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 contains 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...

Administration

- ccrma.stanford.edu/workshops/2008/

- Introductions
  - A little about yourself
  - Background
  - List of your region of interest, and any specific items of interest that you'd like to see covered.

Final projects

Motivation by your personal interests or audio collections

Motivated by examining large audio collections

Examples:

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

- Instrument ID
- Chord recognition
- SFX / Loop searches
- Transcription by classifier
- Speaker ID and characteristics
- Narrator segmentation
- etc.

Example Seed…

Queries and Evaluation

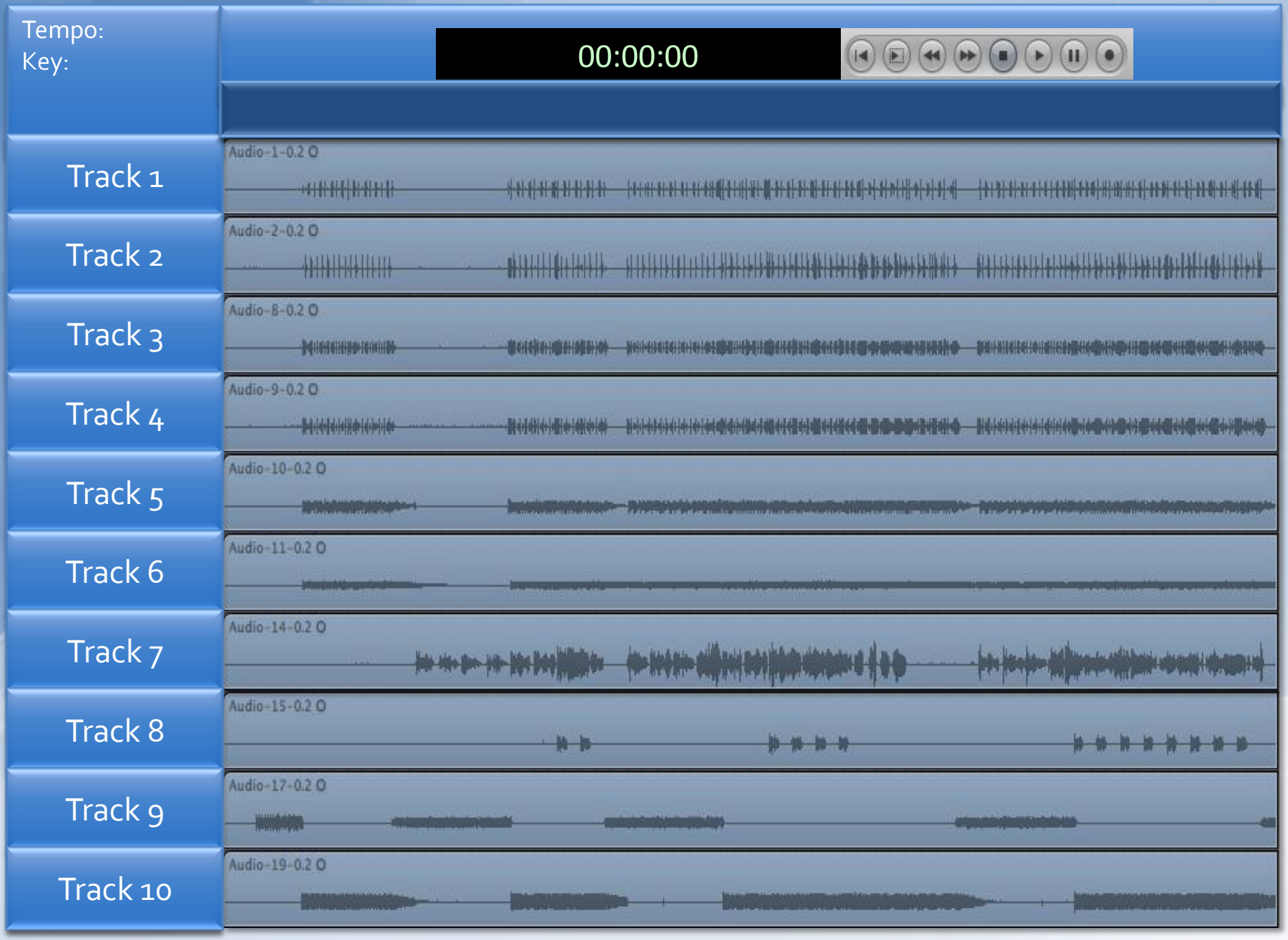
- Query by Humming
  - Lots of academic work
  - Demo with Midomi or Fraunhofer
- Query by audio ID
  - Gracenote ID, Shazam, Audible Magic
  - Noisy audio snippet
- Query by example
  - Find more like this (where “this” has to be specified or inferred)

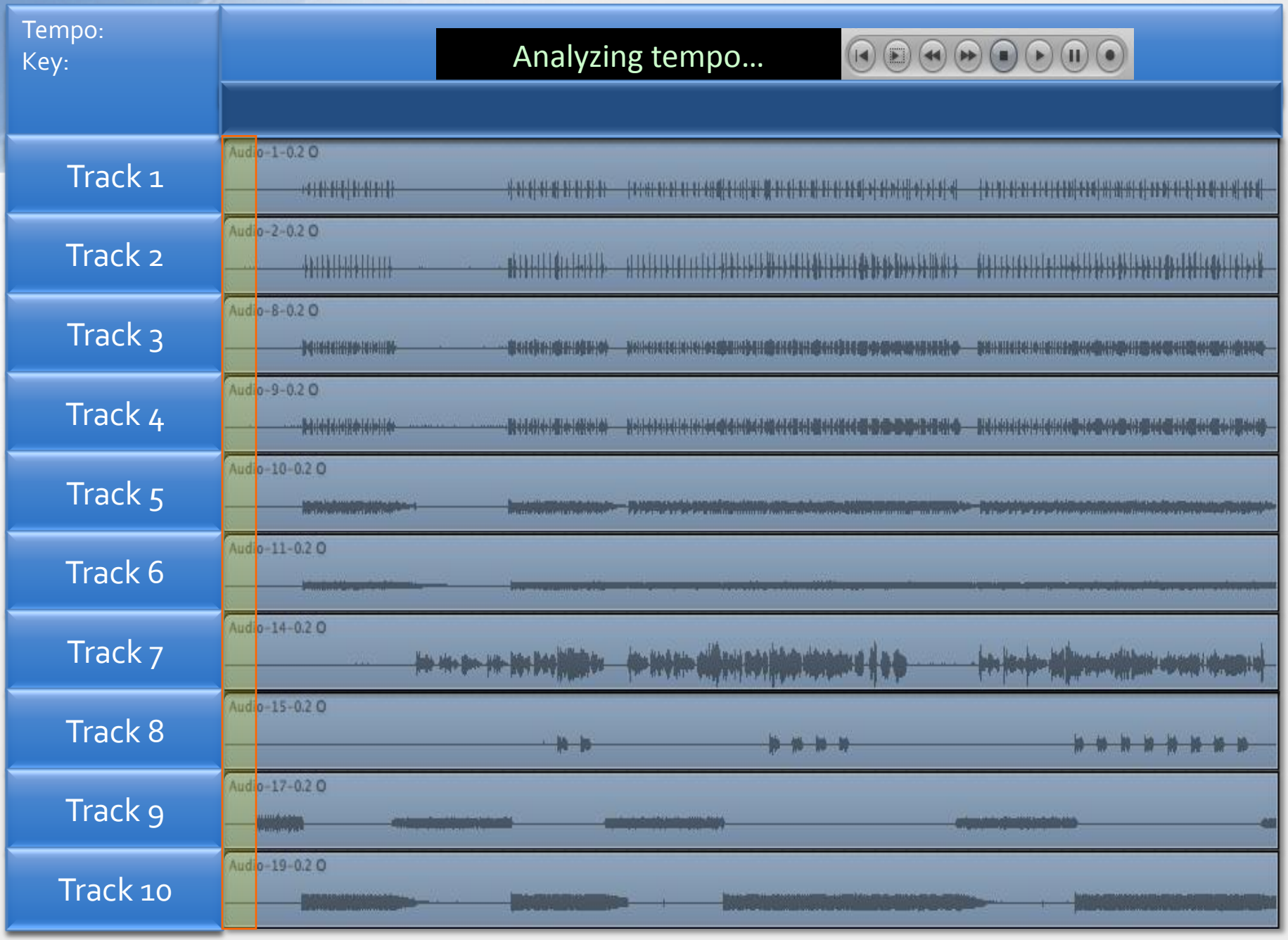
Opportunities

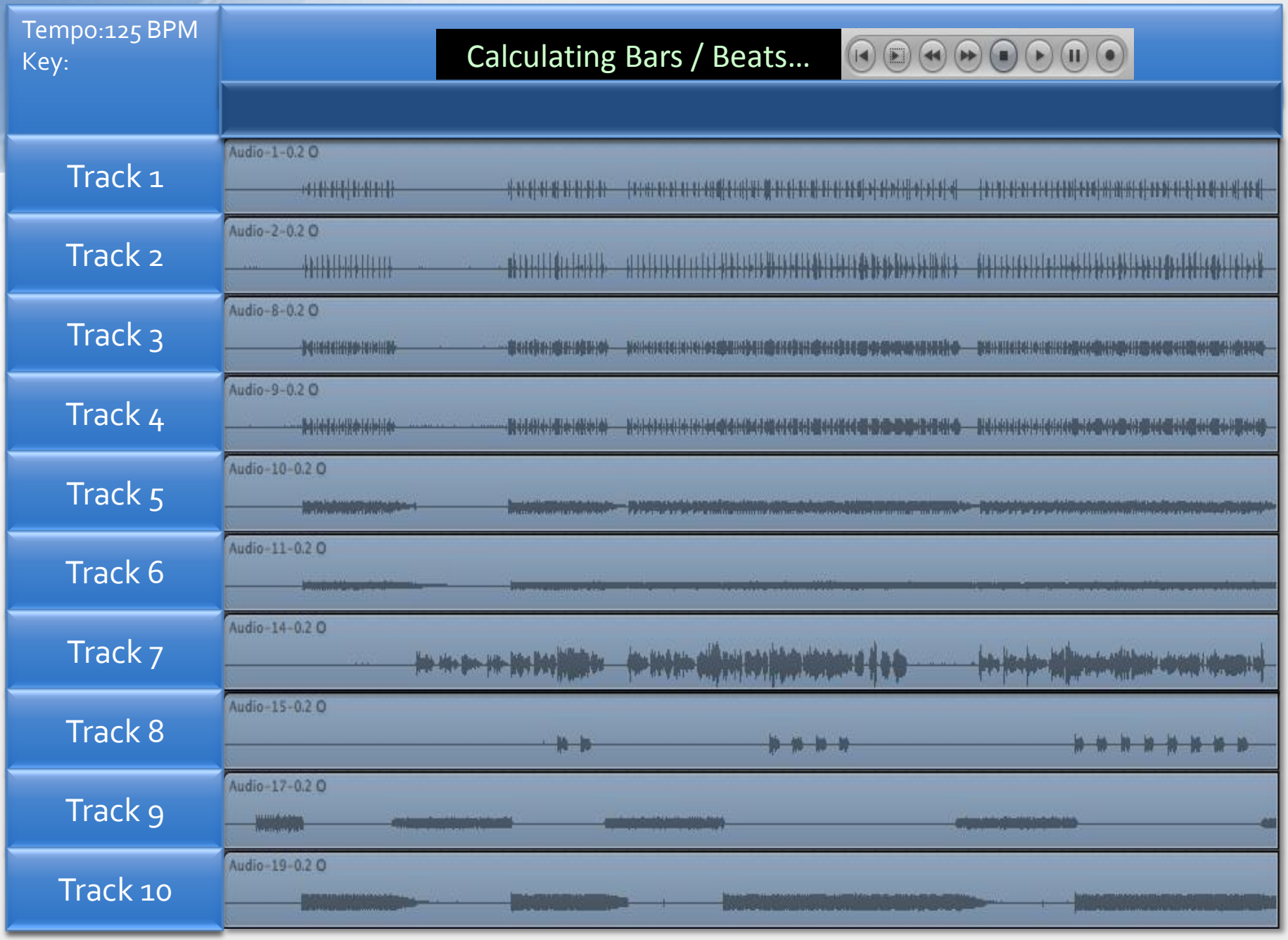
- Genre ID (labels exist, but even humans disagree!)
- Artist classification
  - Tricks: use voice only to improve accuracy to 70% (out of 100 artists)
- Artist similarity
  - Ground truth from : www.musicseer.com
  - Really, what is the similarity?

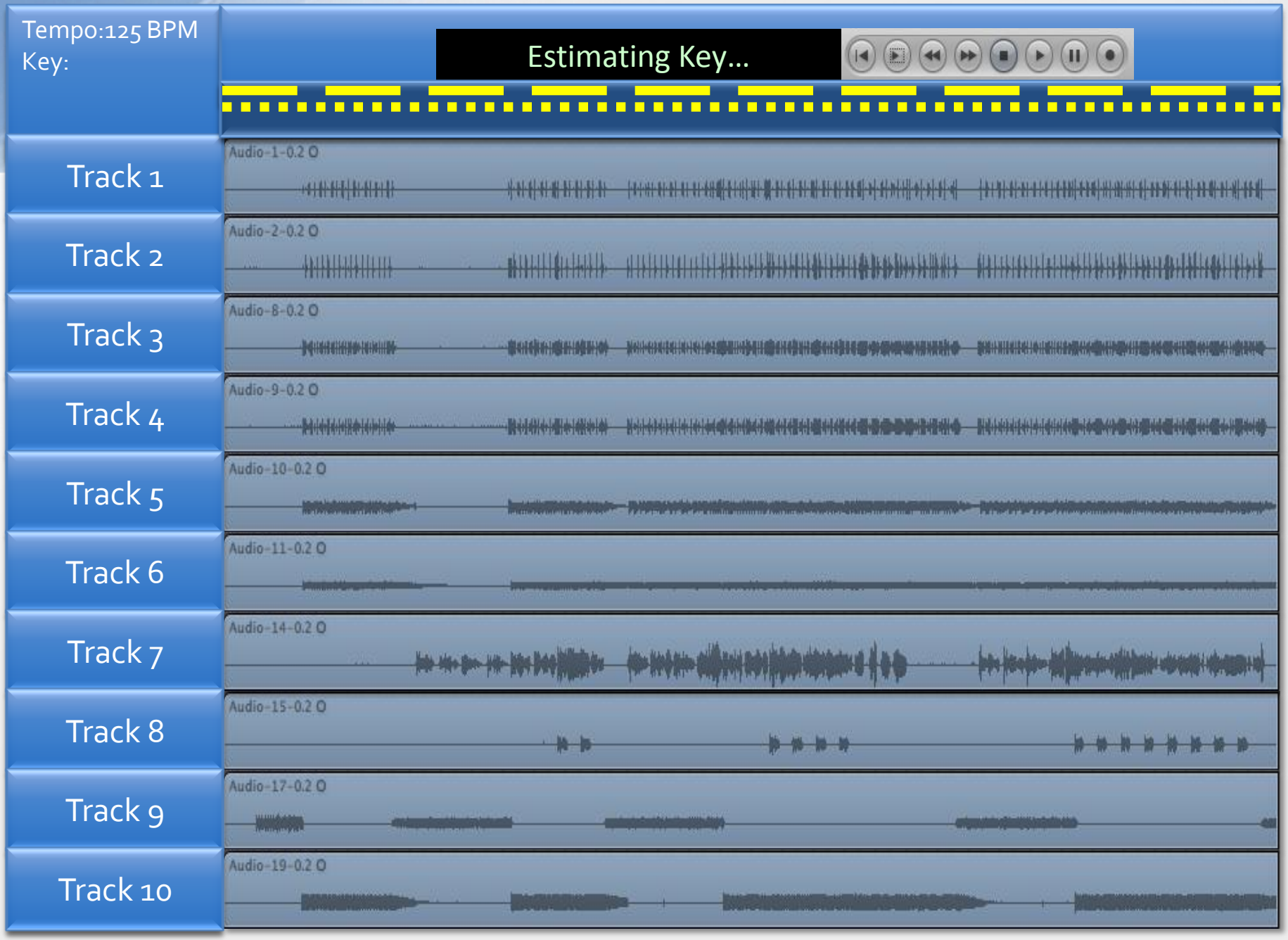
- **But: what is similarity between artists?**
  - pattern recognition systems give a number...

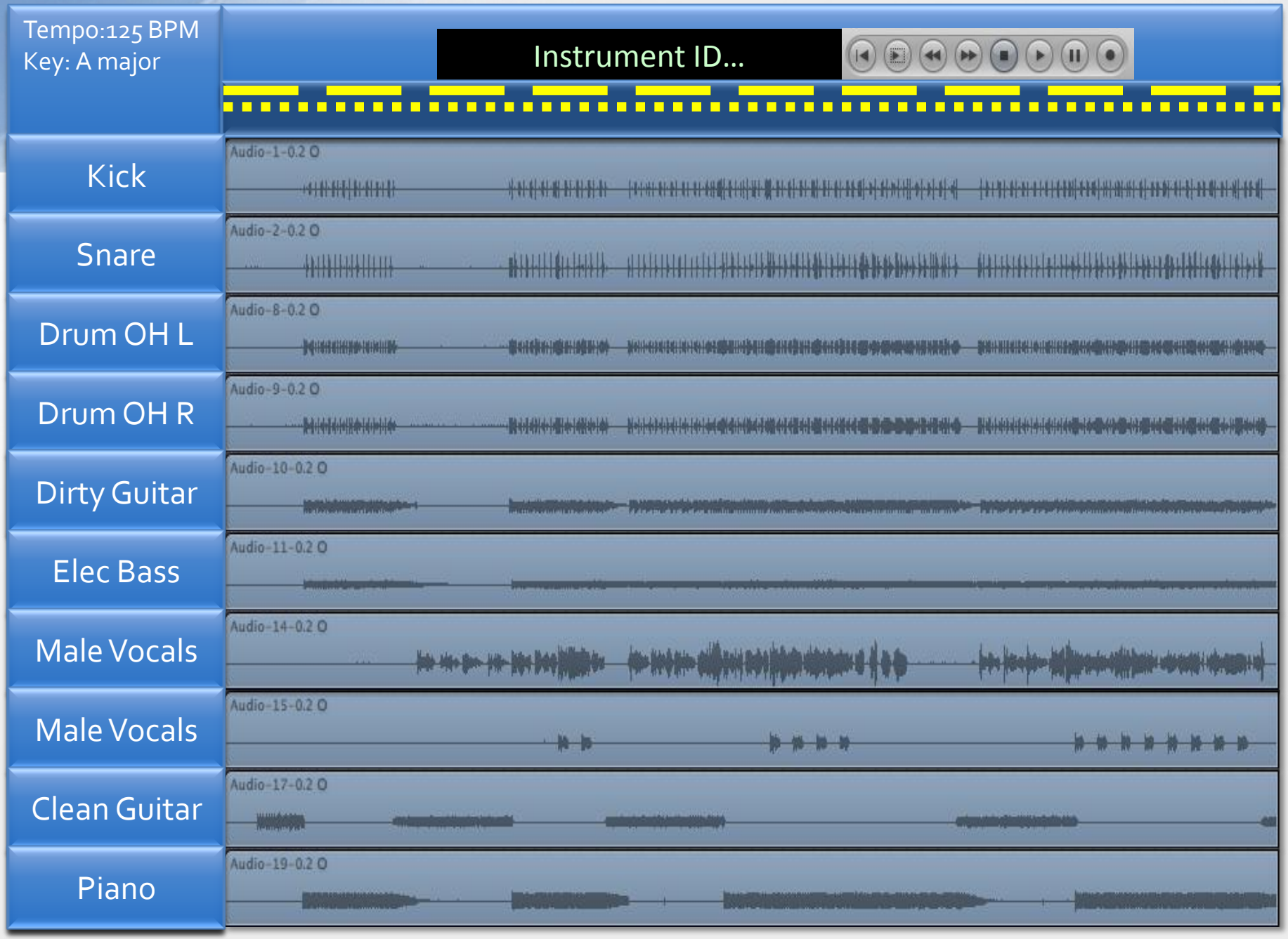
FUTURE SOFTWARE CAPABILITIES

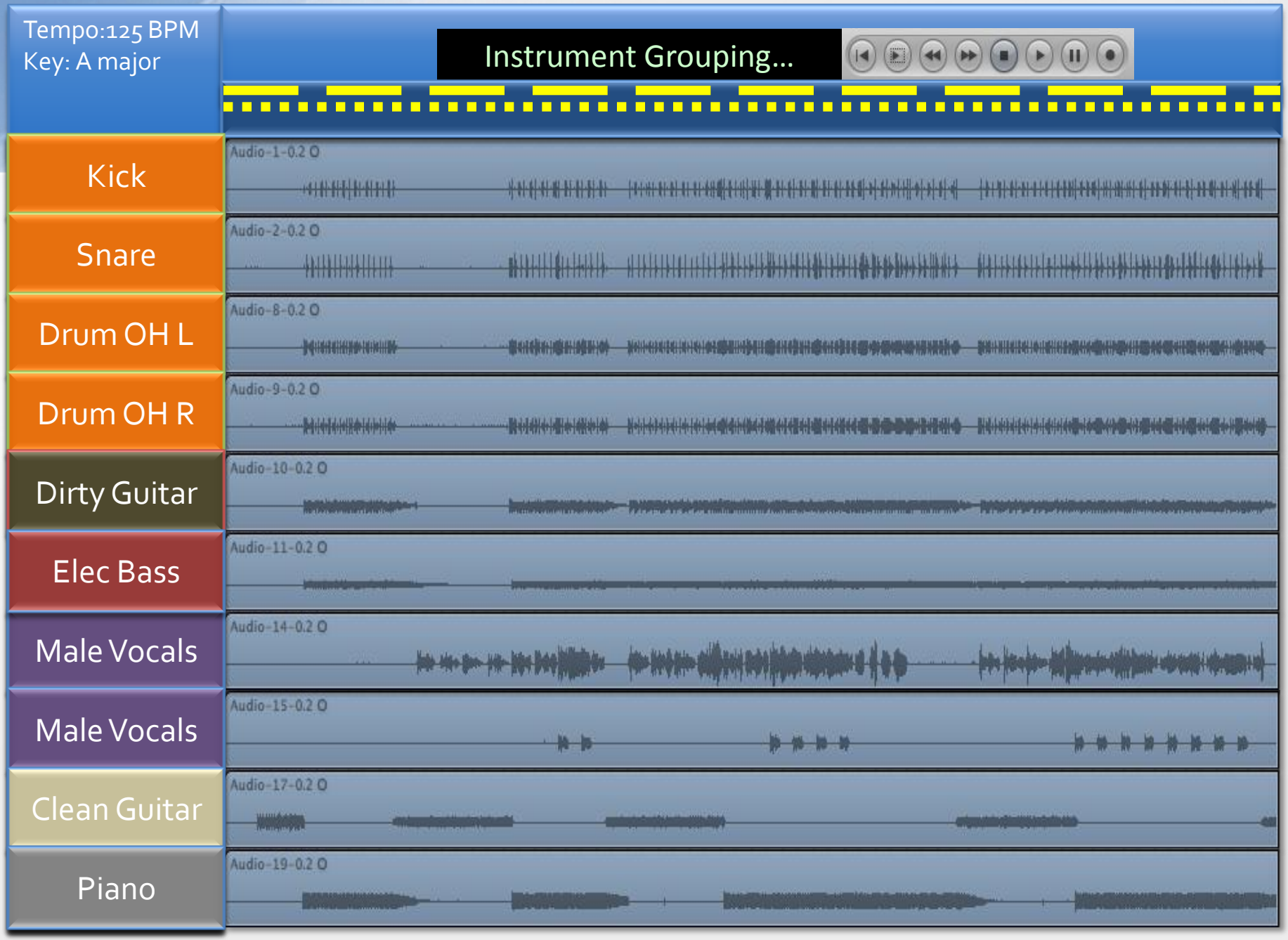

Tempo:
Key:
00:00:00
Track 1
Track 2
Track 3
Track 4
Track 5
Track 6
Track 7
Track 8
Track 9
Track 10
Audio-1-0.2 O
Audio-2-0.2 O
Audio-8-0.2 O
Audio-9-0.2 O
Audio-10-0.2 O
Audio-11-0.2 O
Audio-14-0.2 O
Audio-15-0.2 O
Audio-17-0.2 O
Audio-19-0.2 O

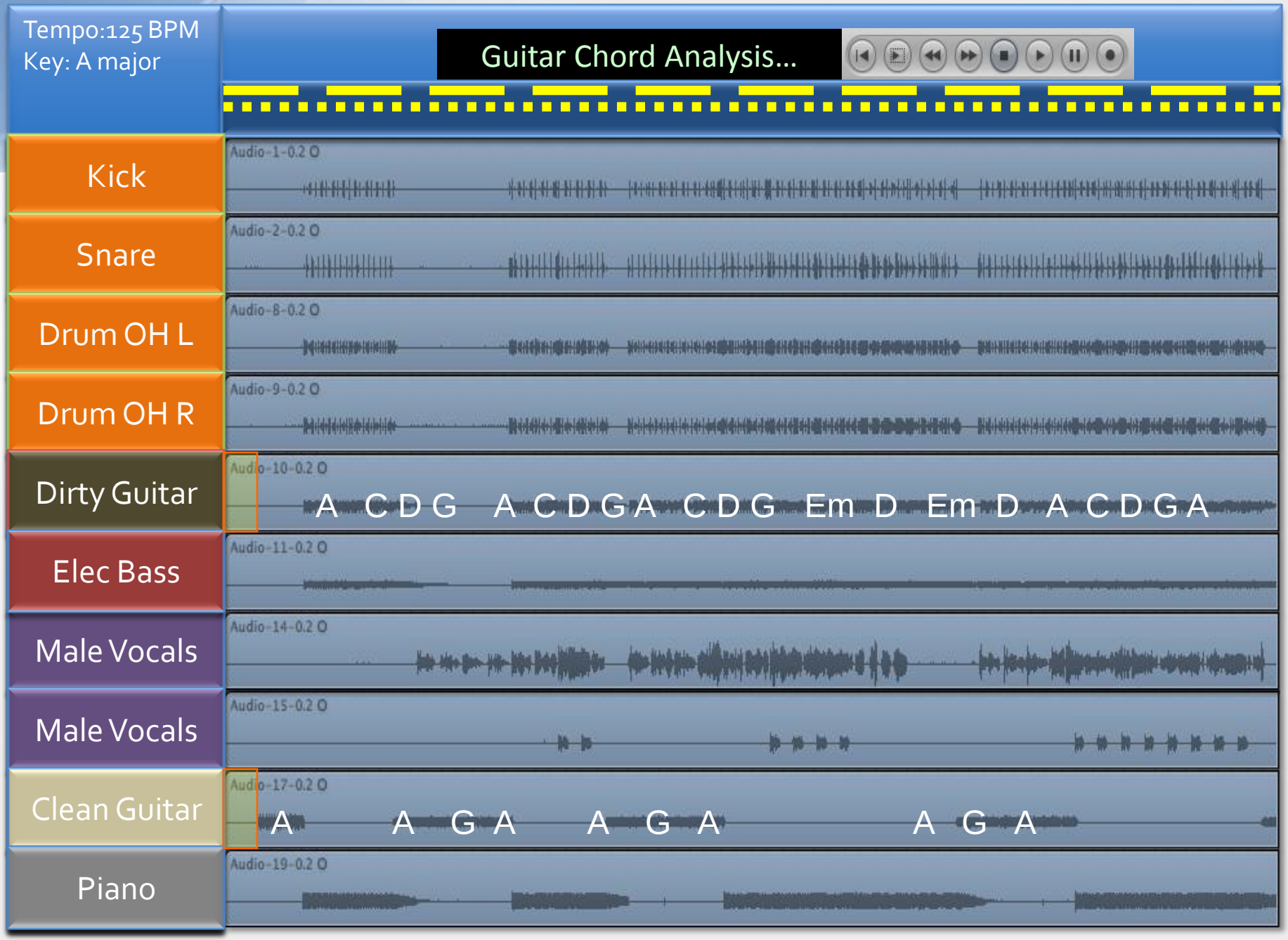

Tempo:
Key:
Analyzing tempo…
Track 1
Track 2
Track 3
Track 4
Track 5
Track 6
Track 7
Track 8
Track 9
Track 10


Tempo:125 BPM
Key:
Calculating Bars / Beats…
Track 1
Track 2
Track 3
Track 4
Track 5
Track 6
Track 7
Track 8
Track 9
Track 10
Audio-1-0.2 O
Audio-2-0.2 O
Audio-8-0.2 O
Audio-9-0.2 O
Audio-10-0.2 O
Audio-11-0.2 O
Audio-14-0.2 O
Audio-15-0.2 O
Audio-17-0.2 O
Audio-19-0.2 O


Tempo:125 BPM
Key:
Estimating Key...
Track 1
Track 2
Track 3
Track 4
Track 5
Track 6
Track 7
Track 8
Track 9
Track 10
Audio-1-0.2 O
Audio-2-0.2 O
Audio-8-0.2 O
Audio-9-0.2 O
Audio-10-0.2 O
Audio-11-0.2 O
Audio-14-0.2 O
Audio-15-0.2 O
Audio-17-0.2 O
Audio-19-0.2 O

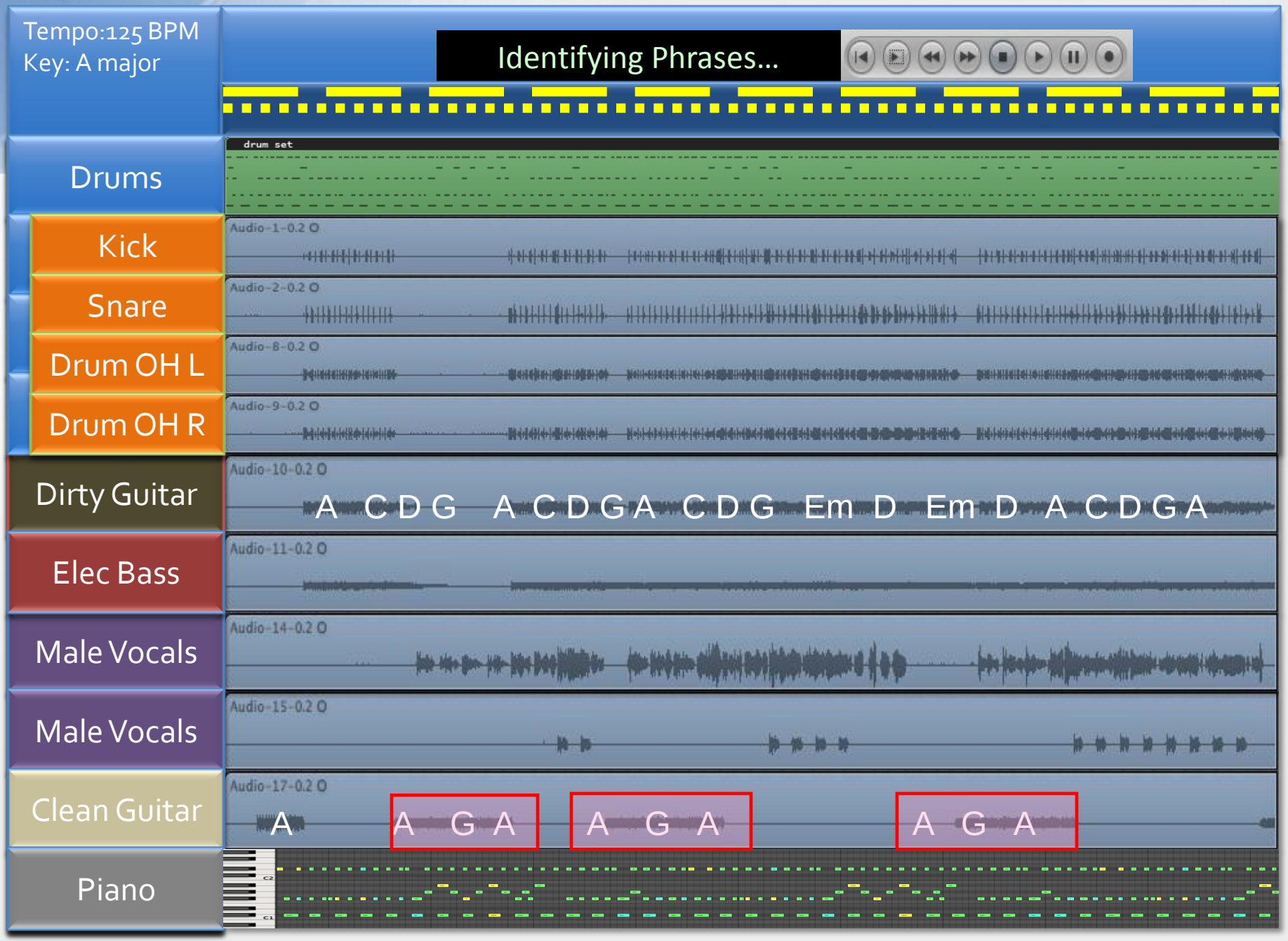

Tempo:125 BPM
Key: A major
Instrument ID...
Kick
Snare
Drum OH L
Drum OH R
Dirty Guitar
Elec Bass
Male Vocals
Male Vocals
Clean Guitar
Piano
Audio-1-0.2 O
Audio-2-0.2 O
Audio-8-0.2 O
Audio-9-0.2 O
Audio-10-0.2 O
Audio-11-0.2 O
Audio-14-0.2 O
Audio-15-0.2 O
Audio-17-0.2 O
Audio-19-0.2 O

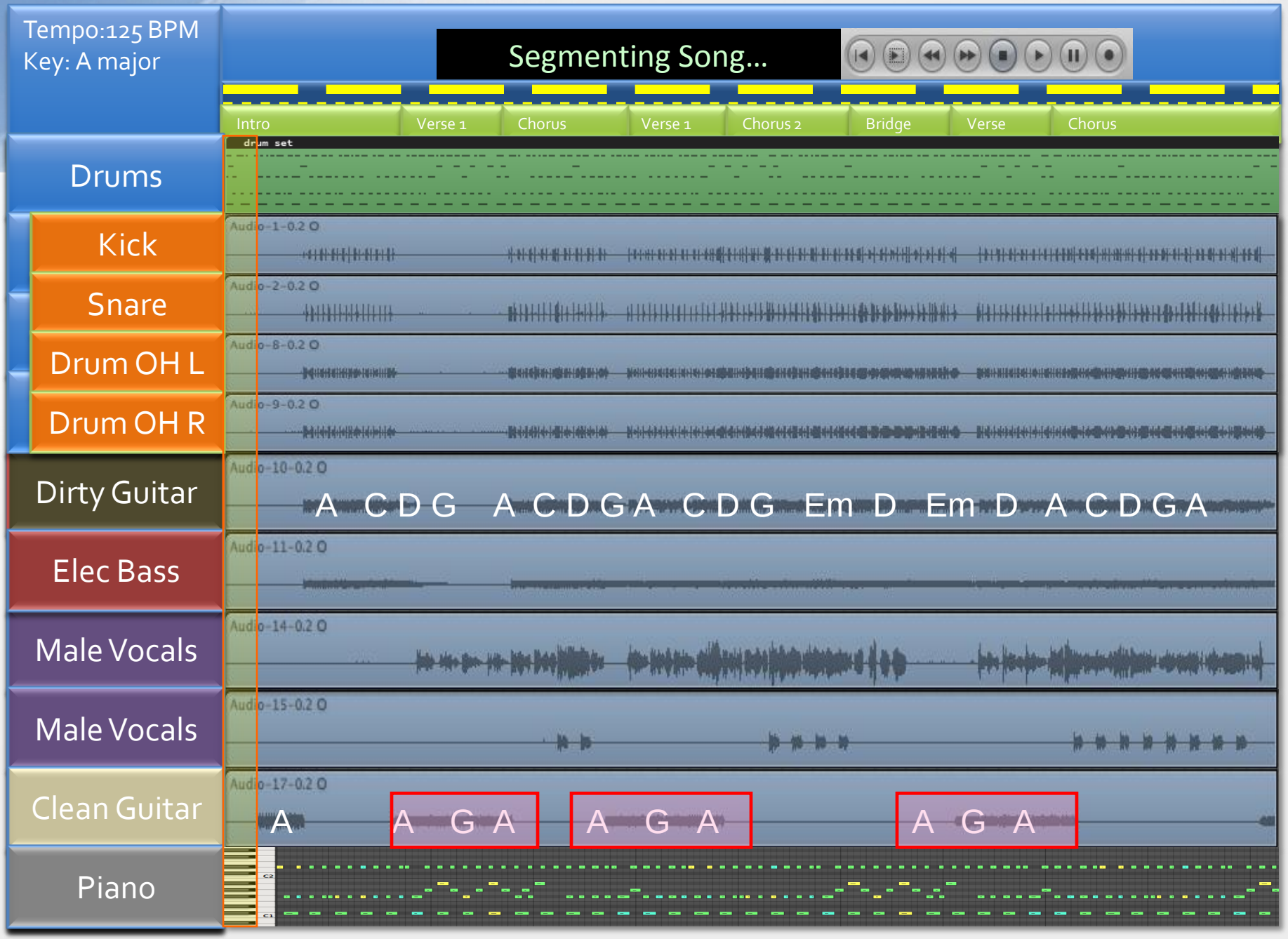

Tempo:125 BPM
Key: A major
Instrument Grouping…
Kick
Snare
Drum OH L
Drum OH R
Dirty Guitar
Elec Bass
Male Vocals
Male Vocals
Clean Guitar
Piano
Audio-1-0.2 O
Audio-2-0.2 O
Audio-8-0.2 O
Audio-9-0.2 O
Audio-10-0.2 O
Audio-11-0.2 O
Audio-14-0.2 O
Audio-15-0.2 O
Audio-17-0.2 O
Audio-19-0.2 O

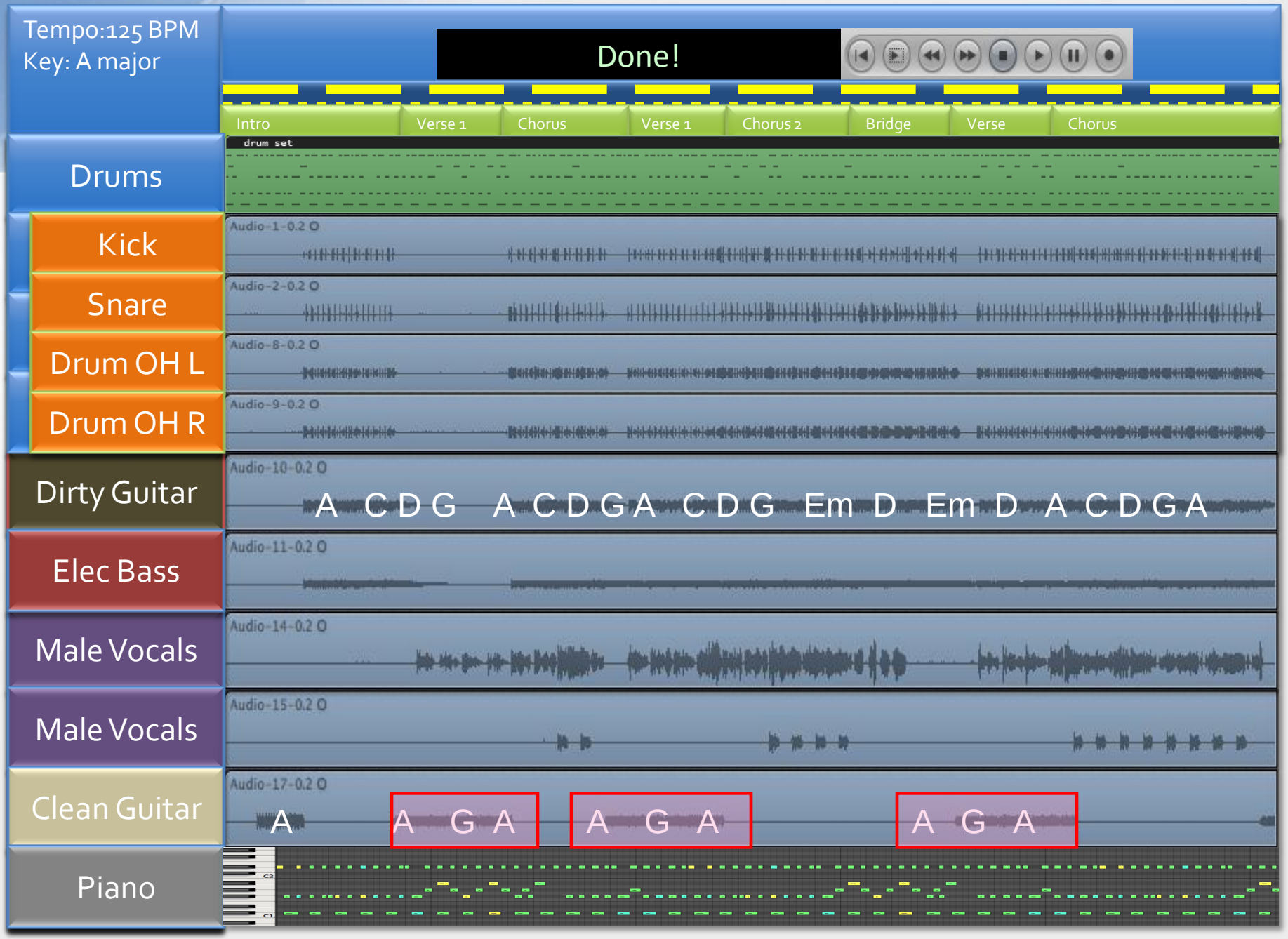

Tempo:125 BPM
Key: A major
Guitar Chord Analysis…
Kick
Snare
Drum OH L
Drum OH R
Dirty Guitar
Elec Bass
Male Vocals
Male Vocals
Clean Guitar
Piano
Audio-1-0.2 O
Audio-2-0.2 O
Audio-8-0.2 O
Audio-9-0.2 O
Audio-10-0.2 O
A C D G A C D G A C D G Em D Em D A C D G A
Audio-11-0.2 O
Audio-14-0.2 O
Audio-15-0.2 O
Audio-17-0.2 O
A A G A A G A A G A
Audio-19-0.2 O


Tempo:125 BPM
Key: A major
Transcribing Drums...
Drums
drum set
Kick
Audio-1-0.2
Snare
Audio-2-0.2
Drum OH L
Audio-8-0.2
Drum OH R
Audio-9-0.2
Dirty Guitar
Audio-10-0.2
A C D G A C D G A C D G Em D Em D A C D G A
Elec Bass
Audio-11-0.2
Male Vocals
Audio-14-0.2
Male Vocals
Audio-15-0.2
Clean Guitar
Audio-17-0.2
A A G A A G A A G A
Piano
Audio-19-0.2


Tempo:125 BPM
Key: A major
Transcribing Piano...
Drums
drum set
Kick
Audio-1-0.2
Snare
Audio-2-0.2
Drum OH L
Audio-8-0.2
Drum OH R
Audio-9-0.2
Dirty Guitar
Audio-10-0.2
A C D G A C D G A C D G Em D Em D A C D G A
Elec Bass
Audio-11-0.2
Male Vocals
Audio-14-0.2
Male Vocals
Audio-15-0.2
Clean Guitar
Audio-17-0.2
A A G A A G A A G A
Piano


Tempo:125 BPM
Key: A major
Identifying Phrases...
Drums
drum set
Kick
Audio-1-0.2
Snare
Audio-2-0.2
Drum OH L
Audio-8-0.2
Drum OH R
Audio-9-0.2
Dirty Guitar
Audio-10-0.2
A C D G A C D G A C D G Em D Em D A C D G A
Elec Bass
Audio-11-0.2
Male Vocals
Audio-14-0.2
Male Vocals
Audio-15-0.2
Clean Guitar
Audio-17-0.2
A
A G A
A G A
A G A
Piano


Tempo:125 BPM
Key: A major
Segmenting Song...
Intro
Verse 1
Chorus
Verse 1
Chorus 2
Bridge
Verse
Chorus
Drums
Kick
Snare
Drum OH L
Drum OH R
Dirty Guitar
Elec Bass
Male Vocals
Male Vocals
Clean Guitar
Piano
drum set
A C D G A C D G A C D G Em D Em D A C D G A
A A G A A G A A G A


Tempo:125 BPM
Key: A major
Done!
Intro
Verse 1
Chorus
Verse 1
Chorus 2
Bridge
Verse
Chorus
Drums
drum set
Kick
Audio-1-0.2
Snare
Audio-2-0.2
Drum OH L
Audio-8-0.2
Drum OH R
Audio-9-0.2
Dirty Guitar
Audio-10-0.2
A C D G A C D G A C D G Em D Em D A C D G A
Elec Bass
Audio-11-0.2
Male Vocals
Audio-14-0.2
Male Vocals
Audio-15-0.2
Clean Guitar
Audio-17-0.2
A
A G A
A G A
A G A
Piano

Why MIR?

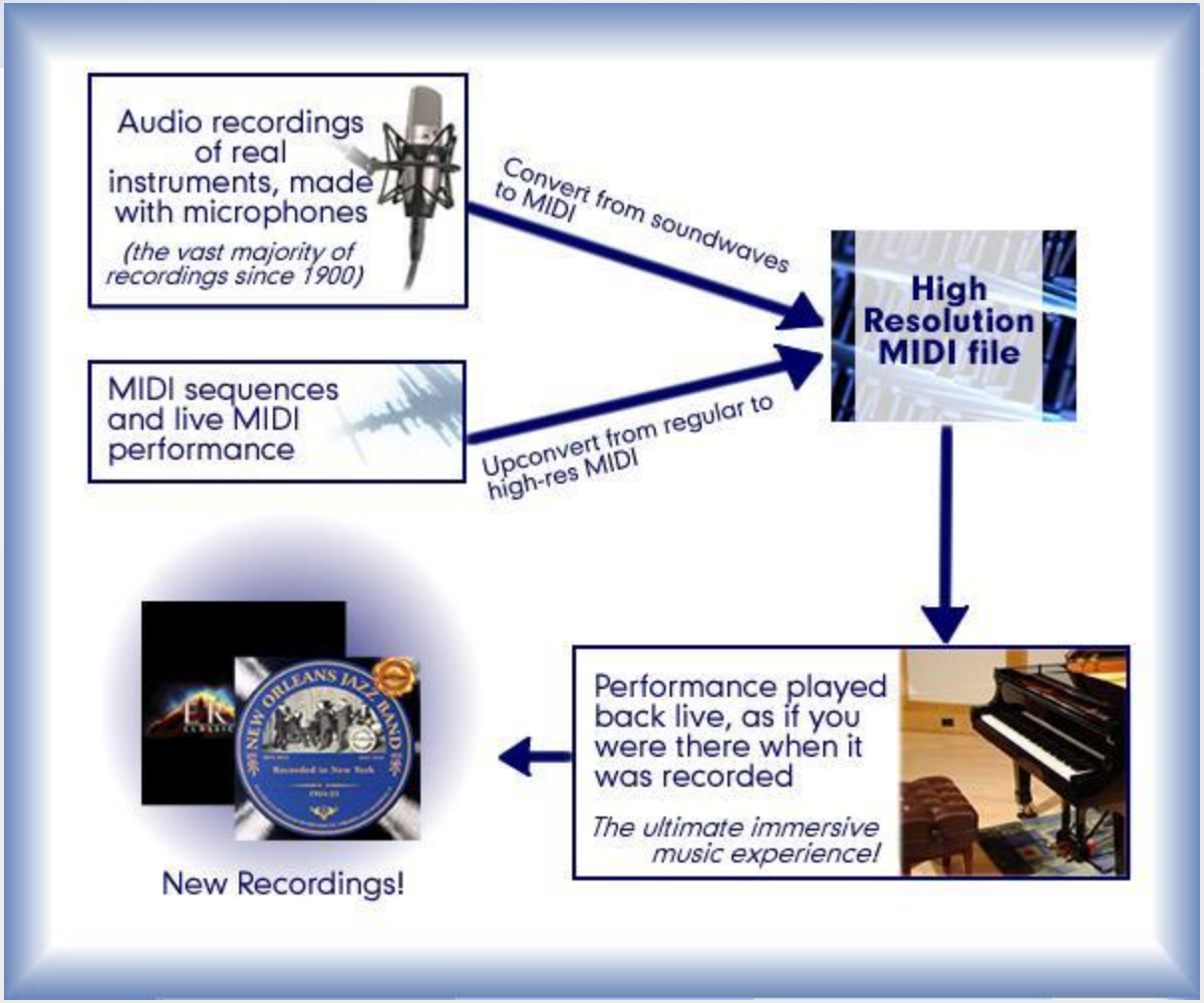
- Find specific item
- Find something vague
- Find something interesting or new
- Google for audio
  - Quaero – French “Google killer" with image and audio "example" queries

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

Motivations / Demos

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


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 ORLEANS JAZZ BAND
Recorded in New York
New Recordings!

Motivations / Demos

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

- More transcription

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

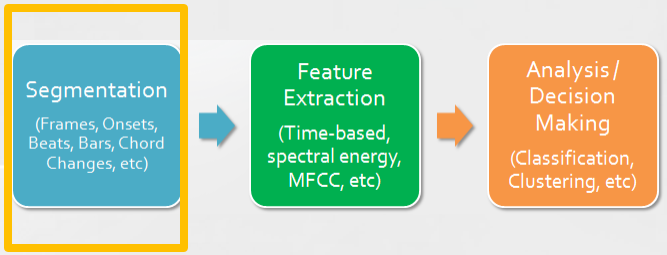
(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction

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

Analysis / Decision Making

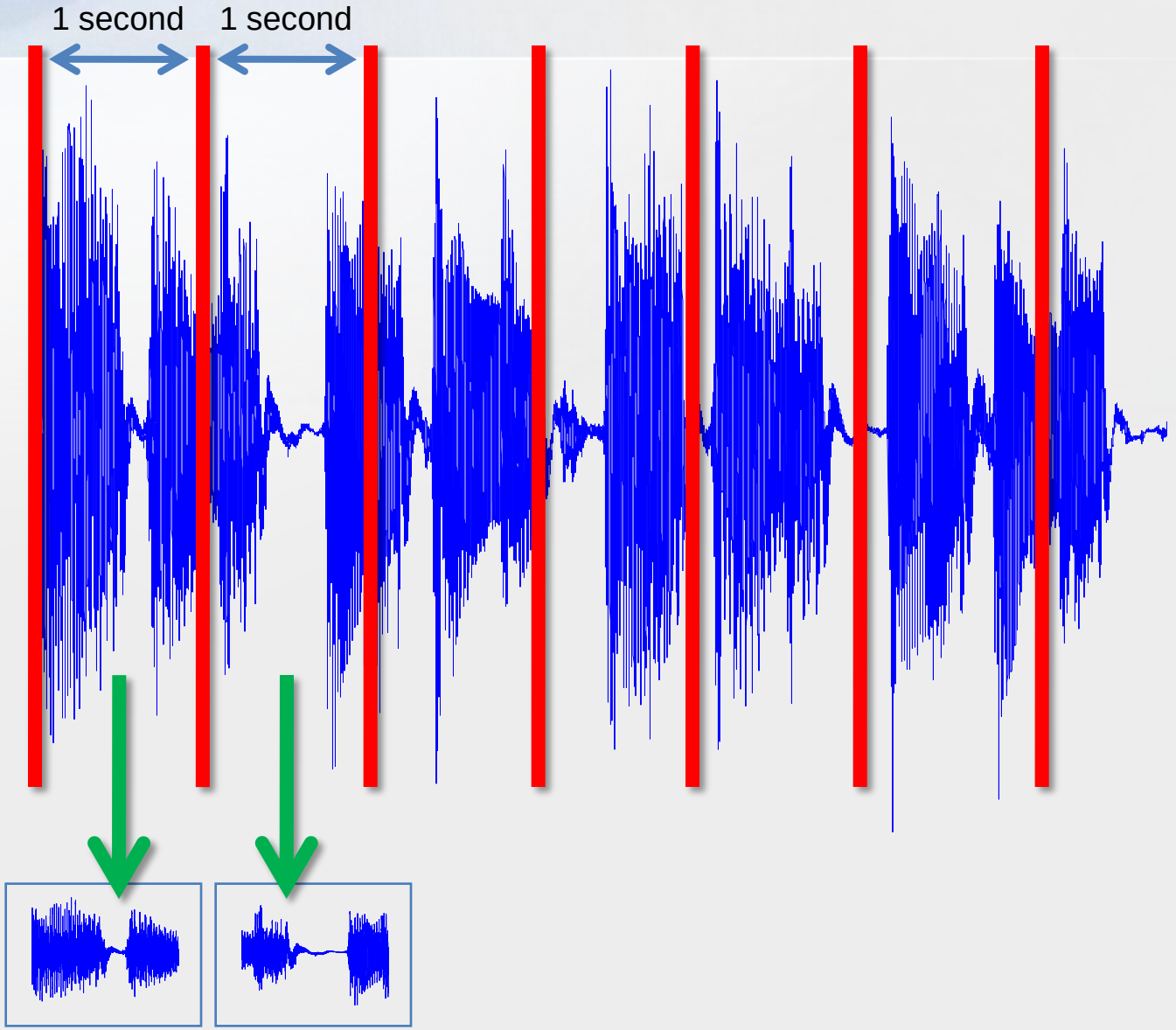
(Classification, Clustering, etc)



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

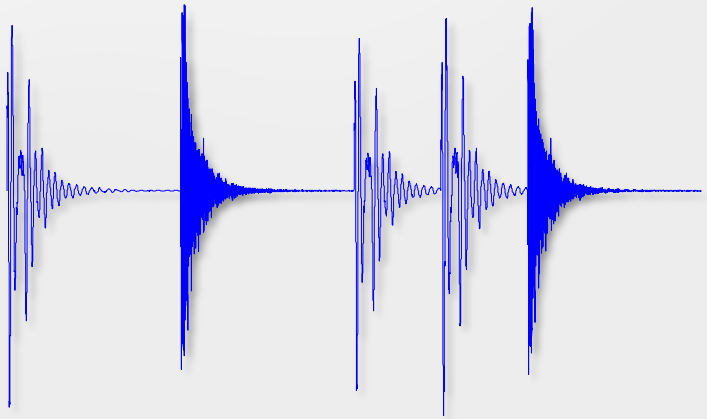

1 second
1 second

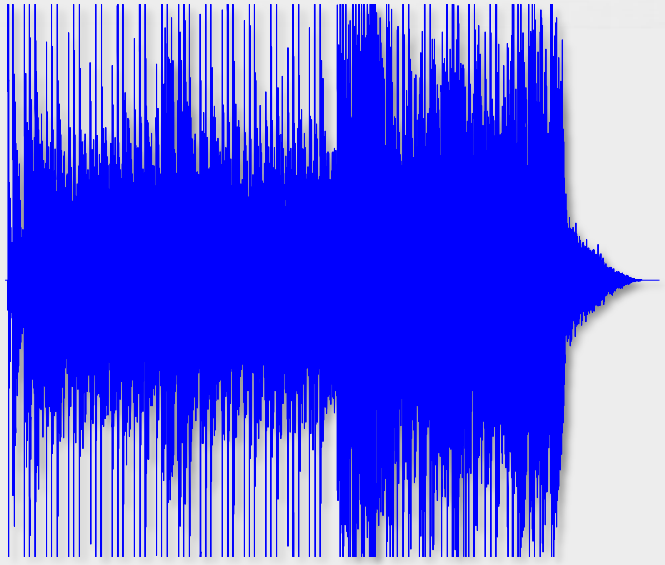
Timing and Segmentation

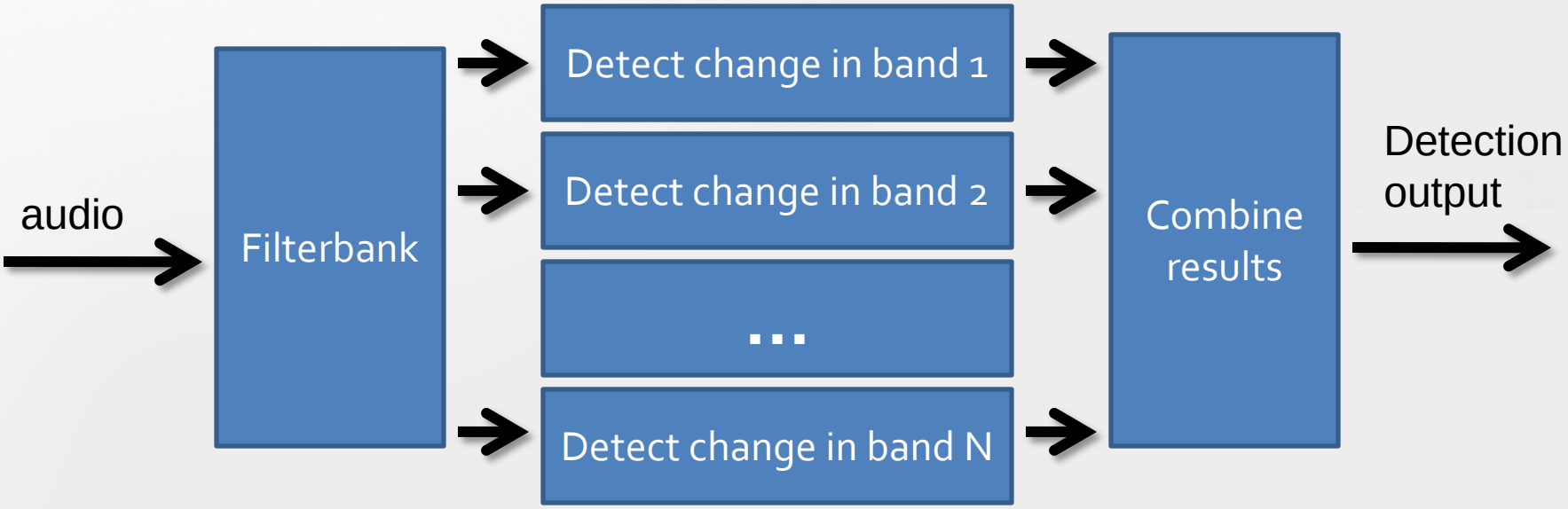
- Slicing up by fixed time slices…
  - 1 second, 80 ms, 100 ms, 20-40ms, etc.
- "Frames"
  - Different problems call for different frame lengths
- Onset detection
- Beat detection
  - Beat
  - Measure / Bar / Harmonic changes
- Segments
  - Musically relevant boundaries
  - Separate by some perceptual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

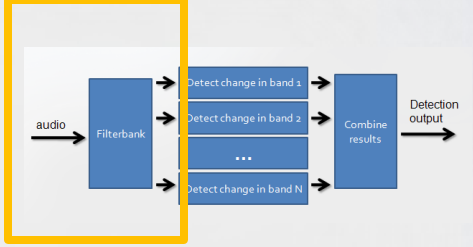
Onset detection

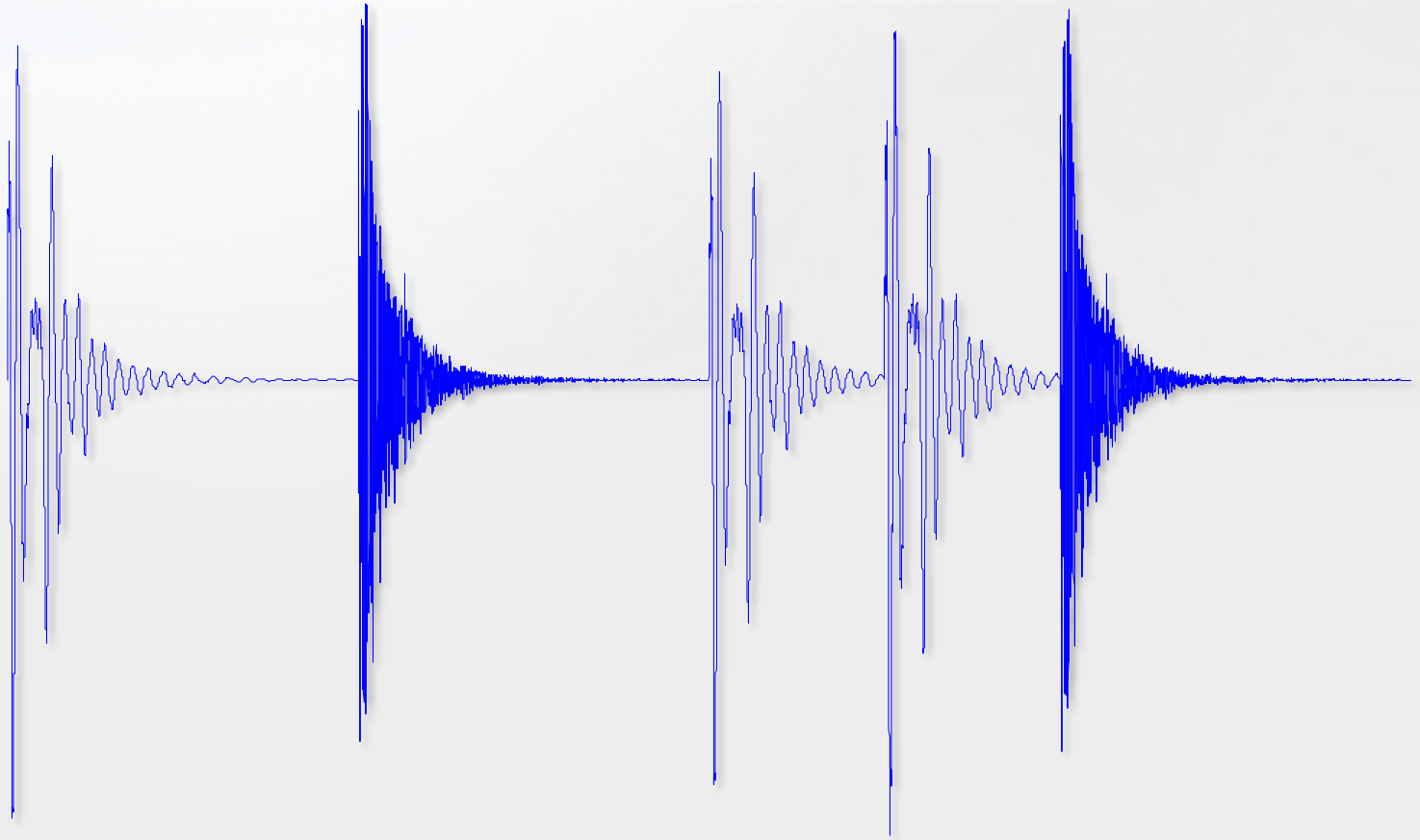
- What is an Onset?
- How to detect?
  - Envelope is not enough
  - Need to examine frequency bands

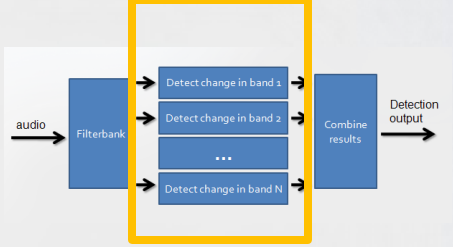





audio
Filterbank
Detect change in band 1
Detect change in band 2
...
Detect change in band N
Combine results
Detection output


audio
Filterbank
Detect change in band 1
Detect change in band 2
...
Detect change in band N
Combine results
Detection output





octave

1

2

3

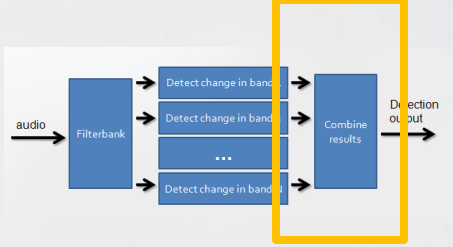
4

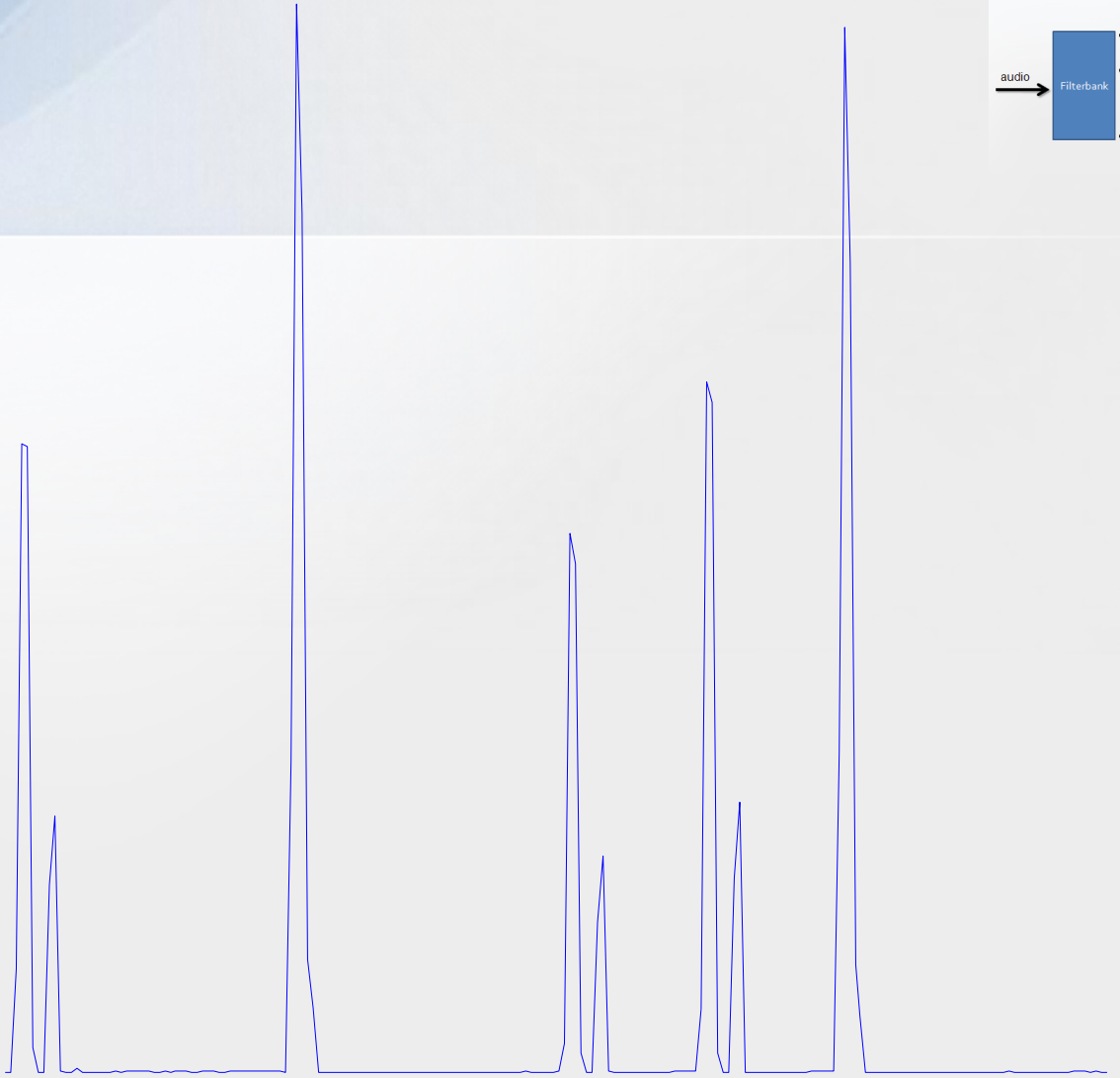
5

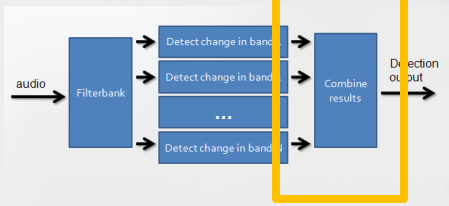
6

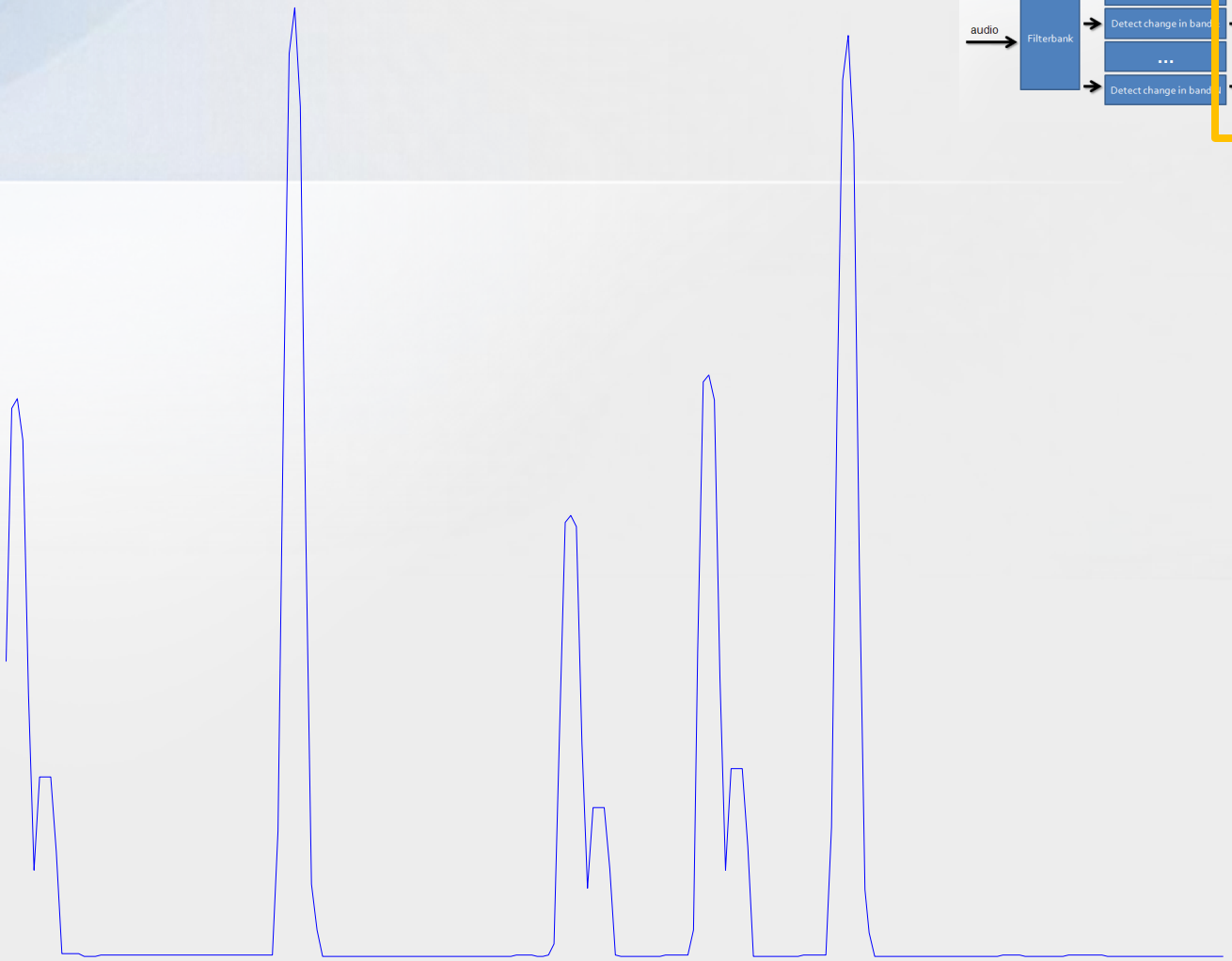
7

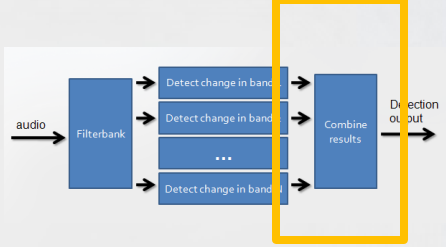
8


audio
Filterbank
Detect change in band
Detect change in band
...
Detect change in band
Combine results
Detection output

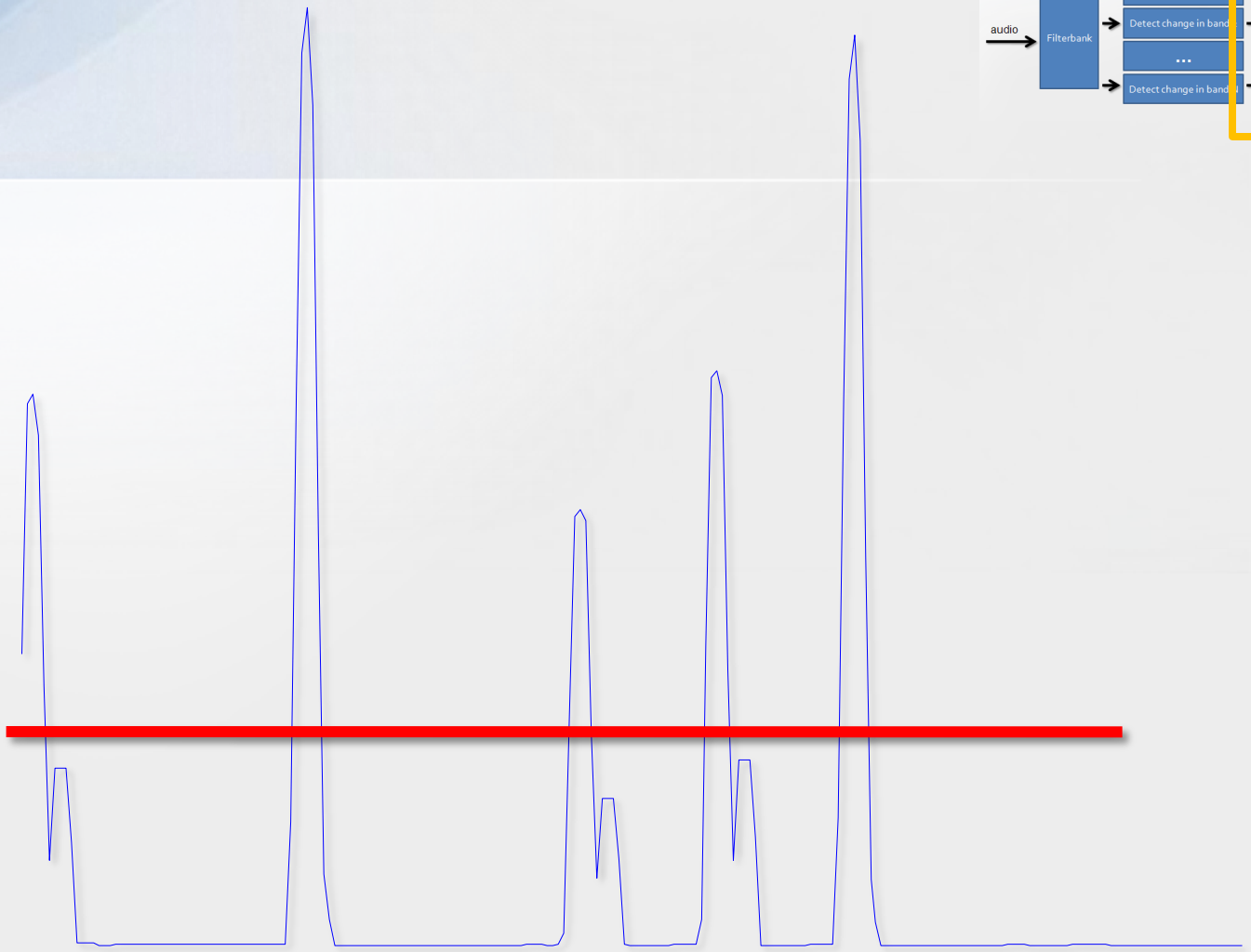


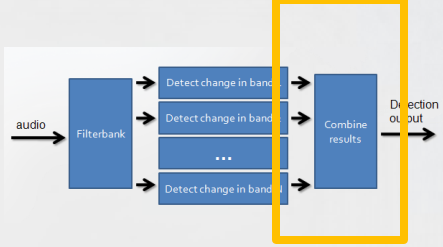

audio
Filterbank
Detect change in band
Detect change in band
...
Detect change in band
Combine results
Detection output



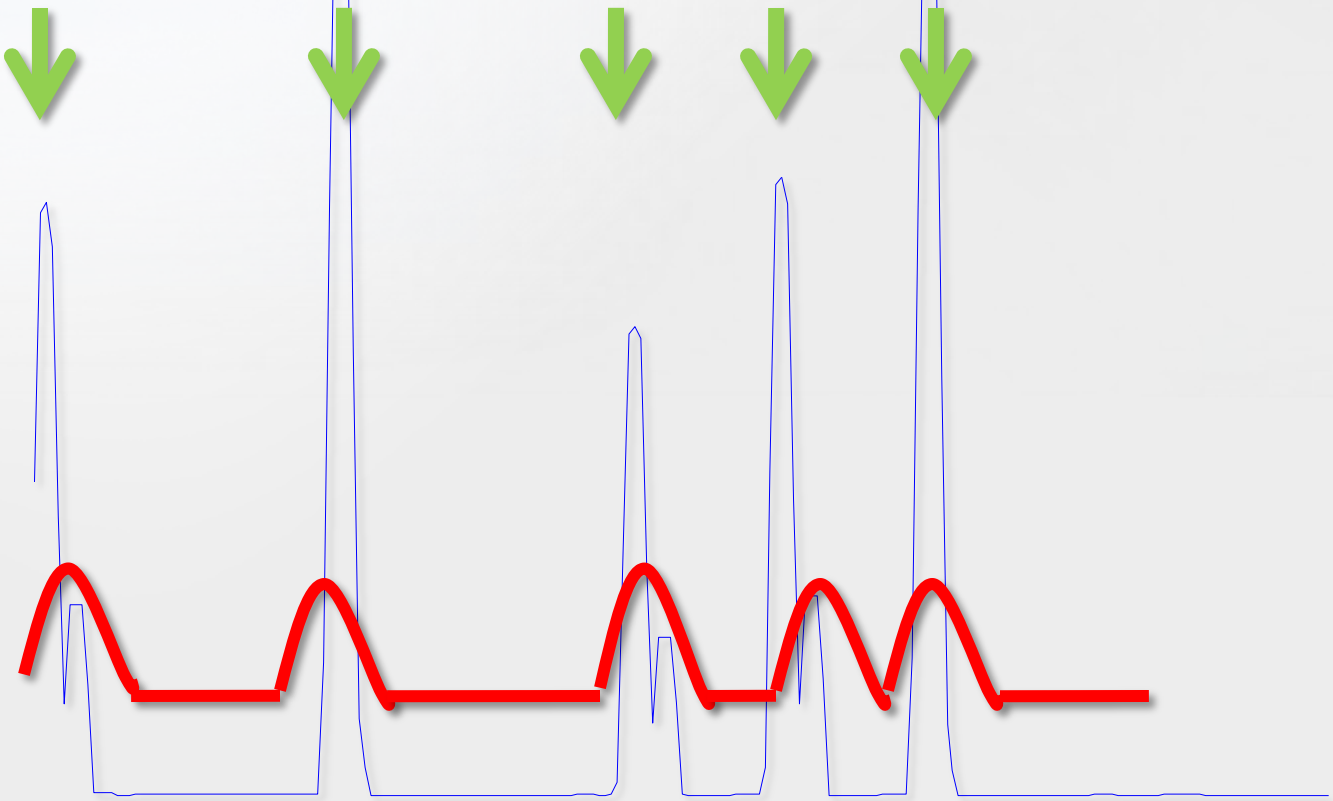

audio
Filterbank
Detect change in band
Detect change in band
...
Detect change in band
Combine results
Detection output

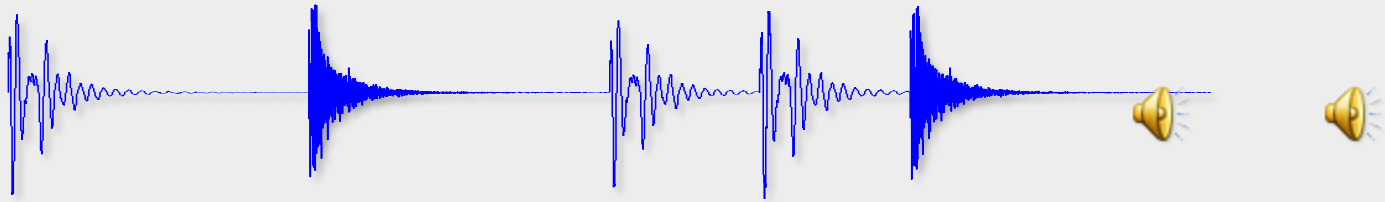
threshold





Adaptive threshold





Beat and Tempo Detection

- Beat detection
  - Tempo (e.g., 125 bpm)
    - Detecting periodicities from the onset detection curve
  - Beat
    - AKA "Tactus" – the "foot tapping rate"
    - Time-frequency analysis -> Resonators -> Probablistic model
  - Measure
    - Musical change rate
    - Harmonic change rate

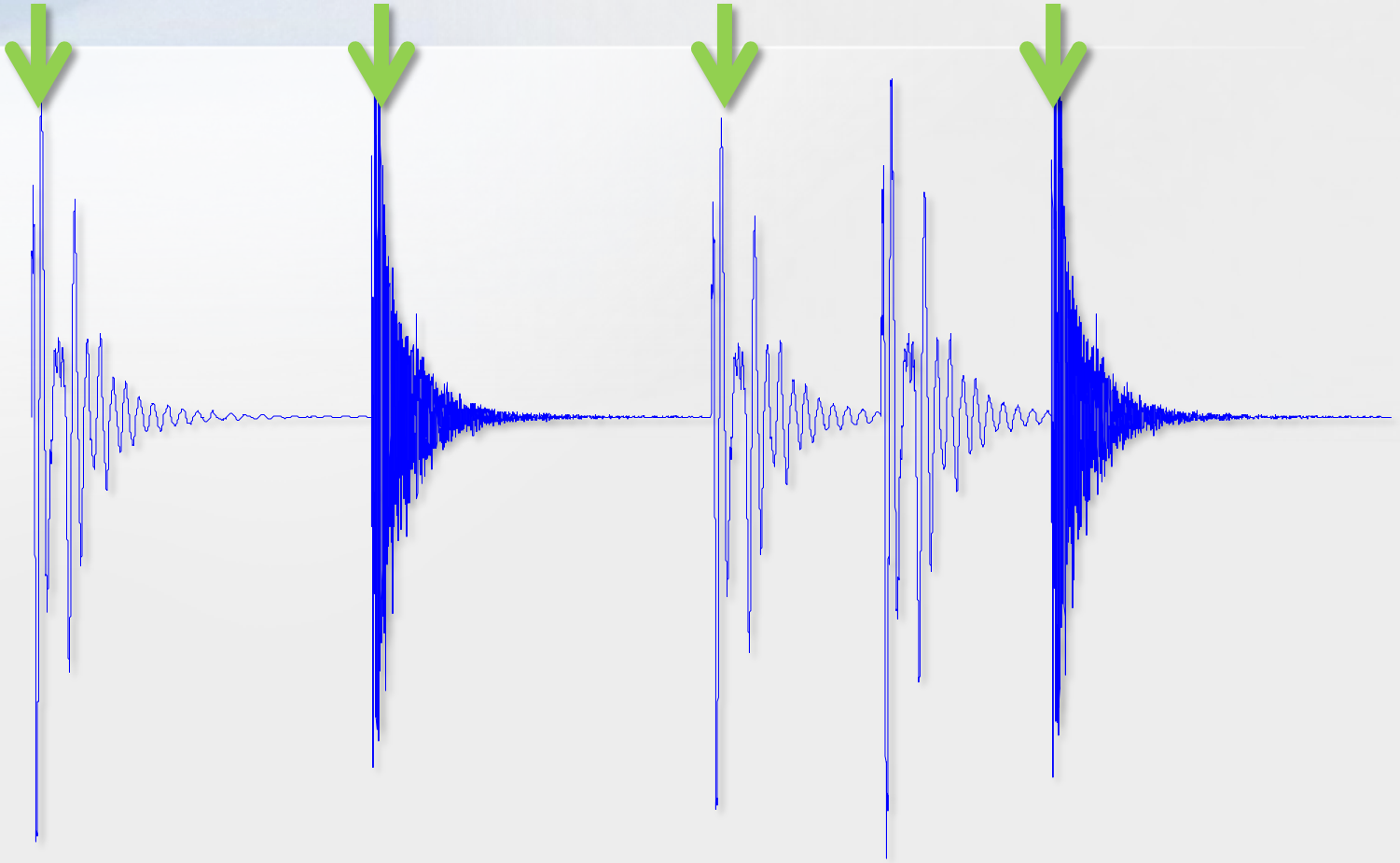
Beat and Tempo Detection

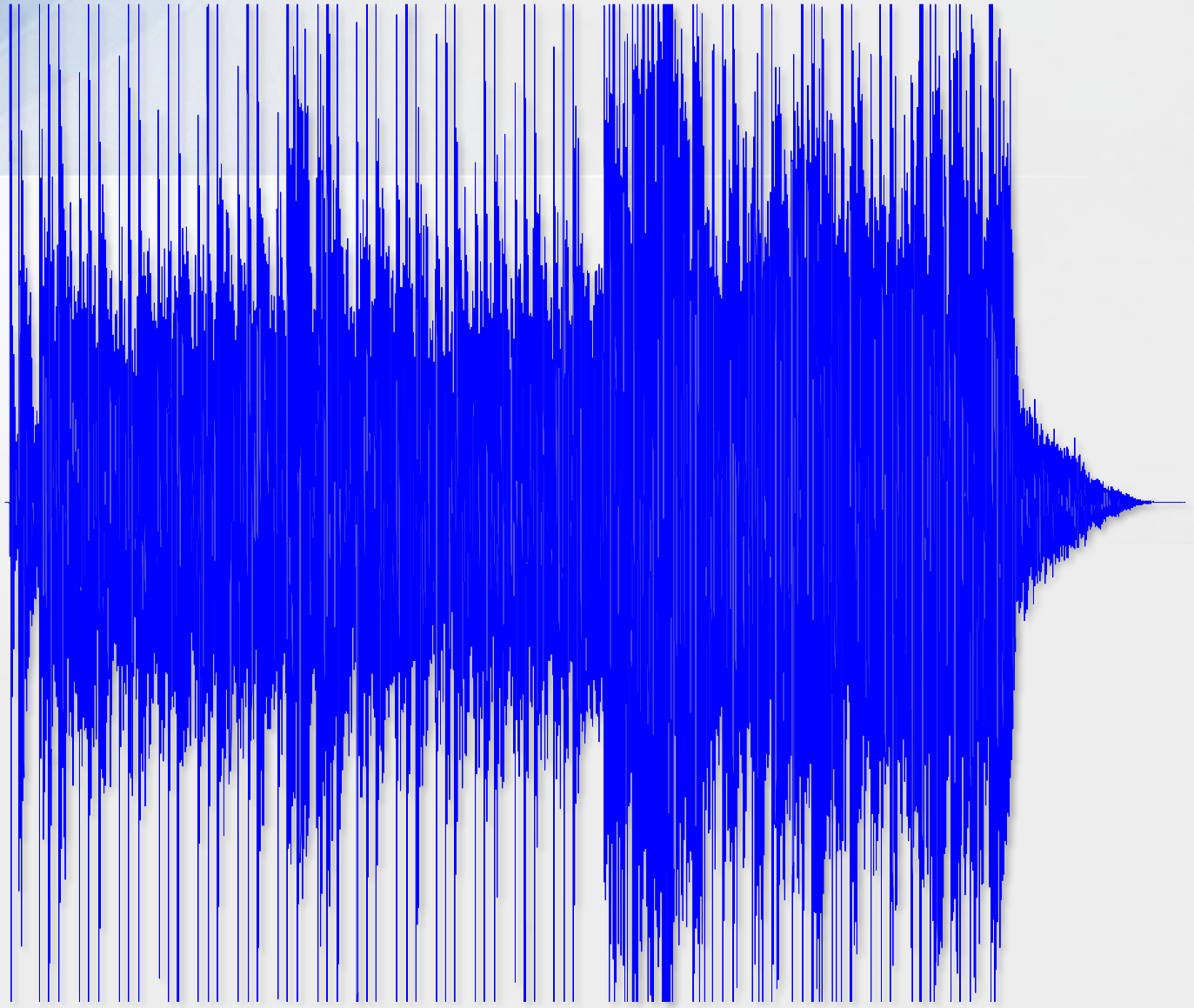
Many, many, many approaches.

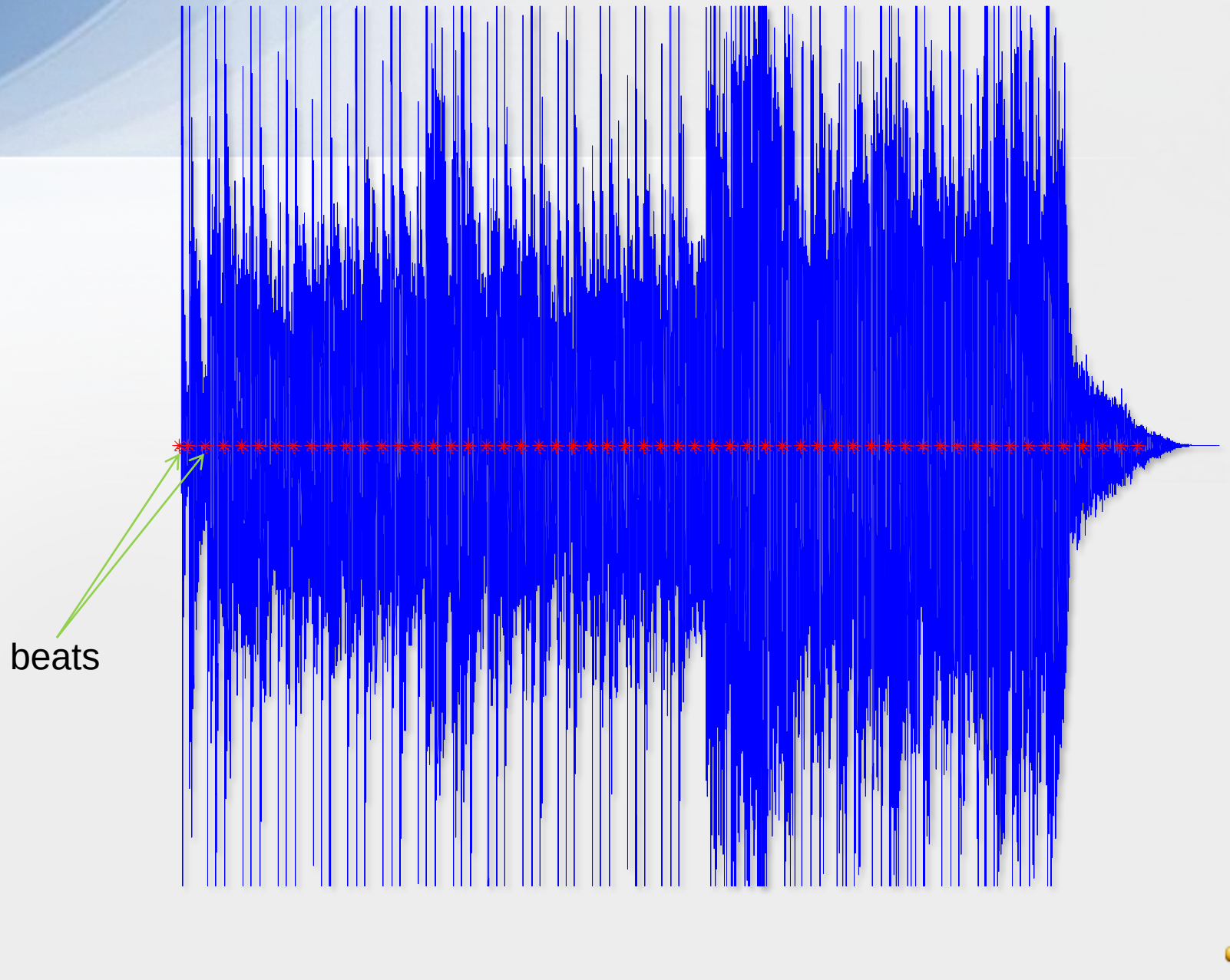
Some simple ones:

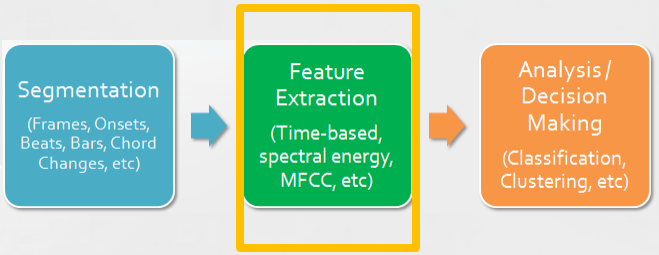
1. Autocorrelation function of the onset detection curve
2. Spectral decomposition of the onset detection curve
3. Combine both strategies: the autocorrelation function is translated into the frequency domain in order to be compared to the spectrum curve - two curves are then multiplied.

Peak picking is applied to the autocorrelation function or to the spectrum representation.

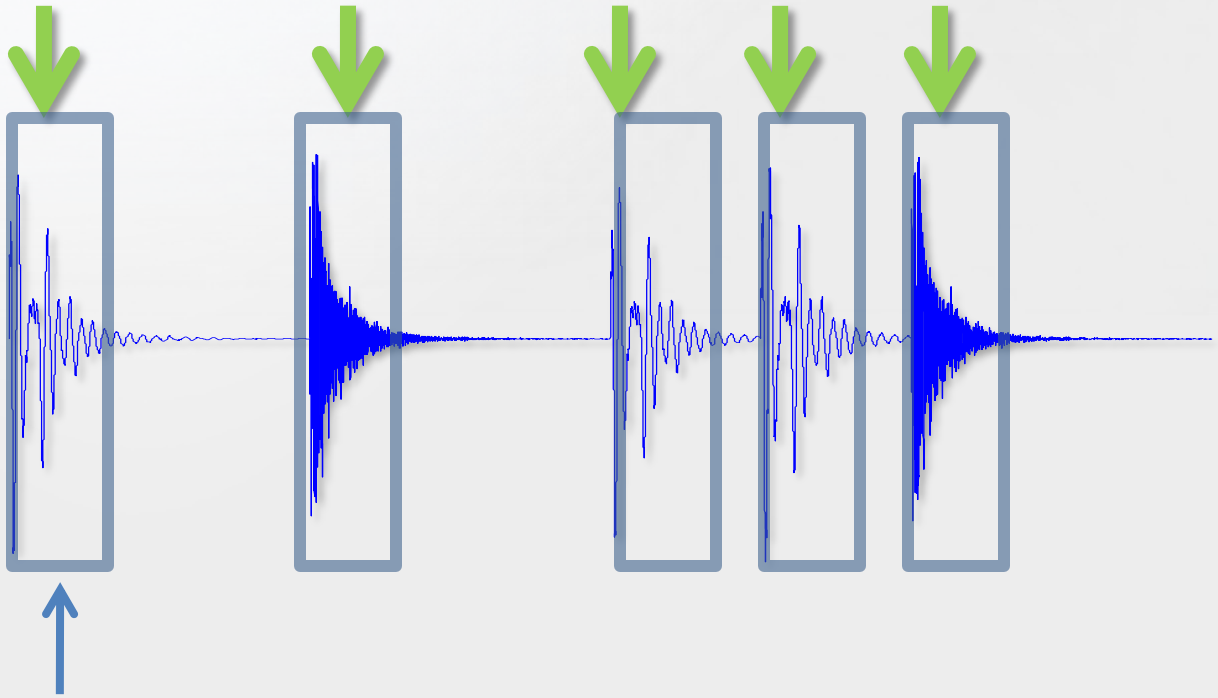





beats



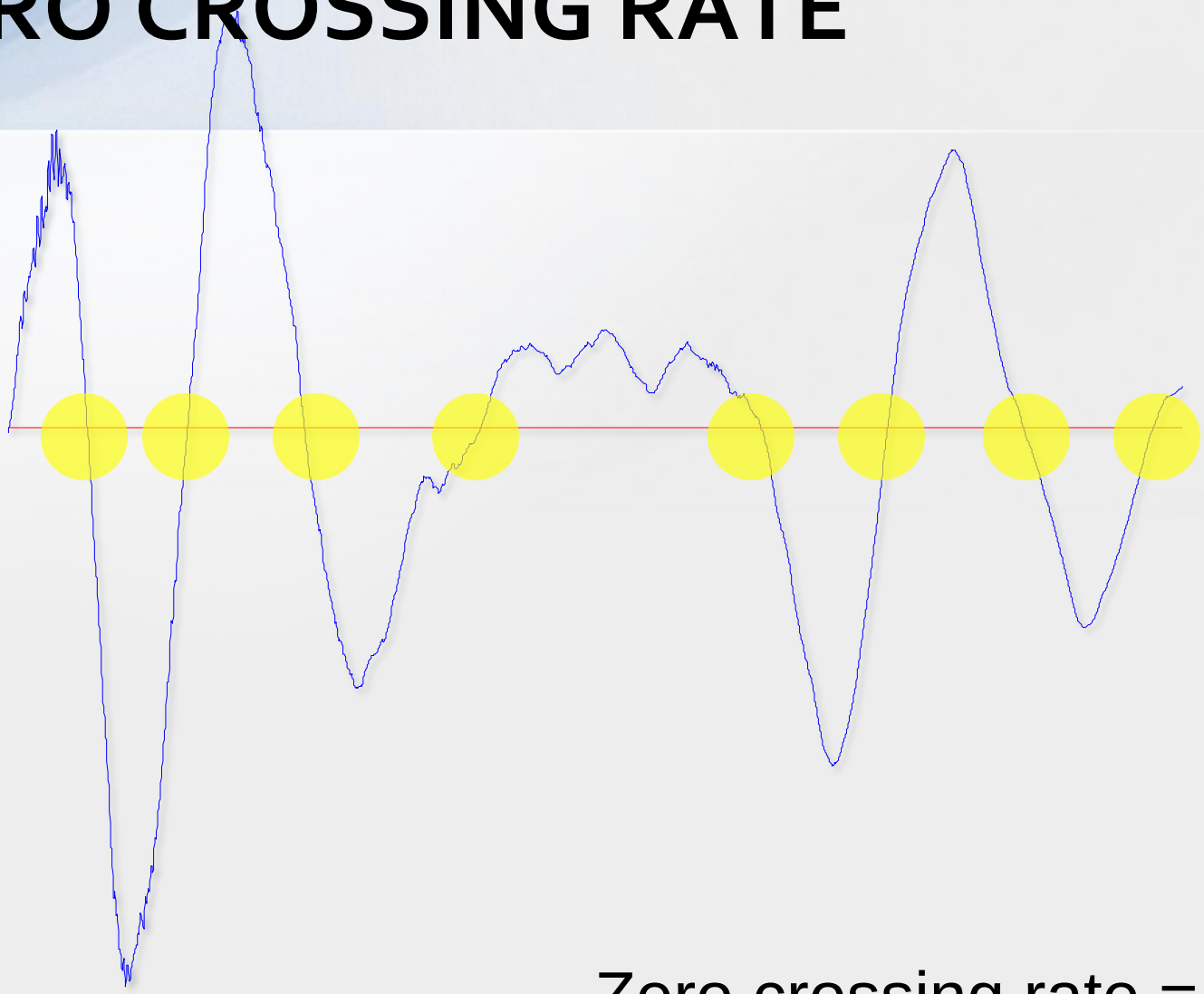
FEATURE EXTRACTION



Frame 1

FRAME 1

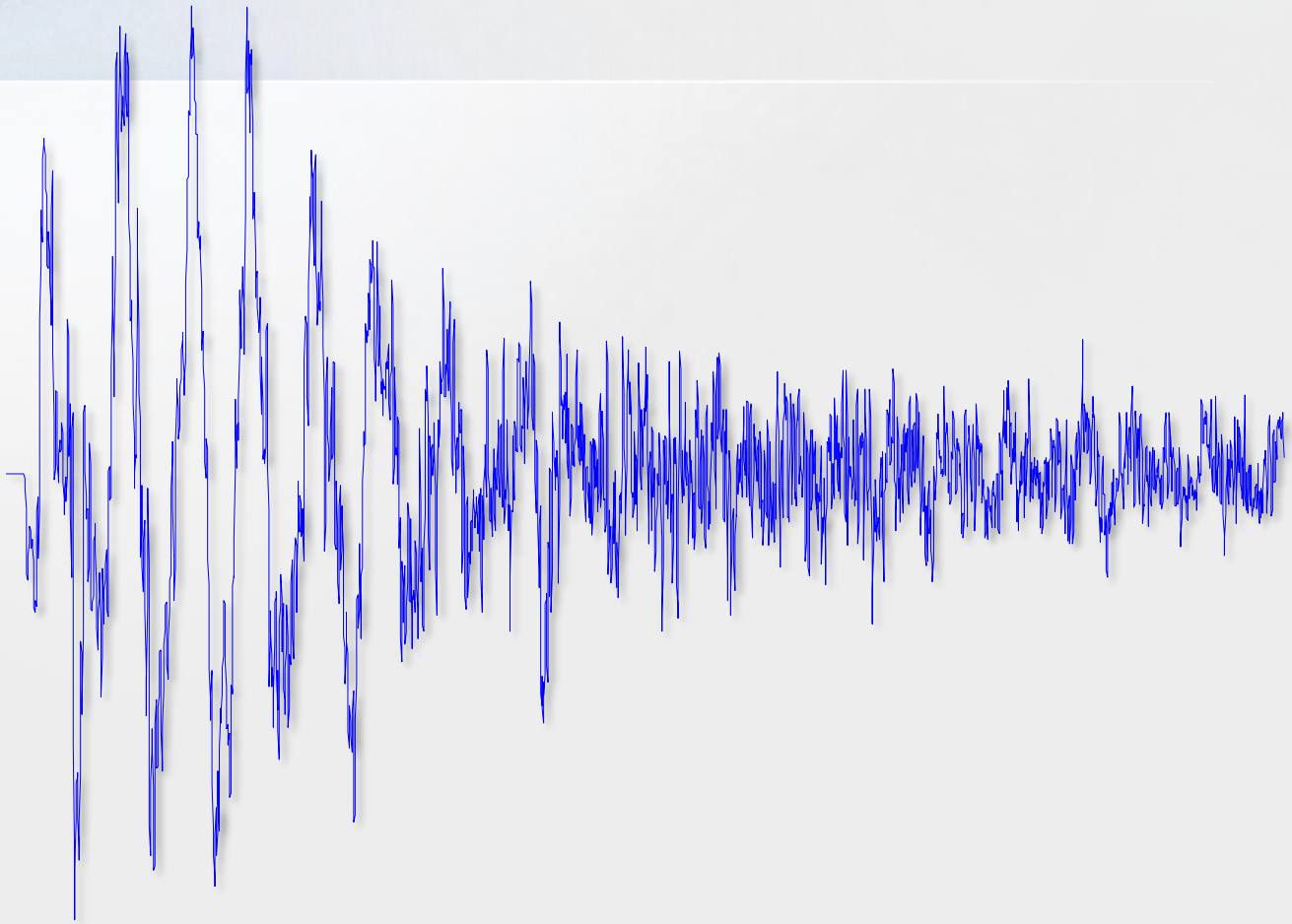
ZERO CROSSING RATE



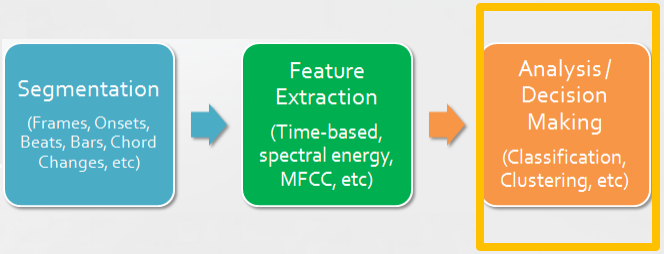
FRAME 1

Zero crossing rate = 9

Frame 2



Zero crossing rate = 423



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