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 contain 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.

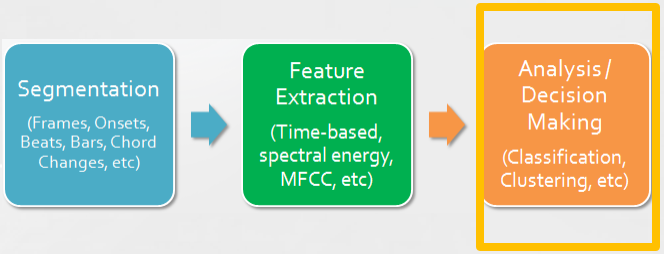
WIKI REFERENCES...

Review from Day 1

- What are the 3 major components of a MIR system?
- Name 3 ways of segmenting audio into frames
- Name 1 feature

- In Matlab, what does frame{1} mean?

- How did the lab go?
- What did you learn from the lab?
- Did you try other audio files?
- Did you do the simple instrument recognition?
- Sound snippet issue



ANALYSIS AND DECISION MAKING

CLASSIFICATION

k-NN

- Explanation…

- Dive into Matlab here for visualization

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)

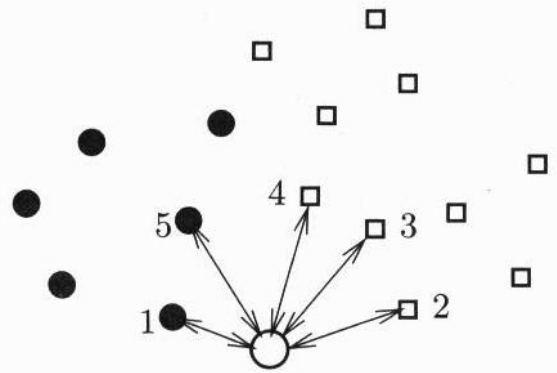
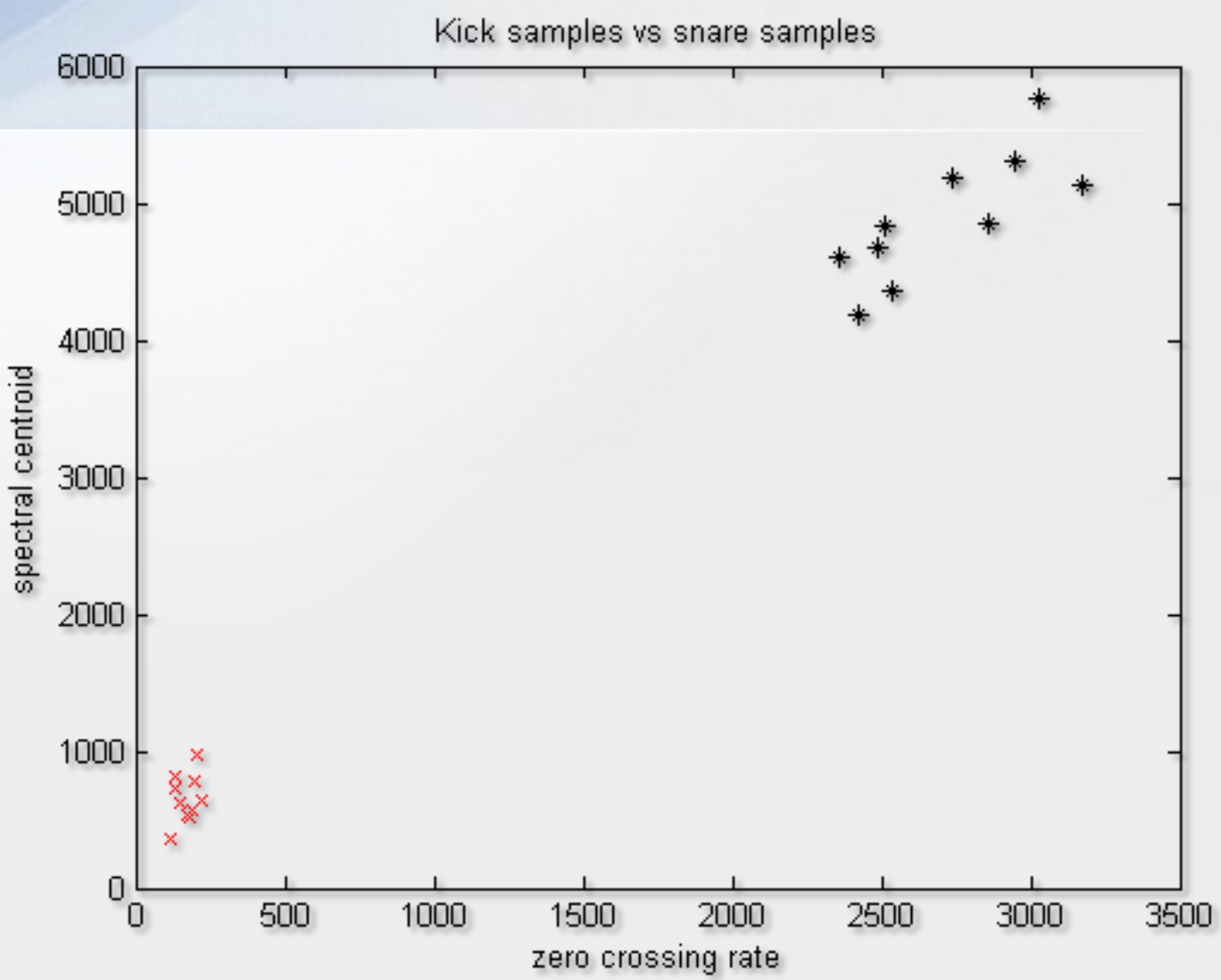


Fig. 2.15. k-nearest neighbours classification of two-dimensional data in the two-class case, with $k = 5$. The new datum $\mathbf{x}$ is represented by a non-filled circle. Elements of the training set (X, Y) are represented with dots (those with label -1) and squares (those with label $+1$). The arrow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

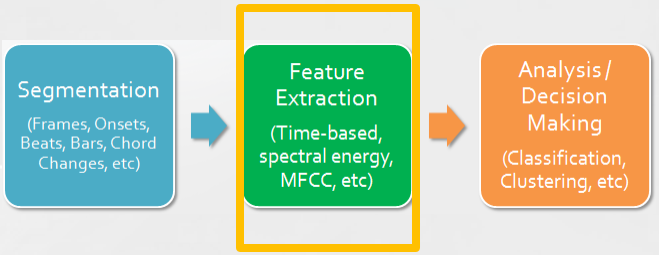
- Instance-based learning – training examples are stored directly, rather than estimate model parameters
- Generally choose k being odd to guarantee a majority vote for a class.

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



FEATURE EXTRACTION

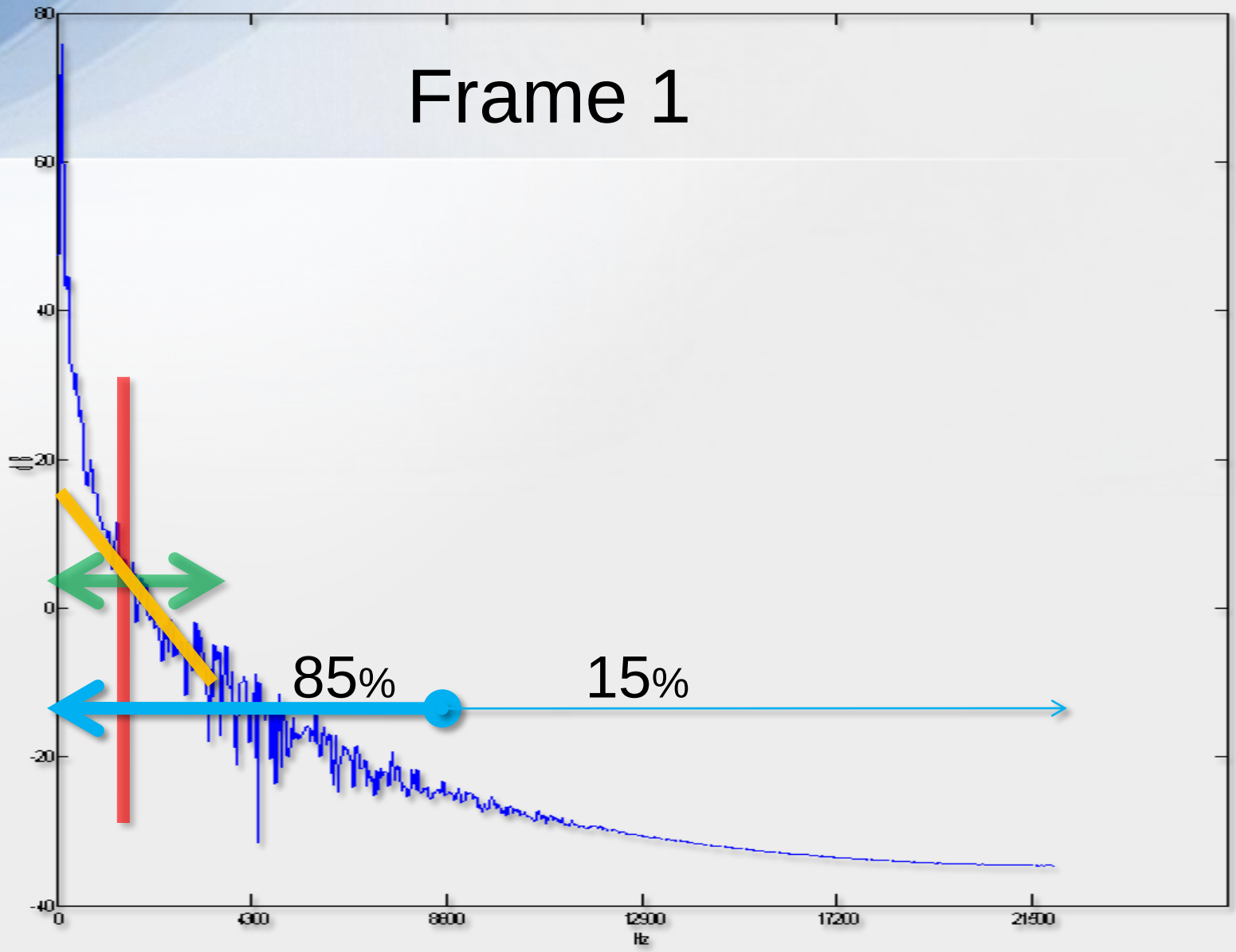
FFT?

Spectral Features

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

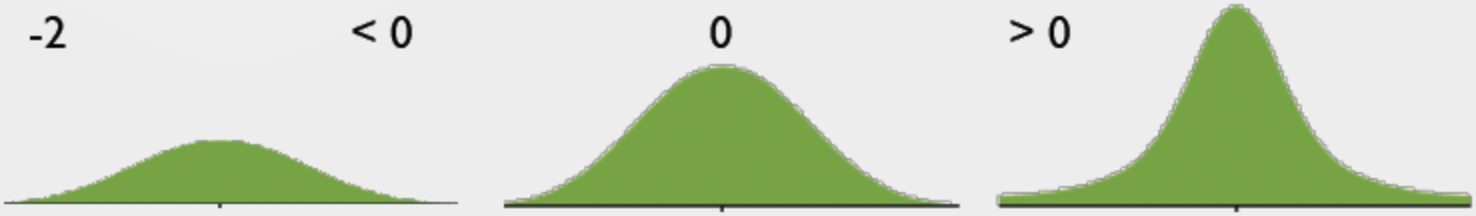
Spectral moments

Frame 1



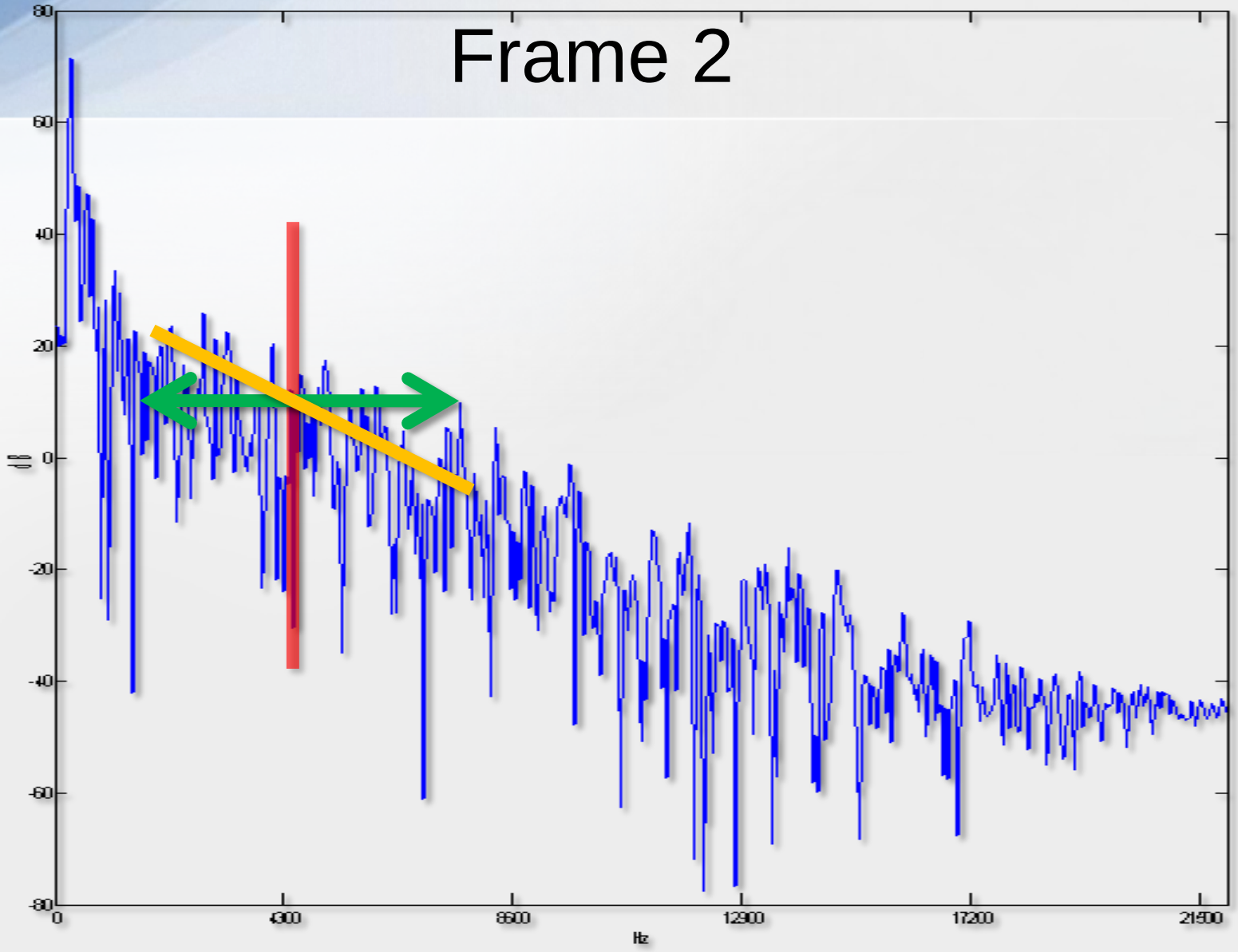
Skewness

Kurtosis



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

Frame 2



FEATURE DEMOS

- Simple re-ordering or slices:
  - Slice up loop into segments and sort via features
  - Play audio
  - Play whole song snippet

Real-world

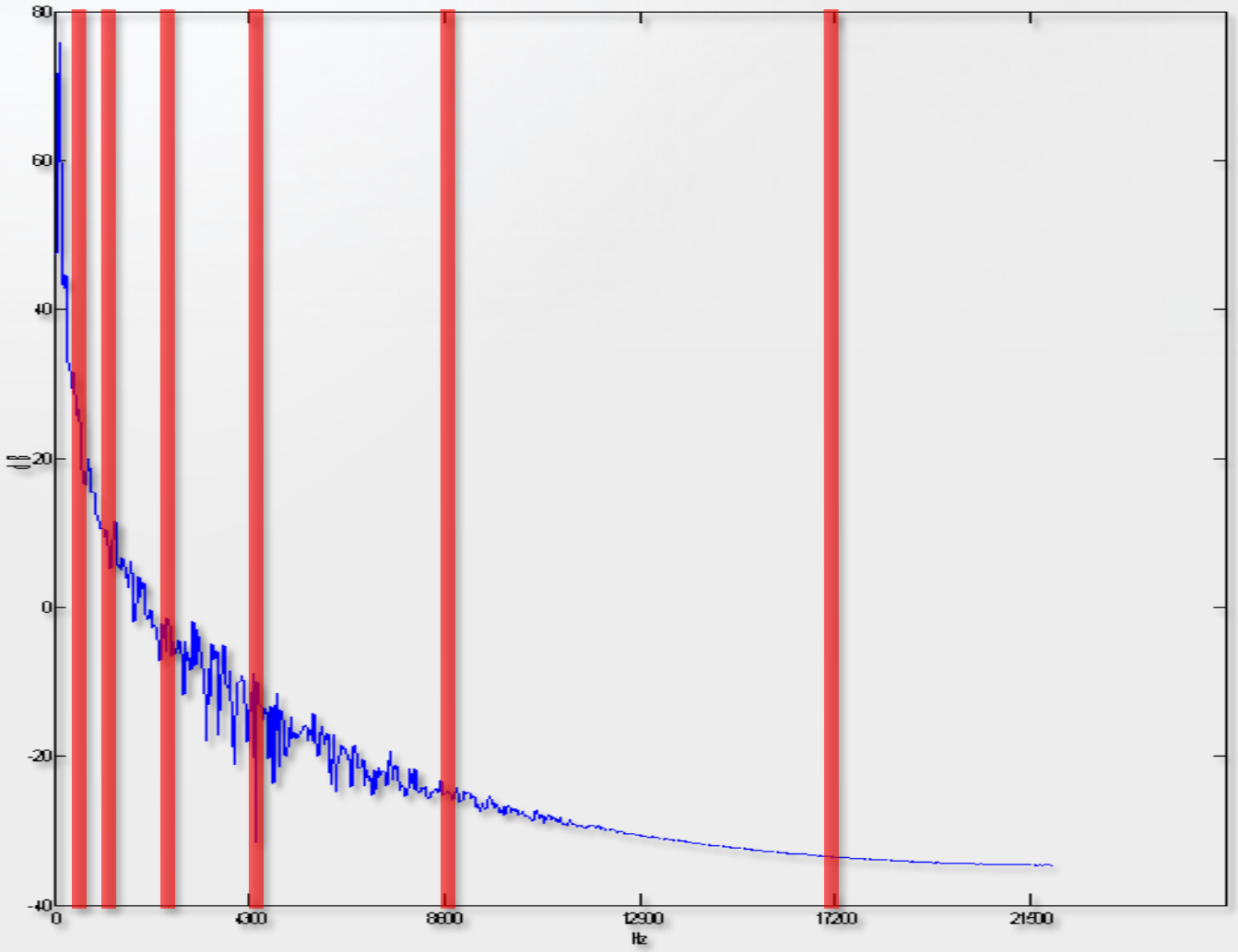
- YouTube uses AudibleMagic's - audio fingerprinting technology, to help identify the audio content of music partners like Warner Music, Sony BMG, and Universal.

- Shazam & Gracenote - "Tagging" - music listening to your phone

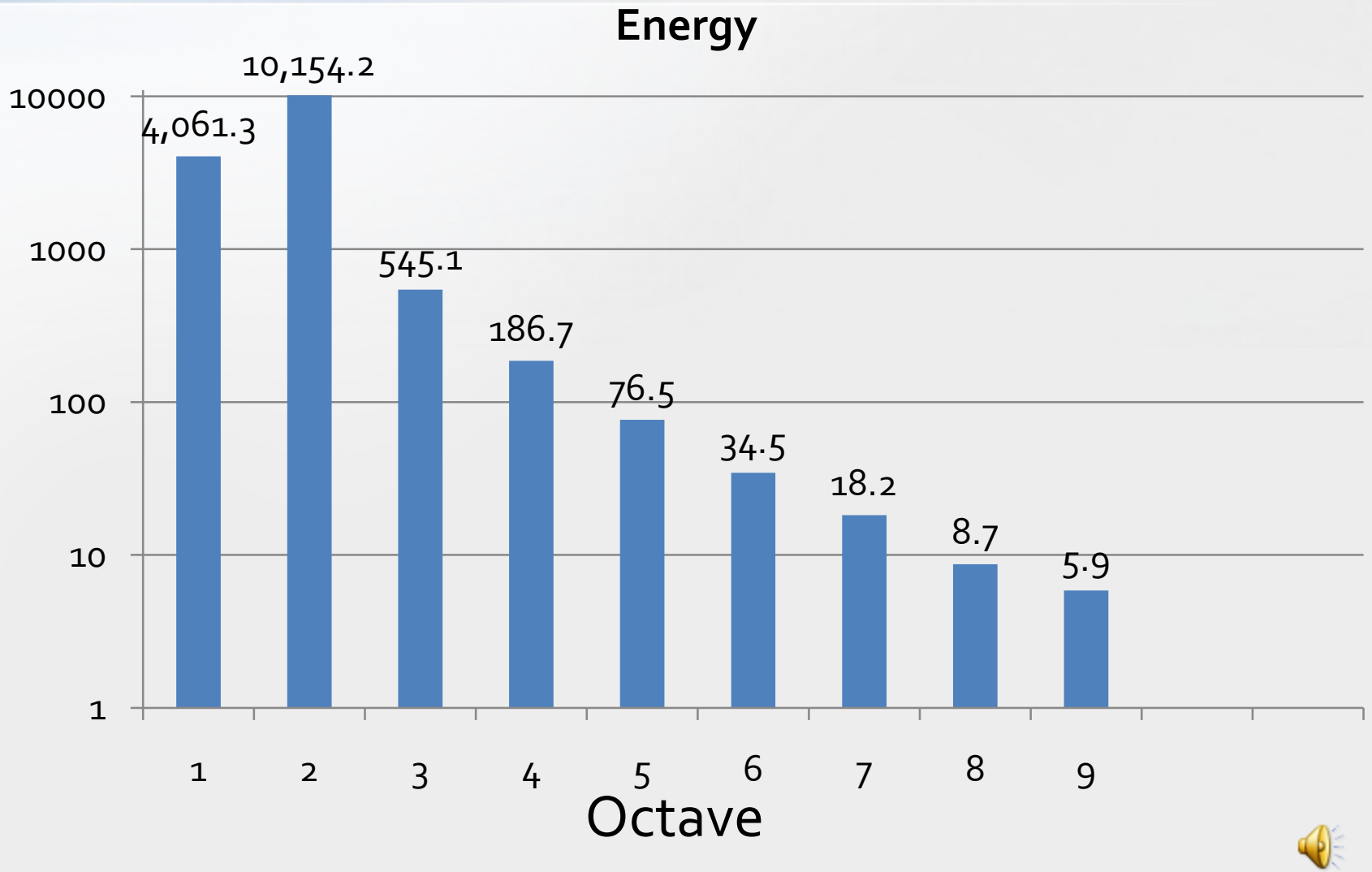
Real-world

- MeapSoft - link

Spectral Bands



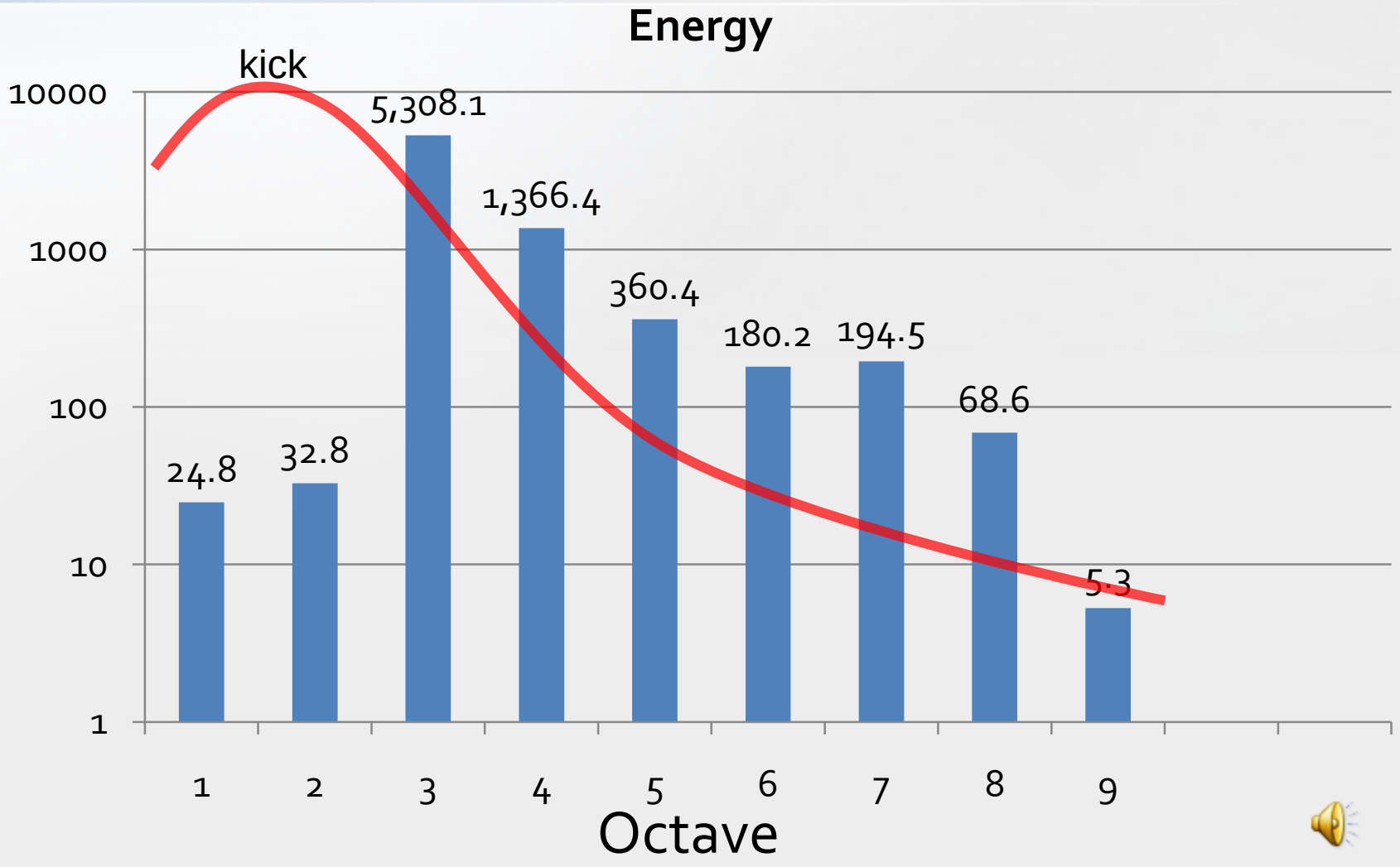
Frame 1


Energy
10000
1000
100
10
1
4,061.3
10,154.2
545.1
186.7
76.5
34.5
18.2
8.7
5.9
1
2
3
4
5
6
7
8
9
Octave

Features – Frame 1

Frame	ZCR	Centroid	BW	Skew	Kurtosis	E1	E2	E3	E4	E5	E6	E7	E8	E9
1	9	2.8kHz	5kHz	2.2	6.7	4000	10100	545	187	77	35	18	9	6

Frame 2


Energy
kick
5,308.1
1,366.4
360.4
180.2
194.5
68.6
5.3
24.8
32.8
10000
1000
100
10
1
1
2
3
4
5
6
7
8
9
Octave

Features : SimpleLoop.wav

Frame	ZCR	Centroid	BW	Skew	Kurtosis	E1	E2	E3	E4	E5	E6	E7	E8
1	9	2.8kHz	5kHz	2.2	6.7	4000	10100	545	187	77	35	18	9
2	423	3.1kHz	4kHz	2	7.2	24	33	5300	1366	360	180	194	68

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

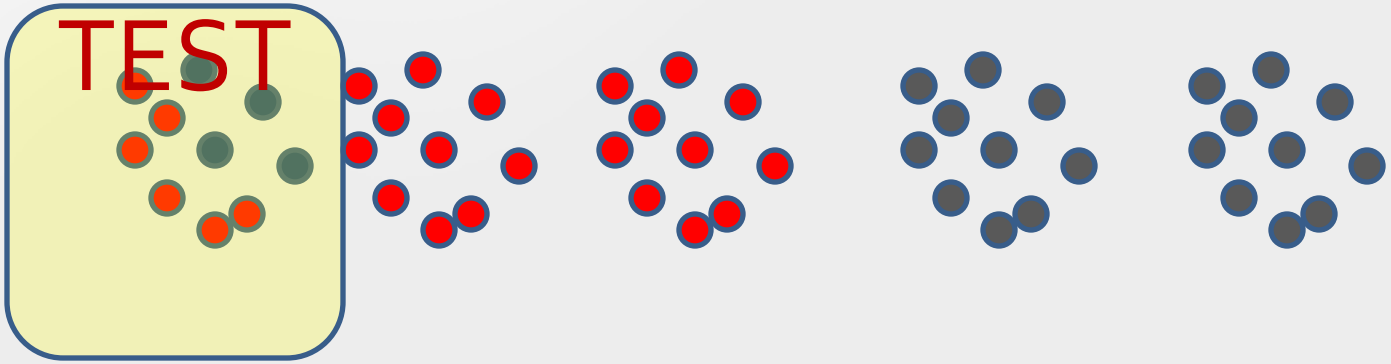
1116 2.6

263 1.45

Training...

TRAINING SET

“1” “0”



Lab 2 Prep – Read it over – and we'll go over

> End Day 2