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 .

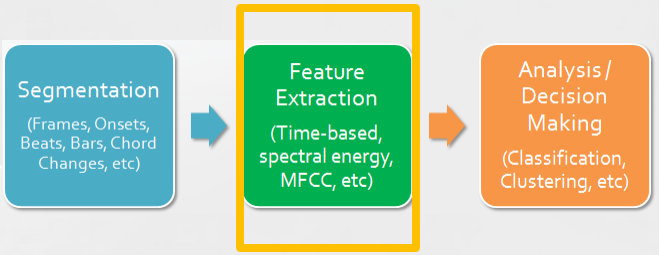
Click on the  symbol on the lower-left corner of a slide to access additional resources.

WIKI REFERENCES...

Review from Day 3

- What does it mean to "wrap" a chromagram?
- True or false – it's important to carefully chose meaningful features
- What are the 3 major components of a MIR system?

- How did the lab go?



FEATURE EXTRACTION

Developing an innate understanding of the features

Visualizing Features

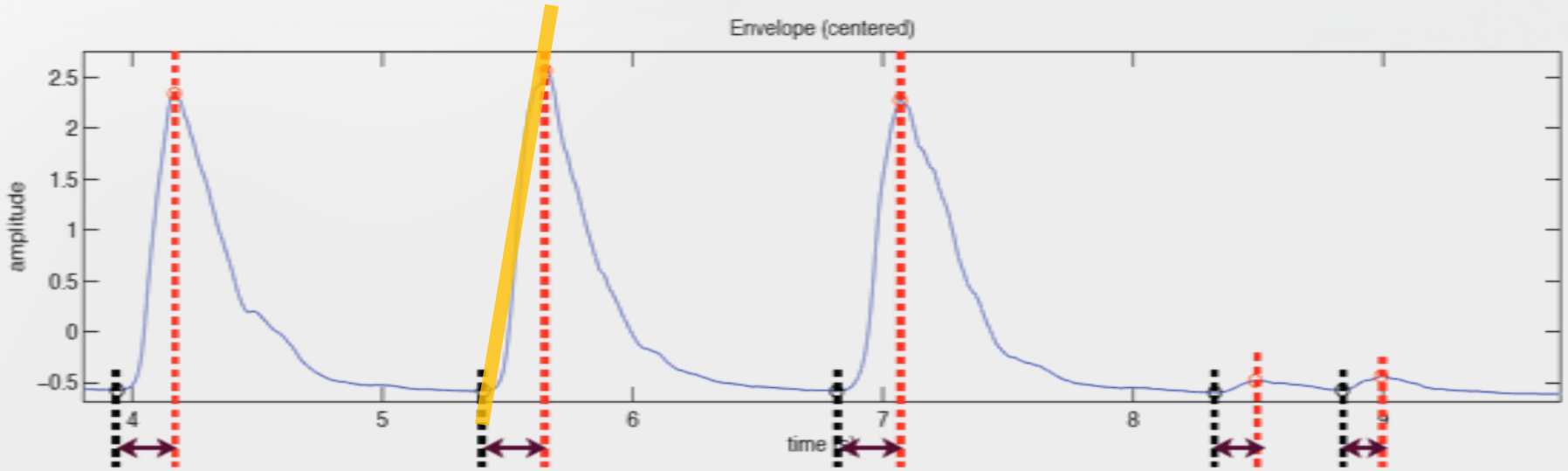
- 2-D visualization
- Matlab
- Weka's Visualization panes
- SOM, IsoMap, mapping multi-D down to 2D/3D

Listening to Features

- Play examples of Matt Hoffman's work (SoundLab)

Temporal Information

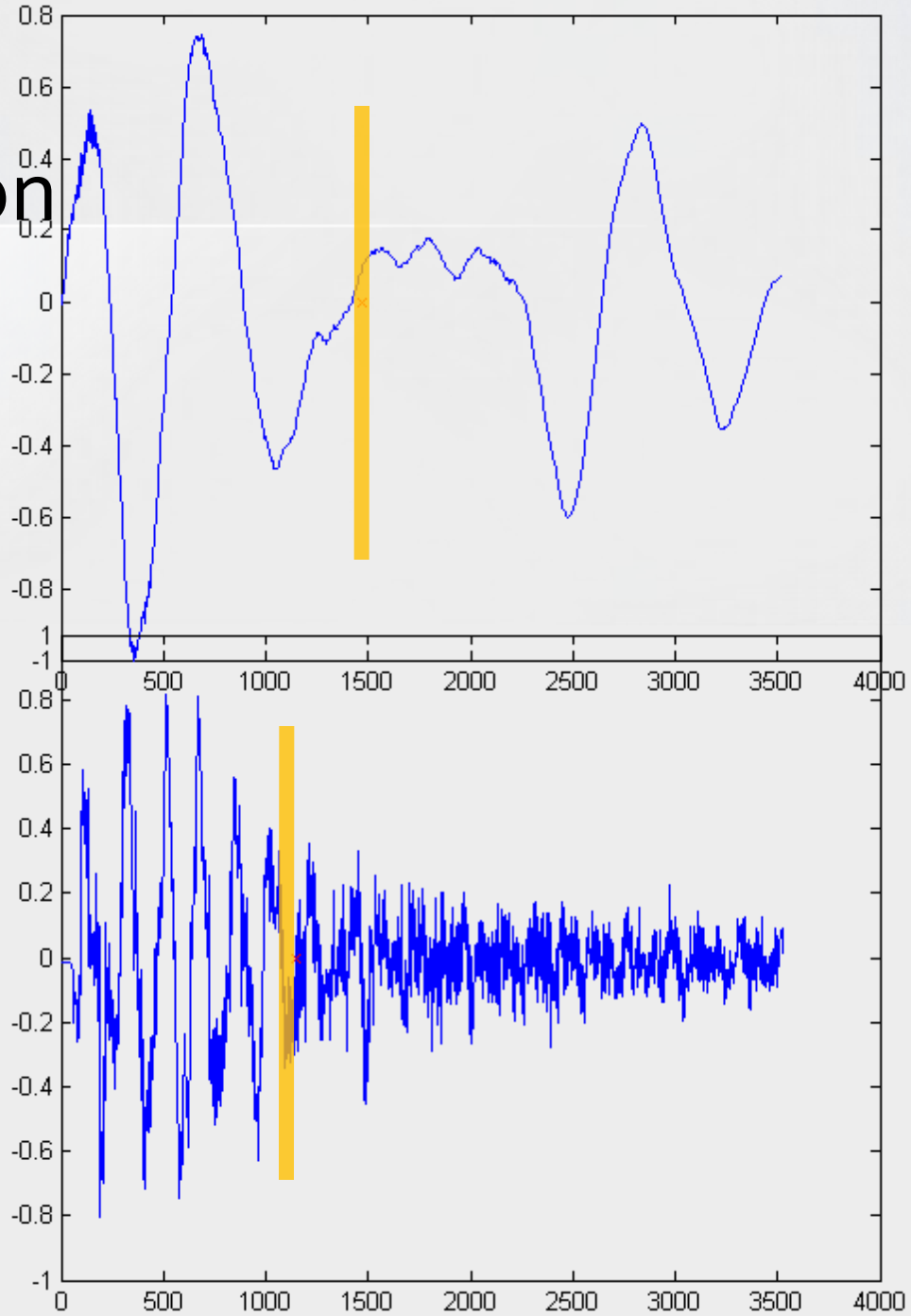
- Rise time or Attack time- time interval between the onset and instant of maximal amplitude
- Attack slope



Picture courtesy: Olivier Lartillot

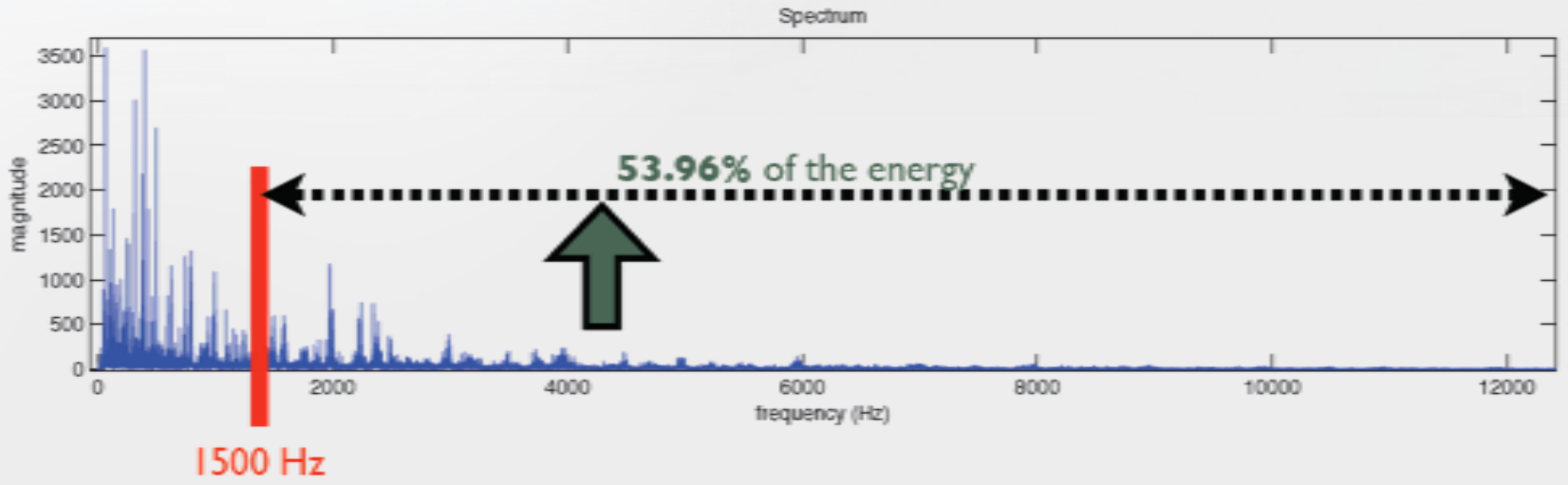
Temporal Information

- Temporal Centroid



Brightness

Amount of energy above a fixed-frequency.
(e.g., 1000, 1500, 3000 Hz)



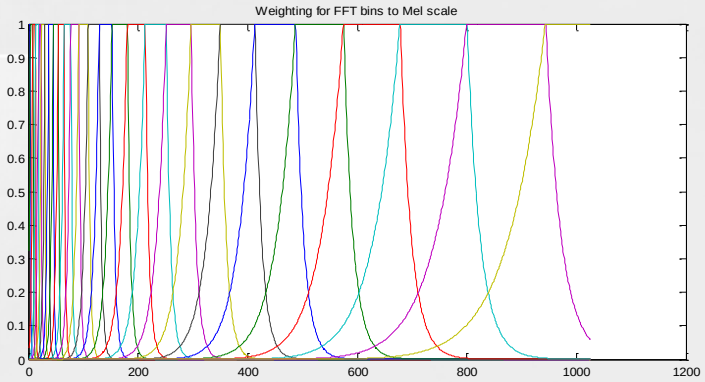
Picture courtesy Olivier Lartillot

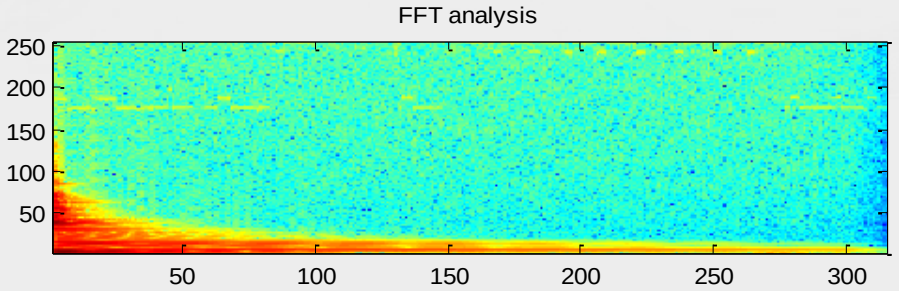
MFCCs

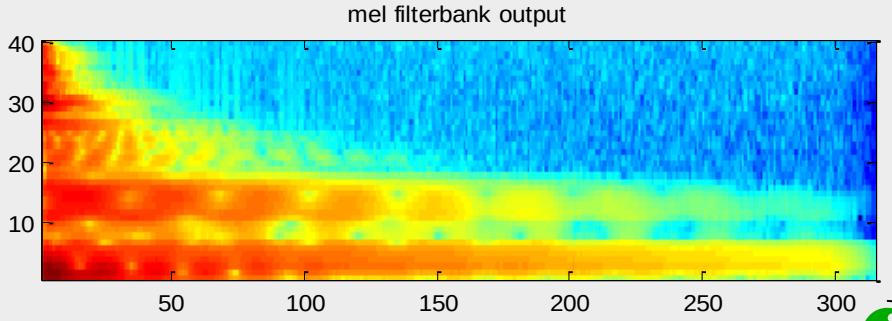
The idea of MFCCs is to capture spectrum in accordance with human perception.

1. STFT
2. log(STFT)
3. Perform mel-scaling to group and smooth coefficients. (perceptual weighting)
4. Decorrelate with DCT

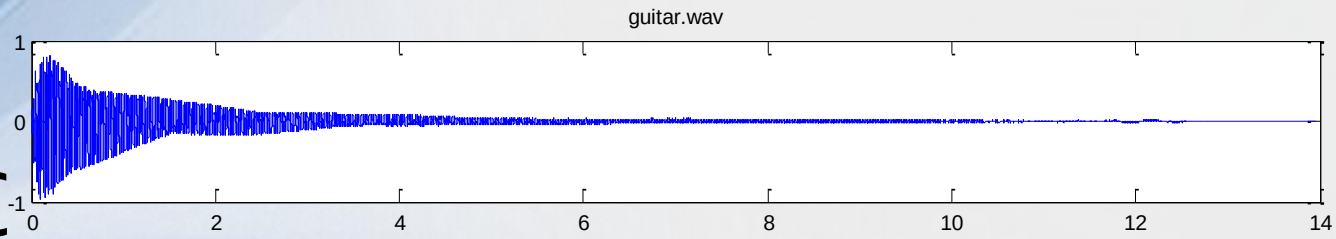
[…continued…]



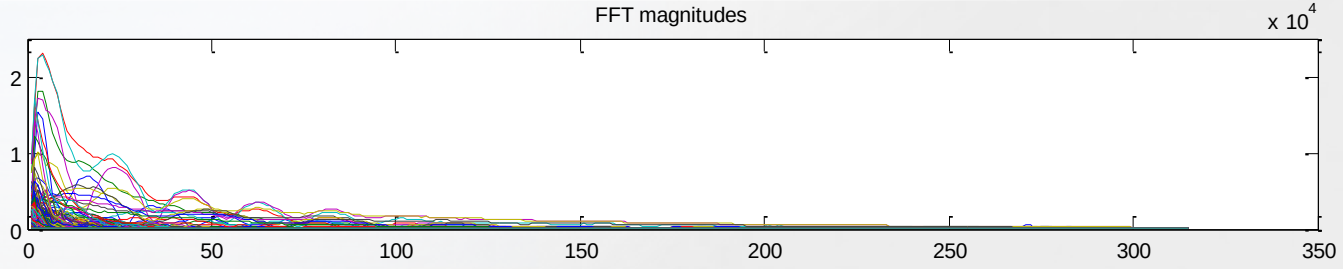




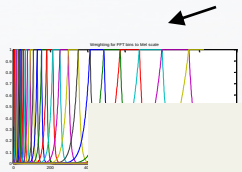
MFCC


guitar.wav
1
0
-1
0
2
4
6
8
10
12
14

1


FFT magnitudes
x 10^4
2
1
0
0
50
100
150
200
250
300
350

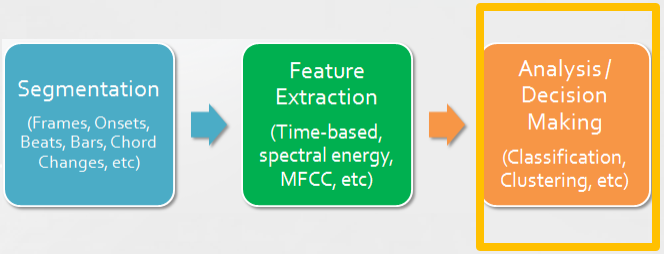
2



3

4

Spectral Energy vs. MFCC



ANALYSIS AND DECISION MAKING

Supervised vs. Unsupervised

- Unsupervised - “clustering”
- Supervised – binary classifiers (2 classes)
- Multiclass is derived from binary

Clustering

- Unsupervised learning – find pockets of data to group together
- Statistical analysis techniques

Clustering

- K = # of clusters

- Choosing the number of clusters – note that choosing the "best" number of clusters according to minimizing total squared distance will always result in same # of clusters as data points.

Clustering

The basic goal of clustering is to divide the data into groups such that the points within a group are close to each other, but far from items in other groups.

Hard clustering – each point is assigned to one and only one cluster.

K-Means

The key points relating to *k-means clustering are:*

- k-means is an automatic procedure for clustering unlabelled data;
- it requires a prespecified number of clusters;
- Clustering algorithm chooses a set of clusters with the minimum within-cluster variance
- Guaranteed to converge (eventually)
- Clustering solution is dependent on the initialization

K-Means

The initialization method needs to be further specified. There are several possible ways to initialize the cluster centers:

- *Choose random data points as cluster centers*
- *Randomly assign data points to K clusters and compute means as initial centers*
- *Choose data points with extreme values*
- *Find the mean for the whole data set then perturb into k means*

>end Day 4