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

Center for Computer Research in Music and Acoustics

Jay LeBoeuf
Leigh Smith
Steve Tjoa

June 2014

Administration

- https://ccrma.stanford.edu/wiki/MIR_workshop_2014
- Daily schedule
- Today's schedule
- Introductions
  - A little about yourself
  - Your area of interest, background with DSP, coding, python, and any specific items of interest that you'd like to see covered.
  - And one interesting fact

Why MIR?

Organize Sound

- Intelligent metadata generation for media collections.
- Discovery and monetization of massive media archives.

Search Sound

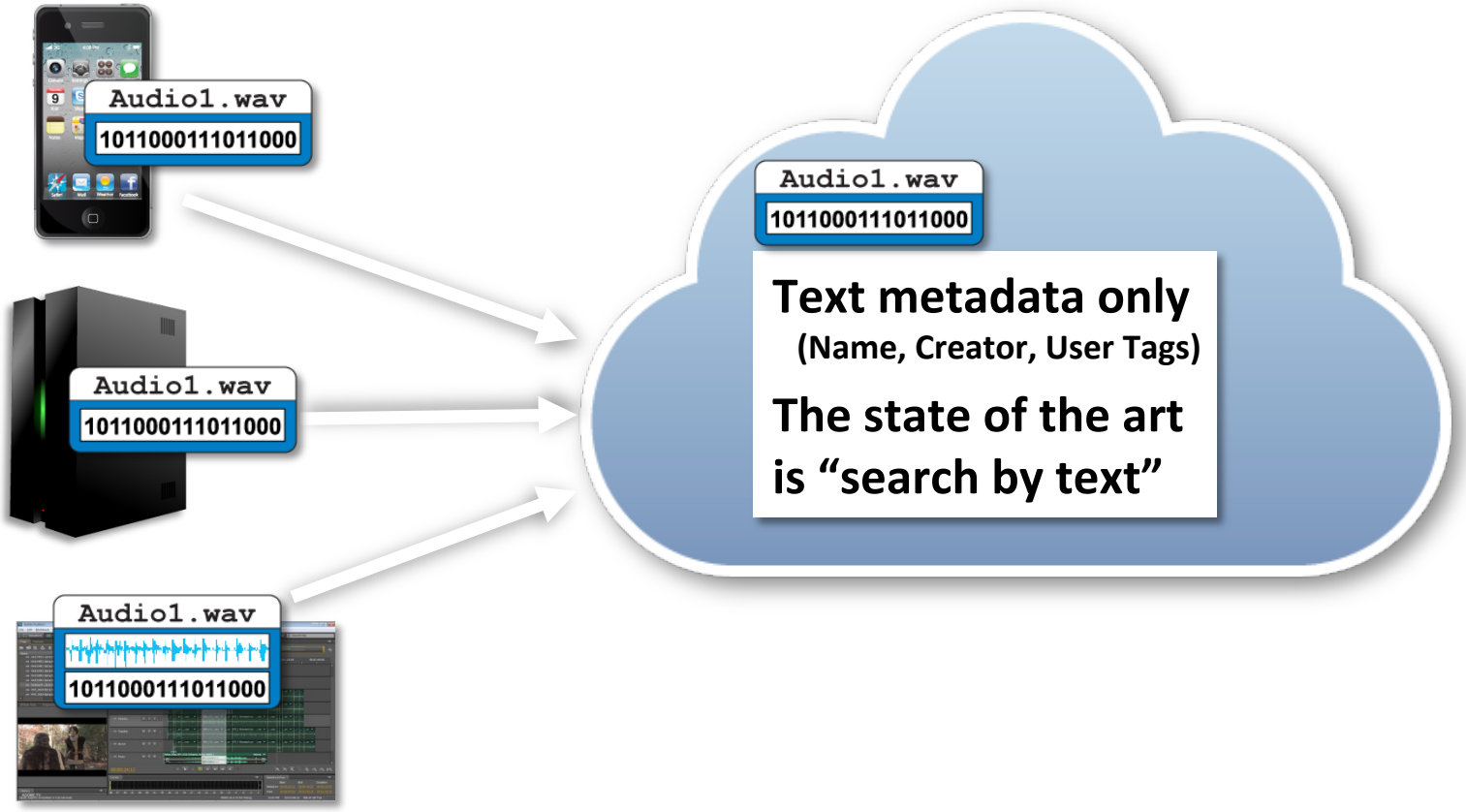
- “Find me something that sounds like this” search engine for audio content (songs, real-world sounds, music loops, speech)
- Music-similarity search

Understand Sound

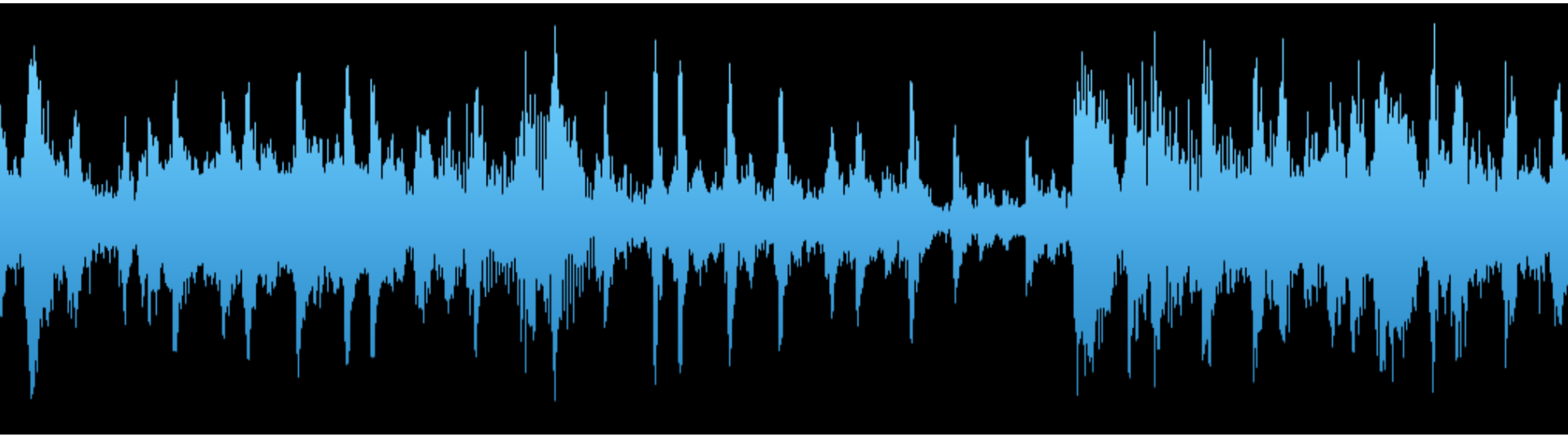
- Revolutionary workflows in consumer products through machine hearing. Automatic control of software, signal processing and mobile devices

Problems

1. Computers are deaf.
2. Content is overwhelming and unsearchable.



Audio Visualization




serato dj intro
SETUP
SYNC
OFF
Valley Dub
BPM
126.00
Subwall
KEY
G#
TIME ELAPSED
01:14.7
PLAYBACK
KEYLOCK
LOOP
REL PITCH
RANGE
SINGLE
CONT
OFF
ON
4
+0.80
±8
Out of It
BPM LOCKED
126.00
01:51.8
0.00
DJ-FX
Samples
FX 1
HPF
LPF
FLANGER
DEPTH
BEATS MULTIPLIER
3/2
TEMPO SOURCE
DECK 1
FX 2
PHASER
ECHO
REVERB
DECK 2
Files
Prepare
All...
A List
Electronic
Drum and Bass
Dub
House
Favourites
Crowd
Personal
International
Local
song artist album length comment bpm
17 Lights Out Solar Flower Moves On 04:42.93 97
23 Out of it Subwall Self Titled 06:21.20 126
35 Valley Dub Subwall Self Titled 06:31.76 125
4 Batman you're Fying Suddenly Loose Northern Lights 04:18.32 103
14 from there to here and back again or not Suddenly Loose Northern Lights 05:40.53 139
16 Lesser Thought Process Taste Foundation Lesser Thought Process 06:54.41 132
31 The Test Taste Foundation 07:30.46 129
22 Nymphs The 88 Get Dancing 05:03.46 136
8 Defend The Guilty 02:52.04 128
9 Dignity Trade Slater My Only Salvation 02:59.10 97
28 St Jervous dance Yesterdays Rebel Moon Mountain Sounds 07:44.88 125
Album Art

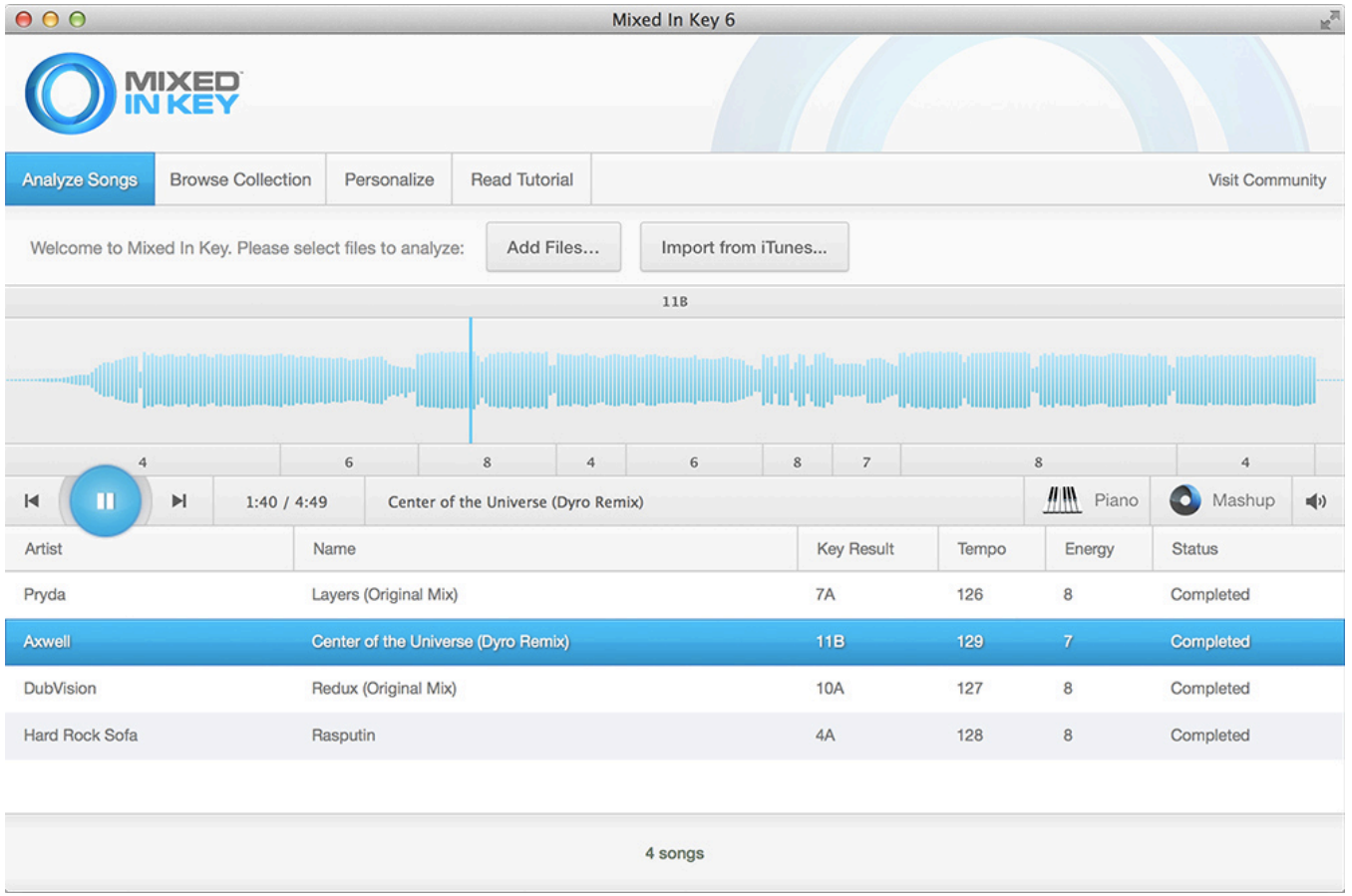
Simplifying User Experience

Detection algorithms help musicians achieve their desired results quicker.

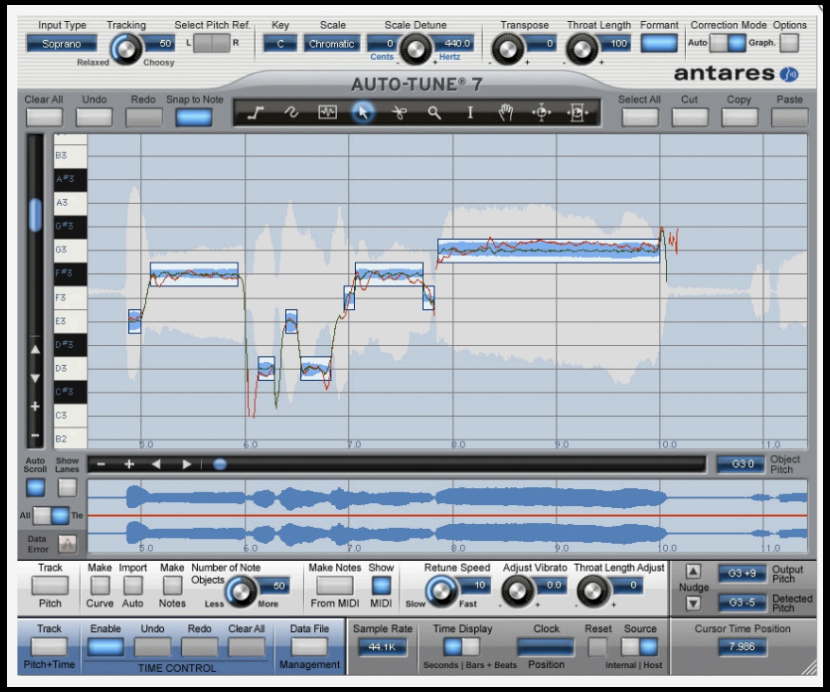
Harmony / Chord / Key Estimation

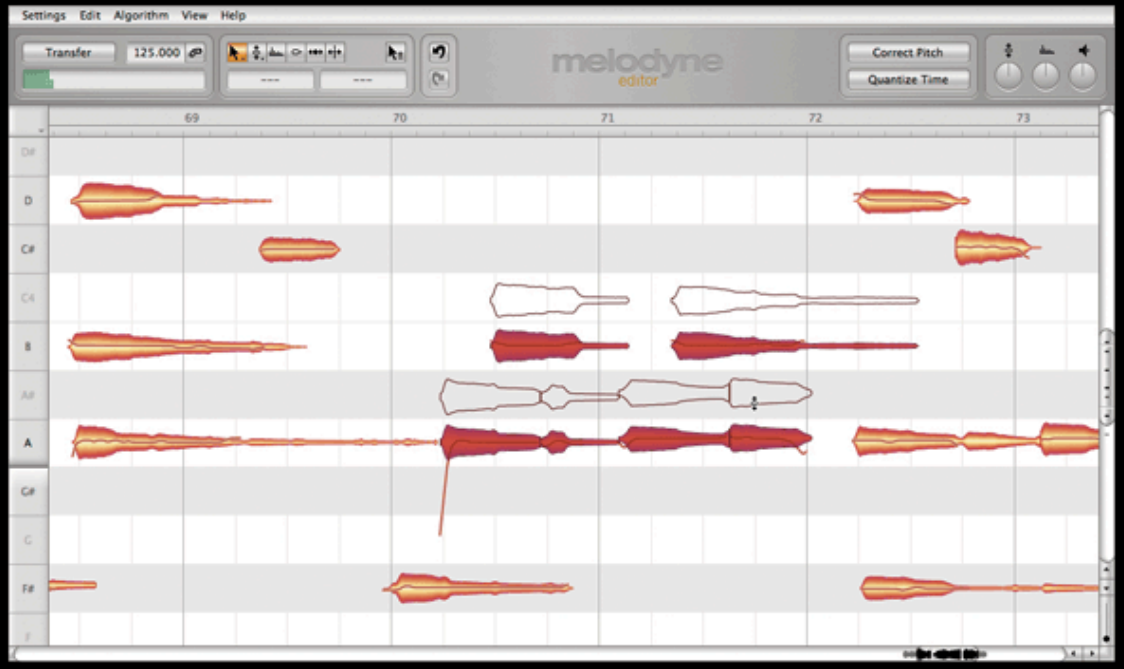

Riffstation | Jimi Hendrix – Hey Joe
Zoom
C
G
D
A
E
Scroll
Jam Master
Riff Builder
Chord Viewer
CHORDS DETECTED
E
A
D
G
C
X O O
1 3 2
D
X O O
2 1 3
A
NEXT CHORD
2
KEY SHIFT
0
TEMPO
164.83
CHORD TYPE
Open
Power
NECK TYPE
CAPO
0
CHORD FINDER
A
Major
X O O
2 1 3
A
METRONOME
1:32.565 / 3:29.813
VOLUME

Mixed In Key


Mixed In Key 6
MIXED IN KEY
Analyze Songs
Browse Collection
Personalize
Read Tutorial
Visit Community
Welcome to Mixed In Key. Please select files to analyze:
Add Files...
Import from iTunes...
11B
4 6 8 4 6 8 7 8 4
1:40 / 4:49
Center of the Universe (Dyro Remix)
Piano
Mashup
Artist | Name | Key Result | Tempo | Energy | Status
Pryda | Layers (Original Mix) | 7A | 126 | 8 | Completed
Axwell | Center of the Universe (Dyro Remix) | 11B | 129 | 7 | Completed
DubVision | Redux (Original Mix) | 10A | 127 | 8 | Completed
Hard Rock Sofa | Rasputin | 4A | 128 | 8 | Completed
4 songs


Manual Pitch Editor


AUTO-TUNE® 7
antares


melodyne
editor


Track
Preset
Auto
Vocals
iZotope Nectar 2
<factory default>
COMPARE
BYPASS
SAFE
Native
MIDI Node: iZotope Nectar 2 MIDI In
MODULES
Pitch
Gate
Compressors
Harmony
Saturation
EQ
De-Esser
Limiter
FX
Reverb
Delay
Voices
Down
Up
Solo Voices
MIDI Control
Variation
Pitch Correction
Amount
Pan & Gain
Pitch & Delay
Scale
Root
Type
Minor
Calibration (A)
440.0 Hz
Key Detection
Start Detection
Suggested Keys
C Major
C Minor
F Minor
Advanced
Mixing Mode
Solo
Bypass
Options
History
Reset
Track Name
Vocals
Presets • Female Vocal
iZotope
Bypass
BEFORE

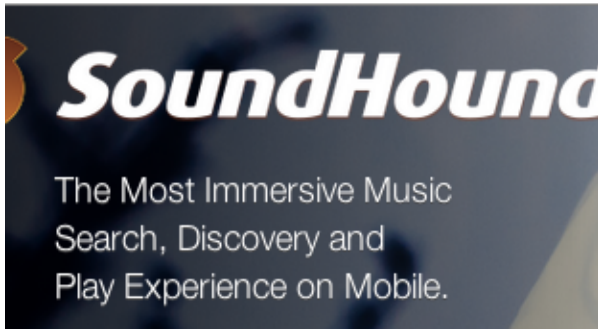
Drum Transcription

	Original	Transcription
Example 1		
Example 2		

Pitch Tracking / Query by Humming

QBH - SoundHound

Madonna - https://soundcloud.com/madonnagr/secret-acapella-lead-vox



https://www.youtube.com/watch?v=QnRt_cWdV8c

https://www.youtube.com/watch?v=IArxakPsPE0

Discovery

Find me sound effects that sound like this?

find similar

BREAKTWEAKER PRESET Zk-Chi Rap SAVE NEW MIDI IN Intensity Gain

SEQUENCER

Tom Smash
Taiko Hits
Wild Hats
Snarez
Big Kick
Epic Bass

1/2X SPEED -1.0 GAIN C PAN

1.1 1.2 1.3 1.4 2.1 2.2

GENERATOR TYPE Sample

SAMPLE

BT-Film Taiko Reverb ... DISCOVER 0 st

START 0 smp END 118125

0 50000

PLAYBACK MODE Forward LOOP 0 smp X-FADE

BT-Film Chase Scene Drum 2.wav
BT-Film Taiko Reverb Rim Pummel 4.wav
BT-Film Taiko 3.wav
BT-Film Cellar Drum 2.wav
BT-Kick Conjunction.wav
BT-Kick Lightweight.wav
BT-Tom Low Free Jazz.wav
JF-Kick Tech 2.wav
BT-Film Cellar Drum 1.wav
BT-Kick Jaxx Pop.wav
Cancel OK

DISTORTION FILTER MIX

New York

LOW PASS
HIGH PASS
BAND PASS

L1 676.2 10023.4 Cutoff

1.0 Res

PRE-DISTORTION FILTER

ENVELOPE LFO

WAVEFORM SAMPLE & HOLD

1 2 3 4

SYNC RESTART INVERT

0.5 RATE 1.00 DEPTH 0.0 PHASE 0 ATTACK

GENERATOR TYPE None ENVELOPE LFO

GENERATOR TYPE None ENVELOPE LFO

SELECT A SINGLE STEP TO SHOW MICROEDIT SETTINGS

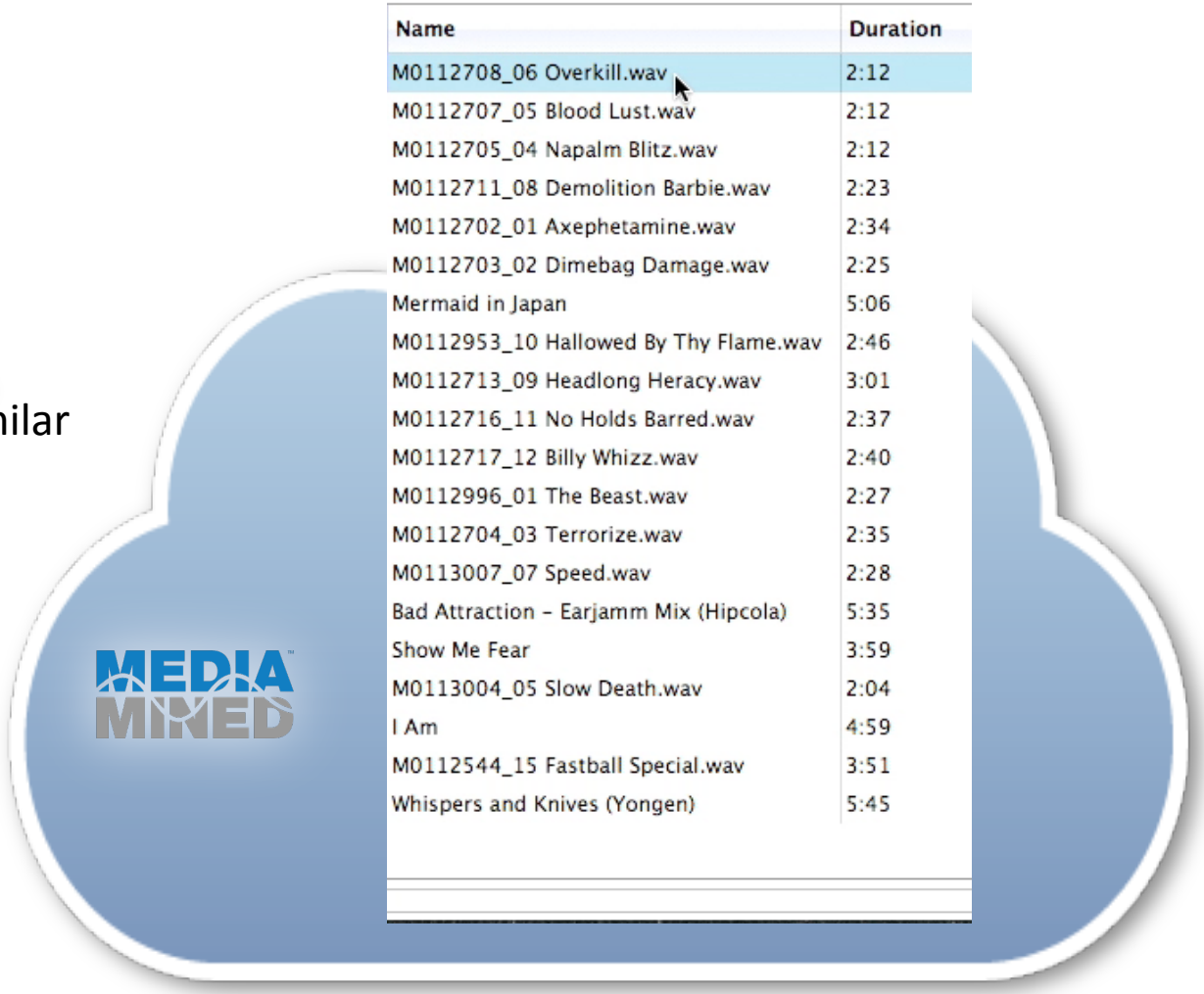
PREVIEW 120.0 SYNC TEMPO GATE LATCH ALWAYS RETRIGGER MIDI PATTERN 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

iZotope | bt


BREAKTWEAKER
PRESET
Zk-Chi Rap
SAVE NEW
MIDI IN
Intensity
Gain
SEQUENCER
1/2X
SPEED
-1.0
GAIN
C
PAN
Tom Smash
Taiko Hits
Wild Hats
Snarez
Big Kick
Epic Bass
GENERATOR TYPE
Sample
SAMPLE
BT-Film Taiko Reverb ...
DISCOVER
0 st
START 0
smp
END 118125
0
50000
PLAYBACK MODE
Forward
LOOP
0 smp
X-FADE
BT-Film Chase Scene Drum 2.wav
BT-Film Taiko Reverb Rim Pummel 4.wav
BT-Film Taiko 3.wav
BT-Film Cellar Drum 2.wav
BT-Kick Conjunction.wav
BT-Kick Lightweight.wav
BT-Tom Low Free Jazz.wav
JF-Kick Tech 2.wav
BT-Film Cellar Drum 1.wav
BT-Kick Jaxx Pop.wav
Cancel
OK
DISTORTION
FILTER
MIX
New York
LOW PASS
HIGH PASS
BAND PASS
L1
676.2
10023.4
Cutoff
1.0
Res
PRE-DISTORTION FILTER
ENVELOPE
LFO
WAVEFORM
SAMPLE & HOLD
SYNC
RESTART
INVERT
0.5
RATE
1.00
DEPTH
0.0
PHASE
0
ATTACK
GENERATOR TYPE
None
GENERATOR TYPE
None
SELECT A SINGLE STEP TO SHOW MICROEDIT SETTINGS
PREVIEW
120.0
SYNC
TEMPO
GATE
LATCH
ALWAYS RETRIGGER
MIDI
PATTERN
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 23 24
iZotope
bt

Music recommendation

find similar



Name	Duration
M0112708_06 Overkill.wav	2:12
M0112707_05 Blood Lust.wav	2:12
M0112705_04 Napalm Blitz.wav	2:12
M0112711_08 Demolition Barbie.wav	2:23
M0112702_01 Axephetamine.wav	2:34
M0112703_02 Dimebag Damage.wav	2:25
Mermaid in Japan	5:06
M0112953_10 Hallowed By Thy Flame.wav	2:46
M0112713_09 Headlong Heracy.wav	3:01
M0112716_11 No Holds Barred.wav	2:37
M0112717_12 Billy Whizz.wav	2:40
M0112996_01 The Beast.wav	2:27
M0112704_03 Terrorize.wav	2:35
M0113007_07 Speed.wav	2:28
Bad Attraction - Earjamm Mix (Hipcola)	5:35
Show Me Fear	3:59
M0113004_05 Slow Death.wav	2:04
I Am	4:59
M0112544_15 Fastball Special.wav	3:51
Whispers and Knives (Yongen)	5:45

Rhythmic Similarity

Create by Example

Here's my seed riff.

Find me a Bass loop that grooves with that.

Find me a Drum loop that grooves with that.

Play them all together.

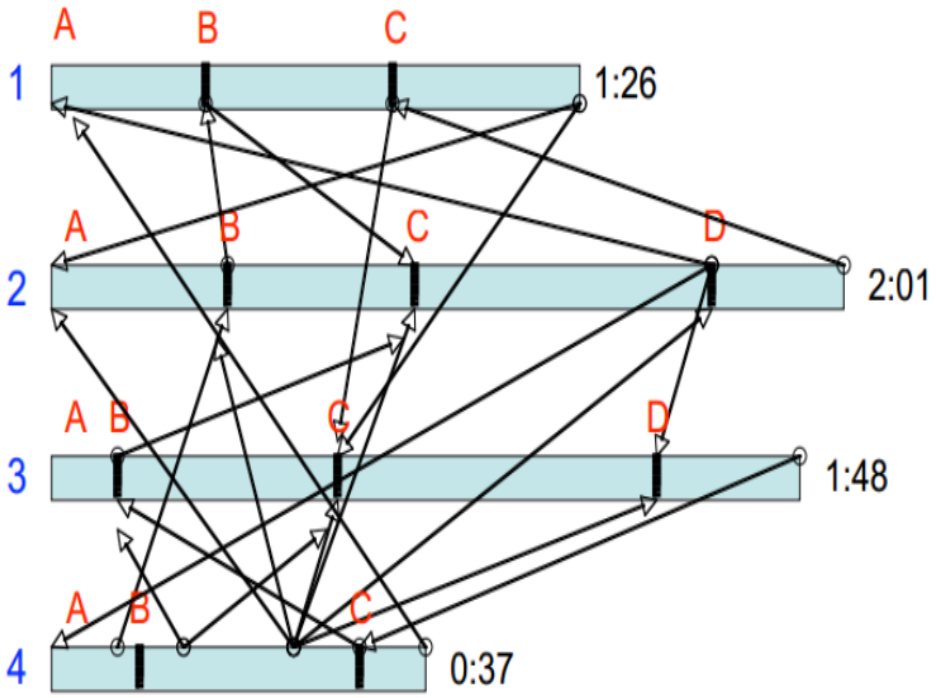
Dynamic Media

James Brown – The Payback

Original

Groove

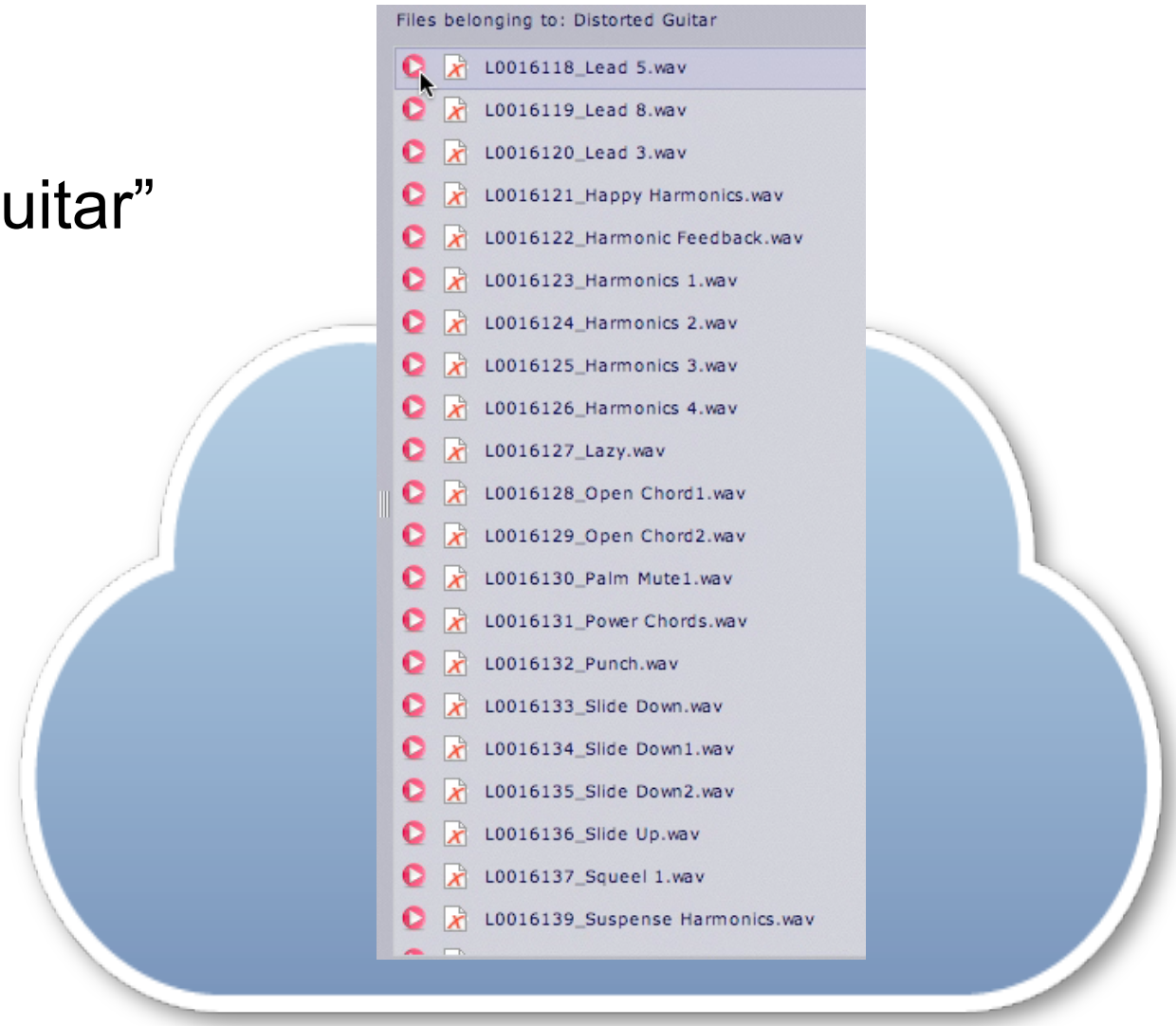
Remixes



Graphic: Clint Bajakian

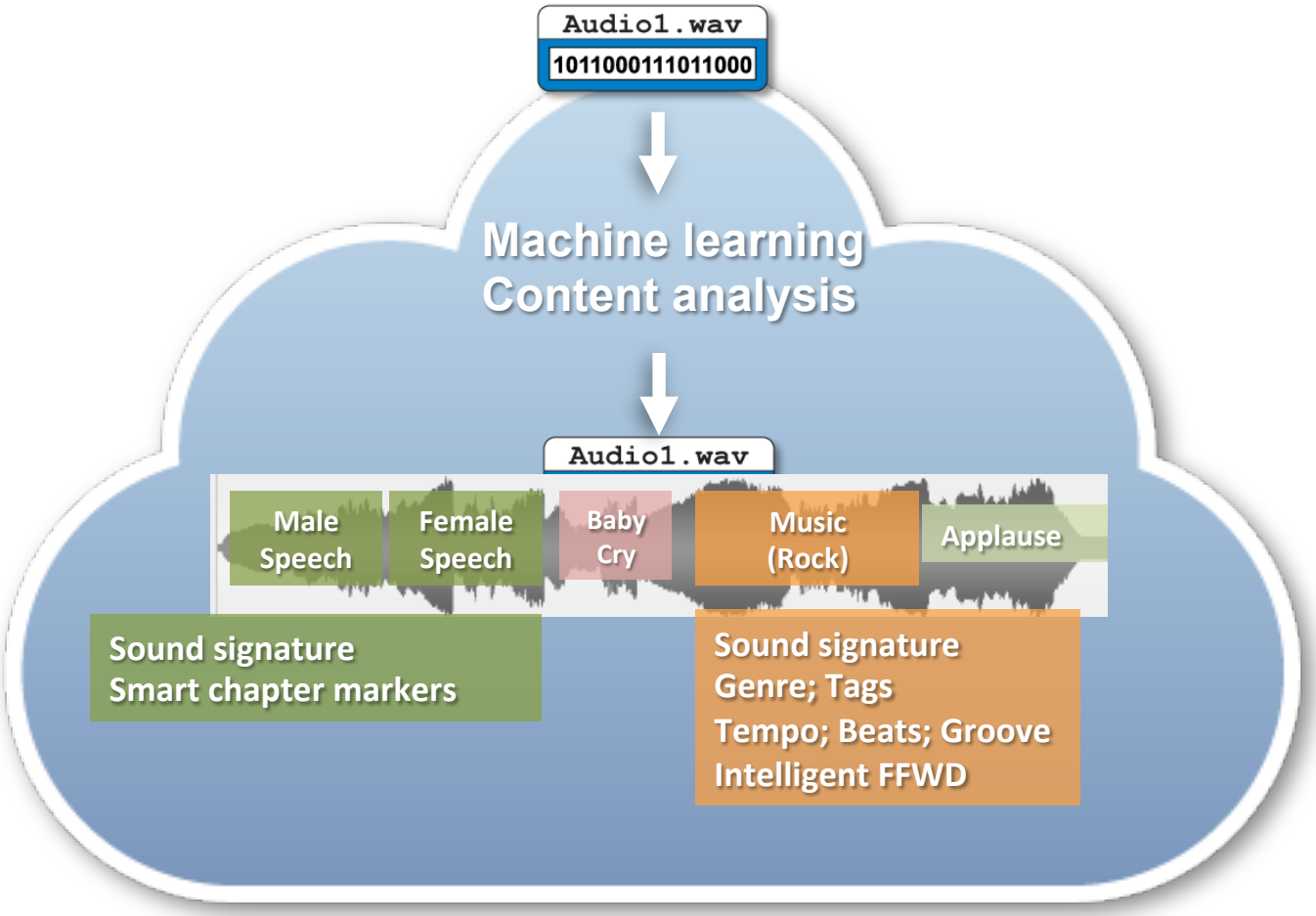
Tagging

“distorted, guitar”



Tagging

Tagging


Audio1.wav
1011000111011000
Machine learning
Content analysis
Audio1.wav
Male Speech
Female Speech
Baby Cry
Music (Rock)
Applause
Sound signature
Smart chapter markers
Sound signature
Genre; Tags
Tempo; Beats; Groove
Intelligent FFWD

Smarter DSP

ROAD TRAFFIC NOISE

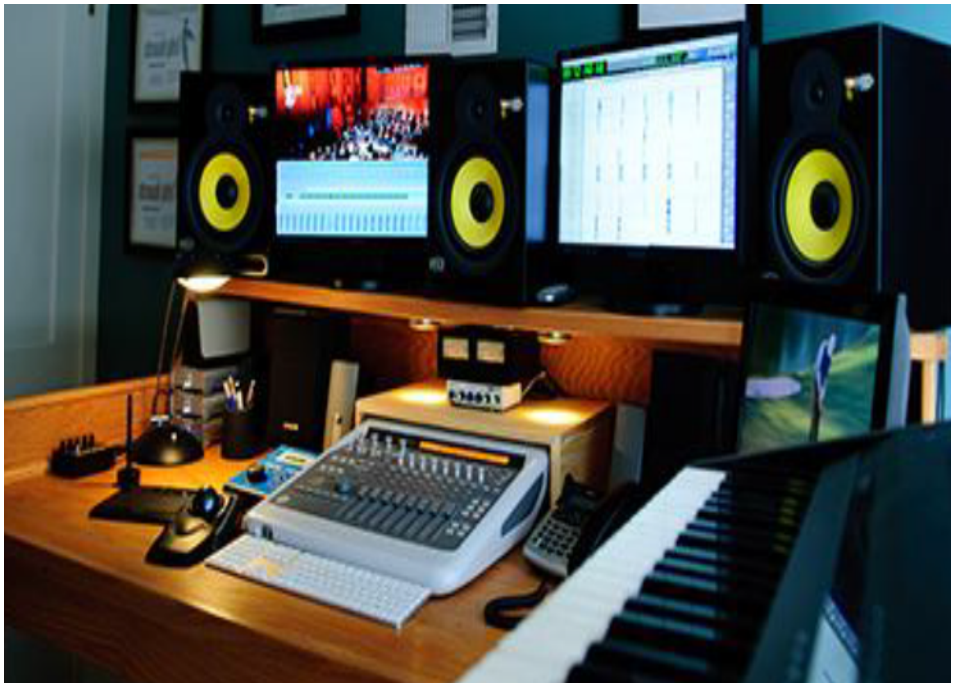
ANR-B

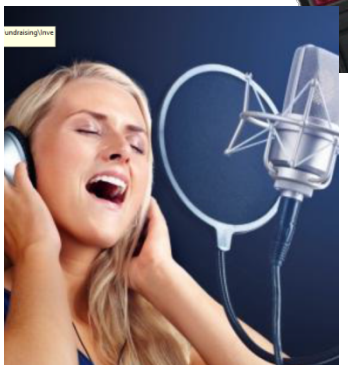
Adaptive real-time noise reduction

DeBreath

http://youtu.be/DZJfURZ12AM?t=7m15

Auto-mixing





Before After

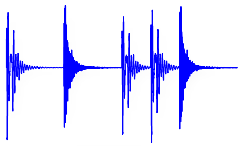
For example: https://www.landr.com/#/

Why MIR?

- content-based querying and retrieval, indexing (tagging, similarity)
- fingerprinting and digital rights management
- music recommendation and playlist generation
- music transcription and annotation
- score following and audio alignment
- automatic classification
- rhythm, beat, tempo, and form
- harmony, chords, and tonality
- timbre, instrumentation
- genre
- emotion, style, and mood analysis
- music summarization

BASIC SYSTEM OVERVIEW

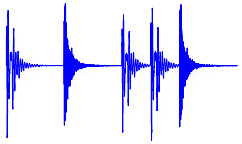
Basic system overview



Segmentation

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

Basic system overview



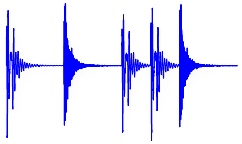
Segmentation

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

Feature Extraction

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

Basic system overview



Segmentation

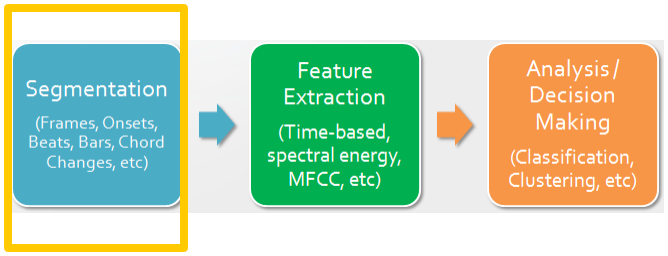
(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction

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

Analysis / Decision Making

(Classification, Clustering, etc)

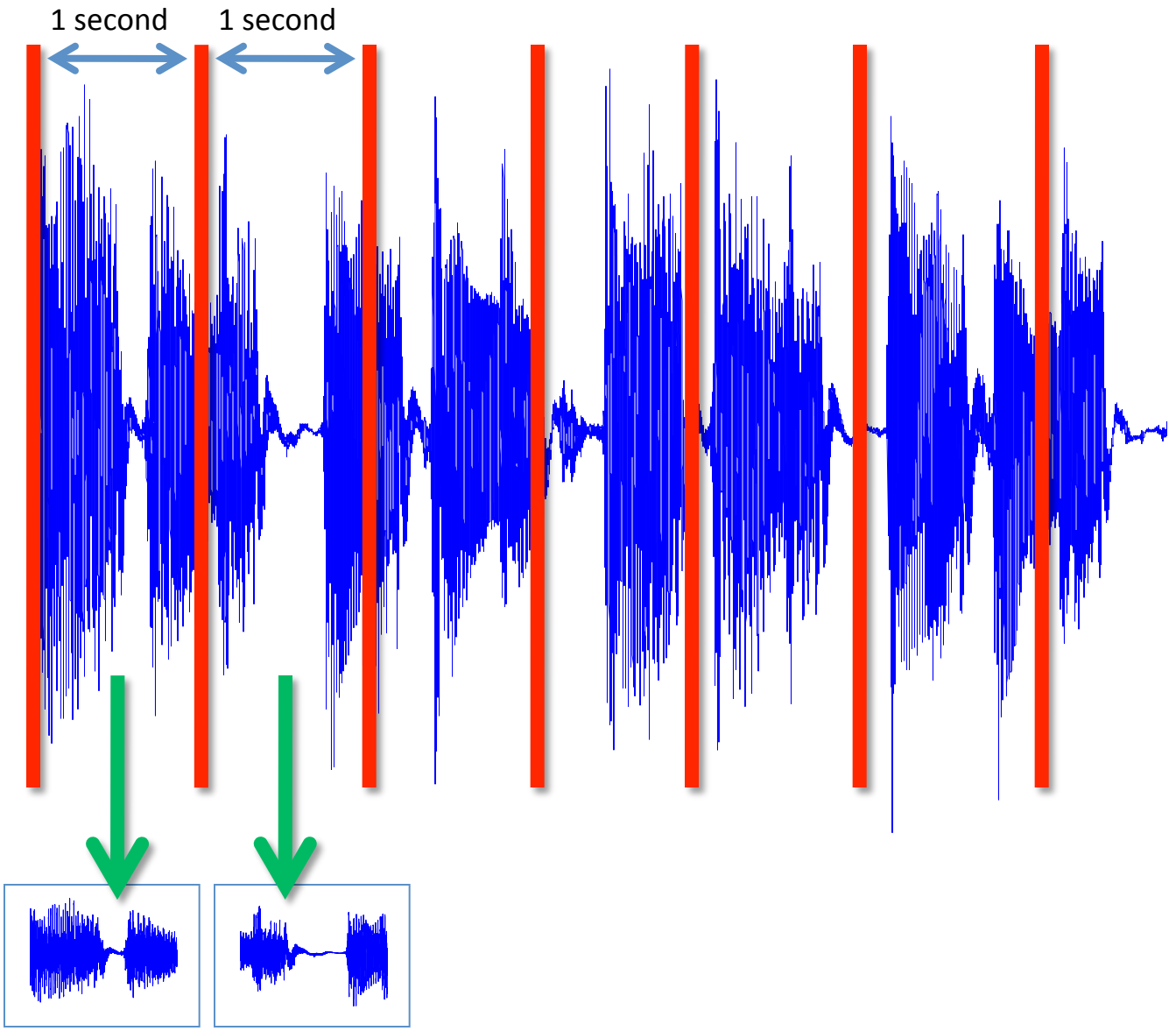


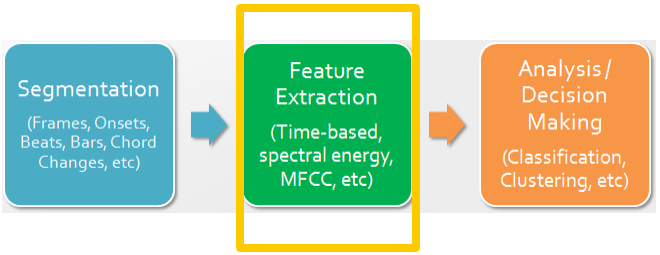
TIMING AND SEGMENTATION

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

Frames

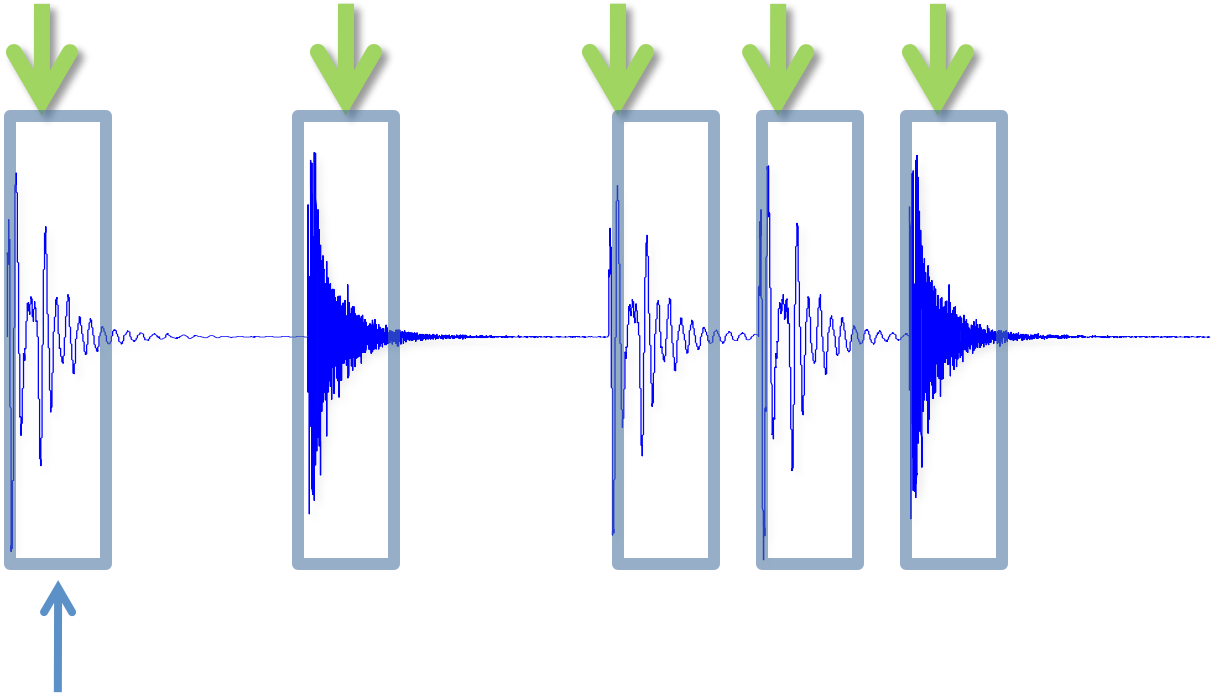




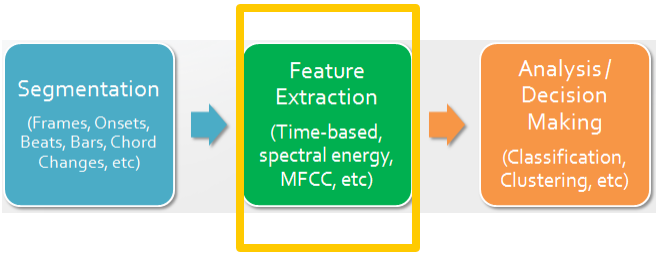
FEATURE EXTRACTION

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

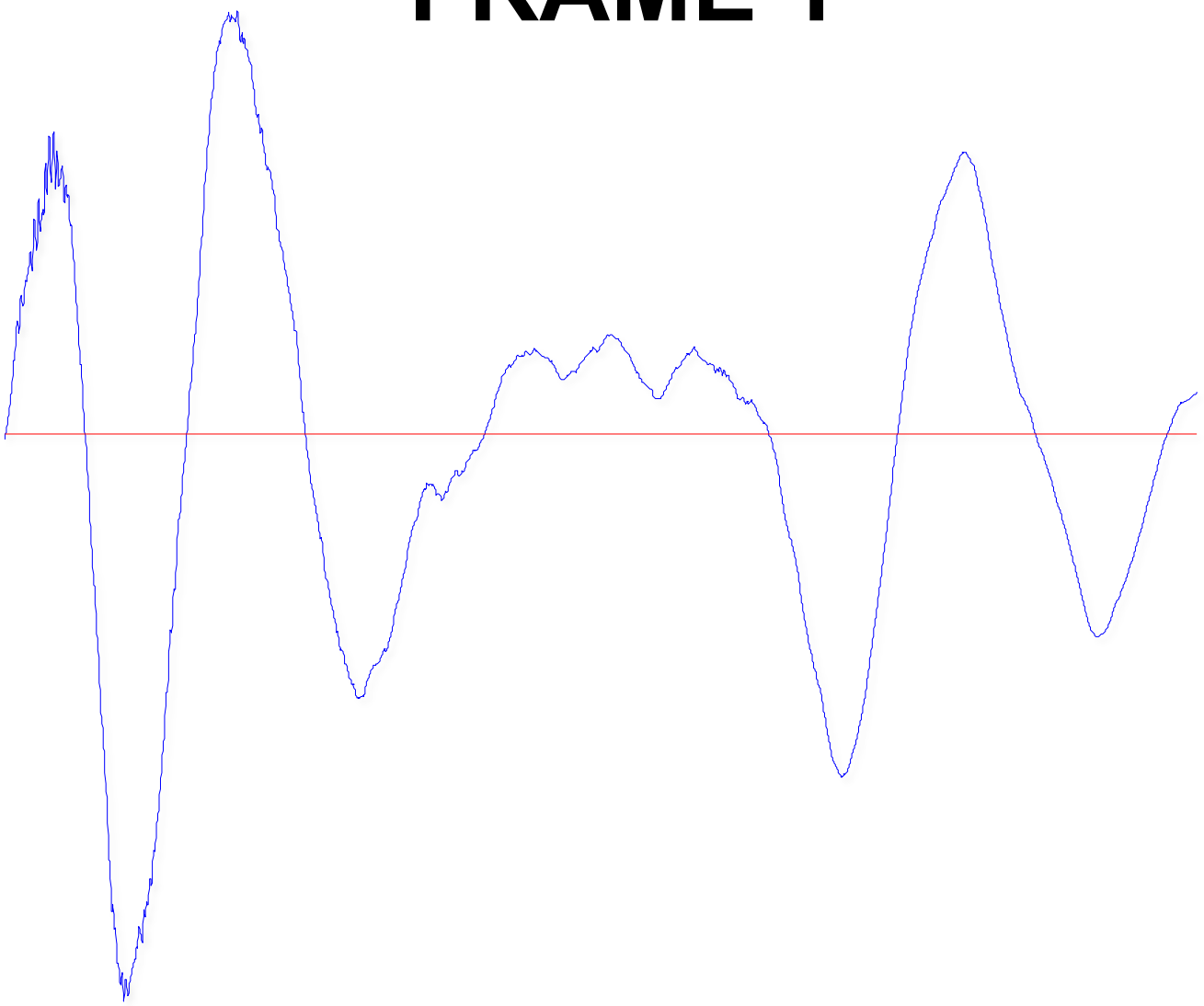


Frame 1

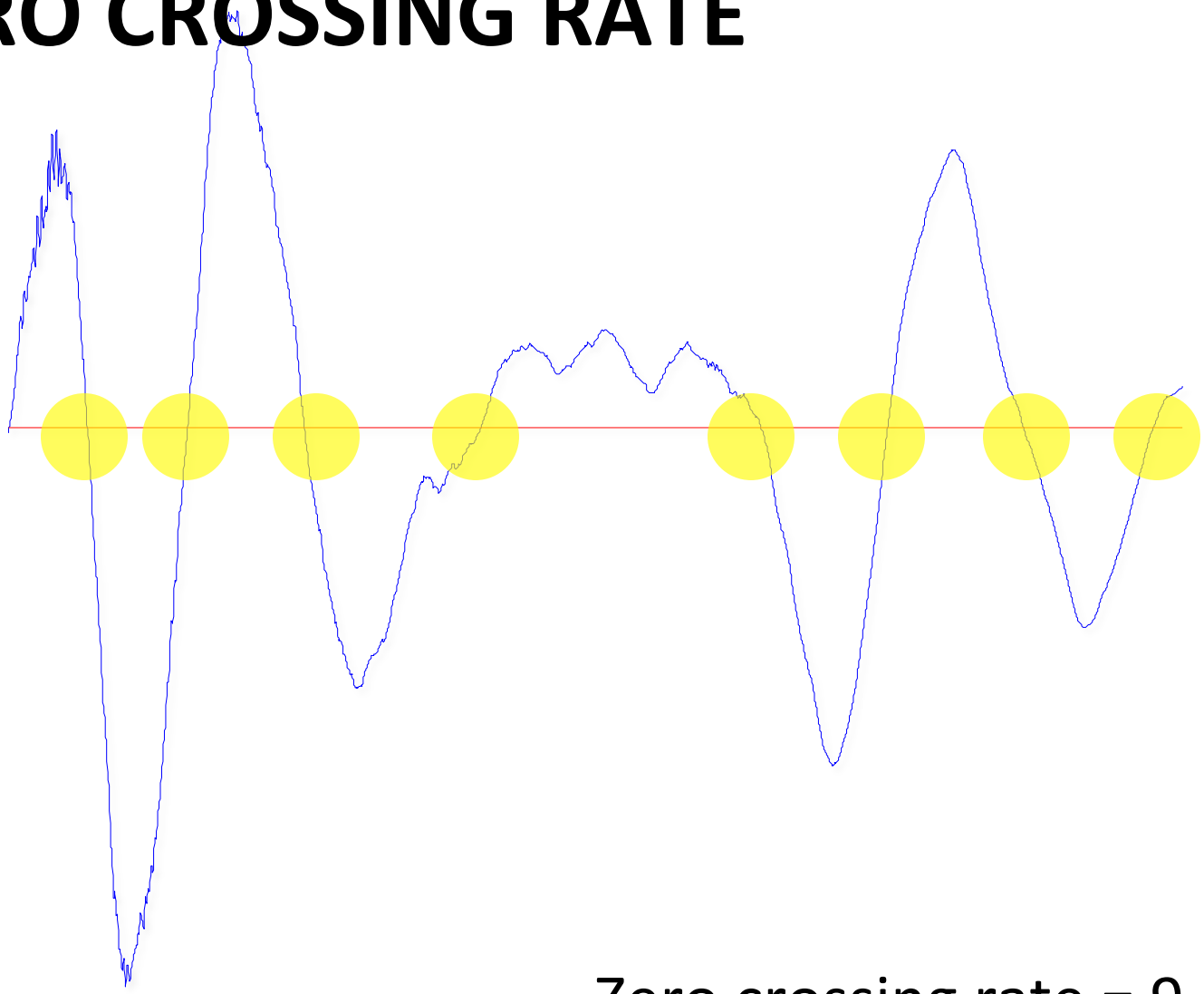


FEATURE EXTRACTION

FRAME 1



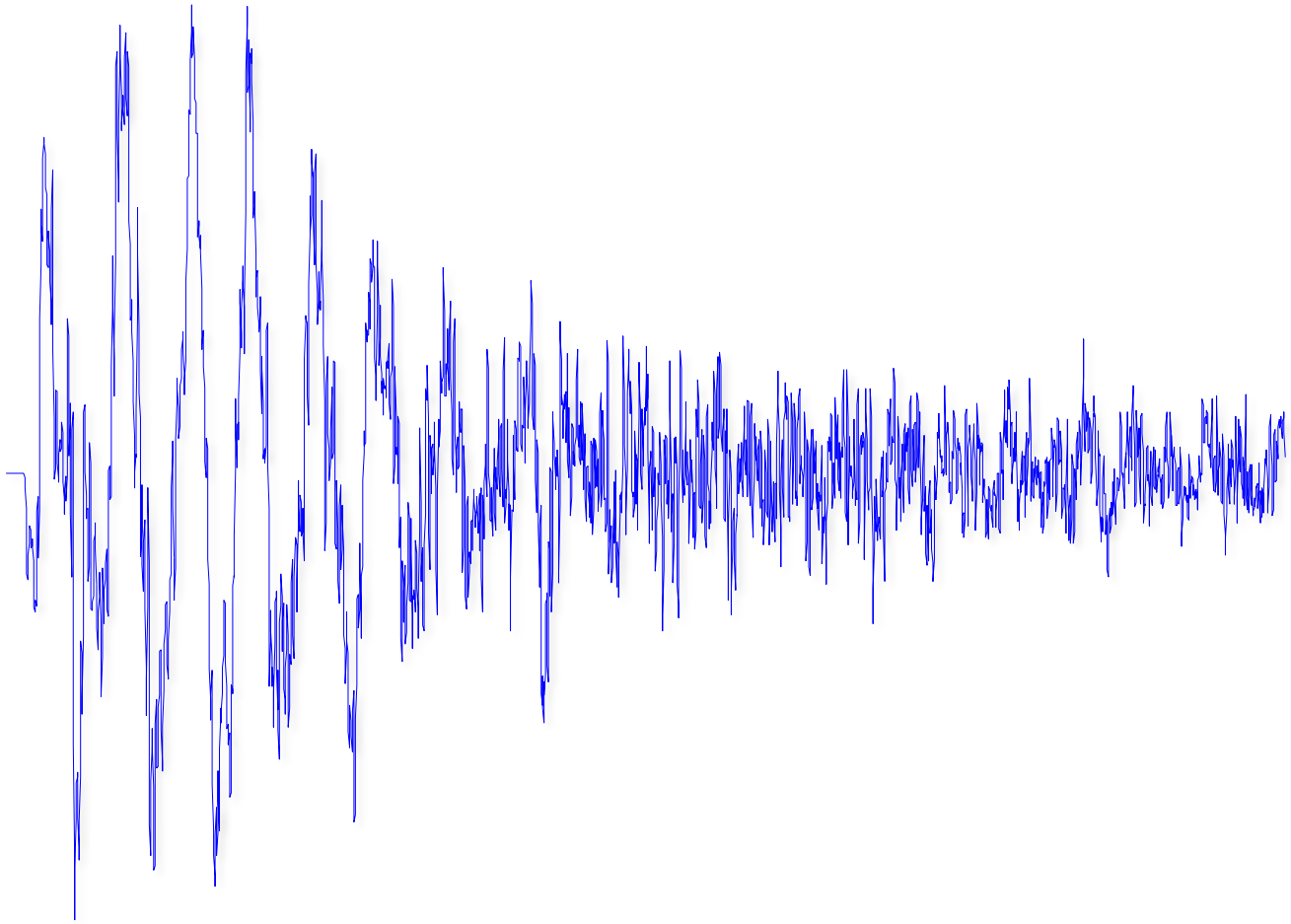
ZERO CROSSING RATE



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

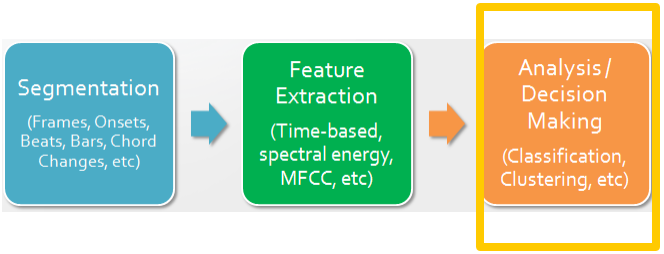
Warning: example results only - not actual results from audio analysis…

Heuristic Analysis

- Use basic thresholds or simple decision tree to form rudimentary transcription of kicks and snares.
- Time for more sophistication…
- Stairway / Stairway sorted

Example Feature Vector

	ZCR	Centroid	Bandwidth	Skew
	1	2	3	4
1	205	982.0780	0.1452	1.3512e+03
2	150	621.0359	0.1042	296.0815
3	120.0000	361.6111	0.0607	263.7817
4	135	809.3978	0.1315	834.4116
5	220	634.7242	0.0906	274.5483
6	175	536.3318	0.0837	188.4155
7	190	567.0412	0.0953	253.0151
8	135	720.2892	0.1153	333.7646
9	195.0000	778.5310	0.1407	1.2328e+03
10	185	514.4315	0.0717	183.0322

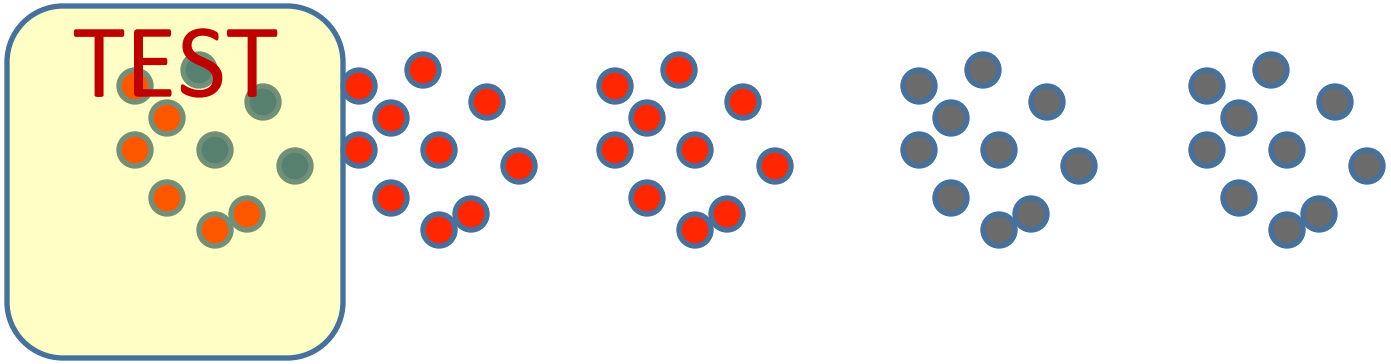


ANALYSIS AND DECISION MAKING
INSTANCE-BASED CLASSIFIERS (K-NN)

Training...

TRAINING SET

“kick drum” “not a kick drum”



k-NN

Explanation…

Advantages:

Training is trivial: just store the training samples very simple to implement and use

Disadvantages

Classification gets very complex with a lot of training data Must measure distance to all training samples; Euclidean distance becomes problematic in high-dimensional spaces;

Can easily be “overfit”

We can improve computation efficiency by storing just the class prototypes.

k-NN

Steps:

- Measure distance to all points.
- Take the k closest
- Majority rules. (e.g., if k=5, then take 3 out of 5)

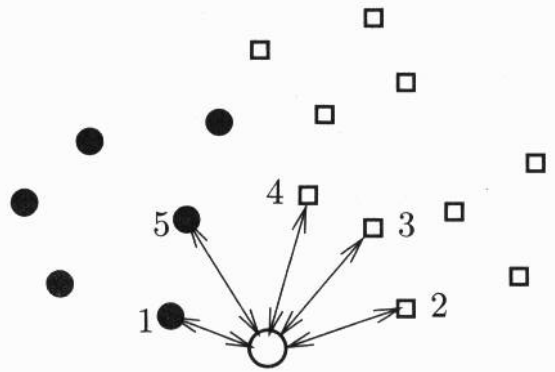
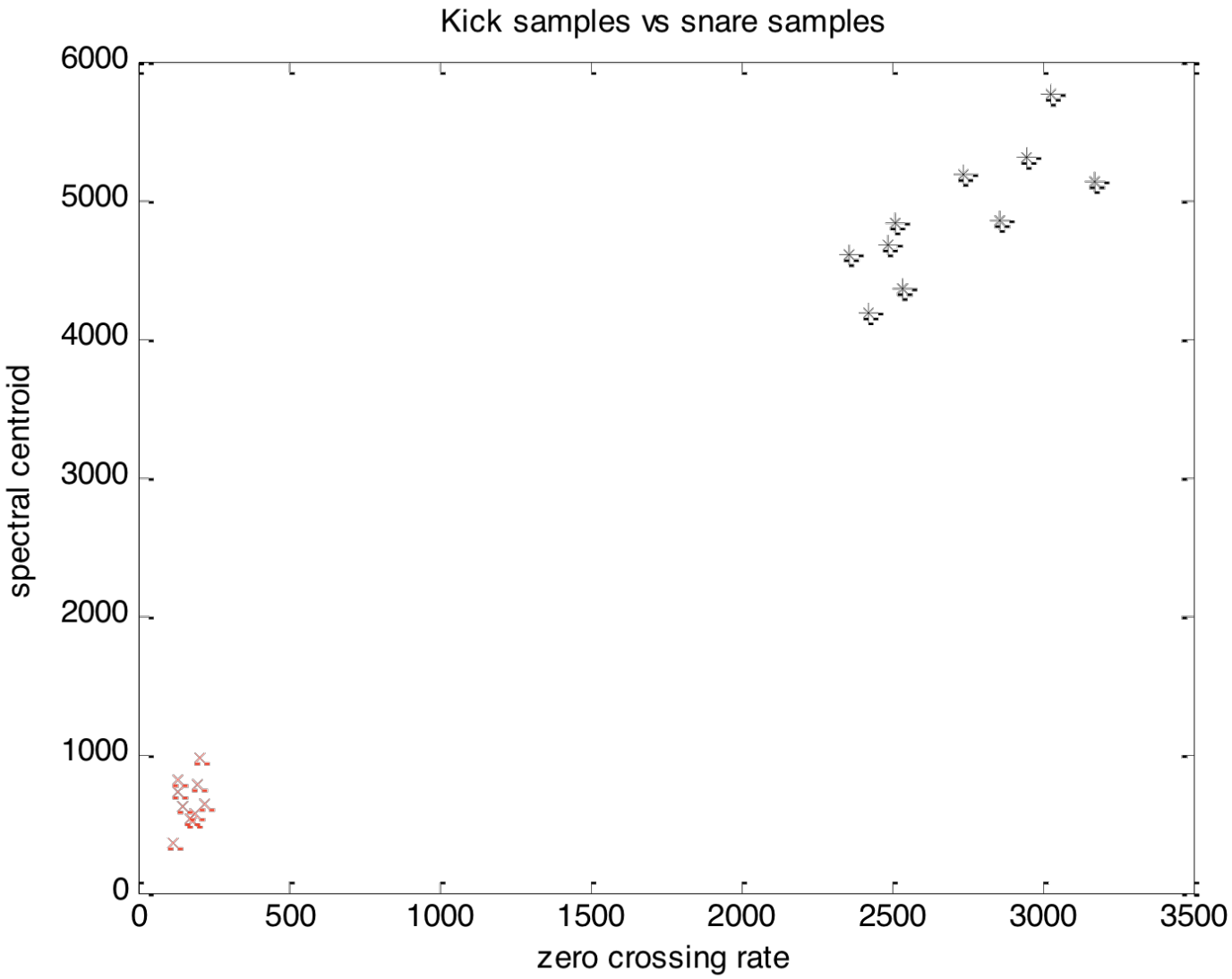


Fig. 2.15. k-nearest neighbours classification of two-dimensional data in the two-class case, with $k = 5$. The new datum $\mathbf{x}$ is represented by a non-filled circle. Elements of the training set (X, Y) are represented with dots (those with label -1) and squares (those with label $+1$). The arrow lengths represent the Euclidean distance between $\mathbf{x}$ and its 5 nearest neighbours. Three of them are squares, which makes $\mathbf{x}$ have the label $\mathsf{y} = +1$.


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

k-NN

- Instance-based learning – training examples are stored directly, rather than estimate model parameters
- Generally choose k being odd to guarantee a majority vote for a class.