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 supplemental material for lectures - 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?
- Did you try other audio files?
- Did you do the simple instrument recognition?
Filterbank

Detect change in band 1

Detect change in band 2

... ...

Detect change in band N

Combine results

Detection output

audio
threshold
Adaptive threshold
Beat and Tempo Detection

• Beat detection
  – Tempo (e.g., 125 bpm)
    • Detecting periodicities from the onset detection curve
  – Beat
    • AKA “Tactus” – the “foot tapping rate”
    • Time-frequency analysis -> Resonators -> Probablistic model
  – Measure
    • Musical change rate
    • Harmonic change rate
Beat and Tempo Detection

Many, many, many approaches.

Some simple ones:

1. Autocorrelation function of the onset detection curve
2. Spectral decomposition of the onset detection curve
3. Combine both strategies: the autocorrelation function is translated into the frequency domain in order to be compared to the spectrum curve - two curves are then multiplied.

Peak picking is applied to the autocorrelation function or to the spectrum representation.
FEATURE EXTRACTION
Spectral Features

- Spectral Flatness Measure
- Spectral Crest Factor
- Spectral Flux
ANALYSIS AND DECISION MAKING
## Scaling!

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

TRAINING SET

“1”    “0”

TEST
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
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 $x$ 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 $x$ and its 5 nearest neighbours. Three of them are squares, which makes $x$ 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 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](#)