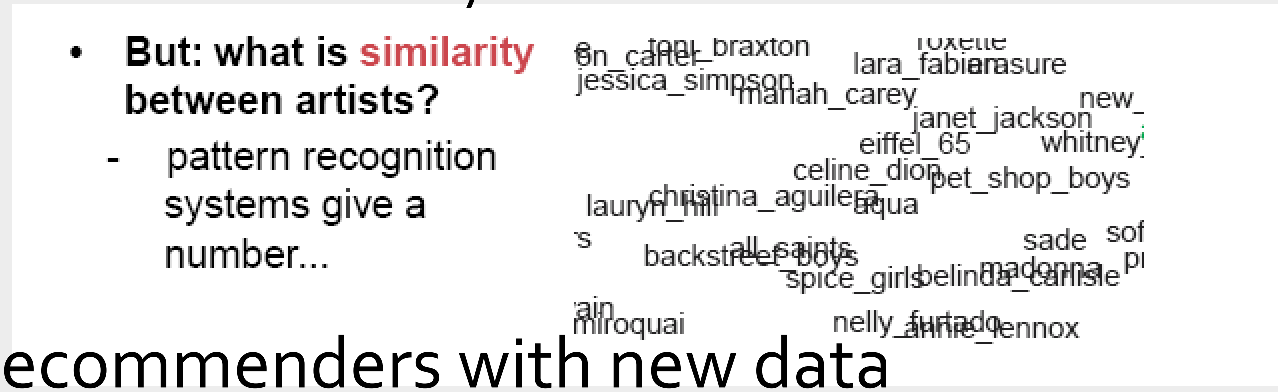
- Retrieval based on similarity (IR and creative applications)
- Live analysis of audio
- Music Discovery / Recommendation
- Query for music
- Assisted Music Transcription
- Audio fingerprint
- Creative applications

Queries

- Query by Humming
  - Lots of academic work
- Query by audio ID
  - Gracenote ID, Shazam, Audible Magic
  - Noisy audio snippet
- Query by example
  - Find more like this (where “this” has to be specified or inferred)

Current “Hot” research areas

- Analysis of commercial music tracks, such as:
- Genre ID (labels exist, but even humans disagree!)
- Artist classification
  - Tricks: use voice only to improve accuracy to 70% (out of 100 artists)
- Artist similarity
  - Really, what is the similarity?



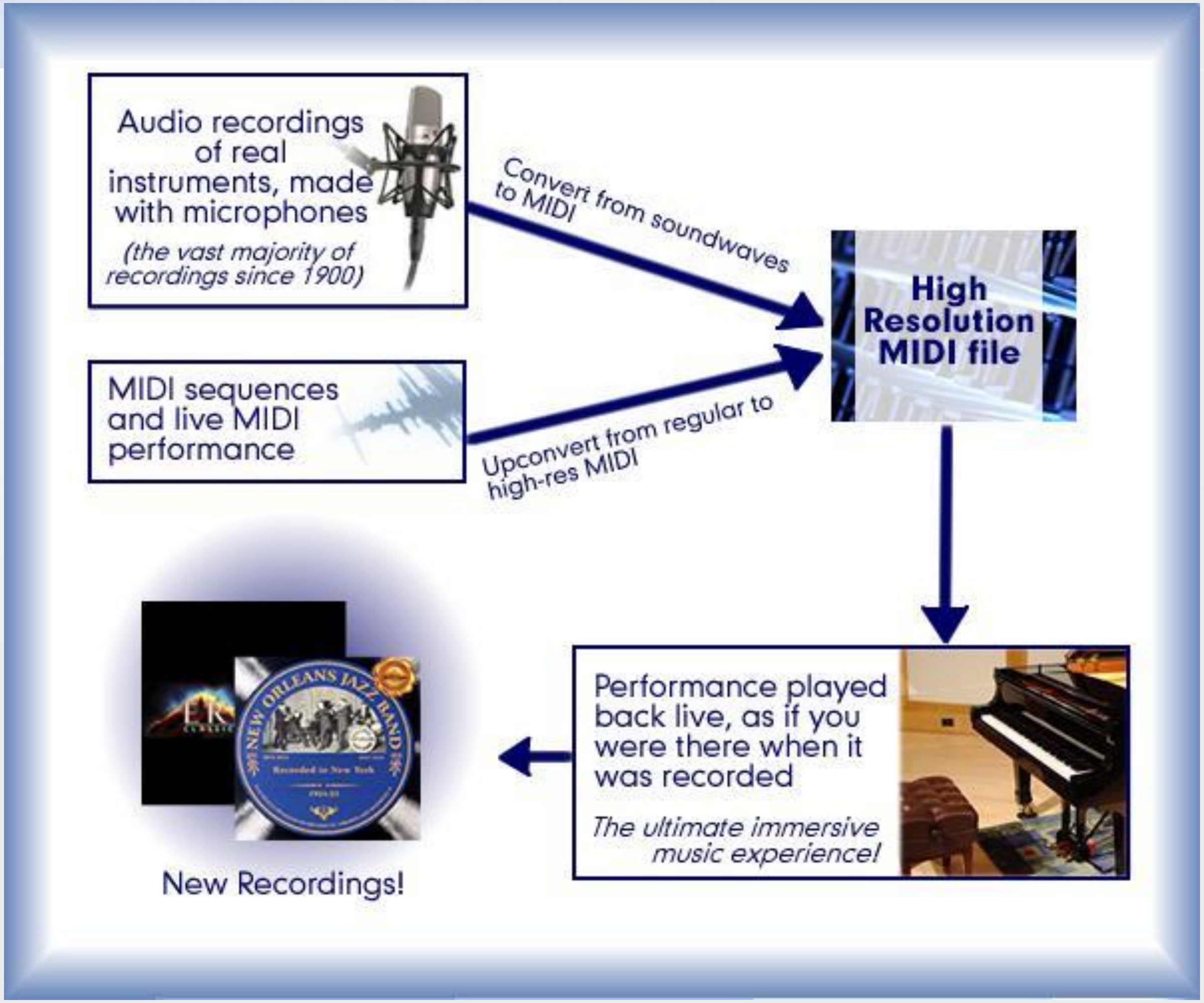
- But: what is **similarity** between artists?
  - pattern recognition systems give a number...

toni_braxton
on_carter
jessica_simpson
mariah_carey
roxette
lara_fabian
erasure
new_
janet_jackson
eiffel_65
whitney
celine_dion
pet_shop_boys
lauryn_hill
christina_aguilera
aqua
sade
sof
all_saints
backstreet_boys
spice_girls
belinda_carlisle
madonna
miroquai
nelly_furtado
annie_lennox

- Augment recommenders with new data

Motivations / Demos

- Transcriptionist vs. Descriptionist approach
  - Music Transcription (restoration) – piano from MIDI



Audio recordings of real instruments, made with microphones

(the vast majority of recordings since 1900)

Convert from soundwaves to MIDI

MIDI sequences and live MIDI performance

Upconvert from regular to high-res MIDI

High Resolution MIDI file

Performance played back live, as if you were there when it was recorded

The ultimate immersive music experience!

New Recordings!

Motivations / Demos

- Transcriptionist vs. Descriptionist approach

  - Music Transcription (restoration) – piano from MIDI
    - http://zenph.com/listen.html
    - More info:
      - http://www.pragprog.com/articles/a-pragmatic-project-live-in-concert/the-methodology

Motivations / Demos

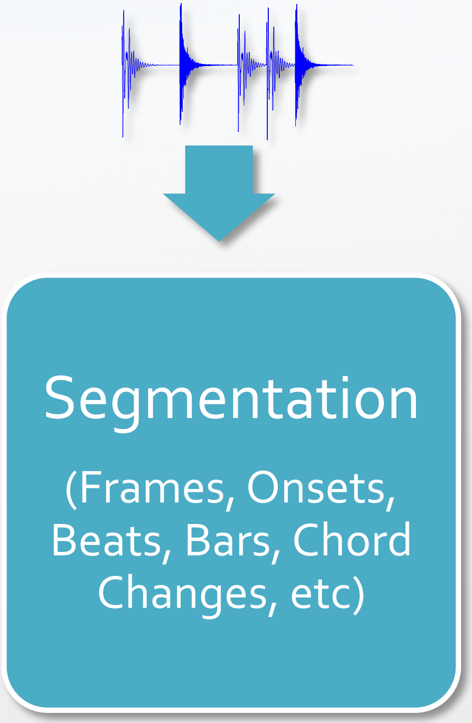
- Transcriptionist vs. Descriptionist approach

  – Music Transcription (restoration) – piano company from MIDI
    - http://zenph.com/listen.html

  - More transcription (drum transcription demo)

BASIC SYSTEM OVERVIEW

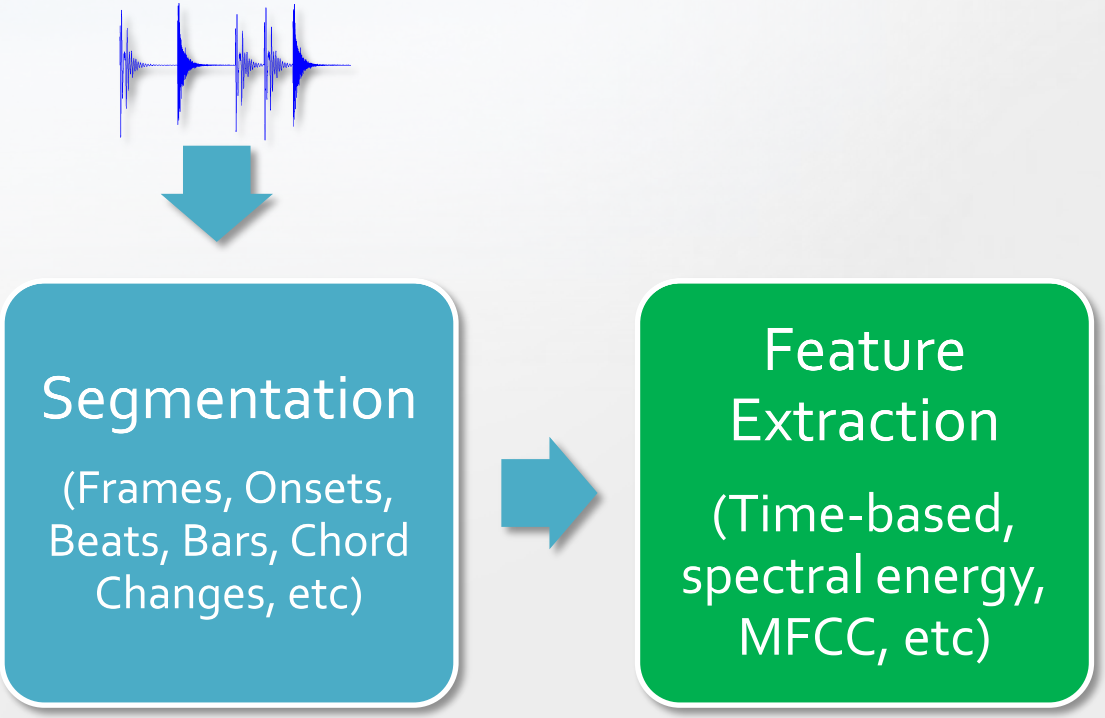
Basic system overview



Segmentation

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

Basic system overview



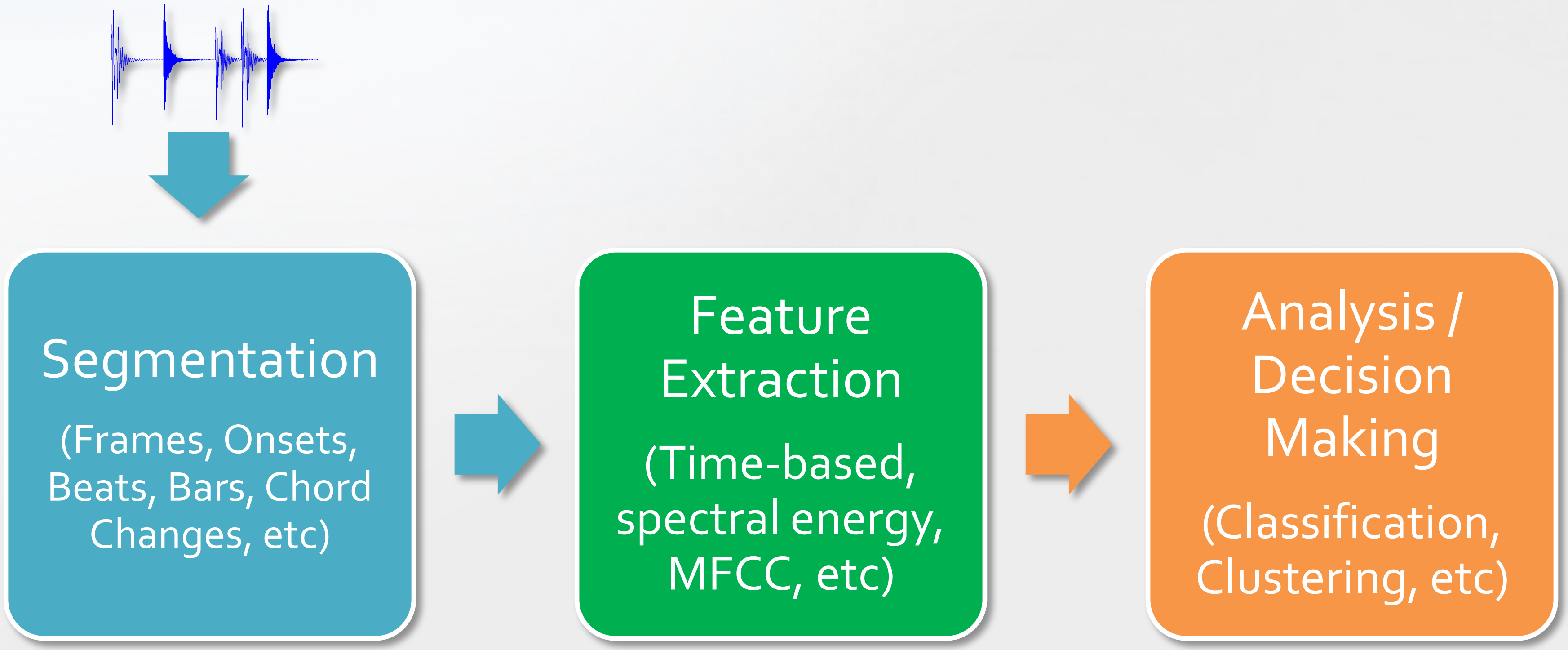
Segmentation

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

Feature Extraction

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

Basic system overview



Segmentation

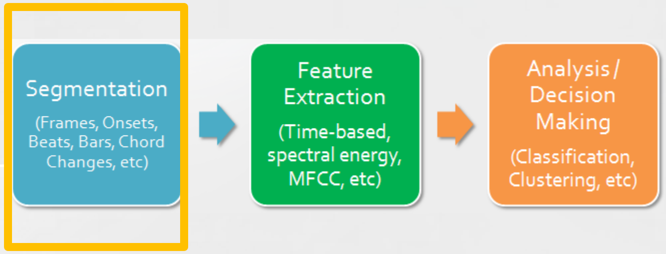
(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction

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

Analysis / Decision Making

(Classification, Clustering, etc)



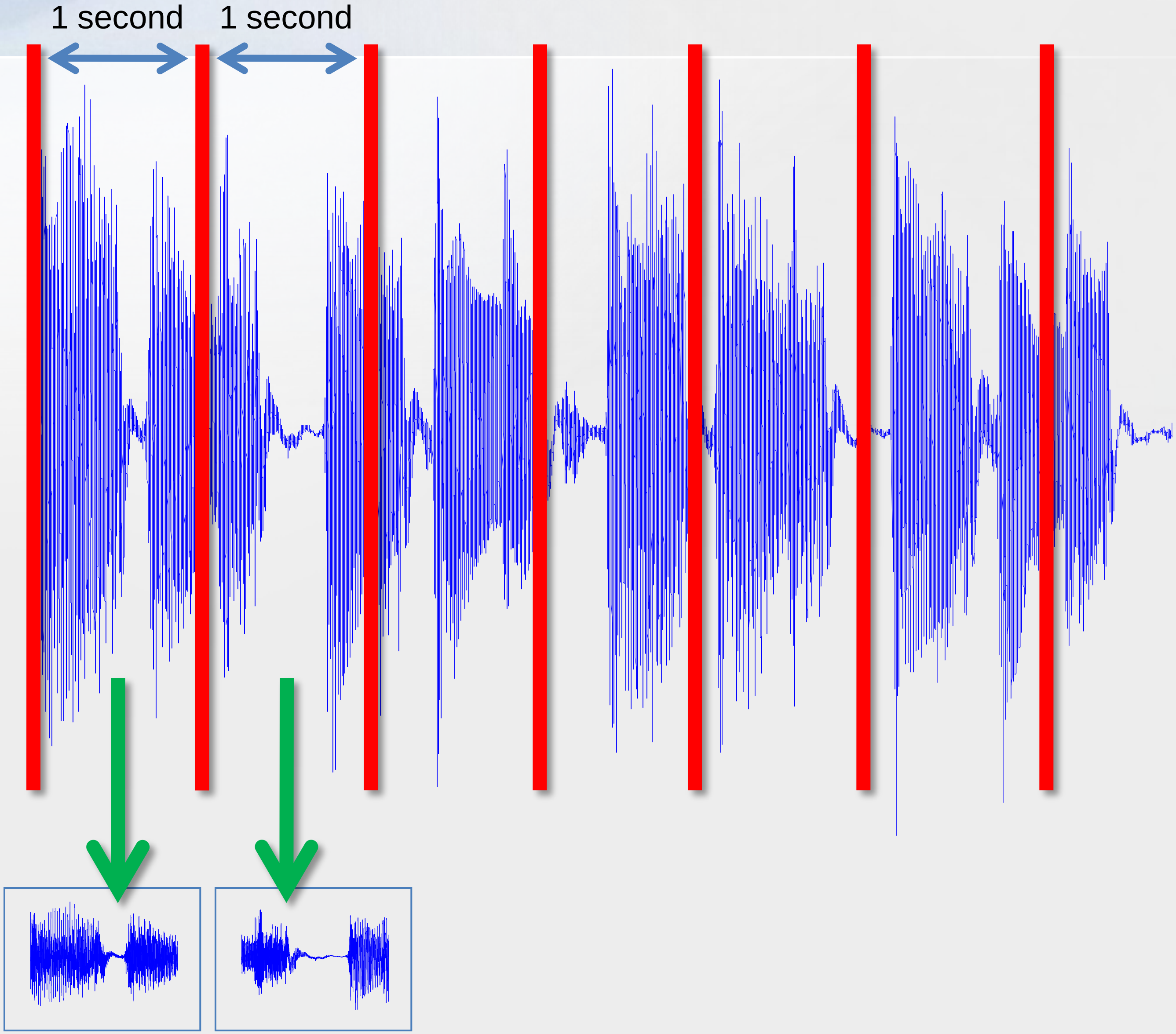
Segmentation
(Frames, Onsets, Beats, Bars, Chord Changes, etc)
Feature Extraction
(Time-based, spectral energy, MFCC, etc)
Analysis / Decision Making
(Classification, Clustering, etc)

TIMING AND SEGMENTATION

Timing and Segmentation

- Slicing up by fixed time slices…
  - 1 second, 80 ms, 100 ms, 20-40ms, etc.
- “Frames”
  - Different problems call for different frame lengths

Frames

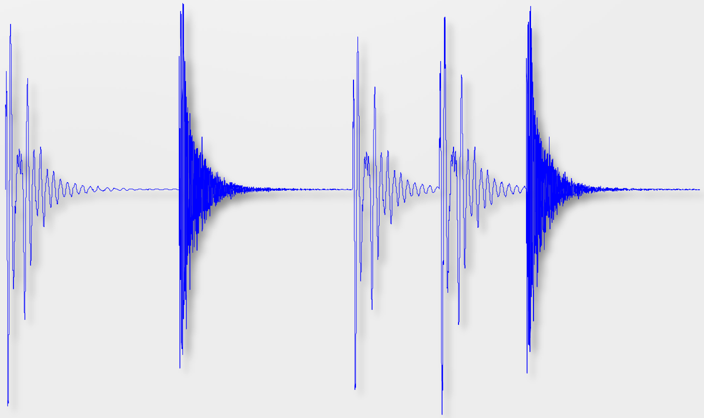


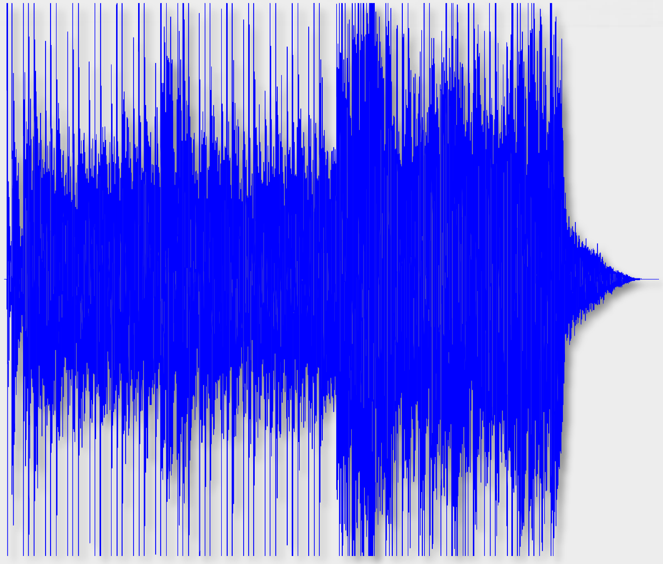
Timing and Segmentation

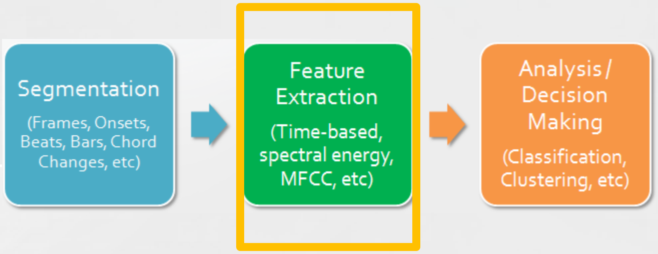
- Slicing up by fixed time slices…
  - 1 second, 80 ms, 100 ms, 20-40ms, etc.
- “Frames”
  - Different problems call for different frame lengths
- Onset detection
- Beat detection
  - Beat
  - Measure / Bar / Harmonic changes
- Segments
  - Musically relevant boundaries
  - Separate by some perceptual cue

Onset detection

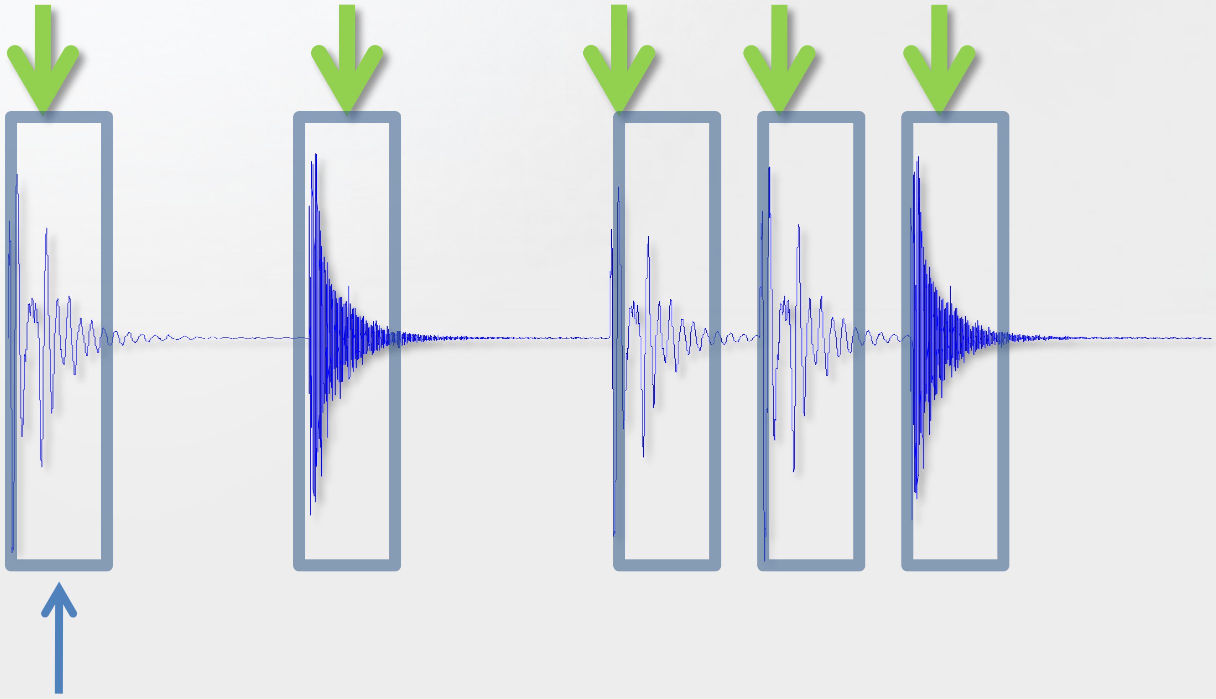
- What is an Onset?
- How to detect?
  - Envelope is not enough
  - Need to examine frequency bands





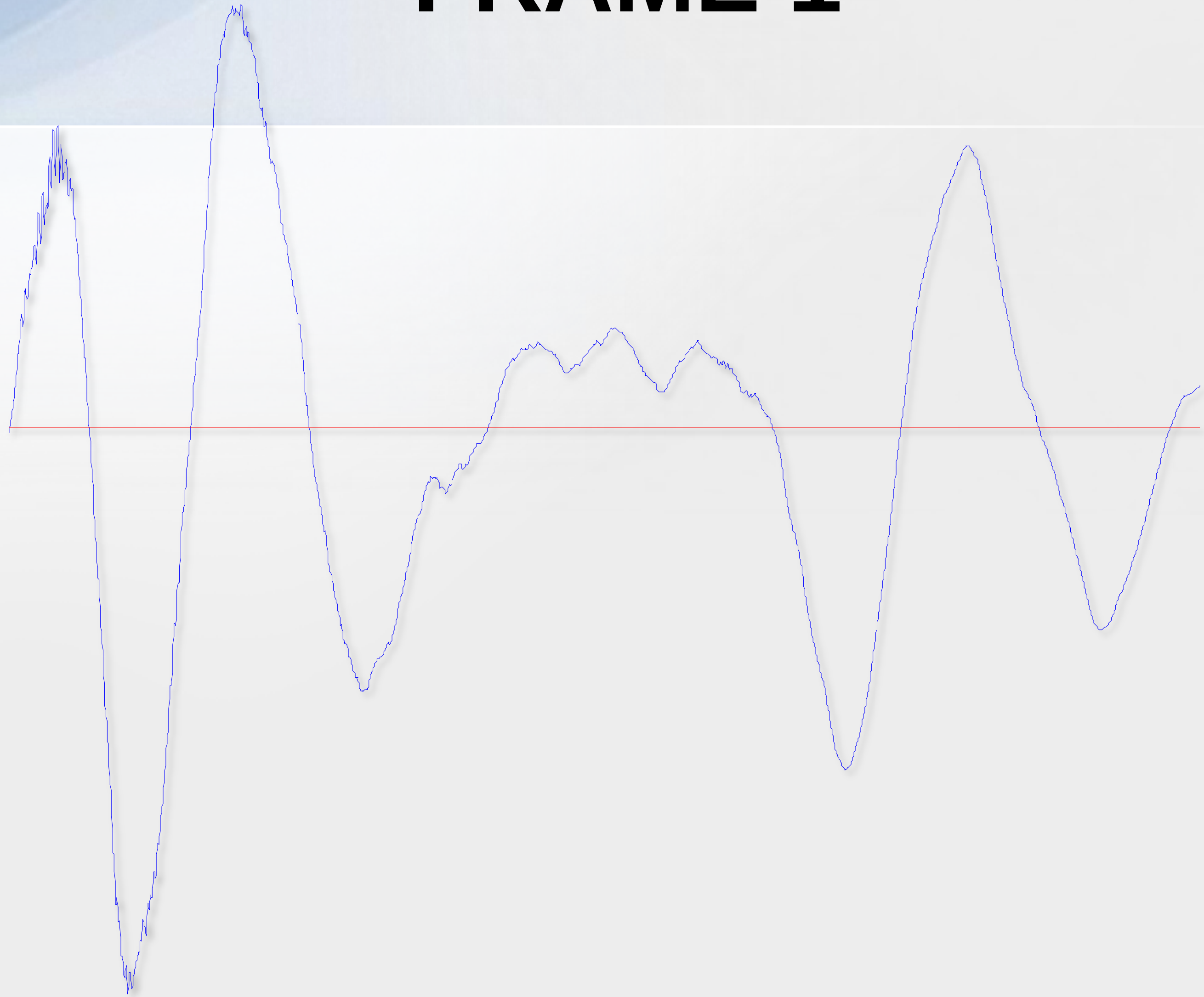

Segmentation
(Frames, Onsets, Beats, Bars, Chord Changes, etc)
Feature Extraction
(Time-based, spectral energy, MFCC, etc)
Analysis / Decision Making
(Classification, Clustering, etc)

FEATURE EXTRACTION

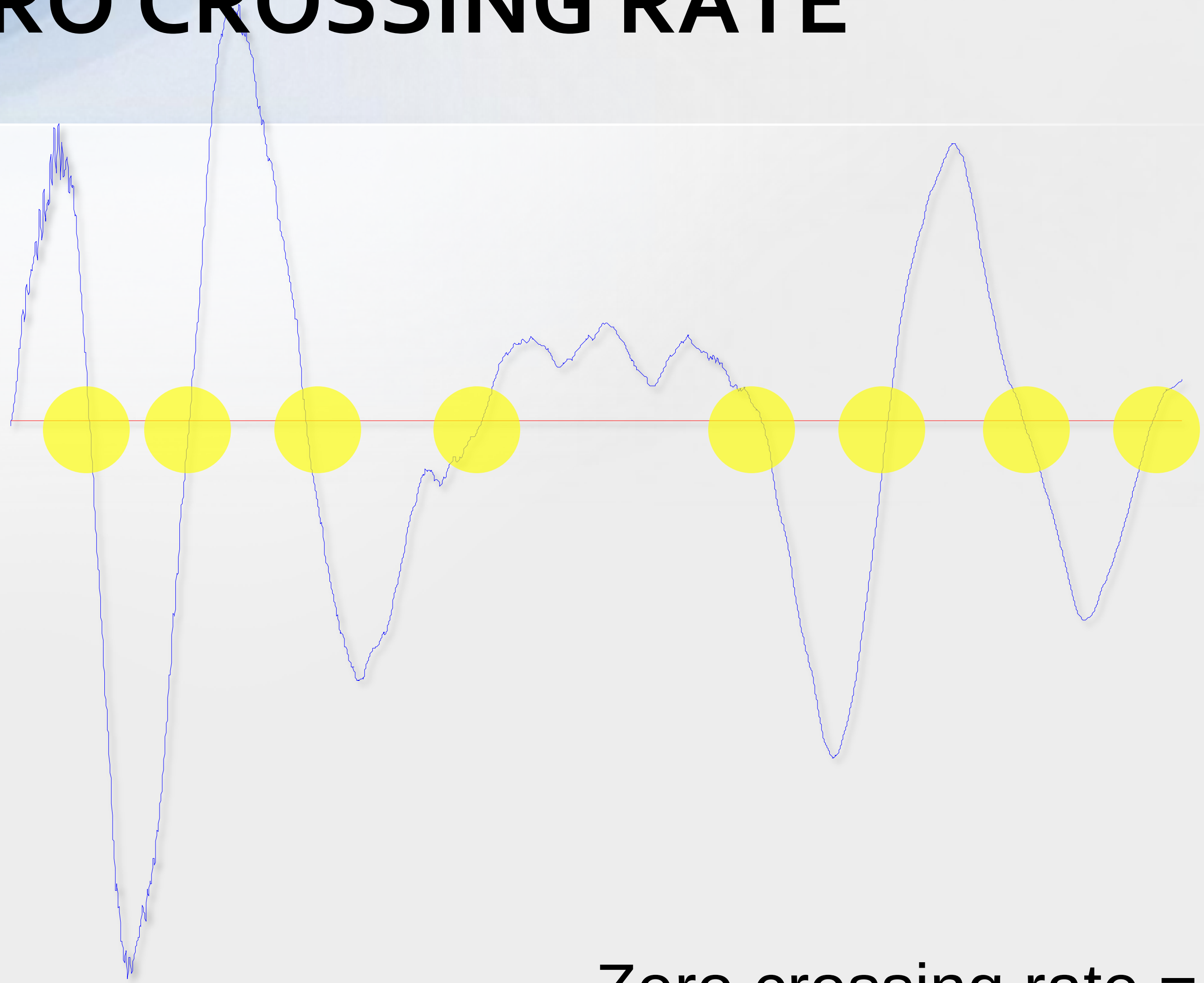


Frame 1

FRAME 1



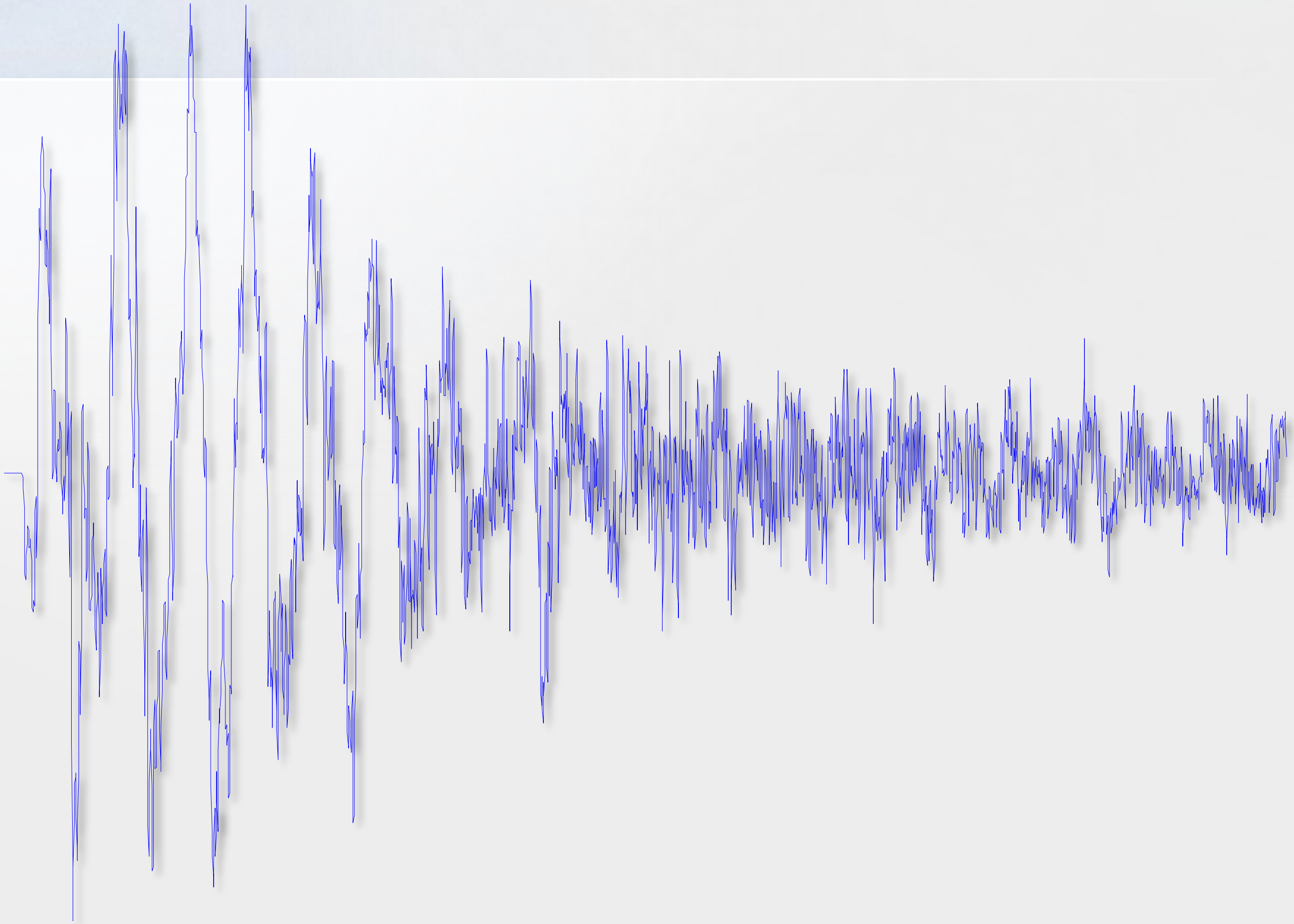
ZERO CROSSING RATE



FRAME 1

Zero crossing rate = 9

Frame 2

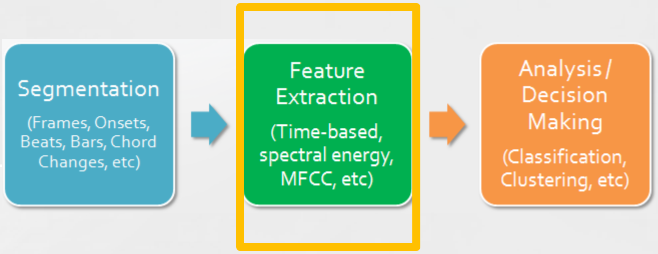


Zero crossing rate = 423

Features : SimpleLoop.wav

Frame	ZCR
1	9
2	423
3	22
4	28
5	390

Warning: example results only - not actual results from audio analysis…


Segmentation
(Frames, Onsets, Beats, Bars, Chord Changes, etc)
Feature Extraction
(Time-based, spectral energy, MFCC, etc)
Analysis / Decision Making
(Classification, Clustering, etc)

FEATURE EXTRACTION

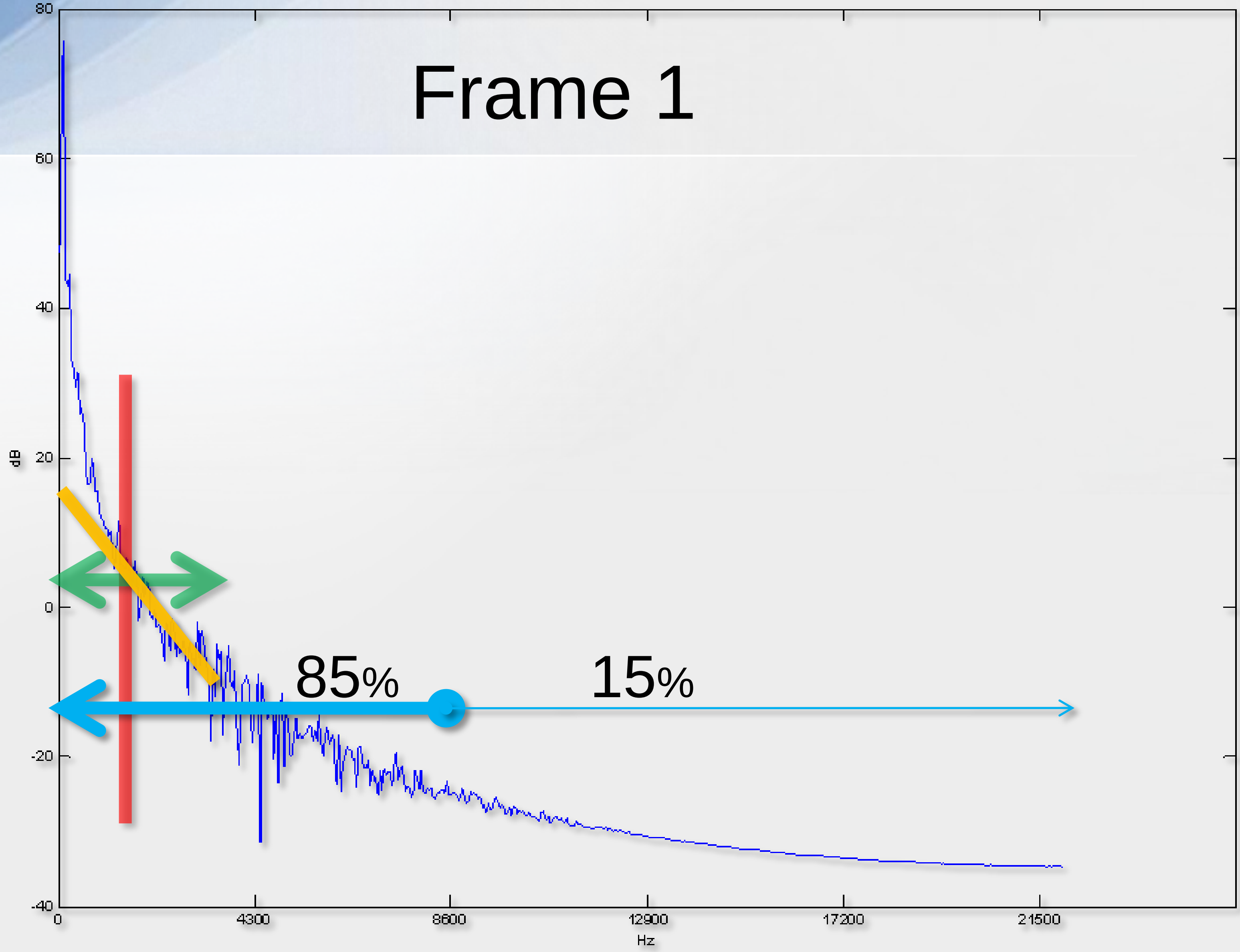
FFT?

Spectral Features

- Spectral Centroid
- Spectral Bandwidth/Spread
- Spectral Skewness
- Spectral Kurtosis
- Spectral Tilt
- Spectral Roll-Off
- Spectral Flatness Measure
- Spectral Crest Factor

Spectral moments

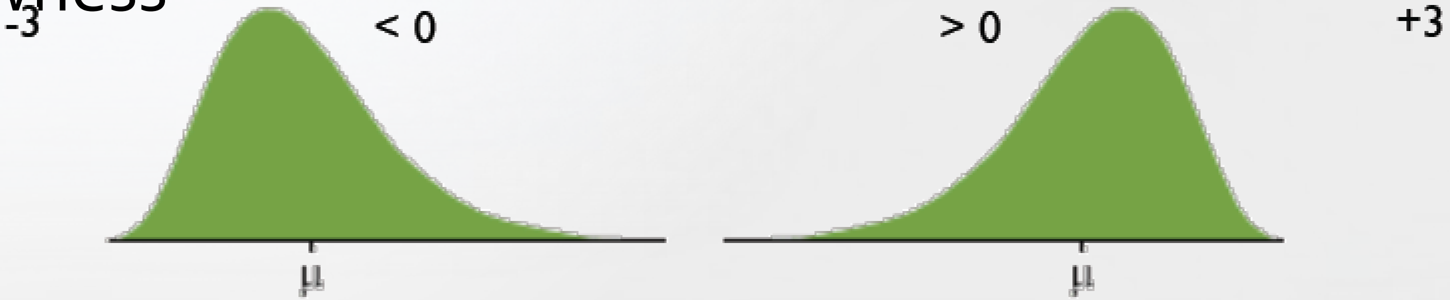
Frame 1



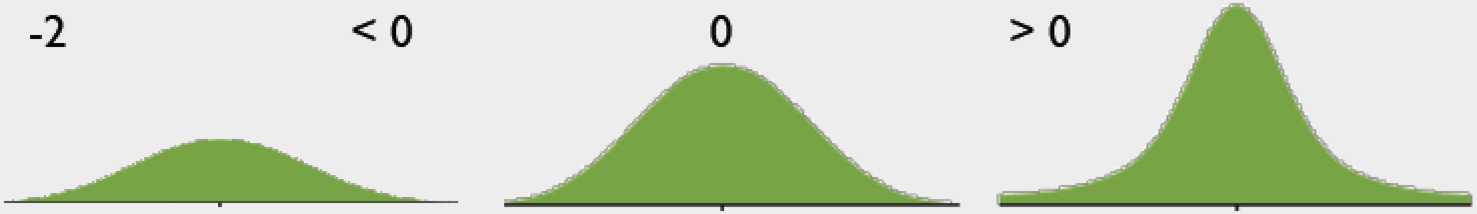
85%

15%

Skewness

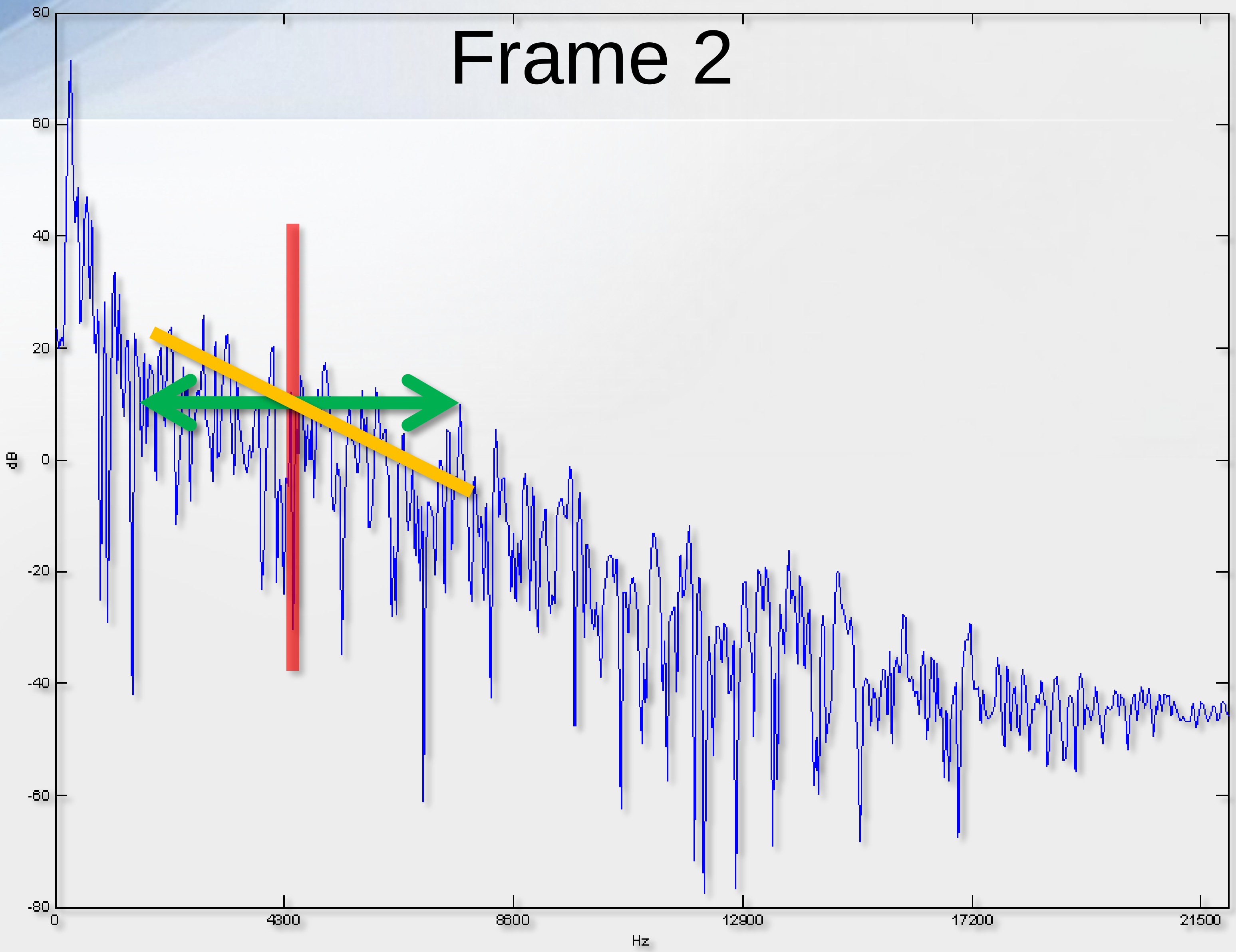


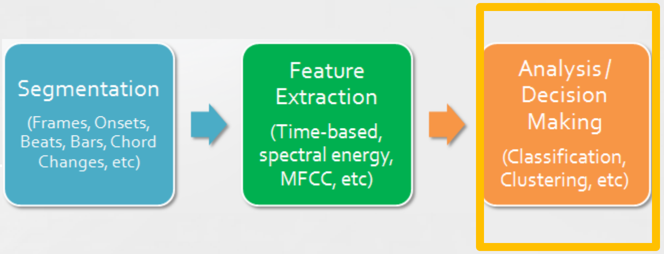
Kurtosis



http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/userguide1.1

Frame 2




Segmentation
(Frames, Onsets, Beats, Bars, Chord Changes, etc)
Feature Extraction
(Time-based, spectral energy, MFCC, etc)
Analysis / Decision Making
(Classification, Clustering, etc)

ANALYSIS AND DECISION MAKING

Heuristic Analysis

- Example: “Cowbell” on just the snare drum of a drum loop. “Simple” instrument recognition!
- Use basic thresholds or simple decision tree to form rudimentary transcription of kicks and snares.
- Time for more sophistication!

> End of Lecture 1