Intelligent Audio Systems:
A review of the foundations and applications of semantic audio analysis and music information retrieval
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 $y = +1$. 
**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
Skewness

Kurtosis

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
## Features – Frame 1

<table>
<thead>
<tr>
<th>Frame</th>
<th>ZCR</th>
<th>Centroid</th>
<th>BW</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
<th>E7</th>
<th>E8</th>
<th>E9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>2.8kHz</td>
<td>5kHz</td>
<td>2.2</td>
<td>6.7</td>
<td>4000</td>
<td>10100</td>
<td>545</td>
<td>187</td>
<td>77</td>
<td>35</td>
<td>18</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>
Frame 2

Octave

Kick

Energy

24.8
32.8
5,308.1
1,366.4
360.4
180.2
194.5
68.6
5.3

1 2 3 4 5 6 7 8 9
# Features: SimpleLoop.wav

<table>
<thead>
<tr>
<th>Frame</th>
<th>ZCR</th>
<th>Centroid</th>
<th>BW</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
<th>E7</th>
<th>E8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>2.8kHz</td>
<td>5kHz</td>
<td>2.2</td>
<td>6.7</td>
<td>4000</td>
<td>10100</td>
<td>545</td>
<td>187</td>
<td>77</td>
<td>35</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>423</td>
<td>3.1kHz</td>
<td>4kHz</td>
<td>2</td>
<td>7.2</td>
<td>24</td>
<td>33</td>
<td>5300</td>
<td>1366</td>
<td>360</td>
<td>180</td>
<td>194</td>
<td>68</td>
</tr>
</tbody>
</table>
## Scaling!

<table>
<thead>
<tr>
<th>ZCR</th>
<th>Centroid</th>
<th>Bandwidth</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>205</td>
<td>982.0780</td>
<td>0.1452</td>
<td>1.3512e+03</td>
</tr>
<tr>
<td>150</td>
<td>621.0359</td>
<td>0.1042</td>
<td>296.0815</td>
</tr>
<tr>
<td>120.0000</td>
<td>361.6111</td>
<td>0.0607</td>
<td>263.7817</td>
</tr>
<tr>
<td>135</td>
<td>809.3978</td>
<td>0.1315</td>
<td>834.4116</td>
</tr>
<tr>
<td>220</td>
<td>634.7242</td>
<td>0.0906</td>
<td>274.5483</td>
</tr>
<tr>
<td>175</td>
<td>536.3318</td>
<td>0.0837</td>
<td>188.4155</td>
</tr>
<tr>
<td>190</td>
<td>567.0412</td>
<td>0.0953</td>
<td>253.0151</td>
</tr>
<tr>
<td>135</td>
<td>720.2892</td>
<td>0.1153</td>
<td>333.7646</td>
</tr>
<tr>
<td>195.0000</td>
<td>778.5310</td>
<td>0.1407</td>
<td>1.2328e+03</td>
</tr>
<tr>
<td>185</td>
<td>514.4315</td>
<td>0.0717</td>
<td>183.0322</td>
</tr>
</tbody>
</table>

1116  2.6  263  1.45
Training...

TRAINING SET

“1”

“0”

TEST
Lab 2 Prep – Read it over – and we’ll go over
> End Day 2