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

Rebecca Fiebrink
Princeton University
fiebrink{at}princeton.edu

July 2011

Administration

- https://ccrma.stanford.edu/wiki/MIR_workshop_2011
- Daily schedule
- Introductions
  - Our background
  - A little about yourself
  - E.g., your area of interest, background with DSP, coding/ programming languages, and any specific items of interest that you'd like to see covered.

Example Seed…

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, style, and mood analysis
- emotion and aesthetics
- 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
  - Zenph - before after

This week…

Day 1	MIR Overview Basic Features ; k-NN Information Retrieval Basics Basic transcription and RT processing
Day 2	Time domain features Frequecy domain features Beat / Onset / Rhythm
Day 3	Segmentation Classification (SVM) Detection in Mixtures
Day 4	Features: Pitch, Chroma Performance Alignment Cover Song ID / Music Collections
Day 5	Auto-Tagging Recommendation Playlisting

A BRIEF HISTORY OF MIR

History: Pre-ISMIR

- Don Byrd @ UMass Amherst + Tim Crawford @ King's College London receive funding for OMRAS (Online Music Recognition and Searching)
  - Sp. 1999: Requested by NSF program director to organize MIR workshop
- J. Stephen Downie + David Huron + Craig Nevill Manning host MIR workshop @ ACM DL / SIGIR 99
- Crawford + Carola Boehm organize MIR workshop at Digital Resources for the Humanities – Sept. '99

ISMIR and MIREX

- 2000: UMass hosts first ISMIR (International Symposium on Music Information Retrieval)
  - Michael Fingerhut (IRCAM) creates music-ir mailing list
- ISMIR run as yearly conference
  - 2001: “Symposium” -> “Conference”
- ISMIR incorporated as a Society in 2008
- MIREX benchmarking contest begun 2005

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

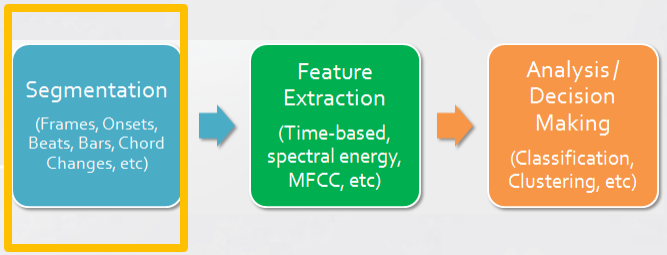
(Frames, Onsets, Beats, Bars, Chord Changes, etc)

Feature Extraction

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

Analysis / Decision Making

(Classification, Clustering, etc)

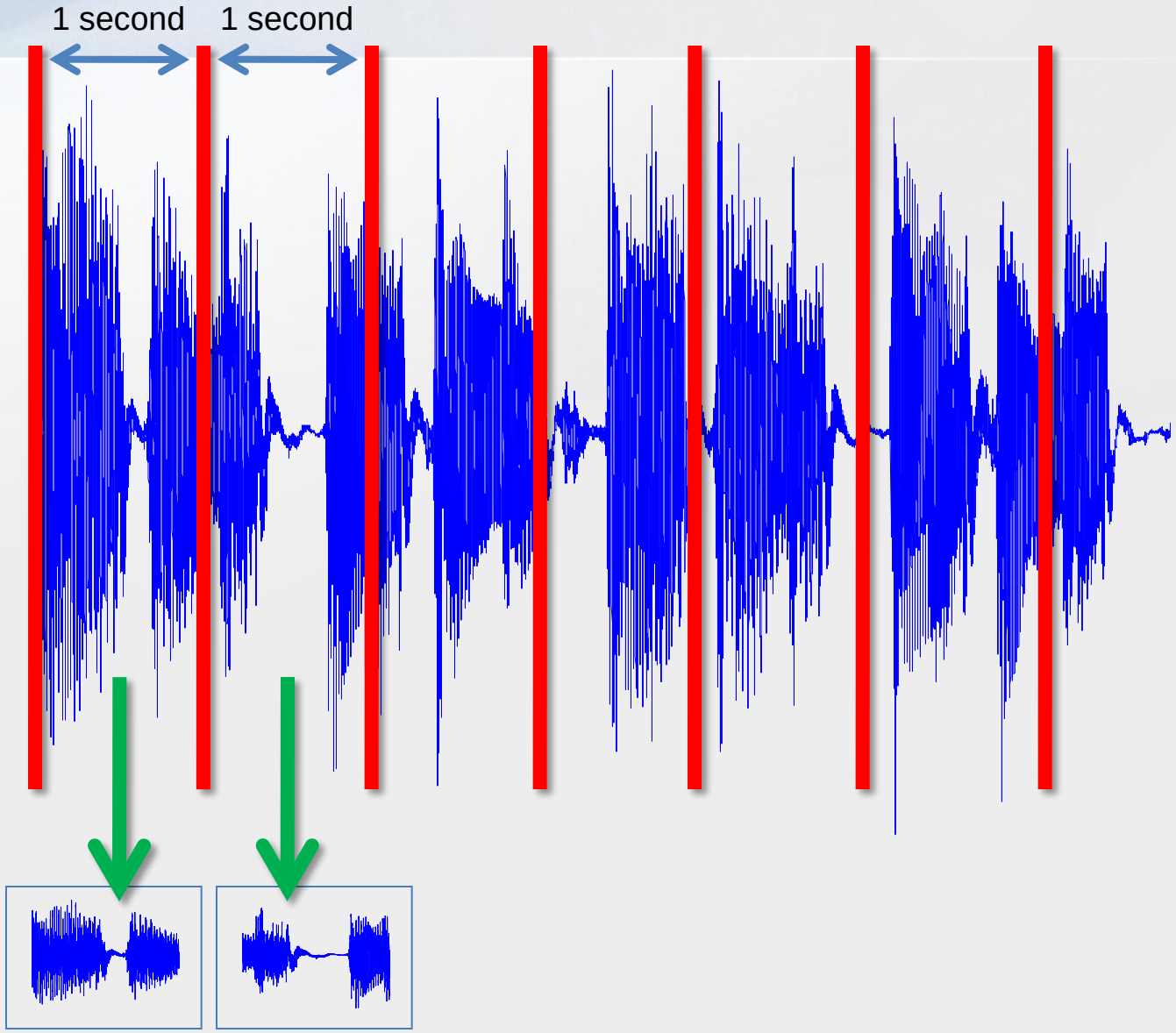


TIMING AND SEGMENTATION

Timing and Segmentation

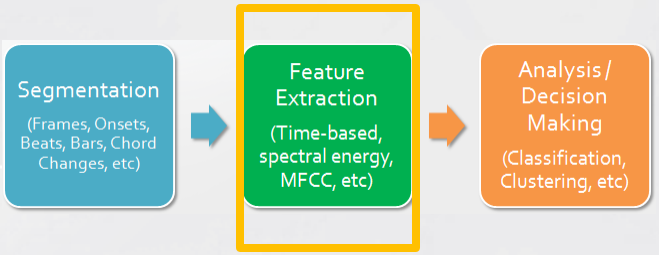
- Slicing up by fixed time slices…
  - 1 second, 80 ms, 100 ms, 20-40ms, etc.
- “Frames”
  - Different problems call for different frame lengths

Frames

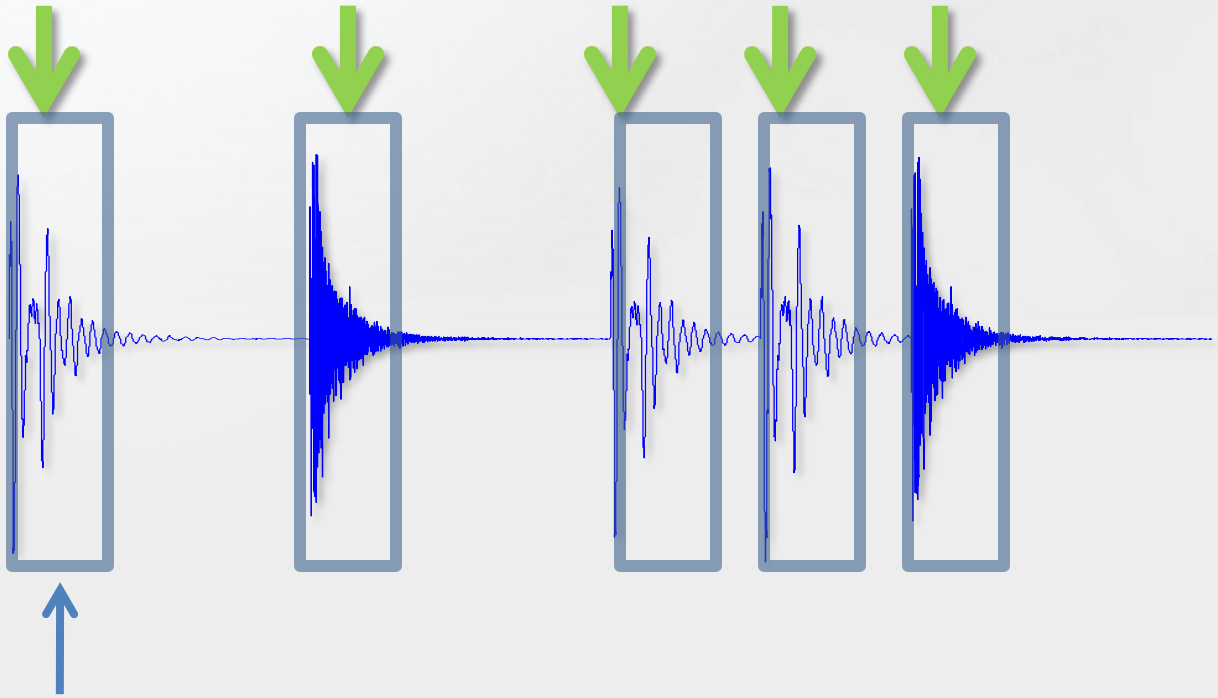


Timing and Segmentation

- 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



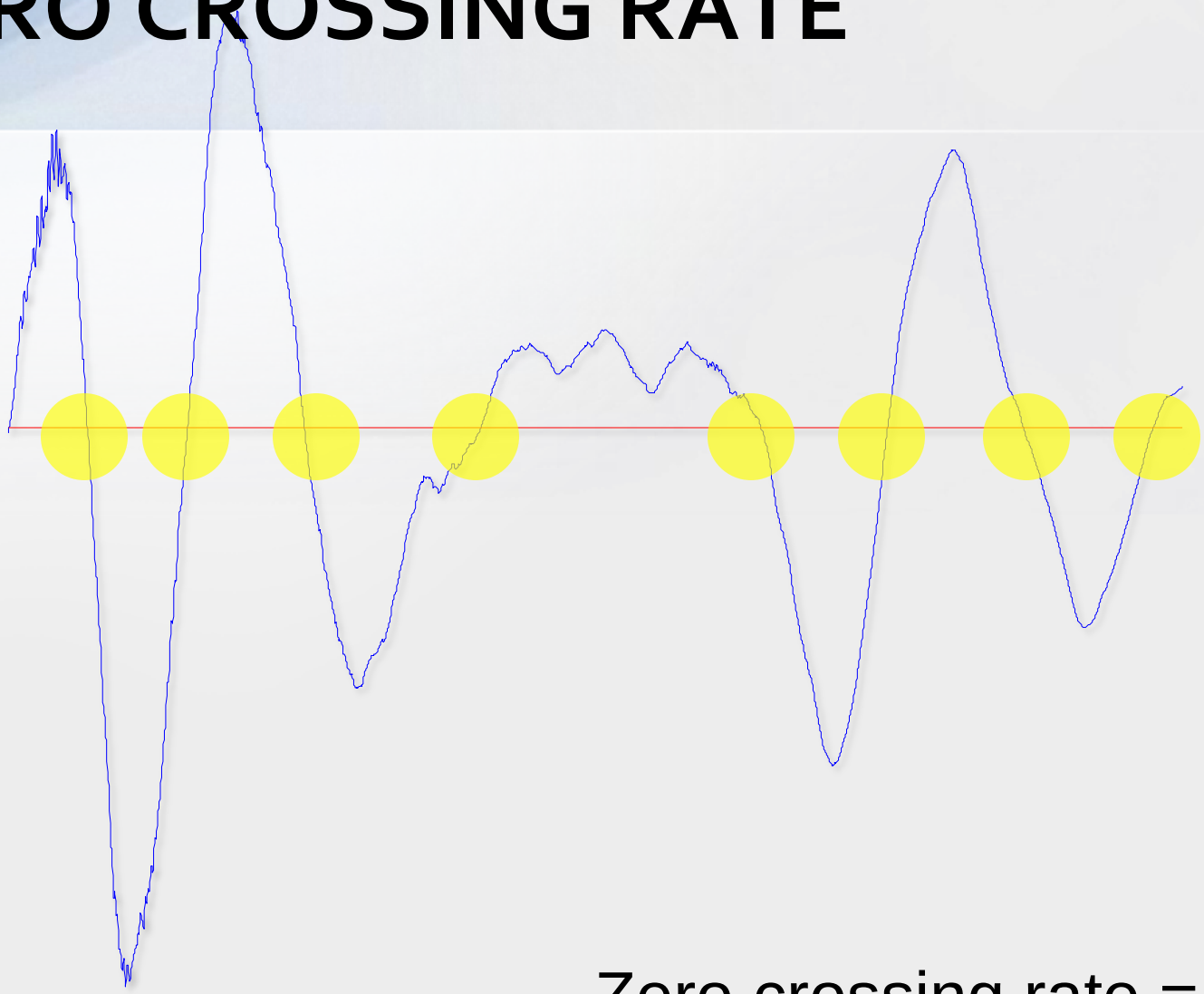
FEATURE EXTRACTION



Frame 1

FRAME 1

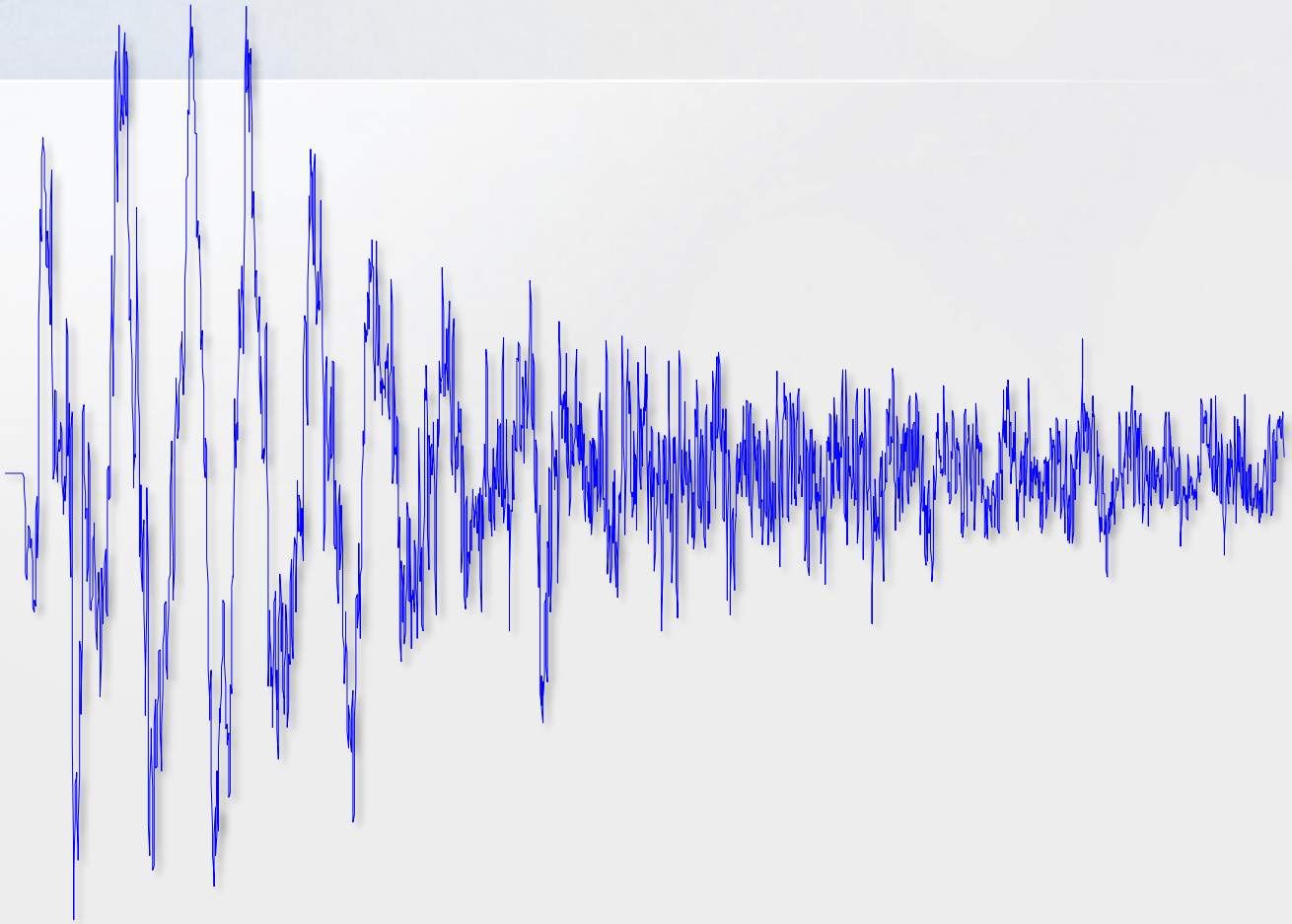
ZERO CROSSING RATE



FRAME 1

Zero crossing rate = 9

Frame 2

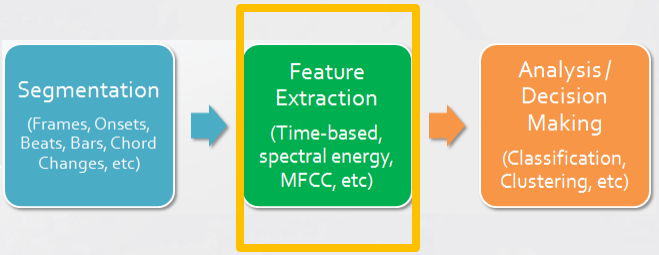


Zero crossing rate = 423

Features : SimpleLoop.wav

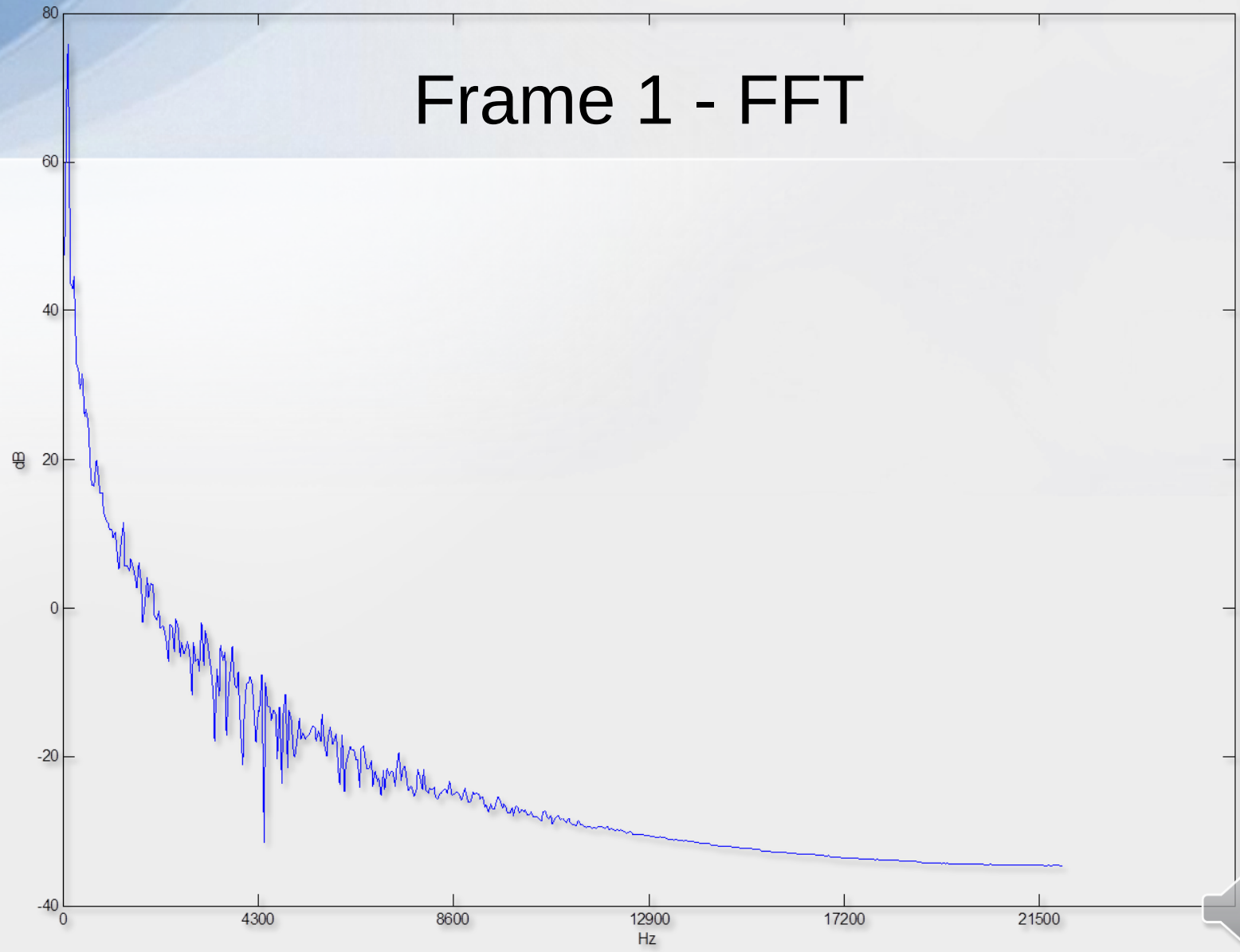
Frame	ZCR
1	9
2	423
3	22
4	28
5	390

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



FEATURE EXTRACTION

Frame 1 - FFT

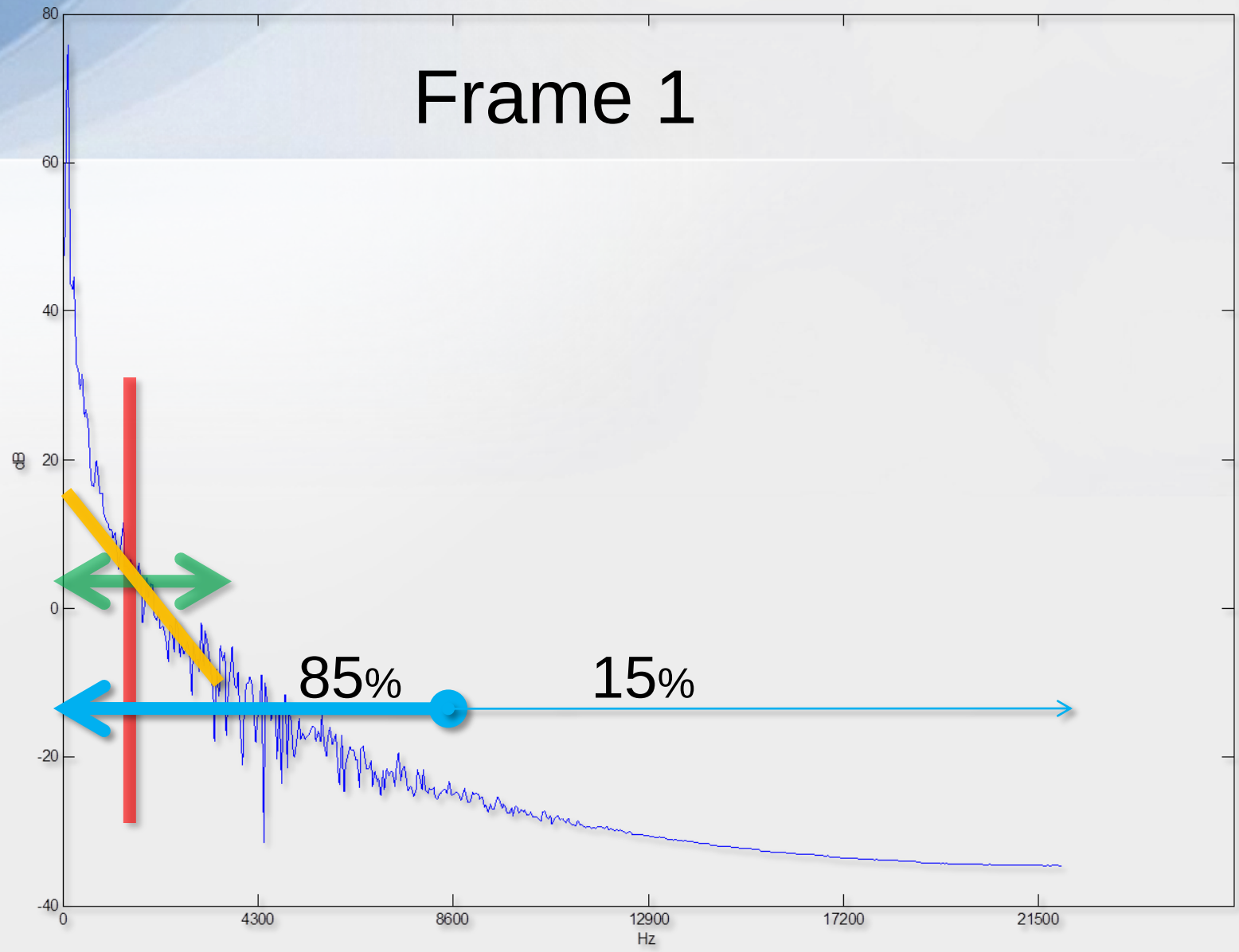


Spectral Features

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

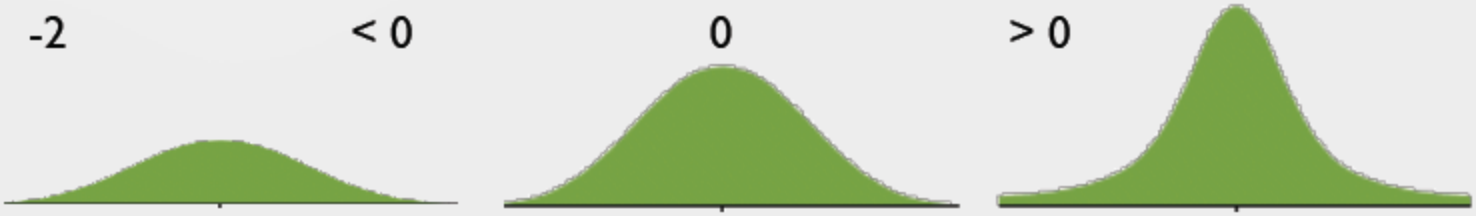
Spectral moments

Frame 1



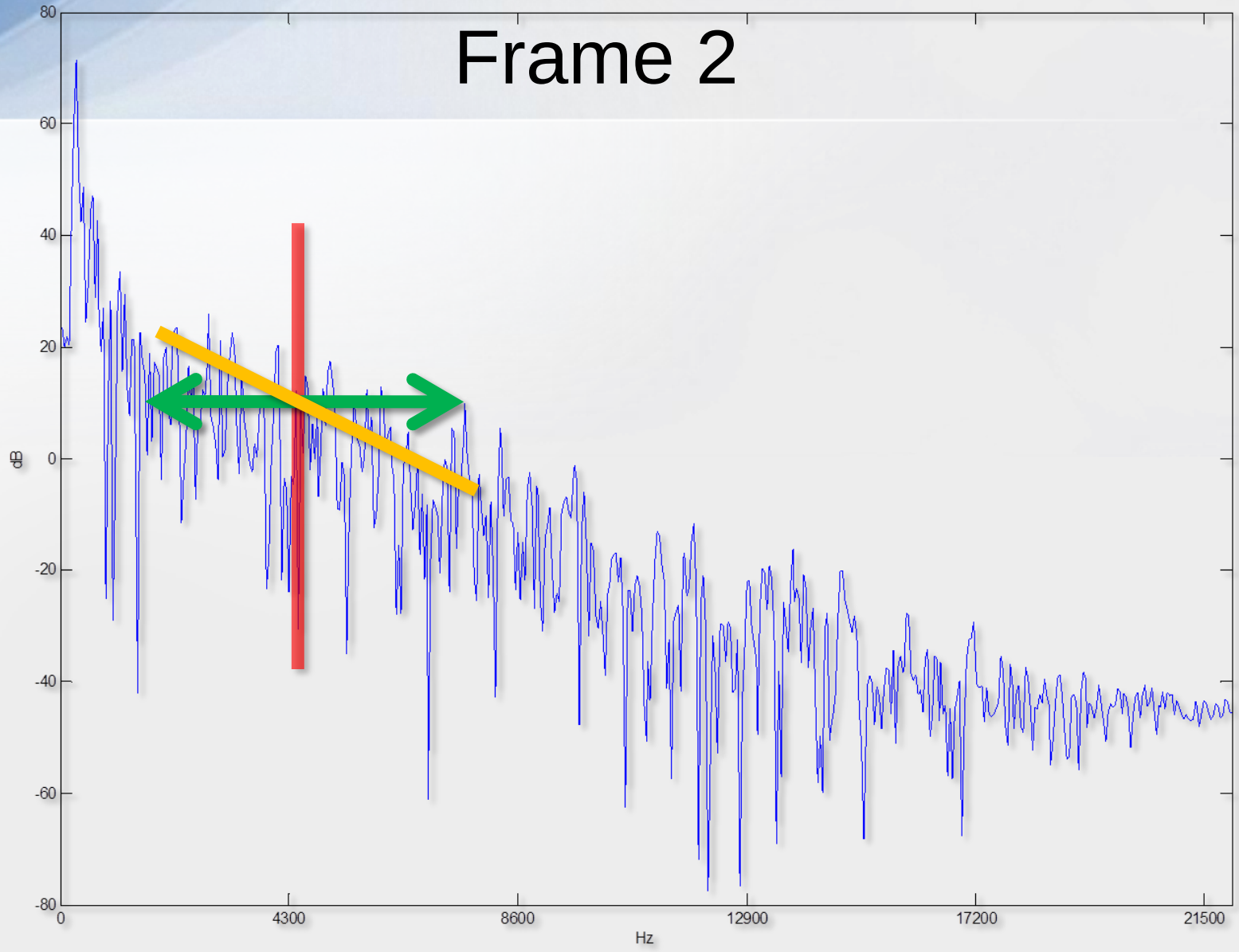
Skewness

Kurtosis



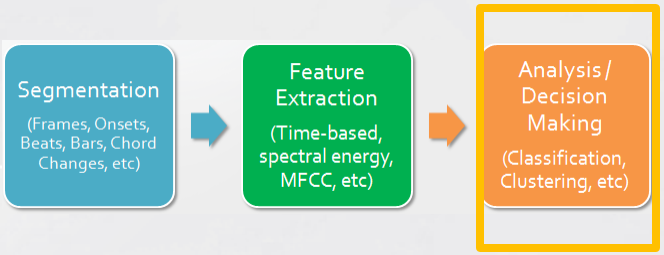
http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/userguide1.1

Frame 2



Example Feature Vector

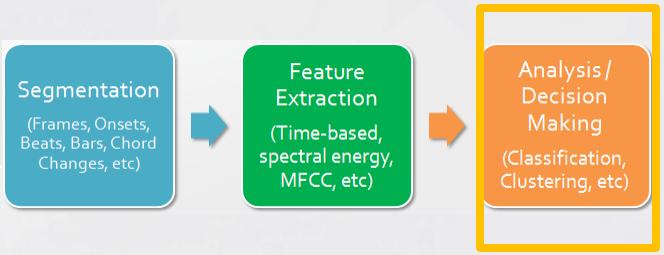
	ZCR	Centroid	Bandwidth	Skew
	1	2	3	4
1	205	982.0780	0.1452	1.3512e+03
2	150	621.0359	0.1042	296.0815
3	120.0000	361.6111	0.0607	263.7817
4	135	809.3978	0.1315	834.4116
5	220	634.7242	0.0906	274.5483
6	175	536.3318	0.0837	188.4155
7	190	567.0412	0.0953	253.0151
8	135	720.2892	0.1153	333.7646
9	195.0000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



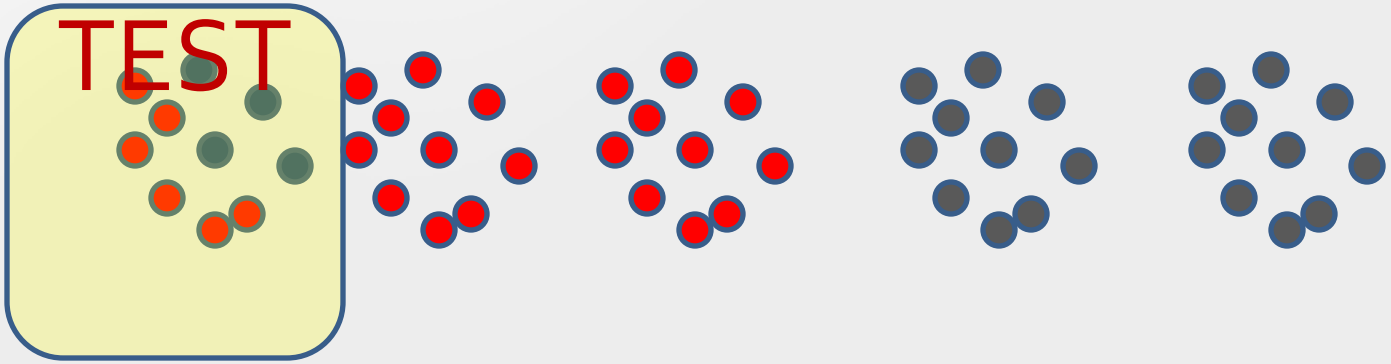
ANALYSIS AND DECISION MAKING
INSTANCE-BASED CLASSIFIERS (K-NN)


LIBERTY
IN GOD WE TRUST
LIBERTY
IN GOD WE TRUST

Training...

TRAINING SET

“1” “0”



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)

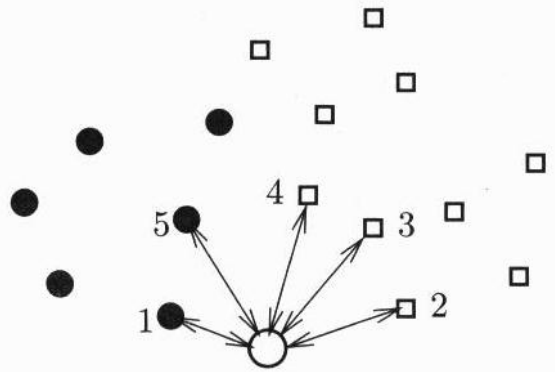
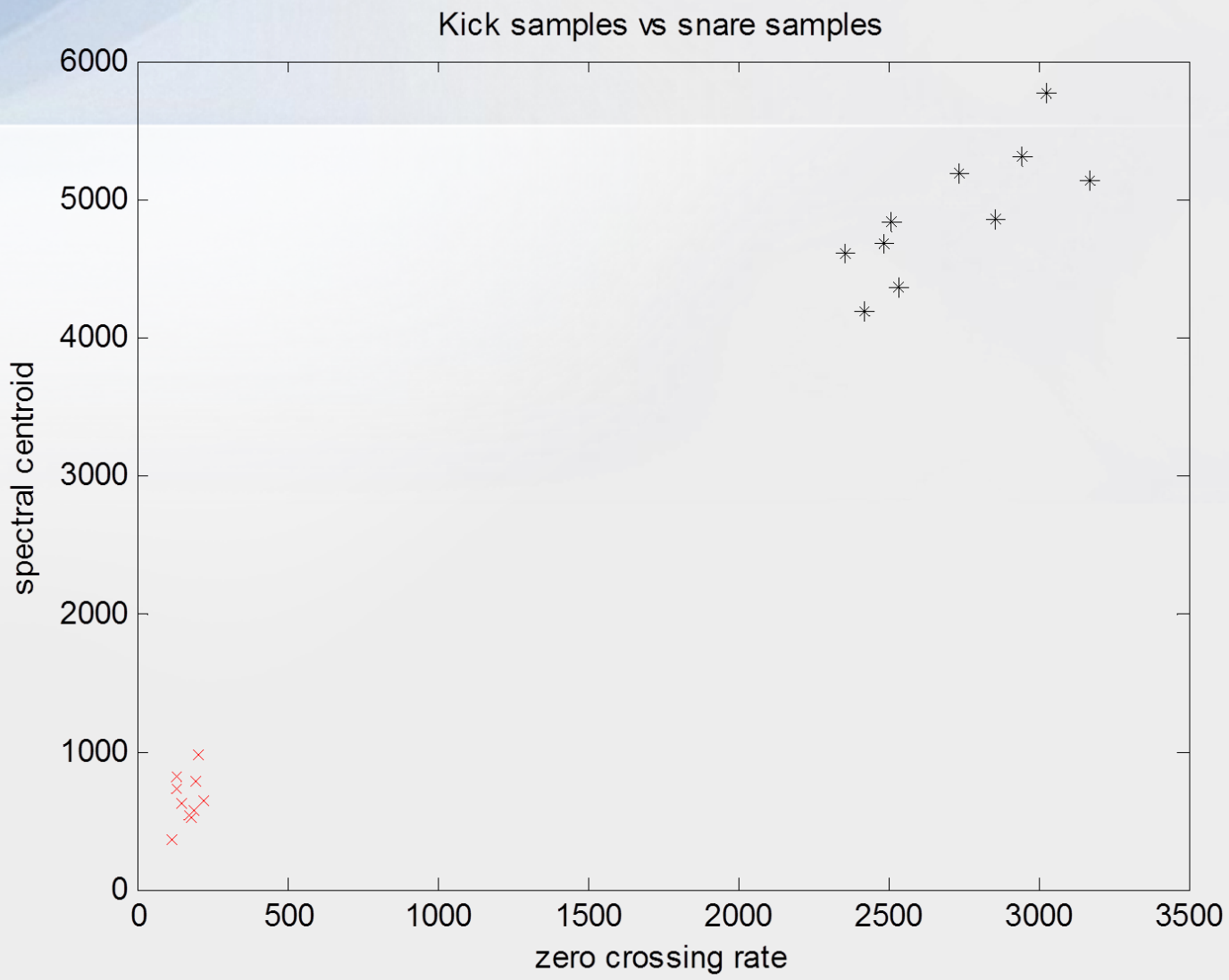


Fig. 2.15. k-nearest neighbours classification of two-dimensional data in the two-class case, with $k = 5$. The new datum $\mathbf{x}$ is represented by a non-filled circle. Elements of the training set (X, Y) are represented with dots (those with label -1) and squares (those with label $+1$). The arrow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

Distance Classification

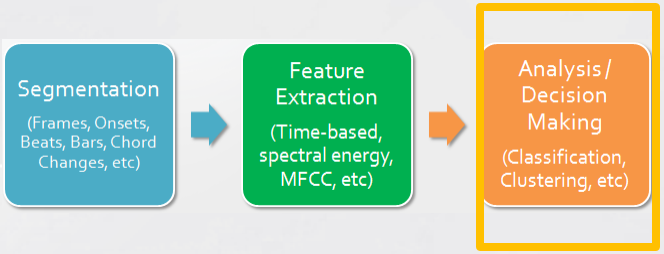
1. Find nearest neighbor
2. Find representative match via class prototype (e.g., center of group or mean of training data class)

Distance metric

Most common: Euclidean distance

Scaling!

	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



EVALUATING ANALYSIS SYSTEMS (the basics)

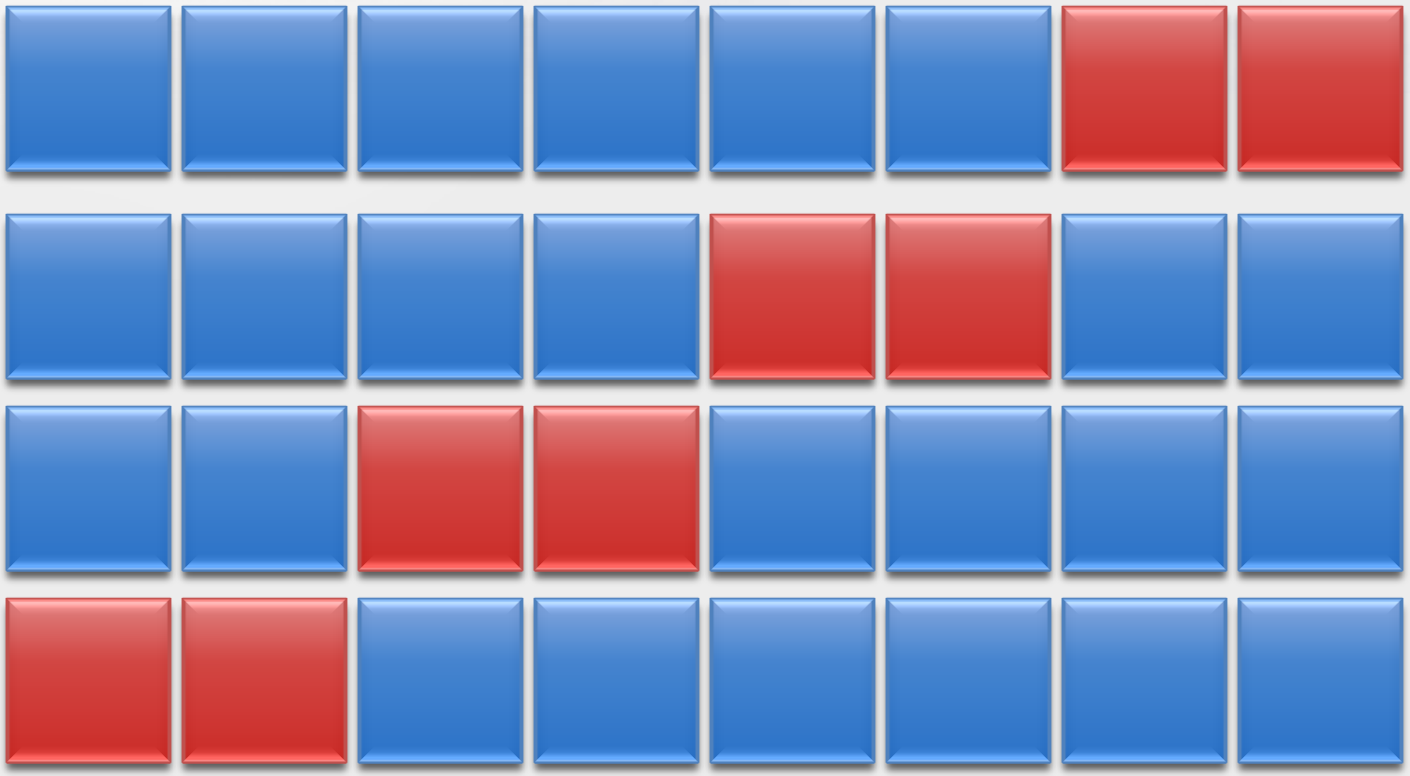
A bad evaluation metric

- "How many training examples are classified correctly?"

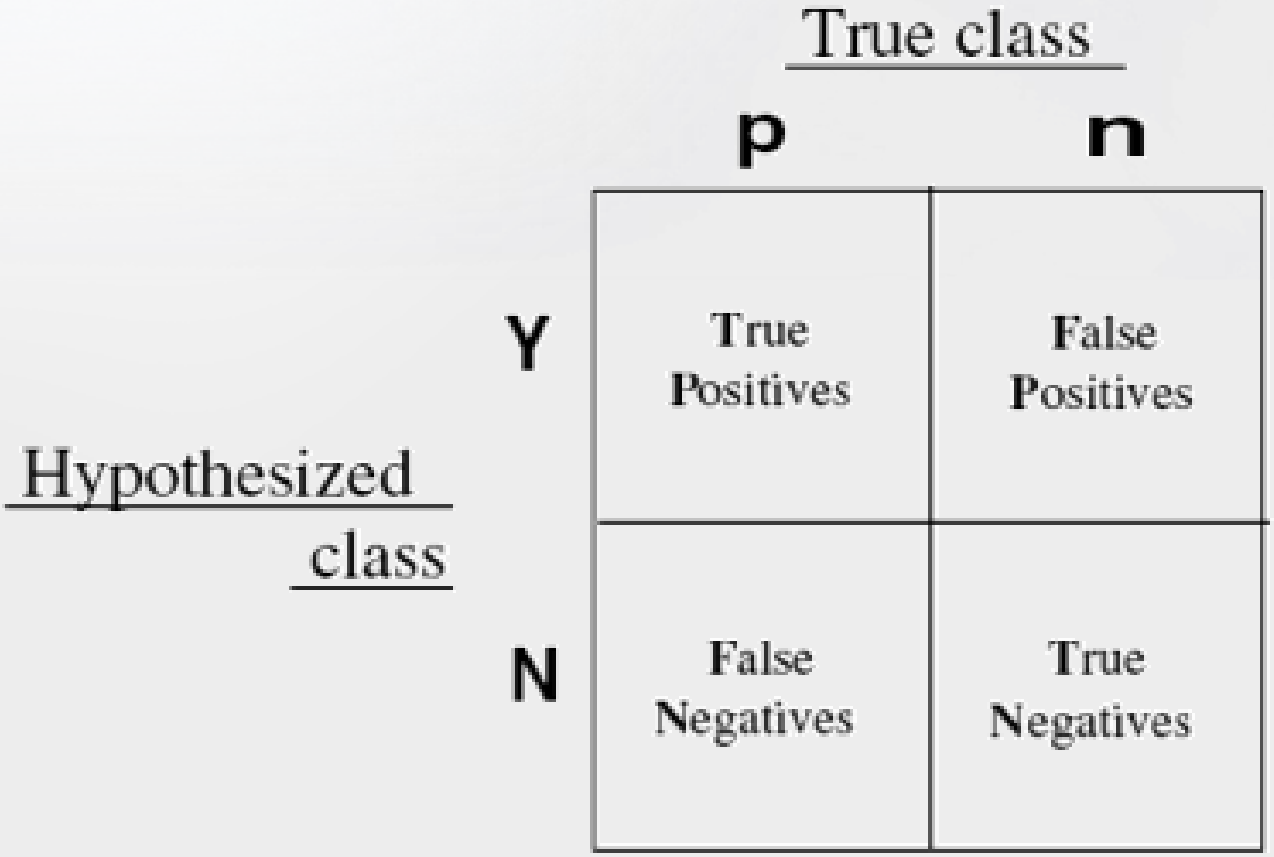
Image from Wikipedia, "Overfitting"

A better evaluation metric

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



Looking beyond accuracy



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

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

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

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

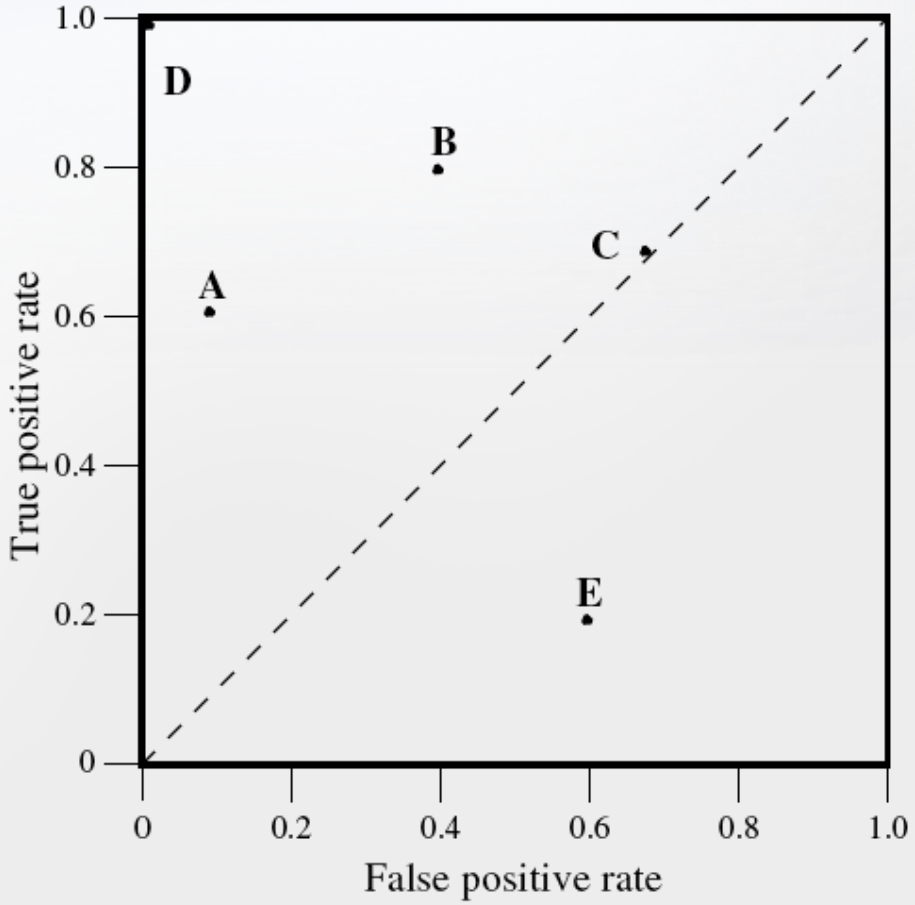
$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

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

ROC Graph

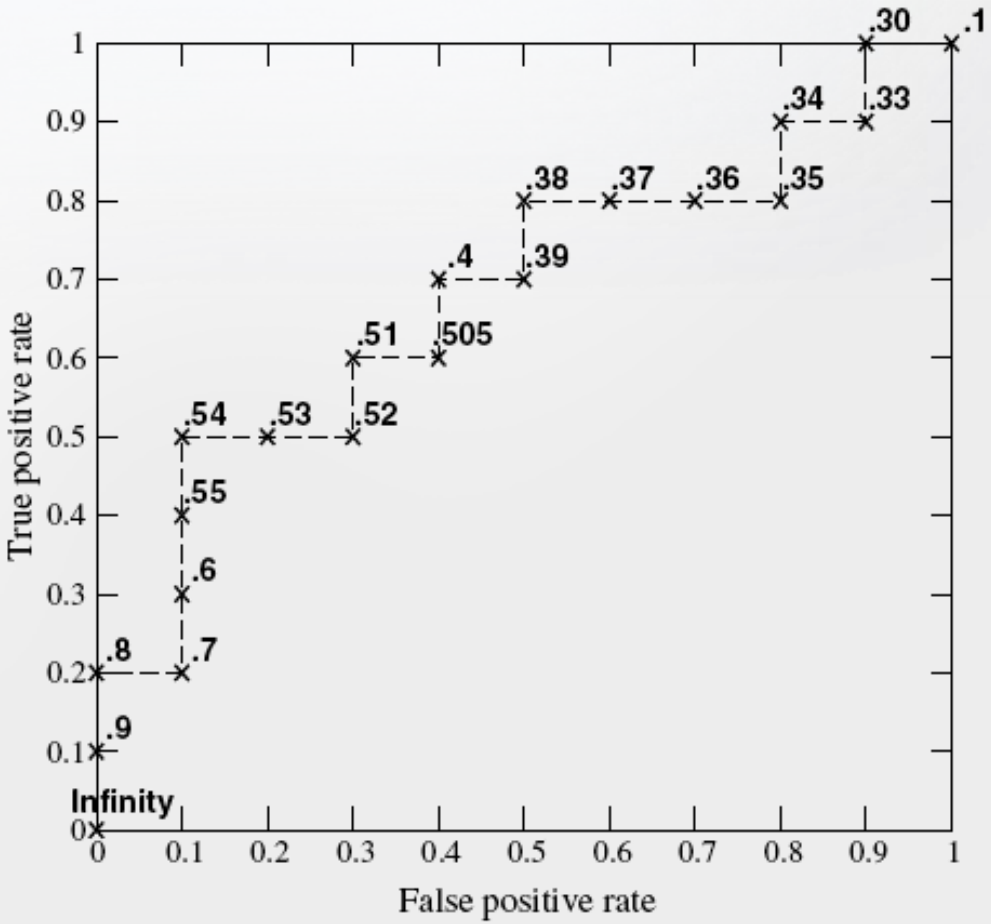
- “Receiver operating characteristics” curve
- A richer method of measuring model performance than classification accuracy
- Plots true positive rate vs false positive rate

ROC plot for discrete classifiers



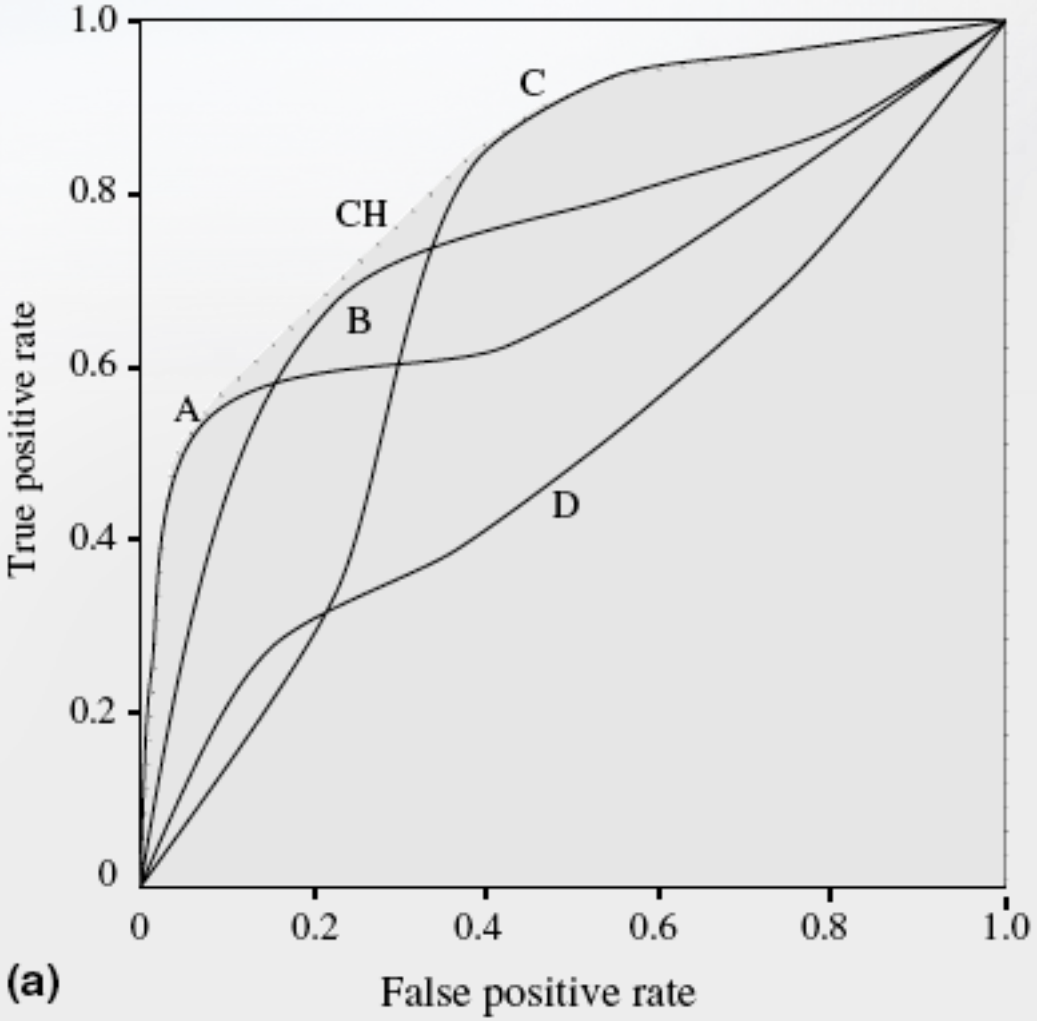
- Each classifier output is either right or wrong
  - Discrete classifier has single point on ROC plot
- The "Northwest" is better!
- Best sub-region may be task-dependent (conservative or liberal may be better)

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

> End of Lecture 1

Onset detection

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