ECEN 4532: Digital Signal Processing Lab

Lecture Notes: Lab 2

Instructor: Prof. Farhad Pourkamali-Anaraki

University of Colorado at Boulder

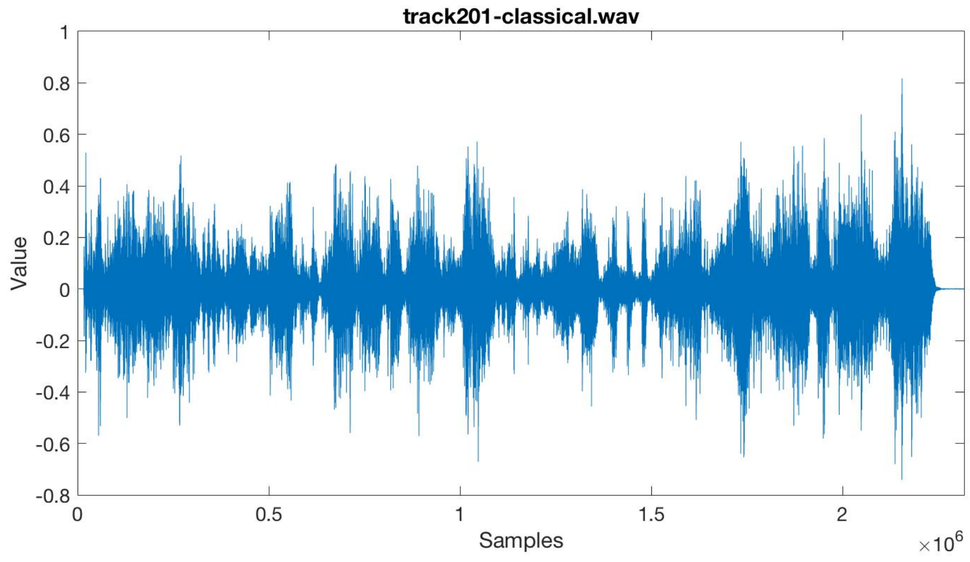
Spring 2017

University of Colorado
Boulder

Lab 2 Overview

- Digital signal processing methods for **content-based** music information retrieval

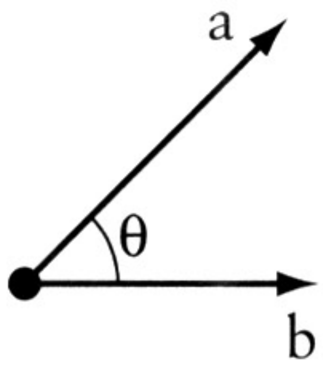
- We will explore **higher** level musical features:
  1. Rhythm
  2. Chroma



Vectors and Matrices (1)

two vectors of attributes or features

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$



measure of similarity

$$\cos(\theta) = \frac{\langle a, b \rangle}{\|a\| \|b\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}}$$

The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 indicating orthogonality, and in-between values indicating intermediate similarity or dissimilarity.

Vectors and Matrices (2)

```
>> a = [1;2;3;4];
>> b = [1;2;3;4];
>> sim = (a' * b) / ( norm(a) * norm(b) )

sim =

     1

>> b = [1;2.2;3;4];
>> sim = (a' * b) / ( norm(a) * norm(b) )

sim =

    0.9994

>> b = [1;2.5;3;4];
>> sim = (a' * b) / ( norm(a) * norm(b) )

sim =

    0.9966
```

Vectors and Matrices (3)

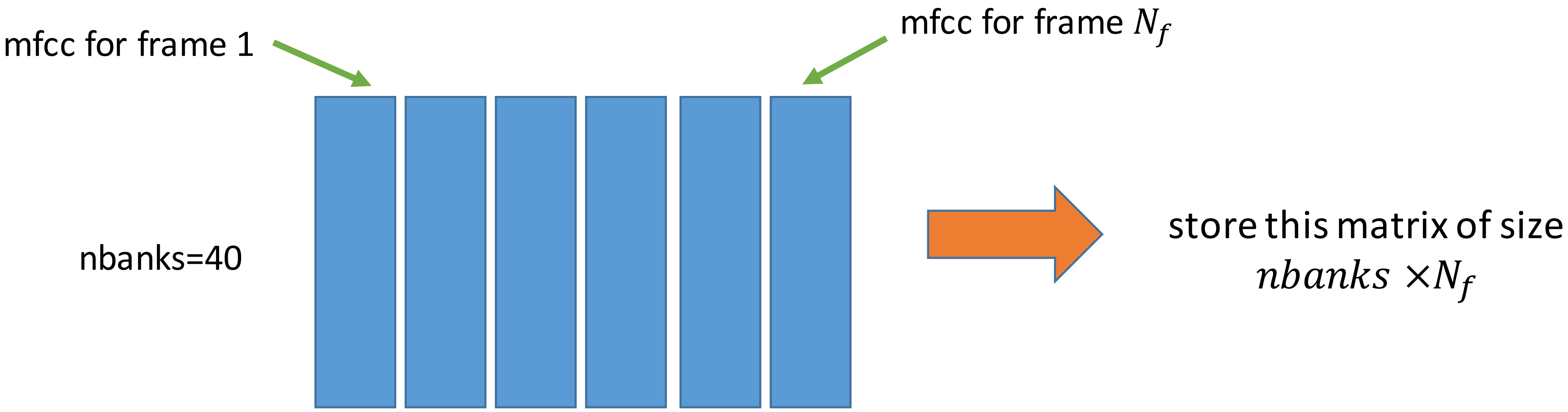
```
>> A = magic(4)

A =

    16     2     3    13
     5    11    10     8
     9     7     6    12
     4    14    15     1

>> a = A(:,4)

a =

    13
     8
    12
     1
```

Spectral Decomposition

- The mfcc measured in **decibels** (dB) is defined by:

$$10\log_{10}(\text{mfcc})$$

- Therefore, you should have a 2D array (matrix):



Similarity Matrix

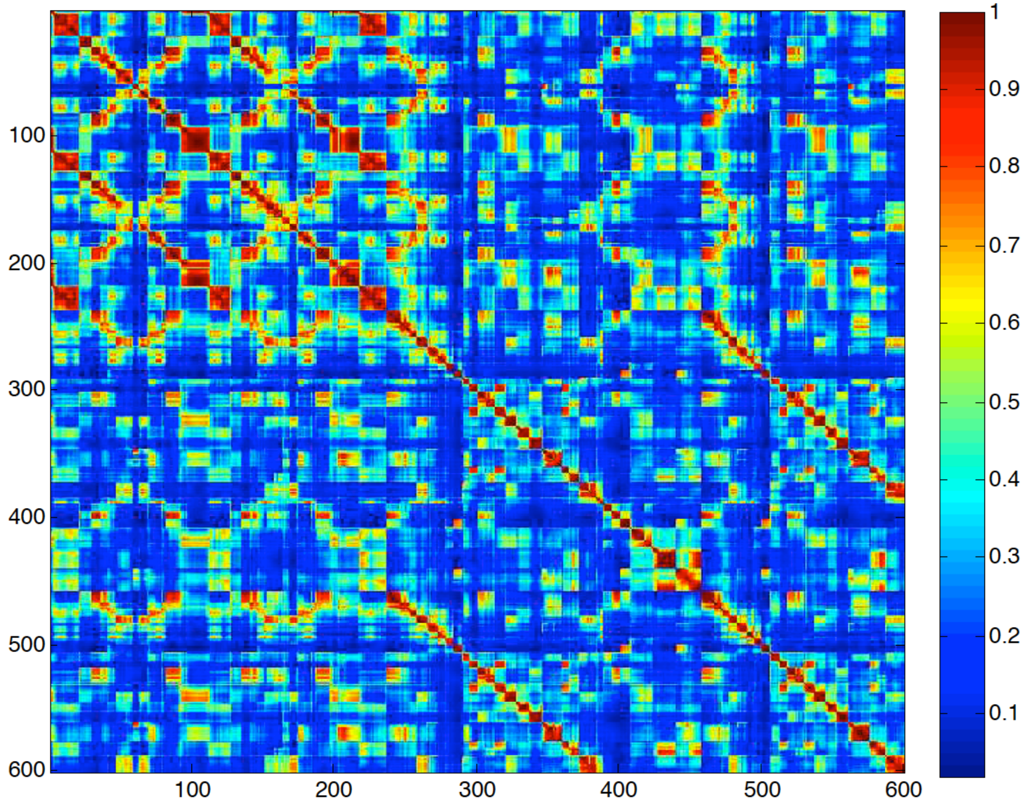
- $S(i, j)$ measures the **cosine** of the angle between the spectral signatures of frames i and j:

$$S(i,j) = \frac{\langle \mathtt{mfcc}(:,i),\ \mathtt{mfcc}(:,j)\rangle}{\|\mathtt{mfcc}(:,i)\| \|\ \mathtt{mfcc}(:,j)\|} = \sum_{k=1}^{\text{nbanks}} \frac{\mathtt{mfcc}(k,i)\ \mathtt{mfcc}(k,j)}{\|\mathtt{mfcc}(:,i)\| \|\ \mathtt{mfcc}(:,j)\|}.$$

- What is $S(i, i)$?
- Is similarity matrix symmetric $S(i, j)$=$S(j, i)$?

Example of a Similarity Matrix

$S(i,j)$ for $i,j = 1, \ldots, 600$



Rhythm

- The presence of **repetitive patterns** in the temporal structure of music.

- We will compute a vector B such that $B(l)$ quantifies the presence of similar spectral patterns for **frames** that are l frames apart.

- The lag associated with the **largest** entry in the array B is a good candidate for the period in the rhythmic structure.

A First Estimate of the Rhythm

$$B(l) = \frac{1}{N_f - l} \sum_{n=1}^{N_f - l} S(n, n+l), \quad l = 0, \ldots, N_f - 1$$

$$\begin{bmatrix} S(1,1) & S(1,2) & S(1,3) & S(1,4) \\ S(2,1) & S(2,2) & S(2,3) & S(2,4) \\ S(3,1) & S(3,2) & S(3,3) & S(3,4) \\ S(4,1) & S(4,2) & S(4,3) & S(4,4) \end{bmatrix}$$

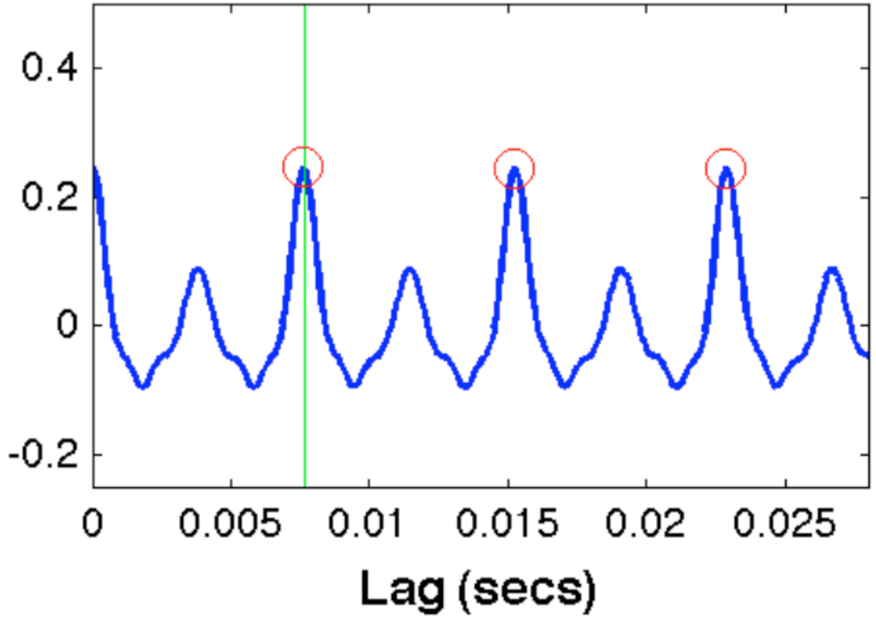
A Better Estimate of the Rhythm

A lag l will be a good candidate for a rhythmic period, if there are many i and j such that if $S(i,j)$ is large then $S(i,j+l)$ is also large.

$$AR(l) = \frac{1}{N_f(N_f - l)} \sum_{i=1}^{N_f} \sum_{j=1}^{N_f - l} S(i,j)S(i,j+l), \quad l = 0, \ldots, N_f - 1.$$

autocorrelation

A Better Estimate of the Rhythm



$l \times \frac{K}{f_s}$ seconds

Figure 2: Rhythm index $AR(l)$as a function of the lag l

Rhythmic Variations Over Time

- Consider short time **windows** formed by 20 frames:

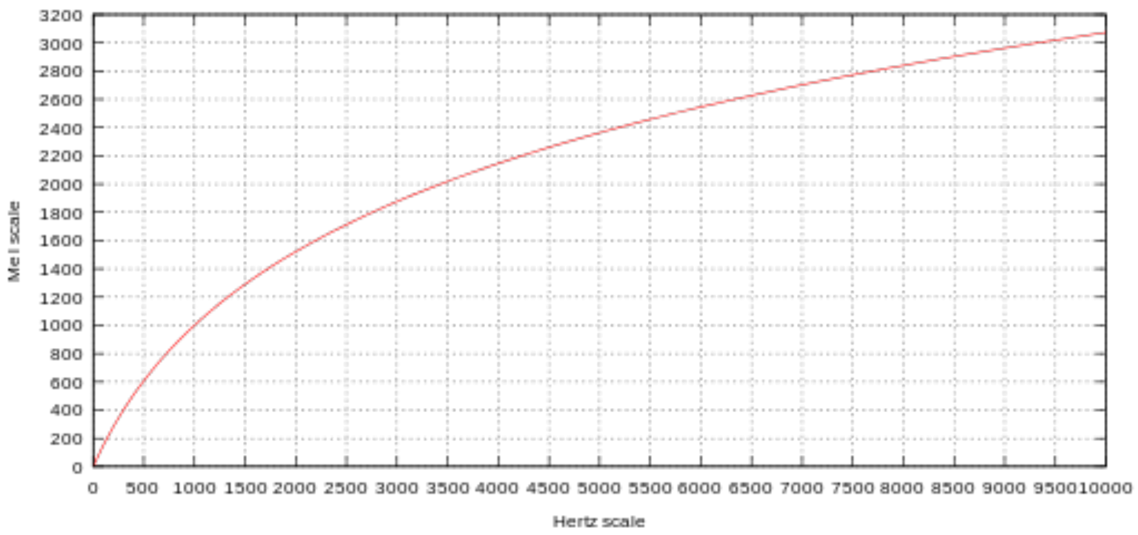
$$AR(l, m) = \frac{1}{20(20-l)} \sum_{i=1}^{20} \sum_{j=1}^{20-l} S(i + m * 20, j + m * 20) S(i + m * 20, j + m * 20 + l),$$

$$\text{for } l = 0, \ldots, 19, \quad \text{and } m = 0, \ldots, N_f/20 - 1.$$

Tonality and Chroma

- The auditory system that is closely related to frequency, also known as pitch, corresponds to a **logarithmic scale** of the frequency.
- Example: Mel scale in Lab 1

$$m = 1127.01048 * \log\left(1 + \omega/700\right)$$



- In this lab, we define another logarithmic scale.

Tempered Scale

- First, we define a reference frequency f_0: Here $f_0 = 27.5Hz$
- The tempered scale introduces 12 frequency intervals between this frequency and the next frequency $2f_0$.
- These frequency intervals known as semitones (sm):

$$f_{sm} = f_0 2^{sm/12}$$

sm semitones away from f_0

Tempered Scale

- Octave: An interval of 12 semitones corresponds to an "octave"

$$2f_0 = 2^{1}.f_0$$

- Another example: two frequencies 440Hz and 27.5Hz

$$440 = 2^{4}.27.5$$

are separated by 4 octaves.

Recall:

$$f_{sm} = f_0 2^{sm/12}$$

Chroma

- The chroma is associated with the relative position of a note inside an octave.

- Therefore, given a frequency f and a reference frequency f_0, we are interested to find chroma c within an octave:

$$\mathrm{sm} = \mathrm{round}(12 \log_2(f/f_0)).$$

$$c = \mathrm{sm} \mod (12)$$

$$\boxed{f_{\mathrm{sm}} = f_0 2^{\mathrm{sm}/12}}$$

Modulo Operation

- Remainder after division

$$c = sm \mod (12)$$

$$sm = 12q + c, \quad \text{with} \quad 0 \leqslant c \leqslant 11, \quad \text{and} \quad q \in \mathbb{N}$$

```
>> mod(13,12)
        ans =    1
```

Problem?

- We don’t know the frequencies of the notes that are being played at a given time.
- Instead, we use the Fourier Transform to find a distribution of the spectral energy.
- We need to find the main peaks in the Fourier Transform

Step 1

- For each frame, find the windowed Fourier Transform:

$$
\begin{aligned}
Y &= \mathrm{FFT}(w. * x_n); \\
K &= N/2 + 1; \\
X_n &= Y(1:K);
\end{aligned}
$$

- We detect the local maxima of the magnitude of the Fourier transform:

$$f_k, k = 1, \ldots, n_{\text{peaks}}$$

Step 2

- Assignment of the peak frequencies to semitones:

$$\mathrm{sm} = \mathrm{round}(12\log_2(f_k/f_0))$$

one of the peak frequencies

$$c = \mathrm{sm} \mod (12)$$

Each peak frequency is assigned to a single semitone sm

Pitch Class Profile (PCP)

- A different approach: For each chroma c=0,…,11, find the weighted sum of all the peak frequencies that are mapped to the note c.

$$\mathrm{PCP}(c) = \sum_{k} w(k,c)|X_n(k)|^2, \quad \text{where } k \text{ is such that } c = \mathrm{round}(12\log_2(f_k/f_0))\bmod 12$$

the spectral energy

Given that the peak frequency is mapped to the note c

Pitch Class Profile (PCP)

- A different approach: For each chroma c=0,…,11, find the weighted sum of all the peak frequencies that are mapped to the note c.

$$\mathrm{PCP}(c) = \sum_{k} \underbrace{w(k,c)|X_n(k)|^2}_{\text{the spectral energy}}, \quad \text{where } k \text{ is such that } c = \mathrm{round}(12\log_2(f_k/f_0)) \bmod 12$$

Given that the peak frequency is mapped to the note c

How much energy is associated with frequencies that are very "close" to the note c across all octaves

Pitch Class Profile (PCP)

$$w(k,c) = \begin{cases} \cos^2(\pi\, r/2) & \text{if } -1 < r = 12\log_2(f_k/f_0) - sm < 1, \\ & \text{where } c = sm \mod (12), \text{ and } sm = \text{round}(12\log_2(f_k/f_0)) \\ 0 & \text{otherwise.} \end{cases}$$

Normalized Pitch Class Profile (NPCP)

We want our definition of chroma to be independent of the loudness of the music

$$\mathrm{NPCP}(\mathbf{c}) = \frac{\mathrm{PCP}(\mathbf{c})}{\sum_{\mathrm{q}=0}^{11} \mathrm{PCP}(\mathbf{q})}$$