DAY 1

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

Jay LeBoeuf
jay{at}izotope.com

June 2012

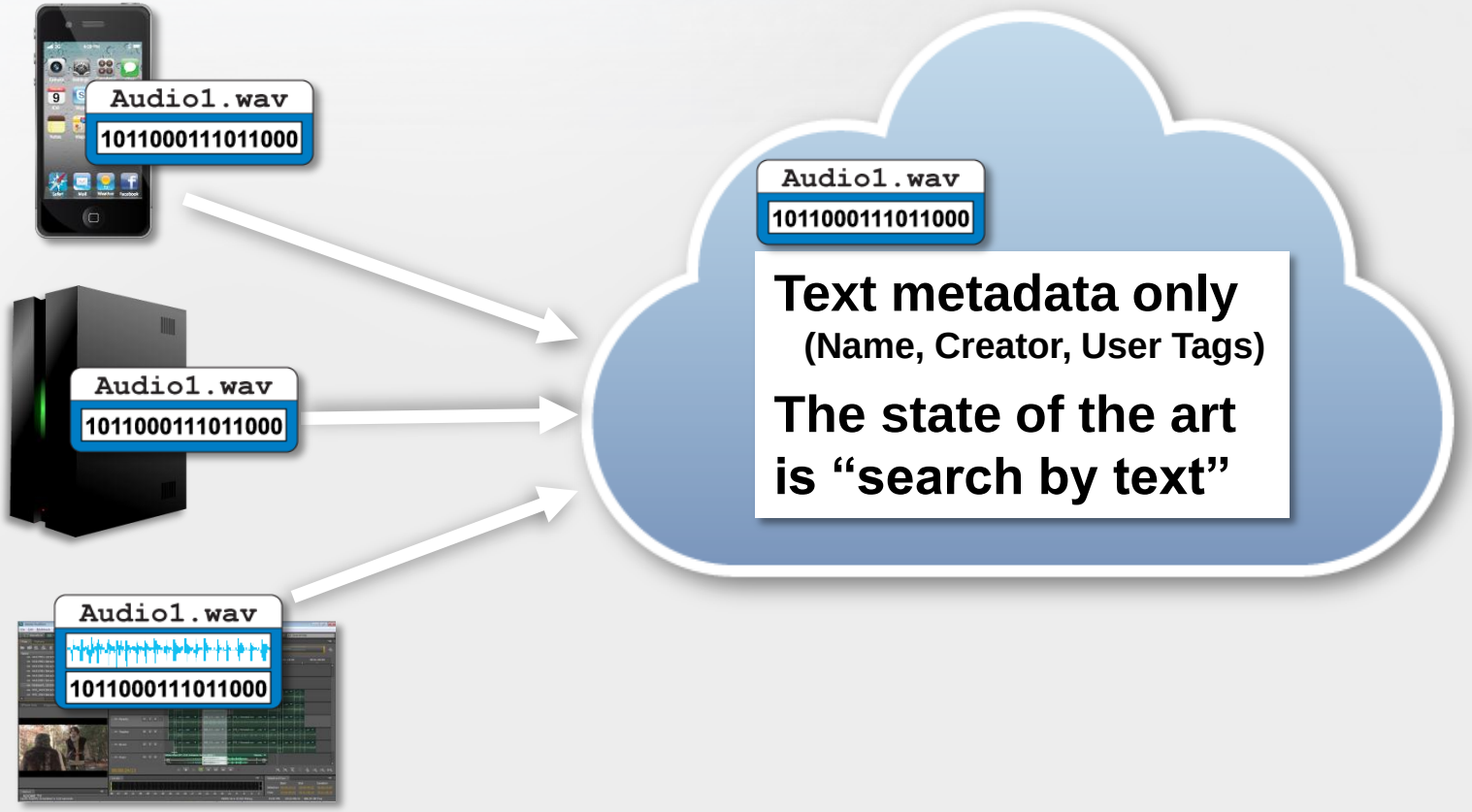
Administration

- https://ccrma.stanford.edu/wiki/MIR_workshop_2012
- Daily schedule
- Introductions
  - Our background
  - A little about yourself
  - Your area of interest, background with DSP, coding?, Matlab?, and any specific items of interest that you'd like to see covered.
  - Will you be using your own laptop for the lab?
    - If so, do you have Matlab? (Mac/PC?)

Example Seed…

Problems?

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



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

Commercial Applications

Pitch and rhythm tracking / analysis
- Algorithms in Guitar Hero / Rock Band
- BMAT's Score
DAW products that include beat/tempo/key/note analysis
- Ableton Live, Melodyne, Mixed In Key
Innovative software for music creation
- Khush, UJAM, Songsmith, VoiceBand
Audio search and QBH (SoundHound)
Music players with recommendation
- Apple Genius, Google Instant Mix
Music recommendation and metadata API
- Gracenote, Echo Nest, Rovi, BMAT, Bach Technology, Moodagent
Broadcast monitoring
- Audible Magic, Clustermedia Labs
Licensable research / software
Imagine Research, Fraunhofer IDMT, ...
Assisted Music Transcription
- Transcribe!, TwelveKeys Music Transcription Assistant
Audio fingerprinting
-SoundHound, Shazam, EchoNest, Gracenote, Civolution, Digimarc

Demos

Assisted Transcription

- drum transcription demo

Audio Search

similarity search

Rhythmic

similarity search

“distorted, guitar”

Imagine Research™

Files belonging to: Distorted Guitar

- L0016118_Lead 5.wav
- L0016119_Lead 8.wav
- L0016120_Lead 3.wav
- L0016121_Happy Harmonics.wav
- L0016122_Harmonic Feedback.wav
- L0016123_Harmonics 1.wav
- L0016124_Harmonics 2.wav
- L0016125_Harmonics 3.wav
- L0016126_Harmonics 4.wav
- L0016127_Lazy.wav
- L0016128_Open Chord1.wav
- L0016129_Open Chord2.wav
- L0016130_Palm Mute1.wav
- L0016131_Power Chords.wav
- L0016132_Punch.wav
- L0016133_Slide Down.wav
- L0016134_Slide Down1.wav
- L0016135_Slide Down2.wav
- L0016136_Slide Up.wav
- L0016137_Squeel 1.wav
- L0016139_Suspense Harmonics.wav

Top 20 matching tracks:

Distorted guitar: 20

MEDIA MINED

Interact with content

Imagine Research™

Find me <u>songs</u> that sound like this?

find similar

MEDIA MINED

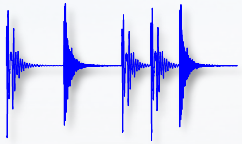
Name	Duration
M0112708_06 Overkill.wav	2:12
M0112707_05 Blood Lust.wav	2:12
M0112705_04 Napalm Blitz.wav	2:12
M0112711_08 Demolition Barbie.wav	2:23
M0112702_01 Axephetamine.wav	2:34
M0112703_02 Dimebag Damage.wav	2:25
Mermaid in Japan	5:06
M0112953_10 Hallowed By Thy Flame.wav	2:46
M0112713_09 Headlong 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

This week…

Day 1	MIR Overview Basic Features ; k-NN Basic classification Time domain features
Day 2	Frequency domain features Beat / Onset / Rhythm
Day 3	Features: Pitch, Chroma Classification (SVM) Detection in Mixtures
Day 4	Music Recommendation Transcription
Day 5	Auto-Tagging Classification

BASIC SYSTEM OVERVIEW

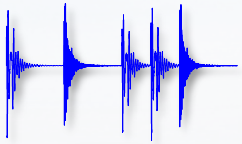
Basic system overview



Segmentation

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

Basic system overview



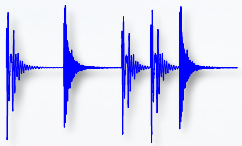
Segmentation

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

Feature Extraction

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

Basic system overview



Segmentation

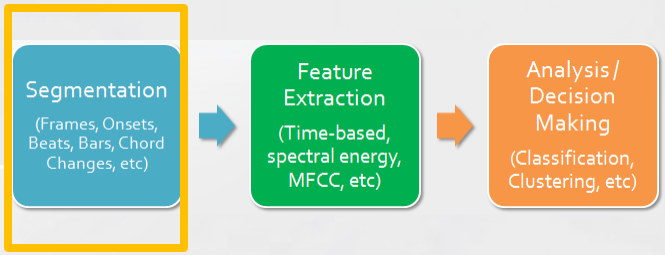
(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction

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

Analysis / Decision Making

(Classification, Clustering, etc)

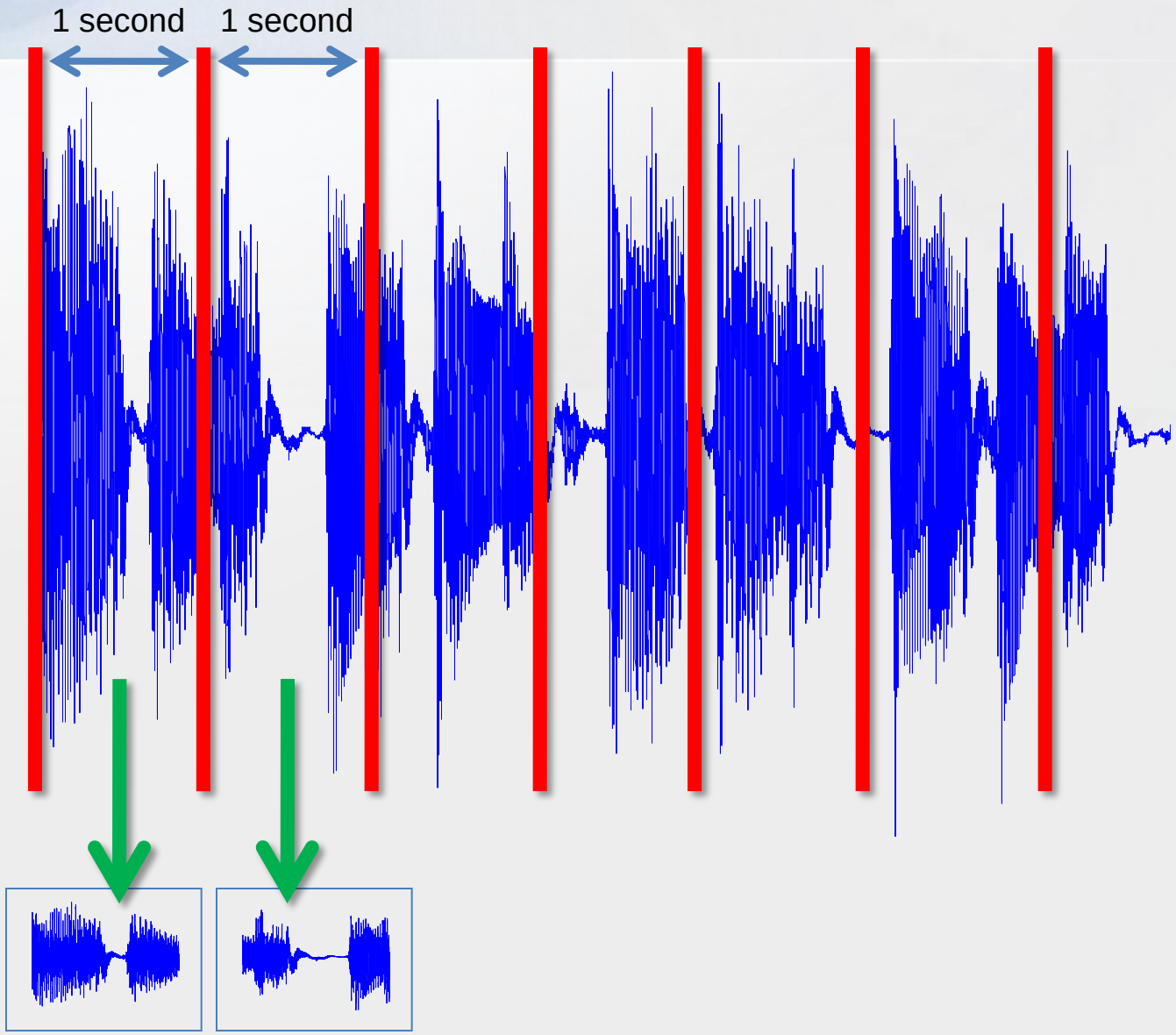


TIMING AND SEGMENTATION

Timing and Segmentation

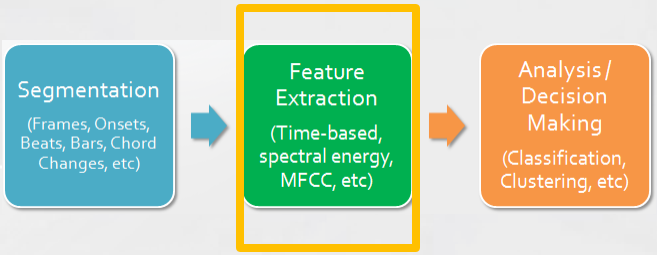
- Slicing up by fixed time slices…
  - 1 second, 80 ms, 100 ms, 20-40ms, etc.
- "Frames"
  - Different problems call for different frame lengths

Frames

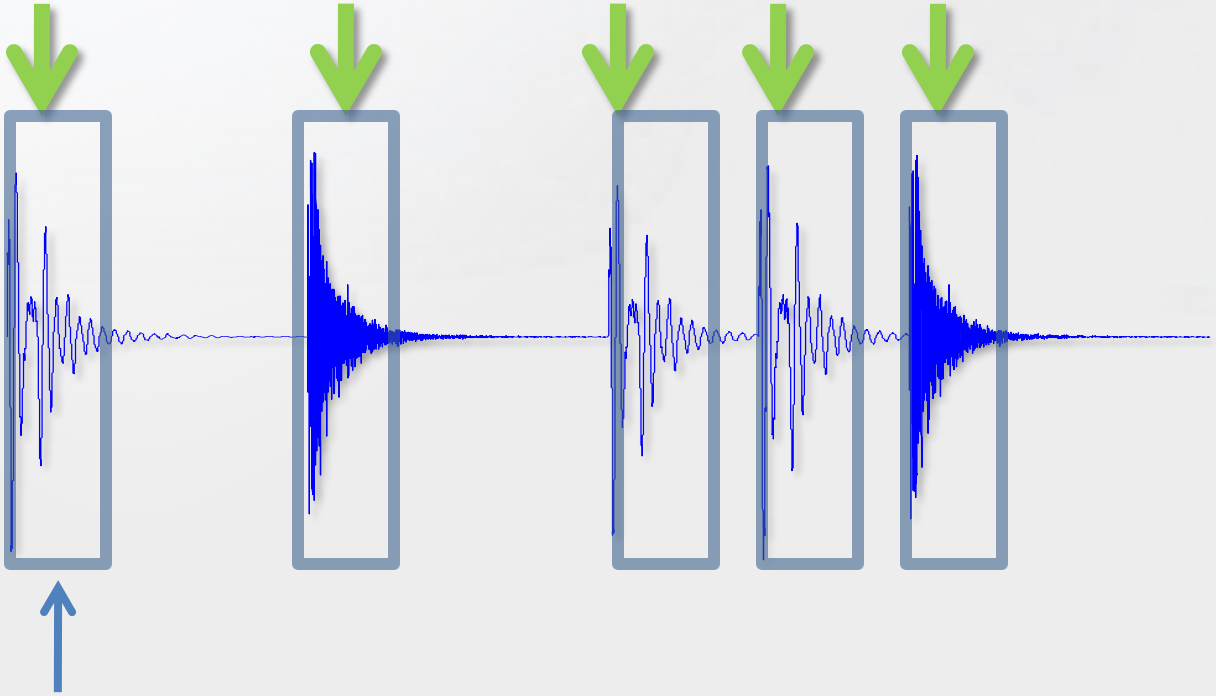


Timing and Segmentation

- 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

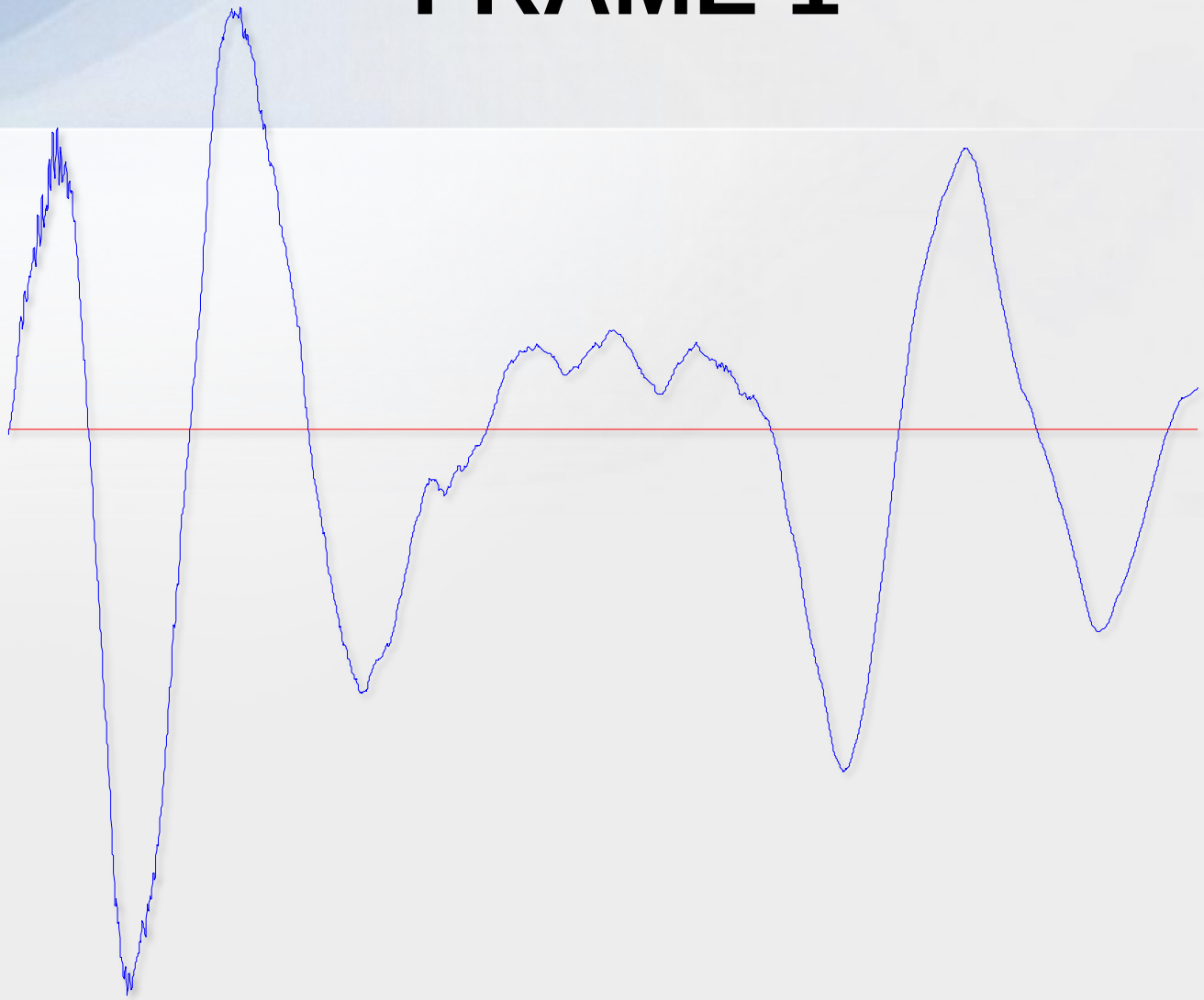


FEATURE EXTRACTION

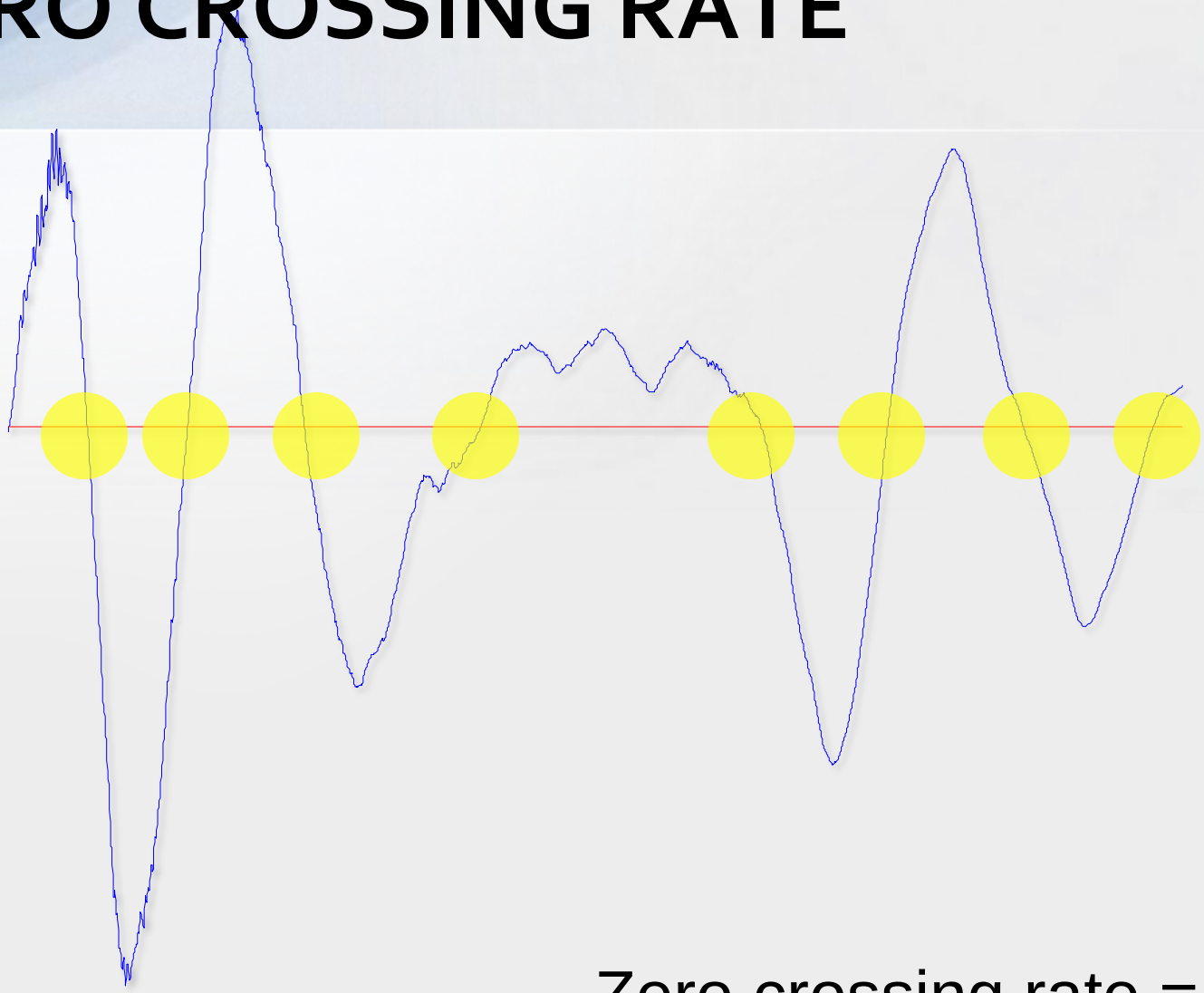


Frame 1

FRAME 1



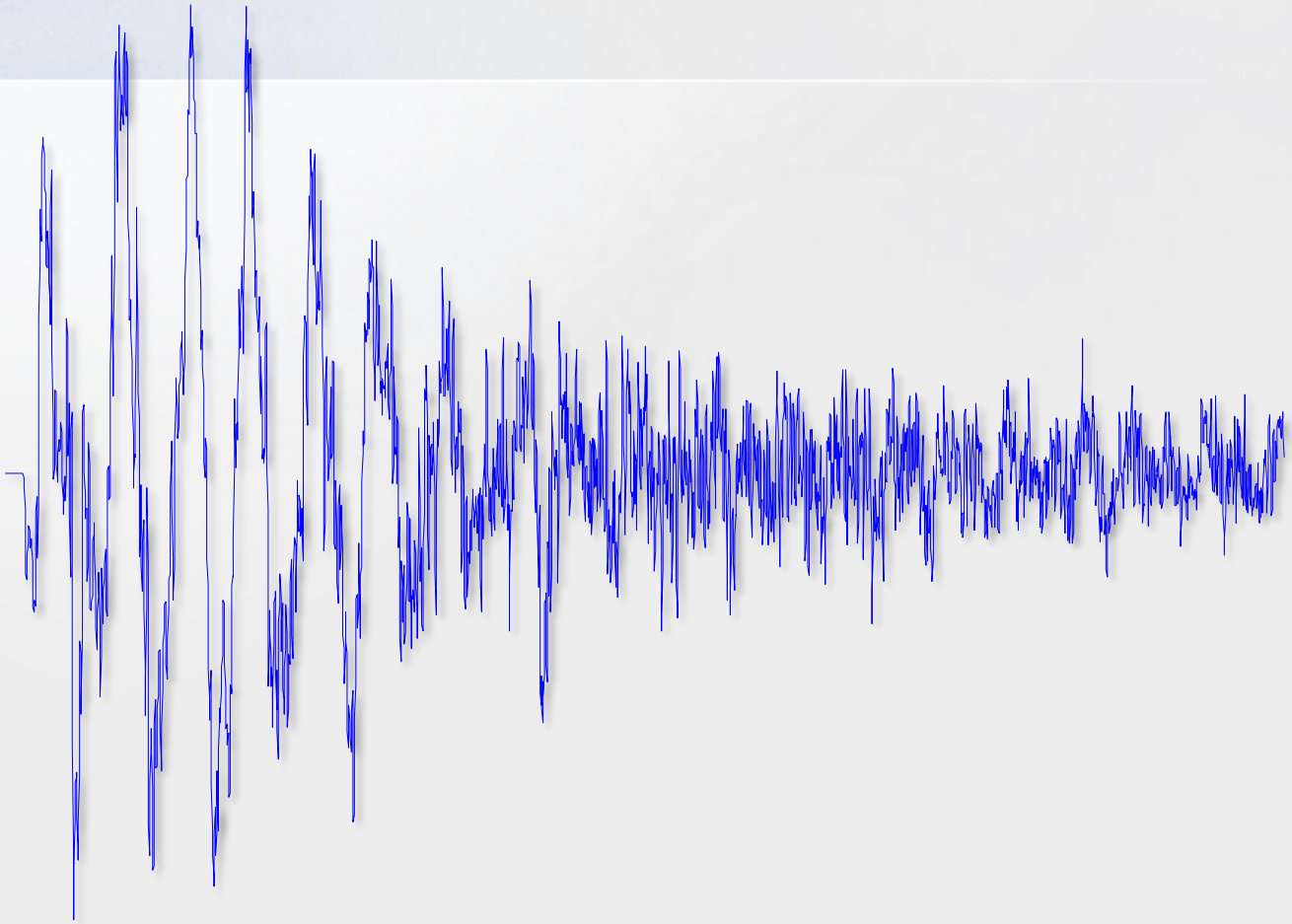
ZERO CROSSING RATE



FRAME 1

Zero crossing rate = 9

Frame 2

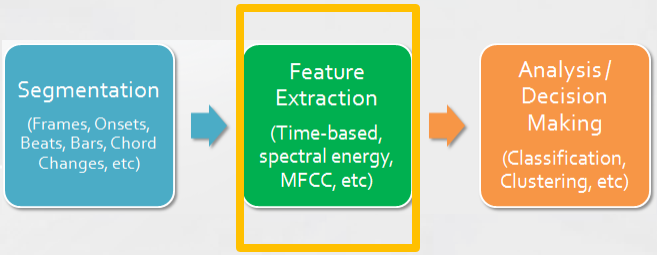


Zero crossing rate = 423

Features : SimpleLoop.wav

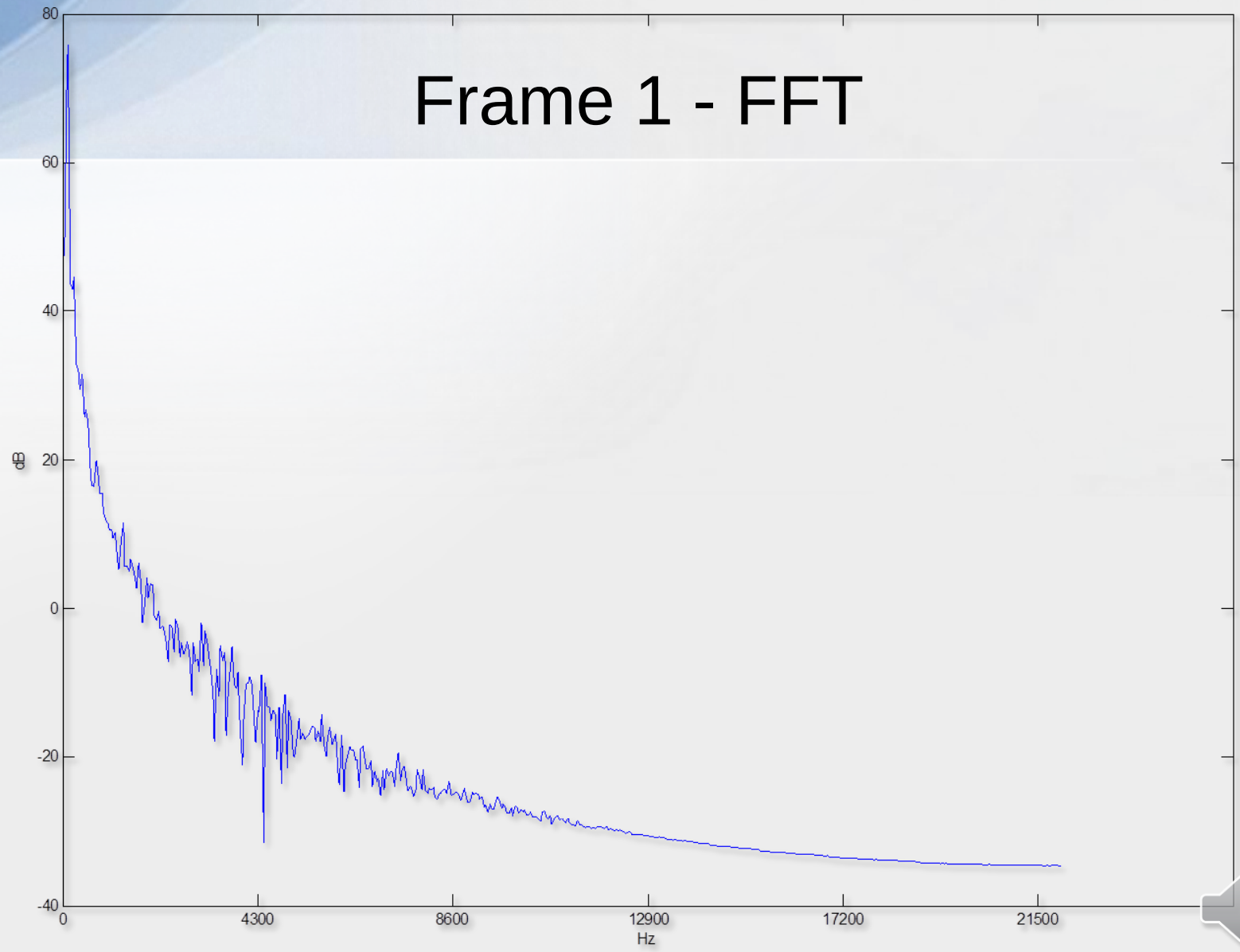
Frame	ZCR
1	9
2	423
3	22
4	28
5	390

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



FEATURE EXTRACTION

Frame 1 - FFT

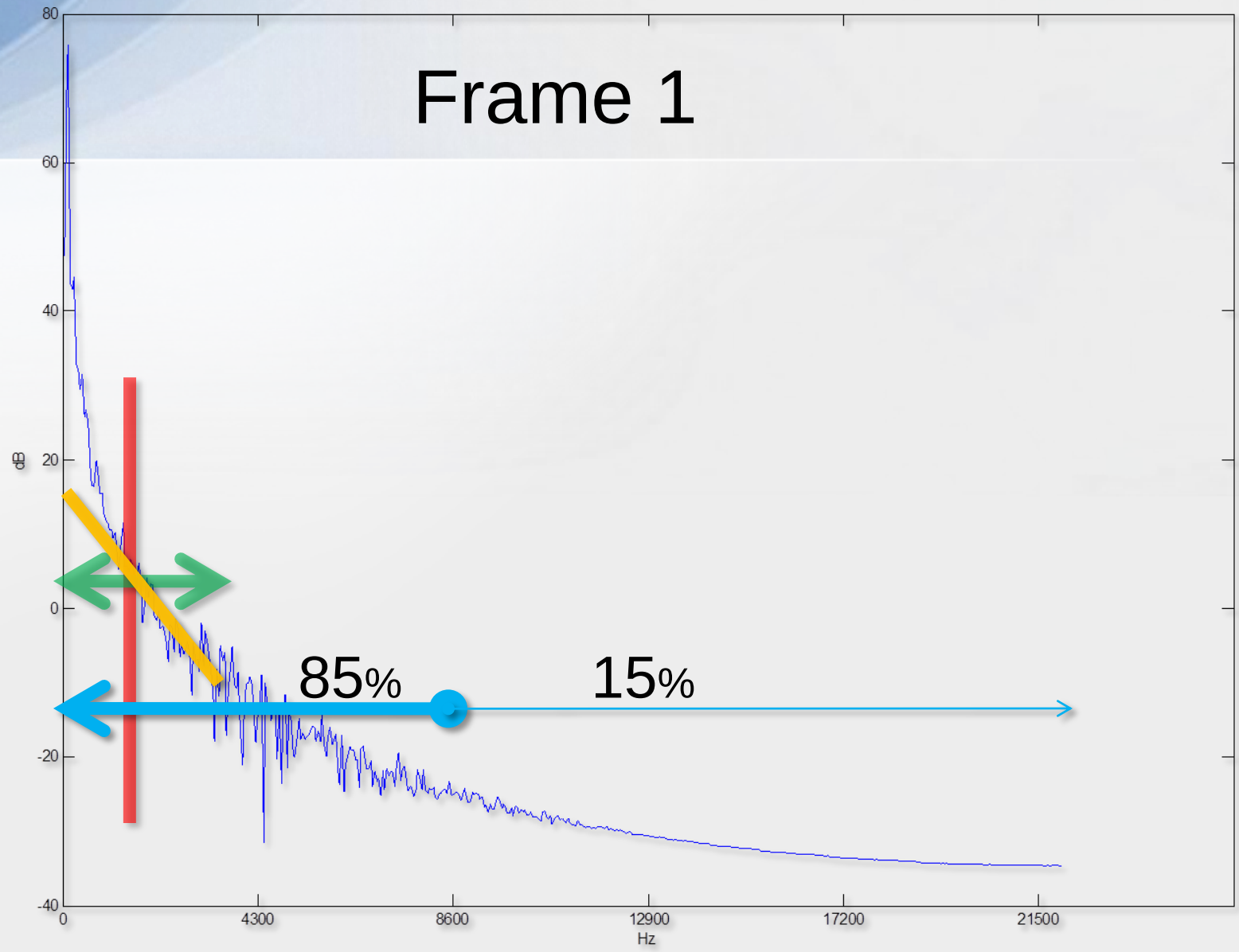


Spectral Features

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

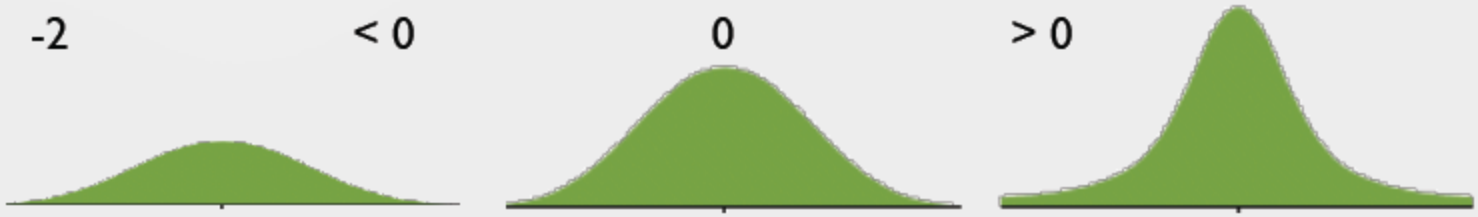
Spectral moments

Frame 1



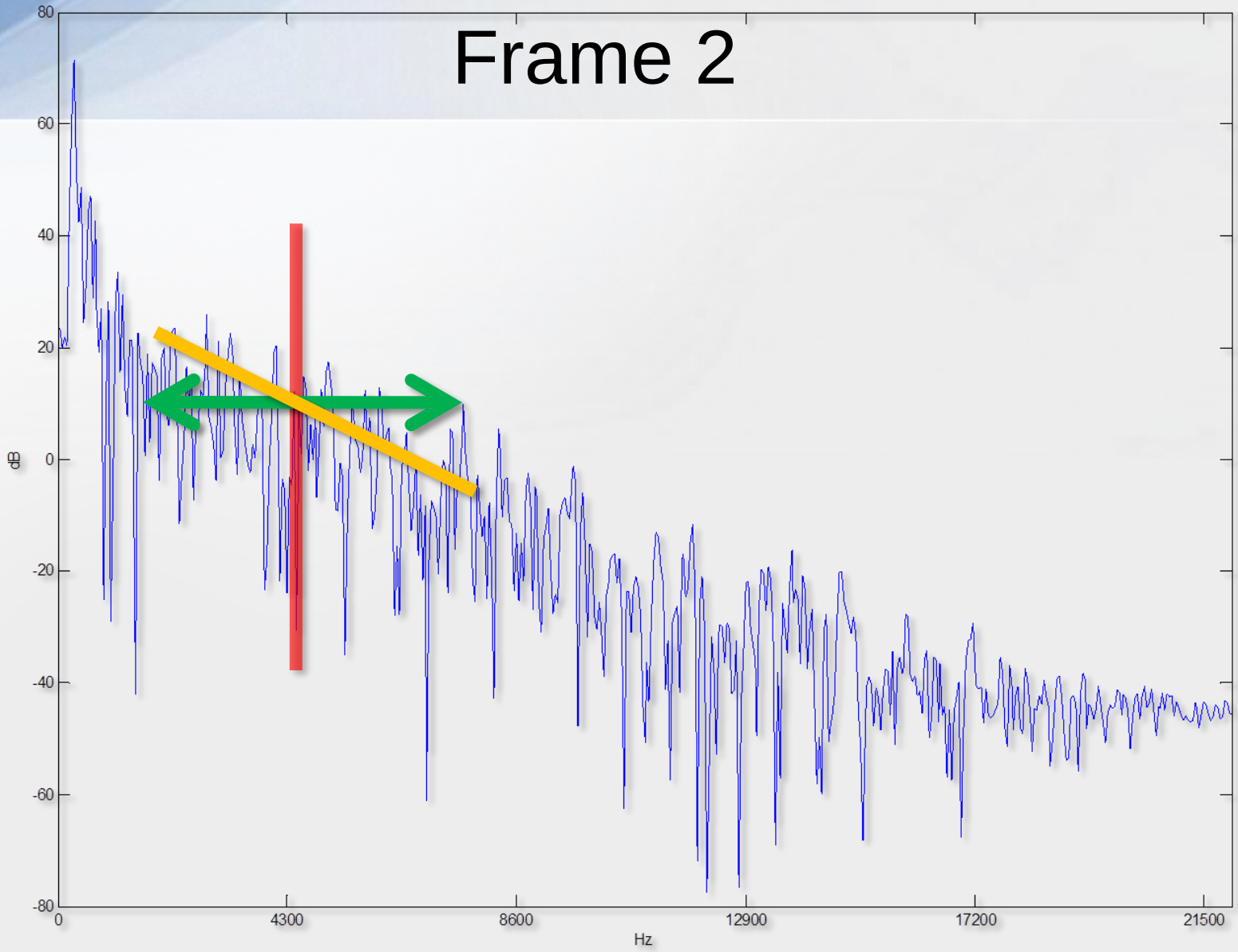
Skewness

Kurtosis



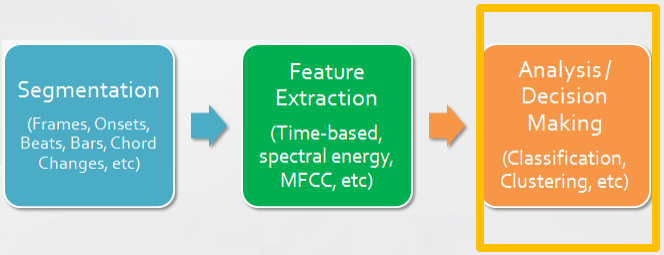
http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/userguide1.1

Frame 2



Example Feature Vector

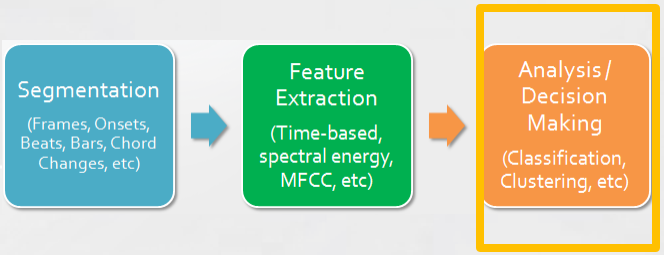
	ZCR	Centroid	Bandwidth	Skew
	1	2	3	4
1	205	982.0780	0.1452	1.3512e+03
2	150	621.0359	0.1042	296.0815
3	120.0000	361.6111	0.0607	263.7817
4	135	809.3978	0.1315	834.4116
5	220	634.7242	0.0906	274.5483
6	175	536.3318	0.0837	188.4155
7	190	567.0412	0.0953	253.0151
8	135	720.2892	0.1153	333.7646
9	195.0000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



ANALYSIS AND DECISION MAKING HEURISTICS

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!



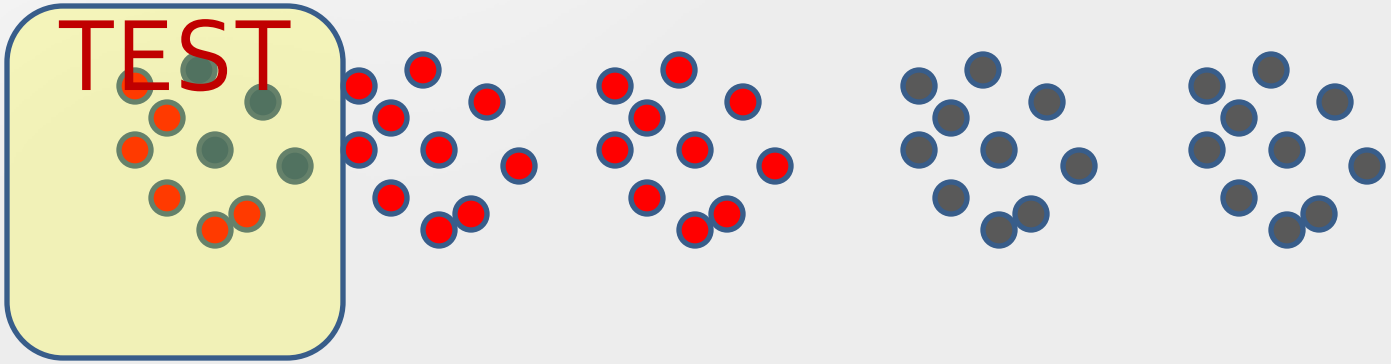
ANALYSIS AND DECISION MAKING INSTANCE-BASED CLASSIFIERS (K-NN)


LIBERTY
LIBERTY
IN GOD WE TRUST

Training...

TRAINING SET

"1" "0"



k-NN

- Explanation…

<u>Advantages:</u>

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

<u>Disadvantages</u>

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)

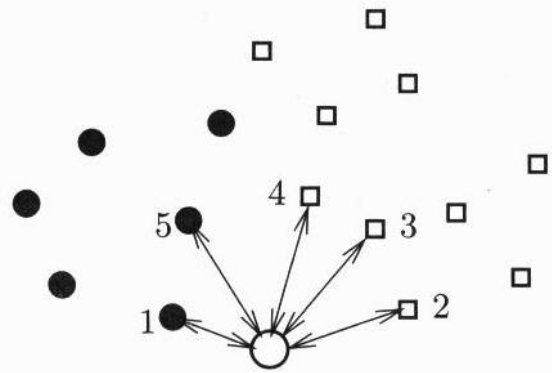
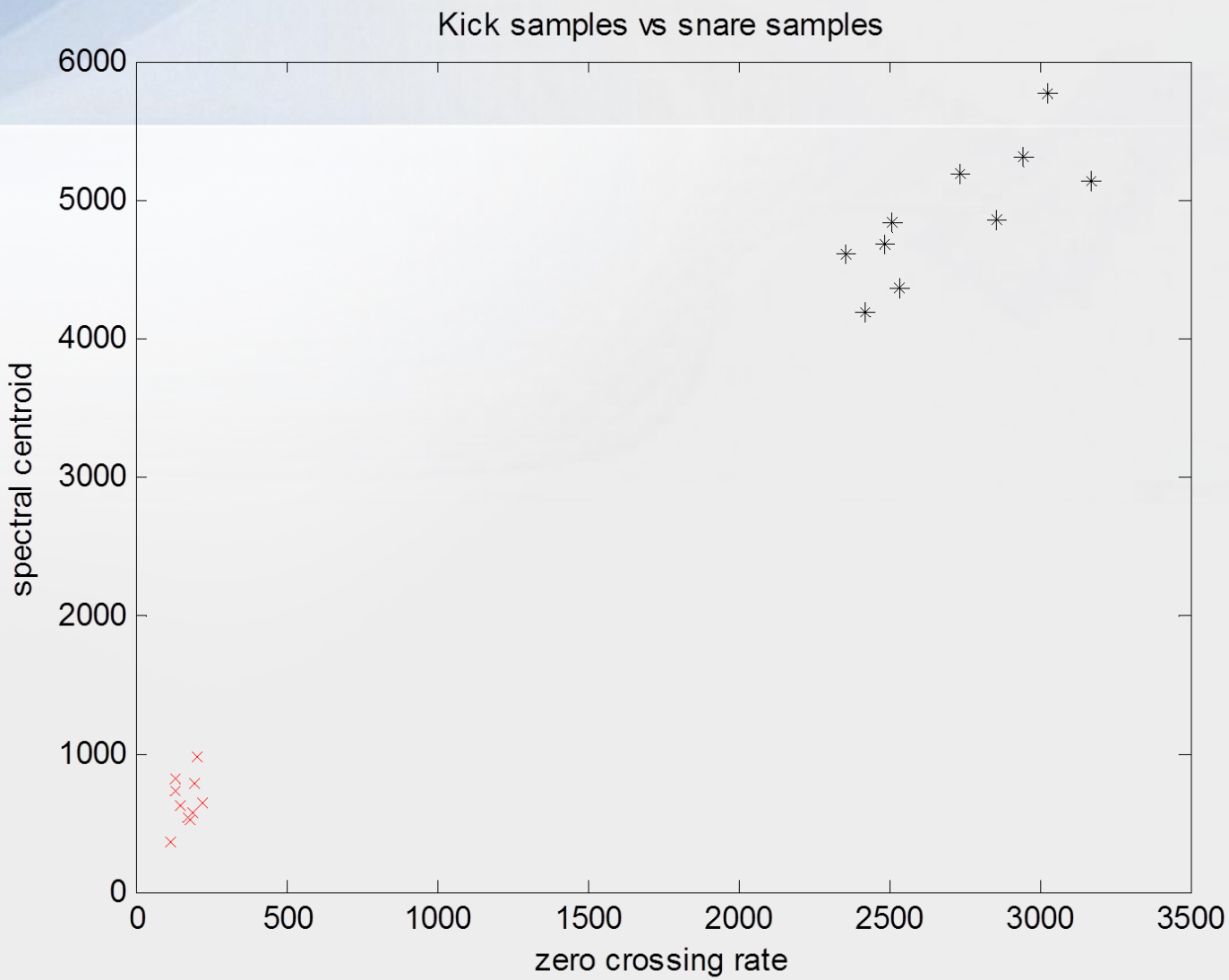


Fig. 2.15. k-nearest neighbours classification of two-dimensional data in the two-class case, with $k = 5$. The new datum $\mathbf{x}$ is represented by a non-filled circle. Elements of the training set (X, Y) are represented with dots (those with label -1) and squares (those with label $+1$). The arrow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

Distance Classification

1. Find nearest neighbor
2. Find representative match via class prototype (e.g., center of group or mean of training data class)

Distance metric

Most common: Euclidean distance

Scaling Example

unscaled				
zcr	centroid	brightness	rolloff	kurtosis
2138.663	4461.262	52.78183	574185.3	1.54E+08
1420.046	3860.736	52.66741	443347.5	1.03E+08
1412.664	4095.468	38.29324	1168014	3.96E+08
1223.236	3551.842	50.74382	435198.1	1.03E+08
2680.428	5266.489	53.54988	444444	1.37E+08
2304.393	4316.17	46.82006	781963.4	2.28E+08
1594.079	4597.727	54.22522	573360.2	1.49E+08
1900.388	4732.717	63.46437	354037.4	83337090
2231.022	4914.268	57.65587	376519.5	1.01E+08
1221.818	4323.917	21.96666	5496309	3.45E+09
1018.673	2169.696	29.44604	3511451	1.48E+09
176.7722	1038.943	37.65943	972874.6	3.18E+08
758.9167	609.0479	21.9855	2146602	1.37E+09
656.2162	1982.091	34.04355	1829962	6.59E+08
563.9213	1309.174	43.58942	1152514	3.71E+08
1104.88	967.4398	37.8372	930845.7	3.04E+08
423.1187	1113.061	27.58021	3217640	1.62E+09
717.5933	1638.758	34.47308	1721541	6.2E+08
242.3024	1256.06	38.70912	891162.7	2.73E+08
474.6985	1240.683	26.55689	3657653	1.75E+09

Scaling Example: linear scale to {-1:1}

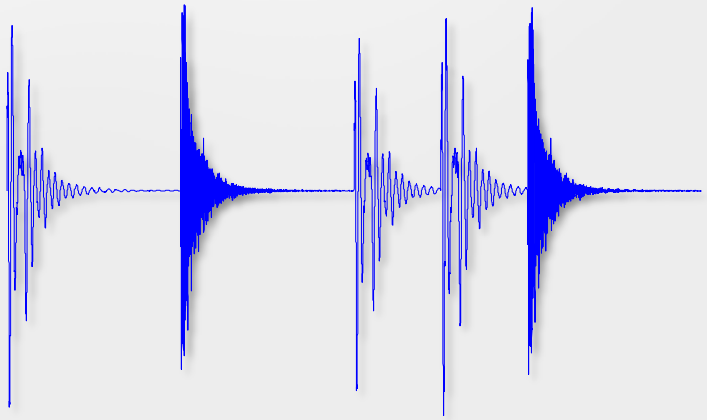
unscaled				
zcr	centroid	brightness	rolloff	kurtosis
2138.663	4461.262	52.78183	574185.3	1.54E+08
1420.046	3860.736	52.66741	443347.5	1.03E+08
1412.664	4095.468	38.29324	1168014	3.96E+08
1223.236	3551.842	50.74382	435198.1	1.03E+08
2680.428	5266.489	53.54988	444444	1.37E+08
2304.393	4316.17	46.82006	781963.4	2.28E+08
1594.079	4597.727	54.22522	573360.2	1.49E+08
1900.388	4732.717	63.46437	354037.4	83337090
2231.022	4914.268	57.65587	376519.5	1.01E+08
1221.818	4323.917	21.96666	5496309	3.45E+09
1018.673	2169.696	29.44604	3511451	1.48E+09
176.7722	1038.943	37.65943	972874.6	3.18E+08
758.9167	609.0479	21.9855	2146602	1.37E+09
656.2162	1982.091	34.04355	1829962	6.59E+08
563.9213	1309.174	43.58942	1152514	3.71E+08
1104.88	967.4398	37.8372	930845.7	3.04E+08
423.1187	1113.061	27.58021	3217640	1.62E+09
717.5933	1638.758	34.47308	1721541	6.2E+08
242.3024	1256.06	38.70912	891162.7	2.73E+08
474.6985	1240.683	26.55689	3657653	1.75E+09

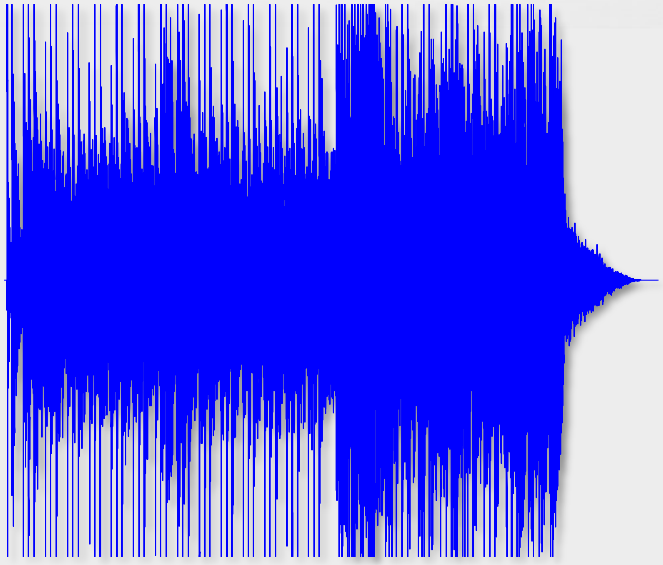
scaled				
zcr	centroid	brightness	rolloff	kurtosis
0.567221	0.654219	0.485151	-0.91438	-0.95797
-0.00683	0.396341	0.479636	-0.96526	-0.98848
-0.01273	0.49714	-0.21313	-0.68342	-0.81426
-0.16405	0.263696	0.386928	-0.96843	-0.98836
1	1	0.522167	-0.96484	-0.96824
0.699611	0.591913	0.19782	-0.83357	-0.91439
0.13219	0.71282	0.554715	-0.9147	-0.96108
0.37688	0.770787	1	-1	-1
0.641	0.848749	0.720057	-0.99126	-0.98934
-0.16518	0.59524	-1	1	1
-0.32746	-0.32983	-0.63953	0.228023	-0.17284
-1	-0.81539	-0.24368	-0.75931	-0.86091
-0.53496	-1	-0.99909	-0.30281	-0.23807
-0.617	-0.41039	-0.41795	-0.42596	-0.6582
-0.69073	-0.69935	0.042118	-0.68945	-0.82931
-0.2586	-0.8461	-0.23511	-0.77566	-0.86889
-0.80321	-0.78357	-0.72945	0.11375	-0.08884
-0.56797	-0.55782	-0.39725	-0.46813	-0.68172
-0.94765	-0.72216	-0.19309	-0.79109	-0.88744
-0.76201	-0.72876	-0.77877	0.284886	-0.01055

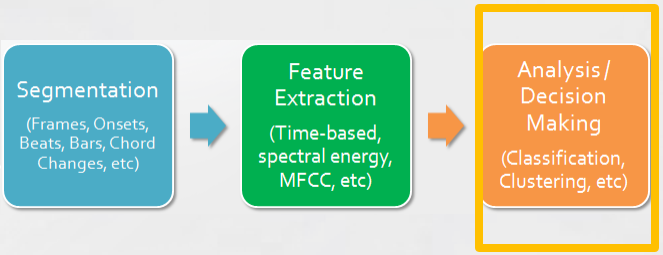
> End of Lecture 1: Part 1

Onset detection

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







EVALUATING ANALYSIS SYSTEMS (the basics)

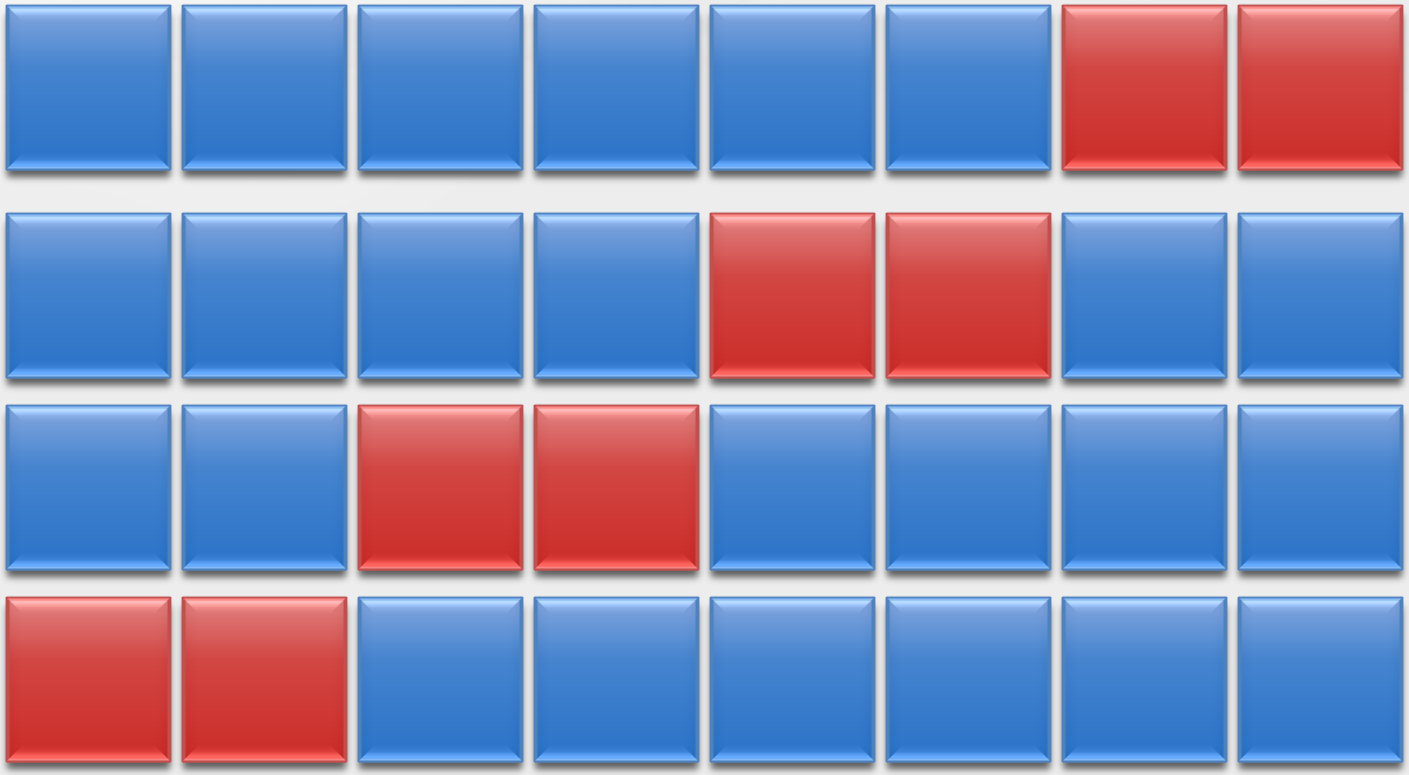
A bad evaluation metric

- "How many training examples are classified correctly?"

Image from Wikipedia, "Overfitting"

A better evaluation metric

- Accuracy on held-out ("test") examples
- Cross-validation: repeated train/test iterations



Looking beyond accuracy

		True class	
		p	n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives

Precision

- Metric from information retrieval: How relevant are the retrieved results?

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

== # TP / (# TP+ # FP)

In MIR, may involve precision at some threshold in ranked results.

Recall

- How complete are the retrieved results?

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

== # TP / (TP + FN)

F-measure

- A combined measure of precision and recall (harmonic mean)
  - Treats precision and recall as equally important

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Accuracy metric summary

Hypothesized class \ True class	p	n
Y	True Positives	False Positives
N	False Negatives	True Negatives
Column totals:	P	N

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

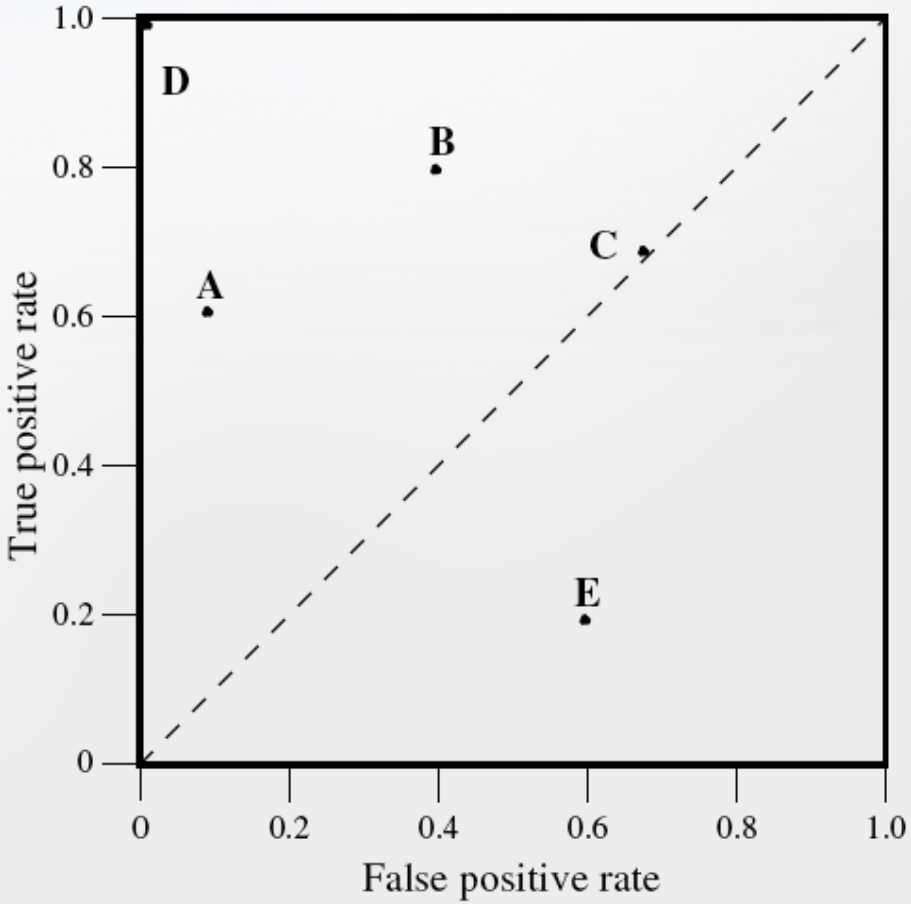
$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

From T. Fawcett, “An introduction to ROC analysis”

ROC Graph

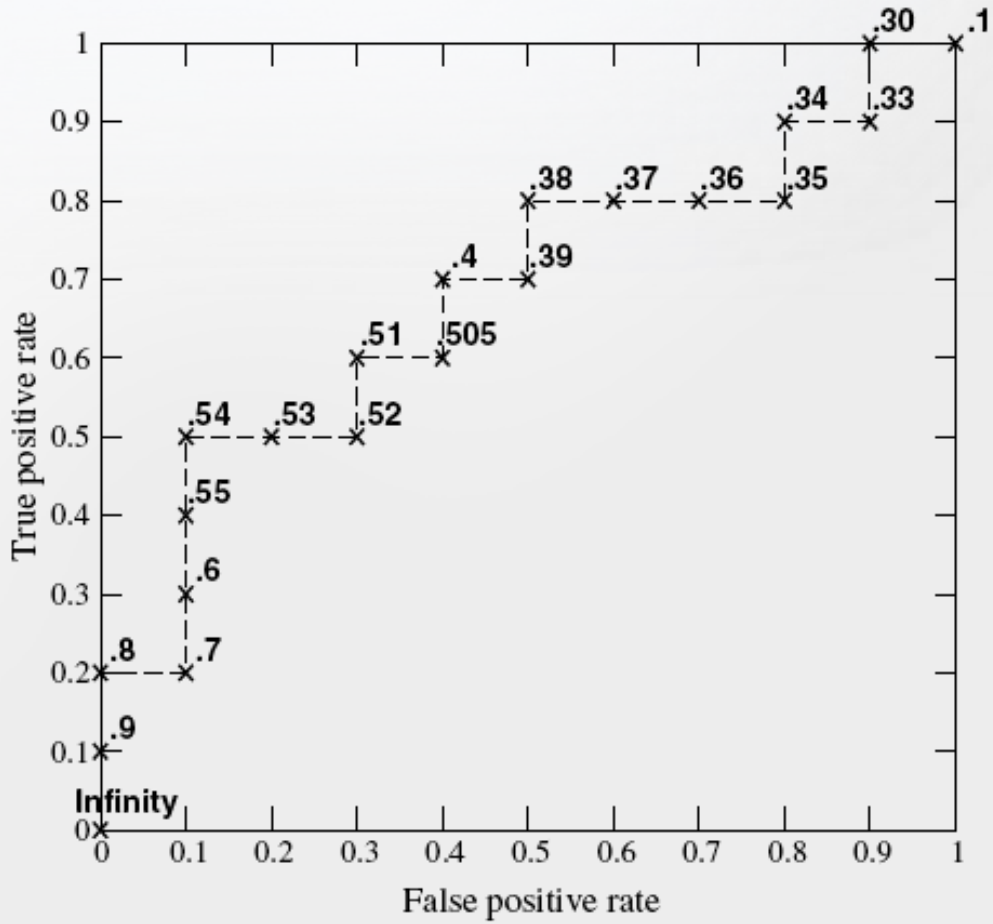
- "Receiver operating characteristics" curve
- A richer method of measuring model performance than classification accuracy
- Plots true positive rate vs false positive rate

ROC plot for discrete classifiers



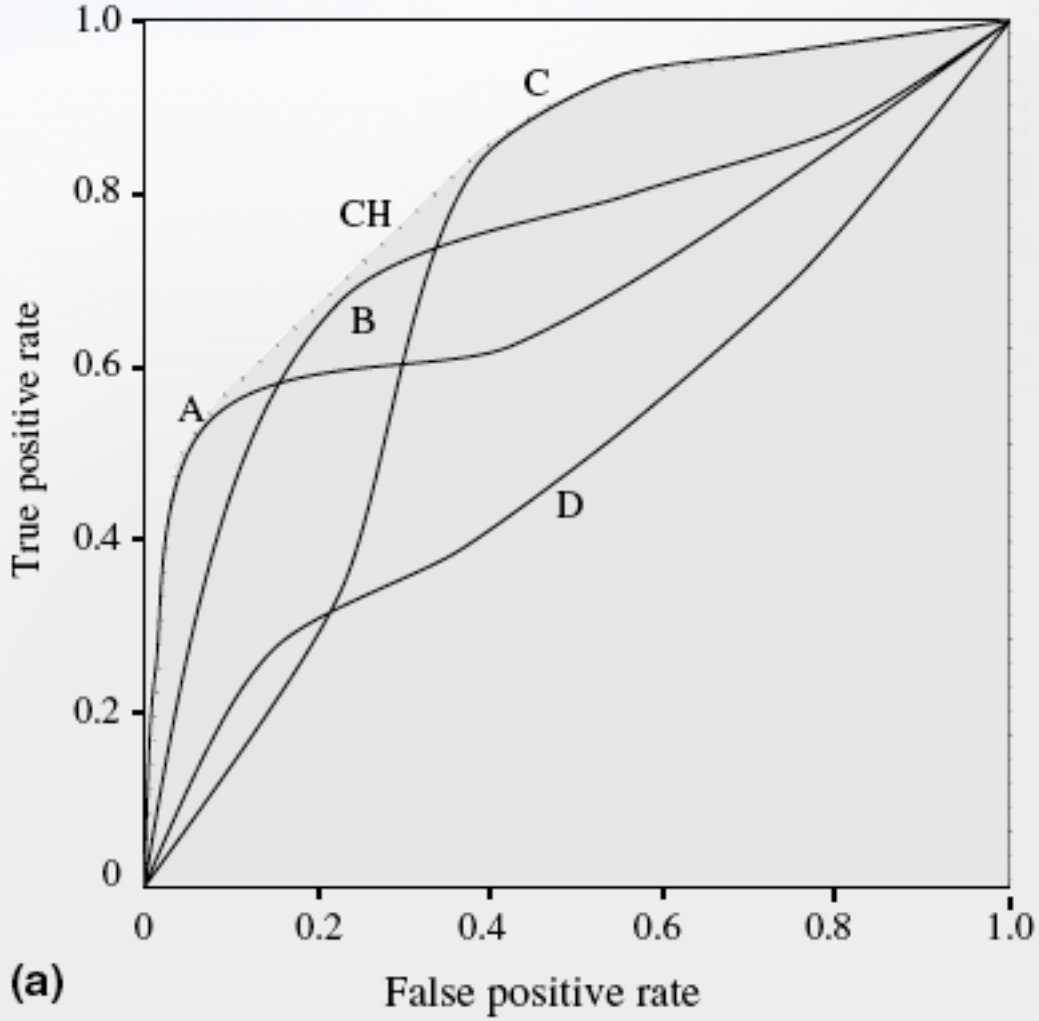
- Each classifier output is either right or wrong
  - Discrete classifier has single point on ROC plot
- The "Northwest" is better!
- Best sub-region may be task-dependent (conservative or liberal may be better)

ROC curves for probabilistic/tunable classifiers



- Plot TP/FP points for different thresholds of **one** classifier
  - Here, indicates that threshold of .5 is not optimal (0.54 is better)

Area under ROC (AUC)



- Compute AUC to compare different classifiers
- AUC = probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.
- AUC not always == "better" for a particular problem