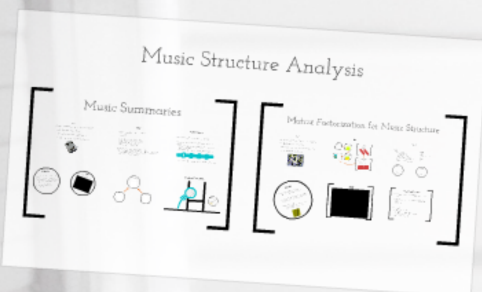
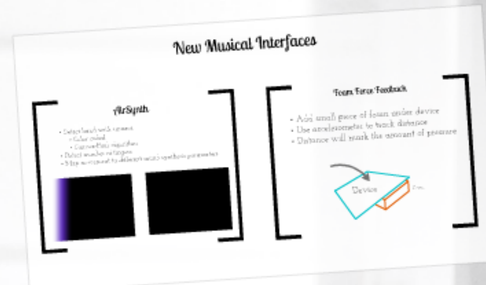
Music Structure Analysis and New Musical Interfaces



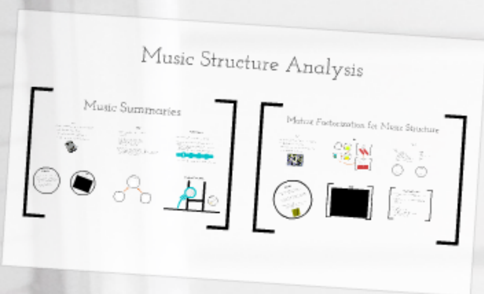


Oriol Nieto
Music and Audio Research Lab
New York University
Jan 10th 2013

MARL
NYU Music and Audio Research Laboratory

the echonest

Music Structure Analysis and New Musical Interfaces



Oriol Nieto
Music and Audio Research Lab
New York University
Jan 10th 2013

MARL
NYU Music and Audio Research Laboratory

the echonest

Music Structure Analysis

Music Summaries

Matrix Factorization for Music Structure

Goal

- To generate audible summaries of music recordings that
    - optimally explain a given track
    - don't contain (much) repeated information
    - compress the entire music piece into a few seconds

Motivation

- To facilitate the navigation of massive digital music collections
- To provide an alternative to audio thumbnailing
  - all the parts of a piece are included in the summary
  - a potential buyer could get a better glimpse of the piece
- To normalize an entire music collection
  - more efficient for MIR techniques

Example

Chopin Mazurka Op. 30 No. 2
Original: A-B-C-B
Summary: A-B-C

How?

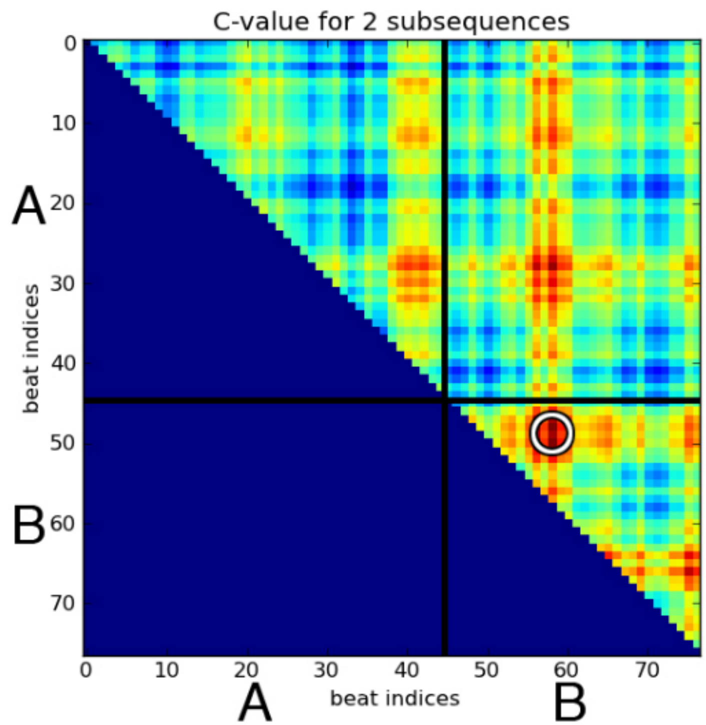
- Extract beat-synchronous harmonic features:
  - Chromas
  - Tonnetz

- Quantize the feature space using k-means
  - reduce space into a few discrete values
  - computationally advantageous
  - loose precision if k is too small

- Exhaustively search the quantized feature space for
  - Best P subsequences of N beats that can compress the entire signal
  - P subsequences of N beats that have the highest amount of disjoint information between them

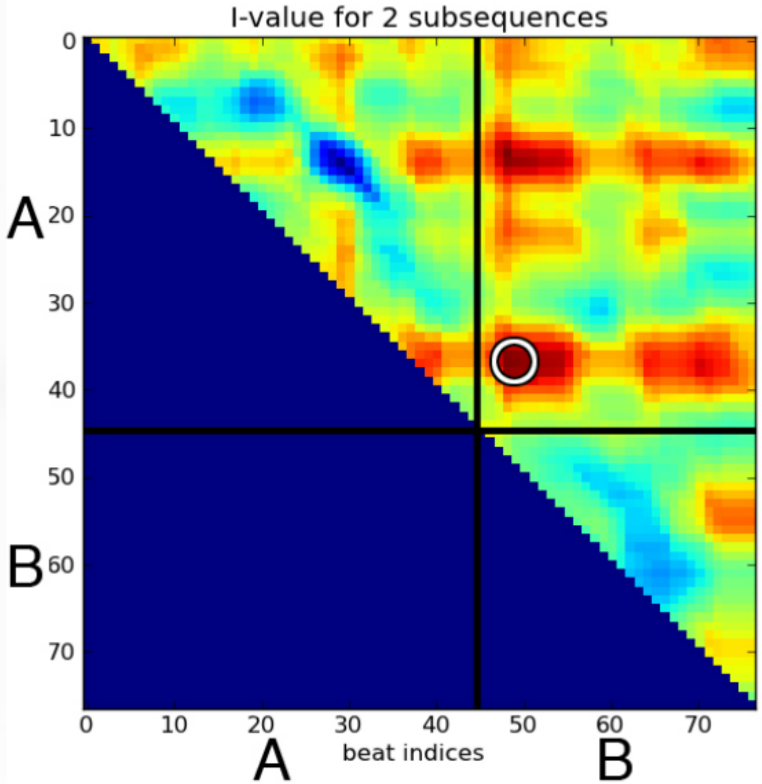
Compression

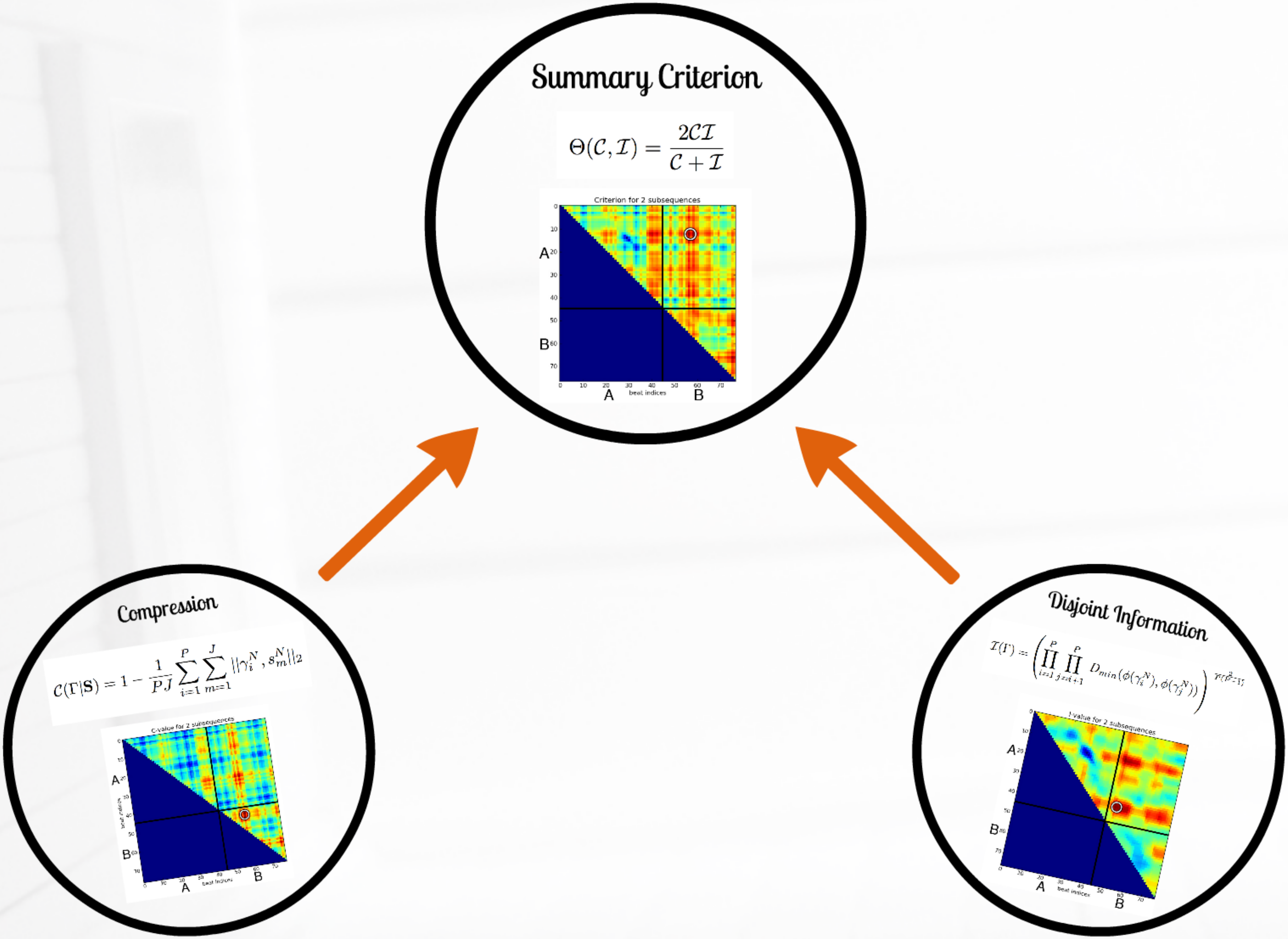
$$\mathcal{C}(\Gamma|\mathbf{S}) = 1 - \frac{1}{PJ}\sum_{i=1}^{P}\sum_{m=1}^{J}||\gamma_i^N, s_m^N||_2$$



Disjoint Information

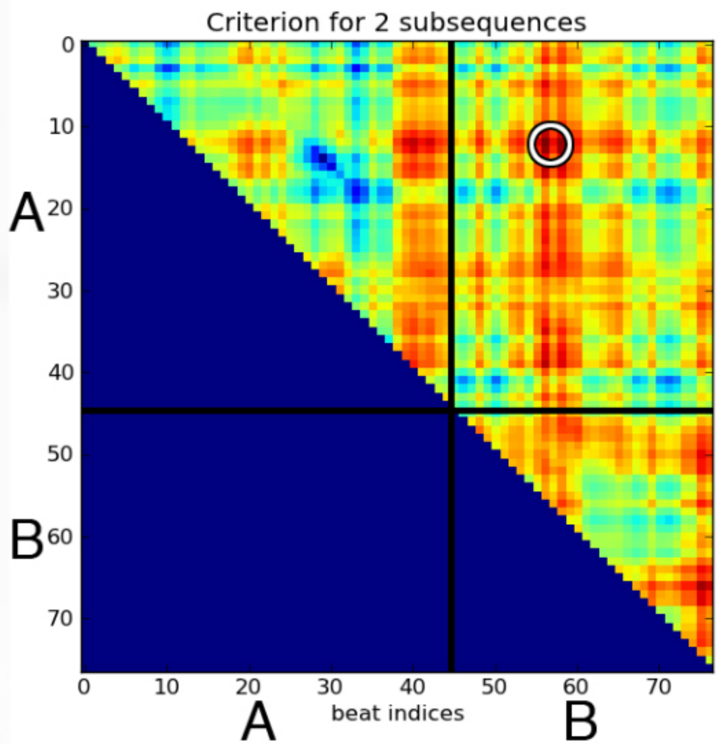
$$\mathcal{I}(\Gamma) = \left(\prod_{i=1}^{P} \prod_{j=i+1}^{P} D_{min}(\phi(\gamma_i^N), \phi(\gamma_j^N)) \right)^{\frac{2}{P(P-1)}}$$




Summary Criterion
$\Theta(\mathcal{C},\mathcal{I}) = \frac{2\mathcal{C}\mathcal{I}}{\mathcal{C}+\mathcal{I}}$
Criterion for 2 subsequences
beat indices
A
B
Compression
$\mathcal{C}(\Gamma|\mathbf{S}) = 1 - \frac{1}{PJ}\sum_{i=1}^{P}\sum_{m=1}^{J} ||\gamma_i^N, s_m^N||_2$
C-value for 2 subsequences
Disjoint Information
I-value for 2 subsequences

Summary Criterion

$$\Theta(\mathcal{C}, \mathcal{I}) = \frac{2\mathcal{C}\mathcal{I}}{\mathcal{C} + \mathcal{I}}$$



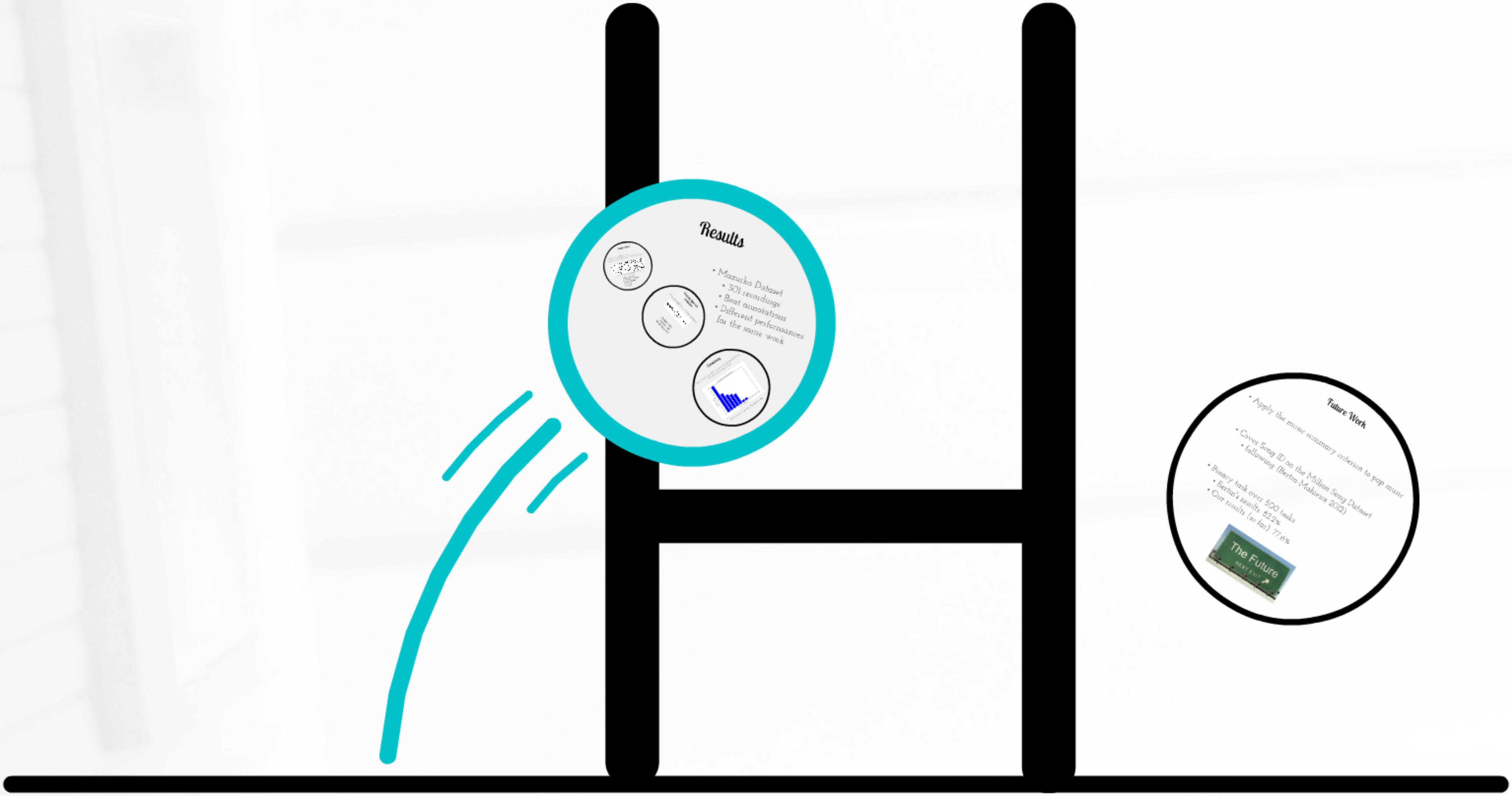
Heuristic Approach

- Brute force method is computationally expensive

- We can approximate the same results with a heuristic approach
  - Initialize P subsequences equally spaced in time across the music recording
  - Find maximum criterion one subsequence at a time

- The problem becomes linear with respect to P instead of exponential

Results and Future Work



Results

- Mazurka Dataset
  - 301 recordings
  - Beat annotations
  - Different performances for the same work

Feature Space

k	GT-C	GT-T	A-C	A-T
50	3.64	3.97	2.71	3.80
100	3.83	4.29	2.68	4.20
200	4.09	4.74	2.87	4.45

Fisher's ratio between recordings from the same work and the rest

k: Dictionary size
GT: Ground Truth Beats
A: Automatic Beats
C: Chromas
T: Tonnetz

Heuristic Approach Evaluation

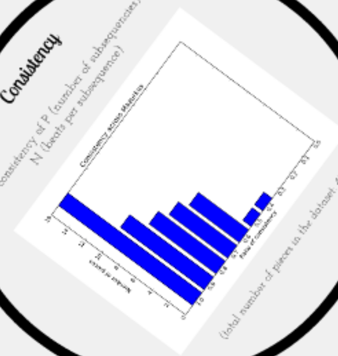
Mean Squared Error between strategies

$$MSE(\Theta) = \frac{1}{S}\sum_{i}^{S}(1 - \Theta_i)^2$$

Random: 21%
Heuristic: 1%
(Brute Force: 0%)

Consistency

Check consistency of P (number of subsequencies) and N (beats per subsequence)



(total number of pieces in the dataset: 49)

Feature Space

- Is automatic beat tracking sufficient?
- Do chroma and Tonnetz features perform equivalently?
- Does performance vary significantly as a function of codebook size?

k	GT-C	GT-T	A-C	A-T
50	3.64	3.97	2.71	3.89
100	3.84	4.29	2.68	4.20
200	4.09	**4.74**	2.87	**4.45**

Fisher's ratio between recordings from the same work and the rest

k: Dictionary size
GT: Ground Truth Beats
A: Automatic Beats
C: Chromas
T: Tonnetz

Heuristic Approach Evaluation

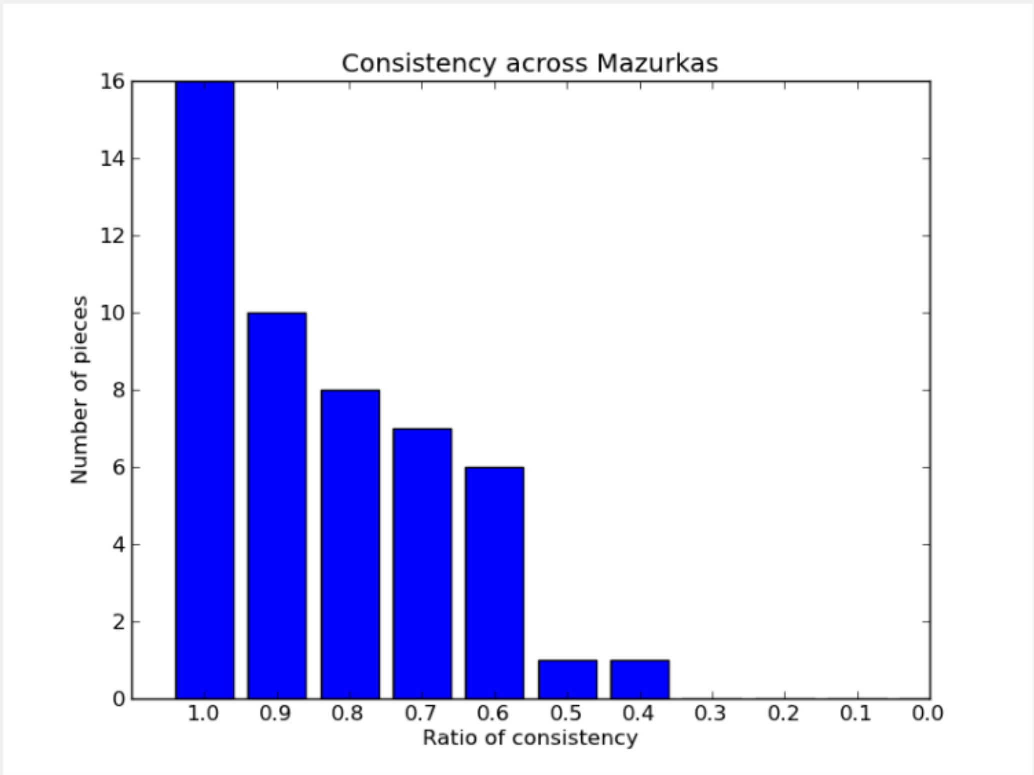
Mean Squared Error between strategies

$$\mathrm{MSE}(\boldsymbol{\Theta}) = \frac{1}{\mathcal{S}} \sum_{i}^{\mathcal{S}} (1 - \boldsymbol{\Theta}_i)^2$$

Random: 21%
Heuristic: 1%
(Brute Force: 0%)

Consistency

Check consistency of P (number of subsequencies) and N (beats per subsequence)



(total number of pieces in the dataset: 49)

Future Work

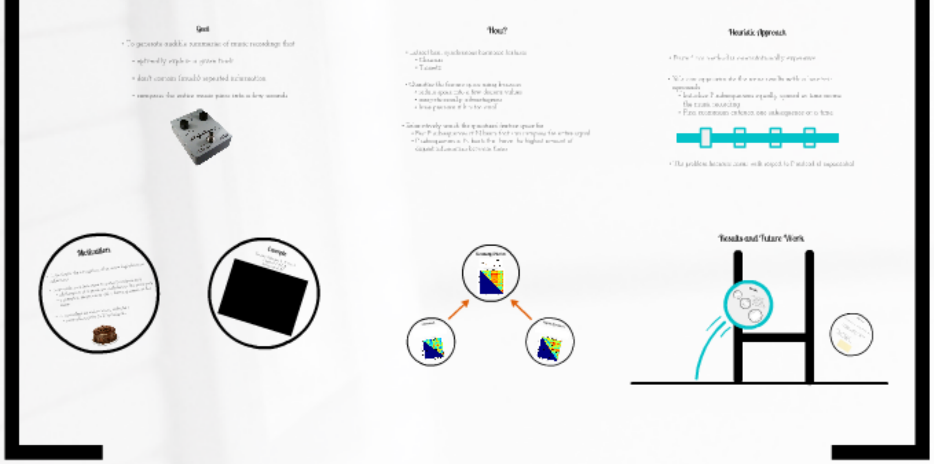
- Apply the music summary criterion to pop music

- Cover Song ID on the Million Song Dataset
    - following (Bertin-Mahieux 2012)

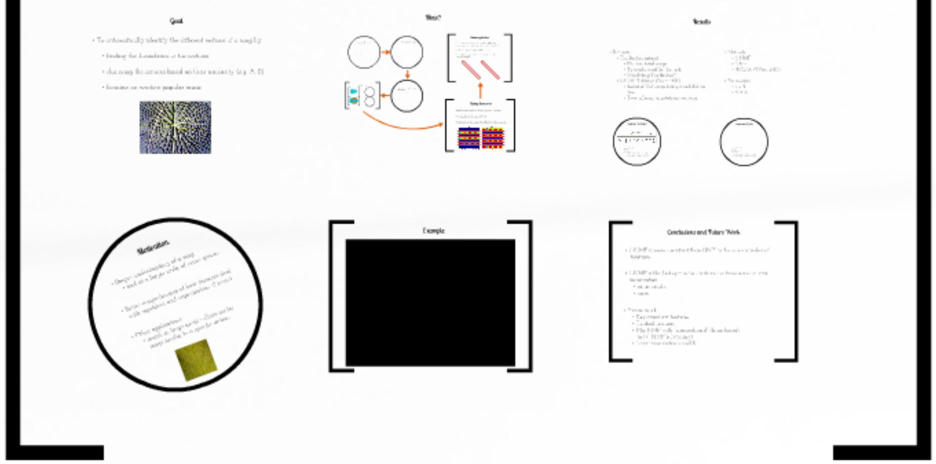
- Binary task over 500 tasks
    - Bertin's results: 82.2%
    - Our results (so far): 77.6%

Music Structure Analysis

Music Summaries



Matrix Factorization for Music Structure

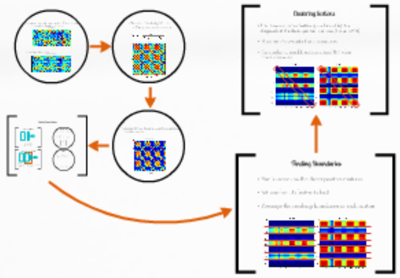


Matrix Factorization for Music Structure

Goal

- To automatically identify the different sections of a song by
  - finding the boundaries of the sections
  - clustering the sections based on their similarity (e.g. A, B)
  - focusing on western popular music

How?



Results

- Datasets:
  - The Beatles dataset
    - 176 annotated songs
    - Typically used for this task
    - Overfitting The Beatles?
  - SALAMI dataset (Smith 2011)
    - Subset of 253 songs freely available on-line
    - Two different annotations per song

- Methods:
  - C-NMF
  - NMF
  - SI-PLCA (Weiss 2010)

- Parameters:
  - r = 2
  - K = 4

Motivation

- Deeper understanding of a song
  - and at a larger scale, of music genres
- Better comprehension of how humans deal with repetition and organization of sound
- Other applications:
  - search in large music collections for songs similar to a specific section

Example

Conclusions and Future Work

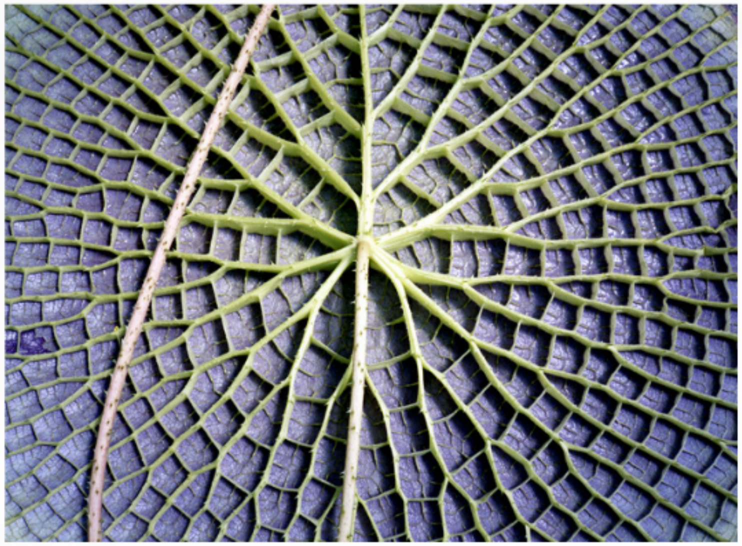
- C-NMF is more consistent than NMF in the same number of iterations

- C-NMF is the best option for clustering sections using matrix factorization
  - better results
  - faster

- Future work:
  - Key-invariant features
  - Timbral features
  - Mix NMF with "checkerboard" (boundaries) and C-NMF (clustering)
  - Learn parameters r and K

Goal

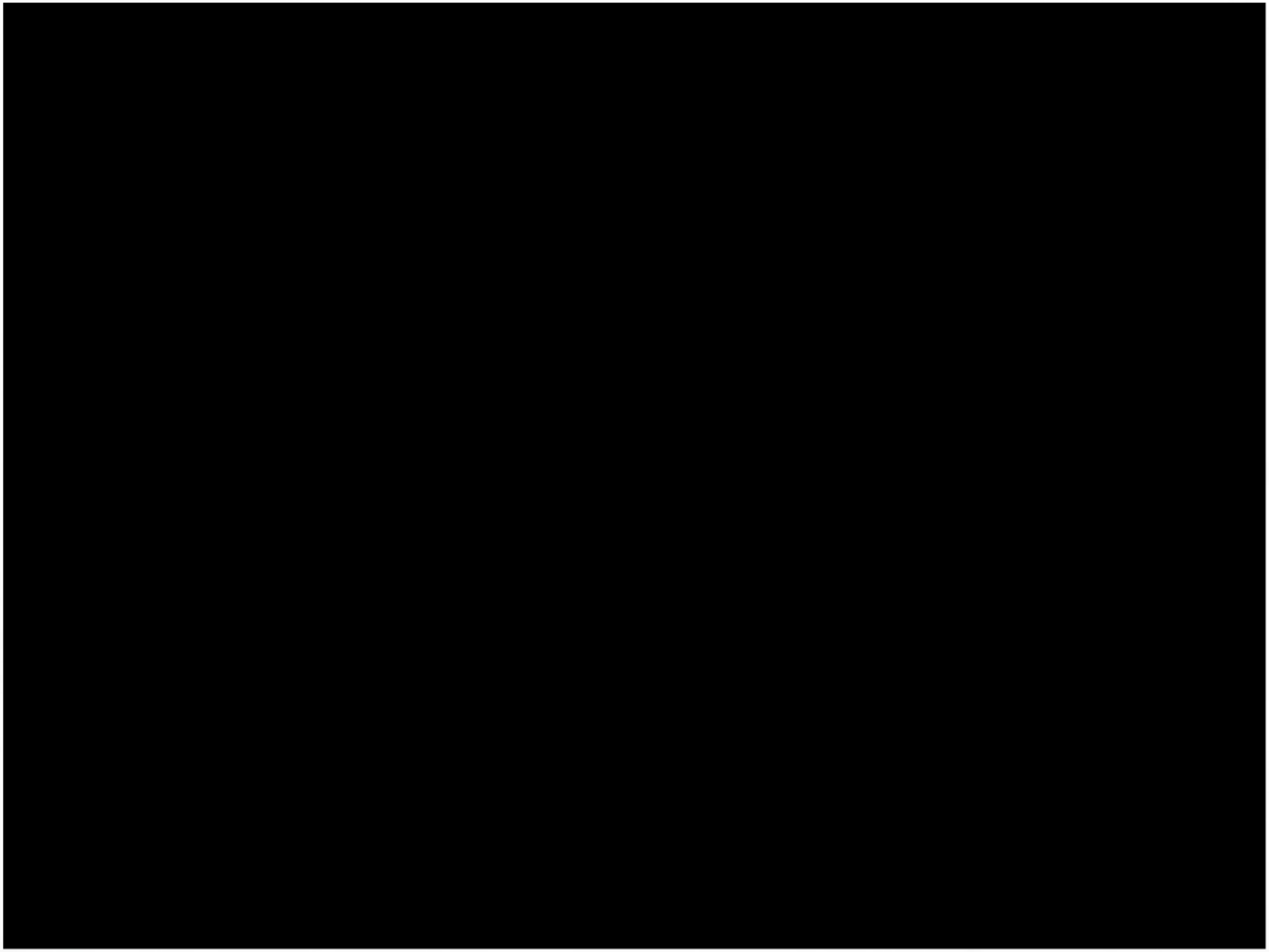
- To automatically identify the different sections of a song by
  - finding the boundaries of the sections
  - clustering the sections based on their similarity (e.g. A, B)
  - focusing on western popular music



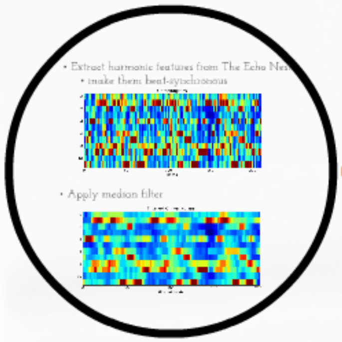
Motivation

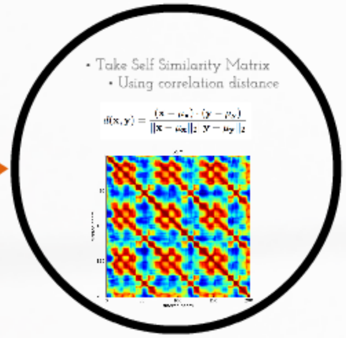
- Deeper understanding of a song
  - and at a larger scale, of music genres
- Better comprehension of how humans deal with repetition and organization of sound
- Other applications:
  - search in large music collections for songs similar to a specific section

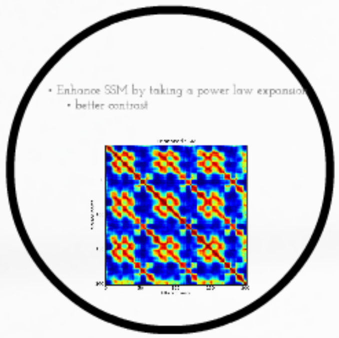
Example

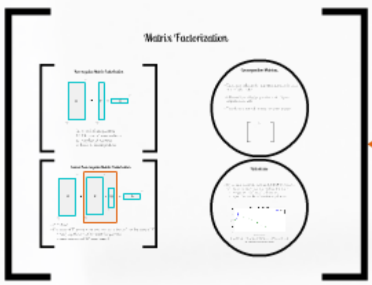


How?



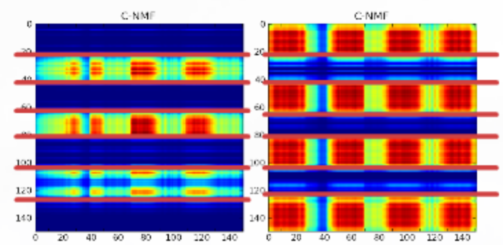






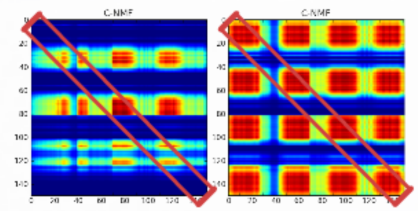
Finding Boundaries

- Run k-means on the decomposition matrices
- Set number of clusters to k=2
- Average the resulting boundaries of each matrix

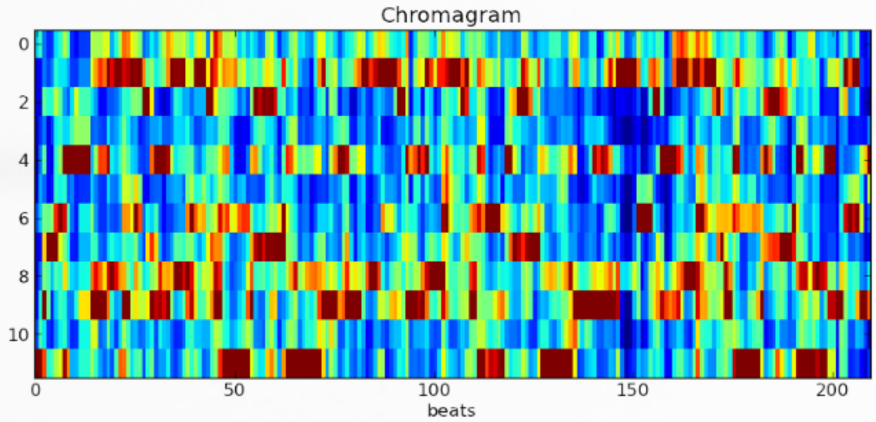


Clustering Sections

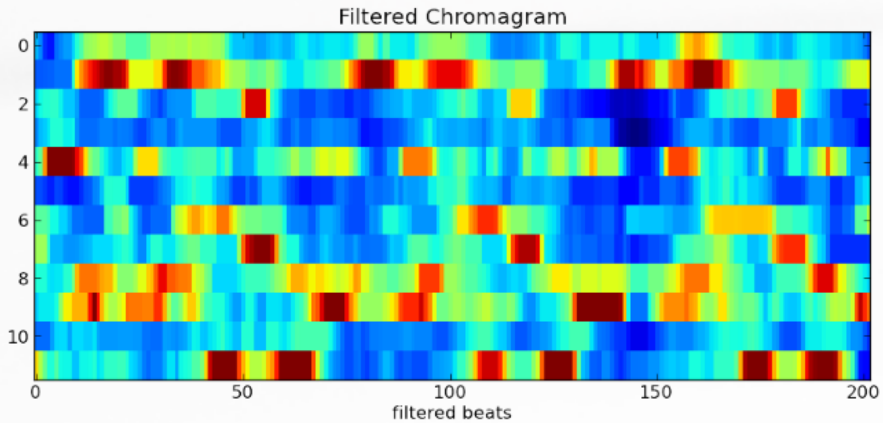
- Run k-means on the feature space formed by the diagonals of the decomposition matrices (Kaiser 2010)
- Make use of previously found boundaries
- The number of possible sections is fixed (K) when running k-means



- Extract harmonic features from The Echo Nest
  - make them beat-synchronous

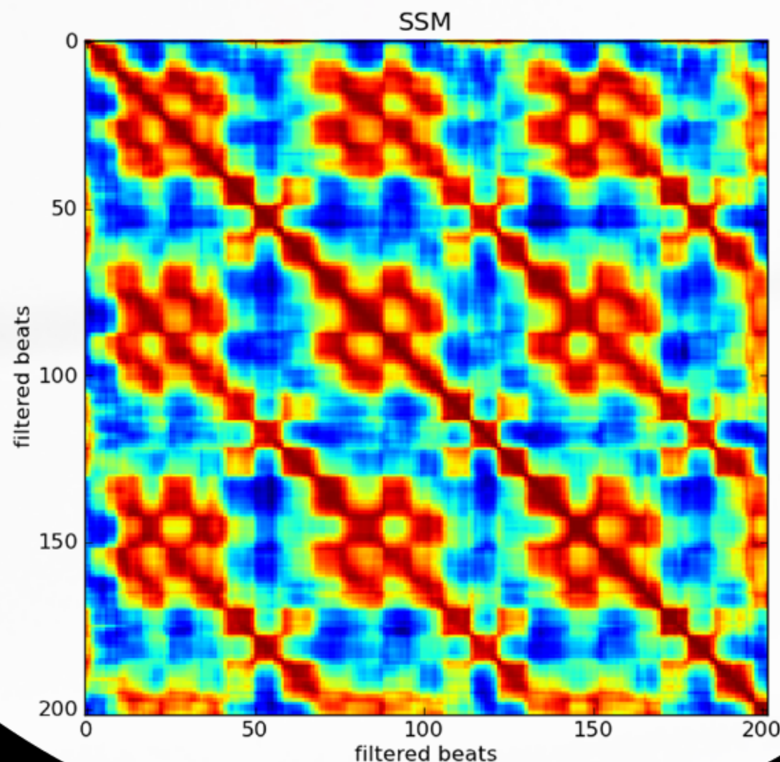


- Apply median filter

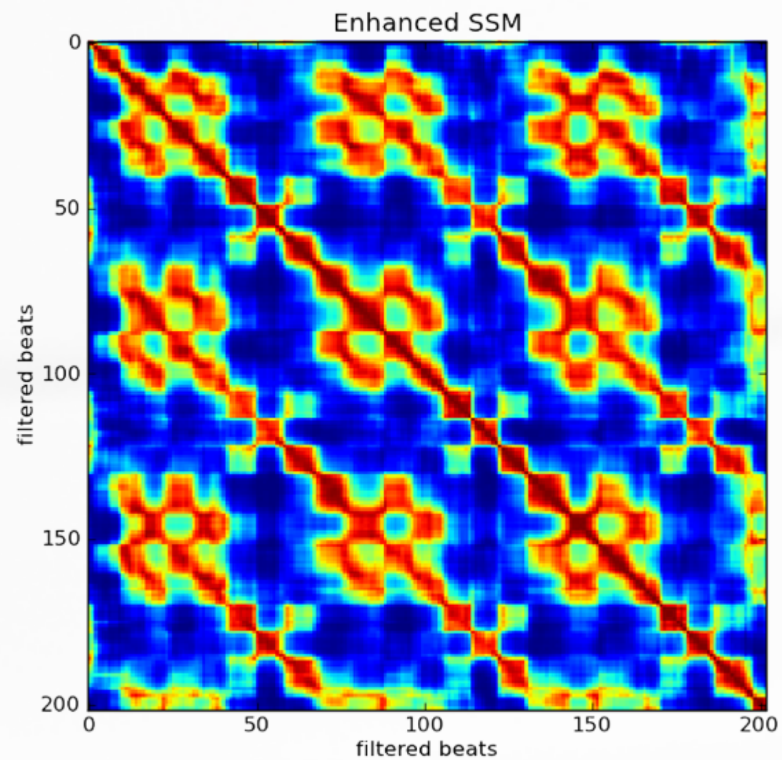


- Take Self Similarity Matrix
    - Using correlation distance

$$d(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x} - \mu_{\mathbf{x}}) \cdot (\mathbf{y} - \mu_{\mathbf{y}})}{||\mathbf{x} - \mu_{\mathbf{x}}||_2 ||\mathbf{y} - \mu_{\mathbf{y}}||_2}$$

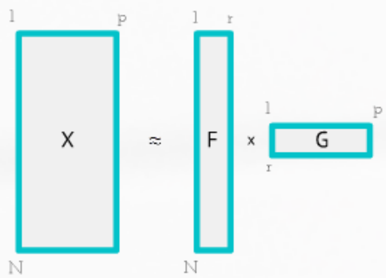


- Enhance SSM by taking a power law expansion
    - better contrast



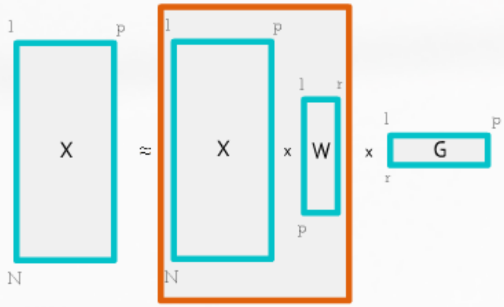
Matrix Factorization

Non-negative Matrix Factorization



X, F, and G are positive
N: Number of observations
p: Number of features
r: Rank of decomposition

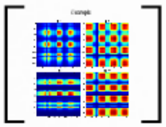
Convex Non-negative Matrix Factorization



- F = XW
- Columns of F become convex combinations of the features of X
  - Each coefficient of W must be positive
  - Each column of W must sum 1

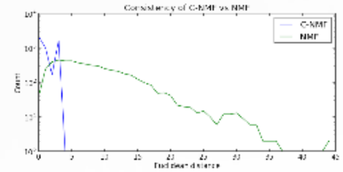
Decomposition Matrices

- There are r different decomposition matrices for each matrix factorization
- Obtained by multiplying a column of F by its respective row of G
- They have a key role in music structure analysis



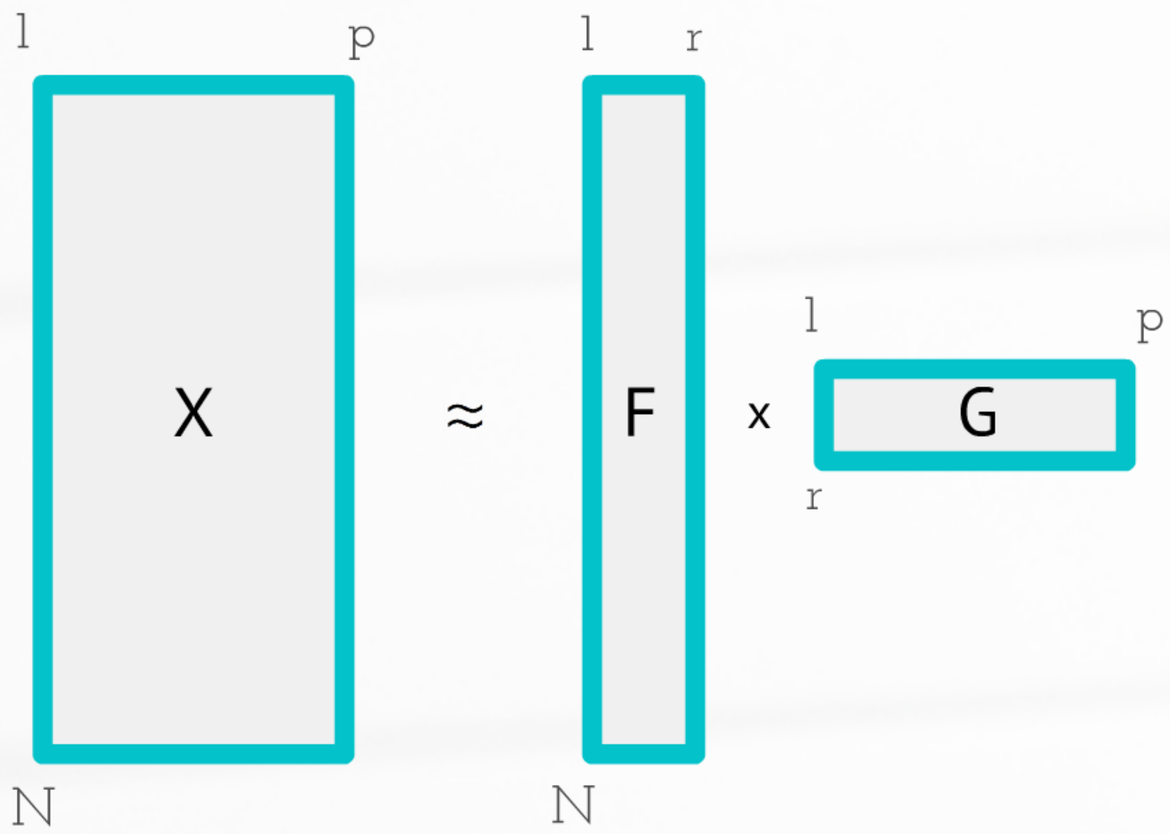
Robustness

- By adding a convex constrain to NMF we tend to find more consistent solutions in less iterations
  - Less prone to fall into local minima
  - Lower the number of iterations -> Faster



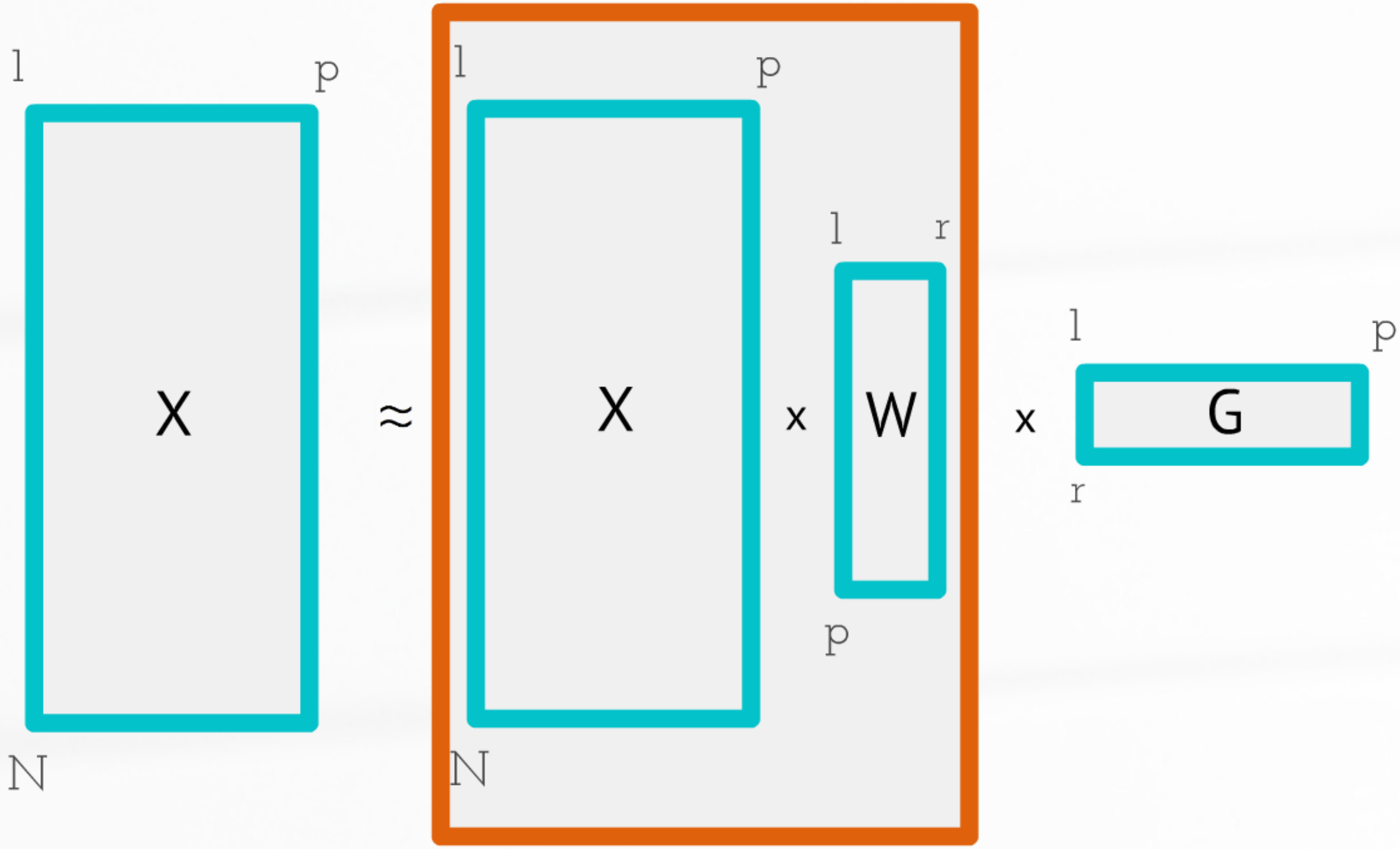
Logarithmic histogram of distances between 100 sets of decomposition matrices

Non-negative Matrix Factorization



X, F, and G are positive
N: Number of observations
p: Number of features
r: Rank of decomposition

Convex Non-negative Matrix Factorization

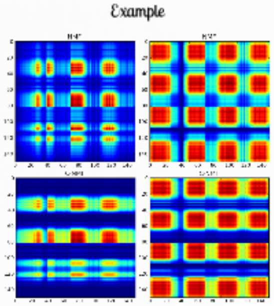


- F = XW
- Columns of F become convex combinations of the features of X
  - Each coefficient of W must be positive
  - Each column of W must sum 1

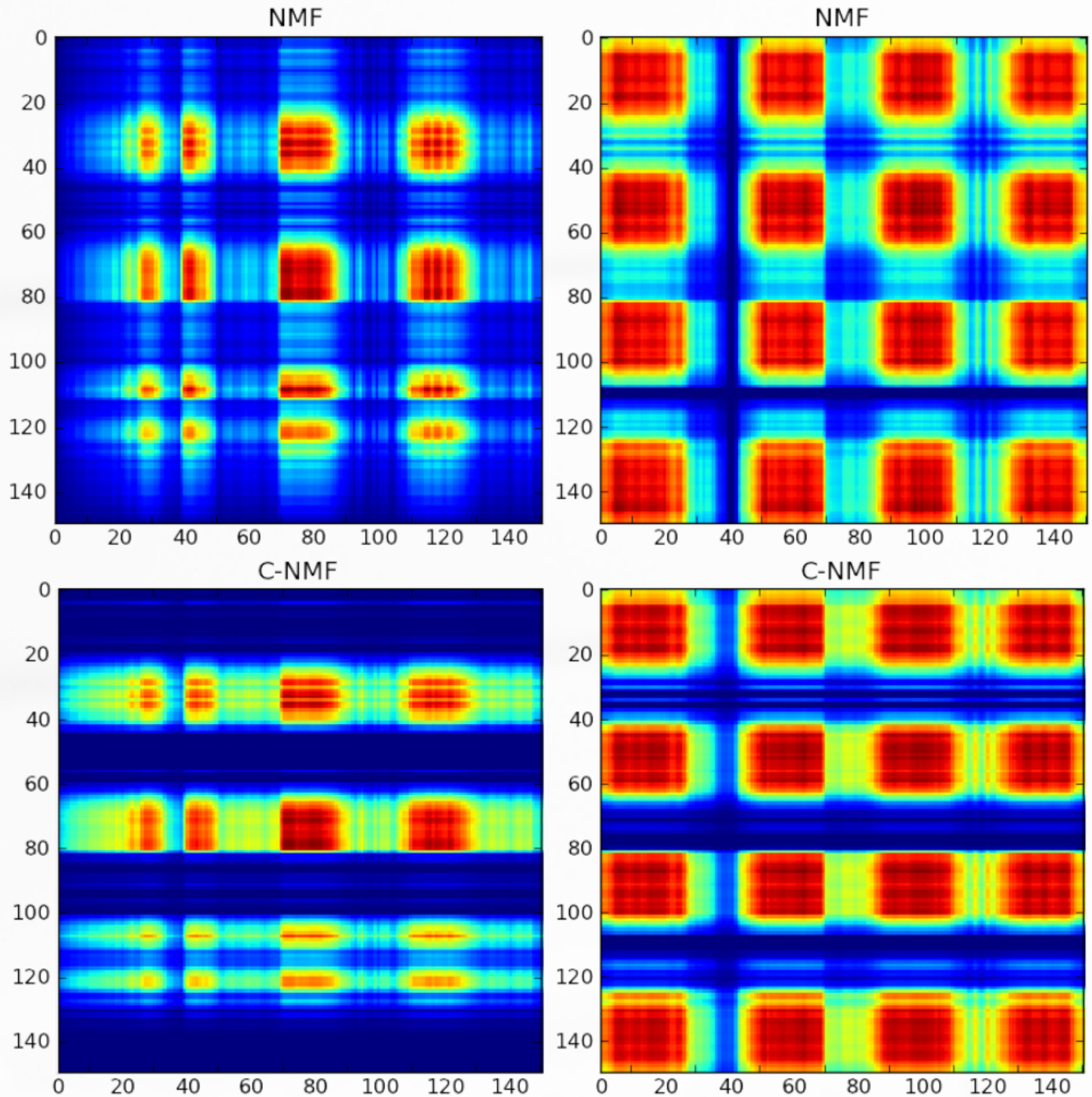
Decomposition Matrices

- There are r different decomposition matrices for each matrix factorization
- Obtained by multiplying a column of F by its respective row of G
- They have a key role in music structure analysis

Example

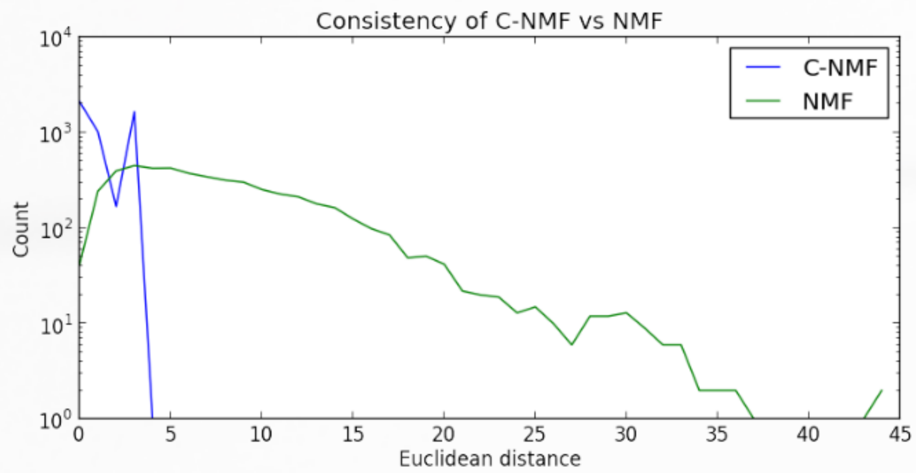


Example



Robustness

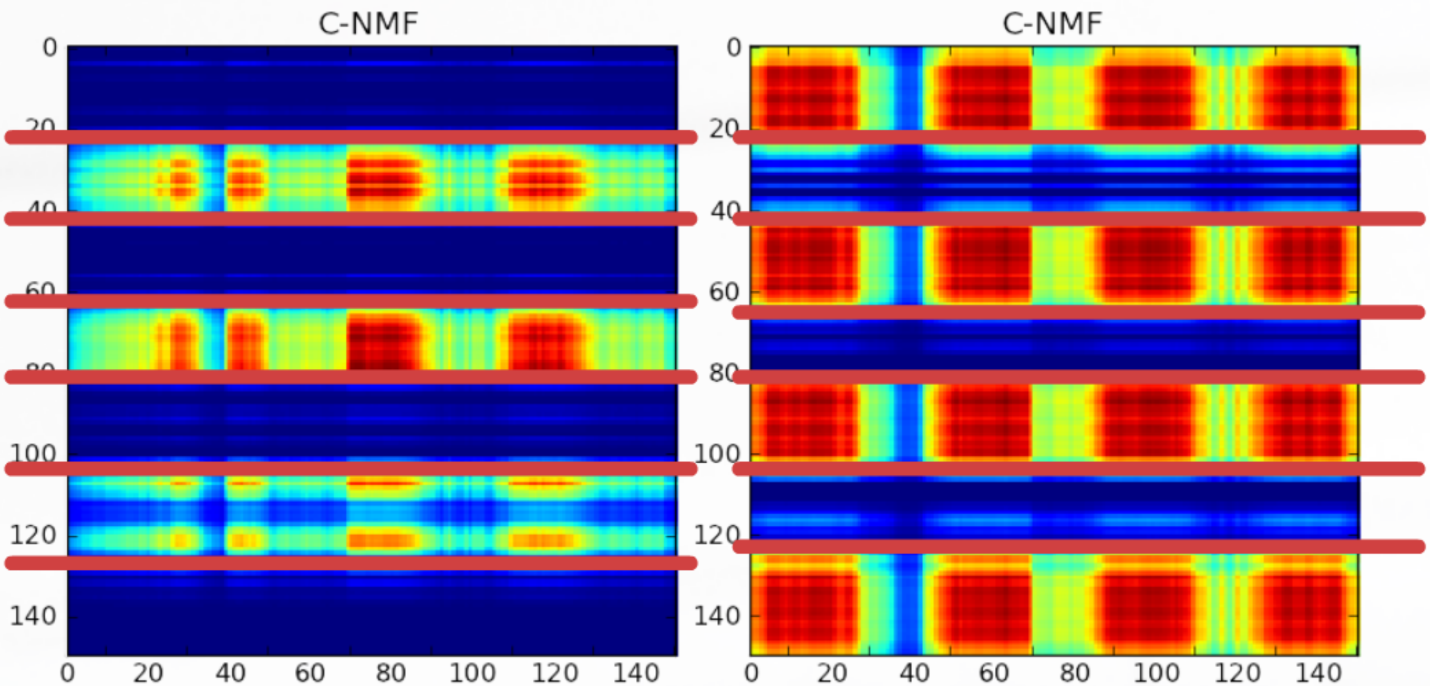
- By adding a convex constrain to NMF we tend to find more consistent solutions in less iterations
  - Less prone to fall into local minima
  - Lower the number of iterations -> Faster



Logarithmic histogram of distances between 100 sets of decomposition matrices

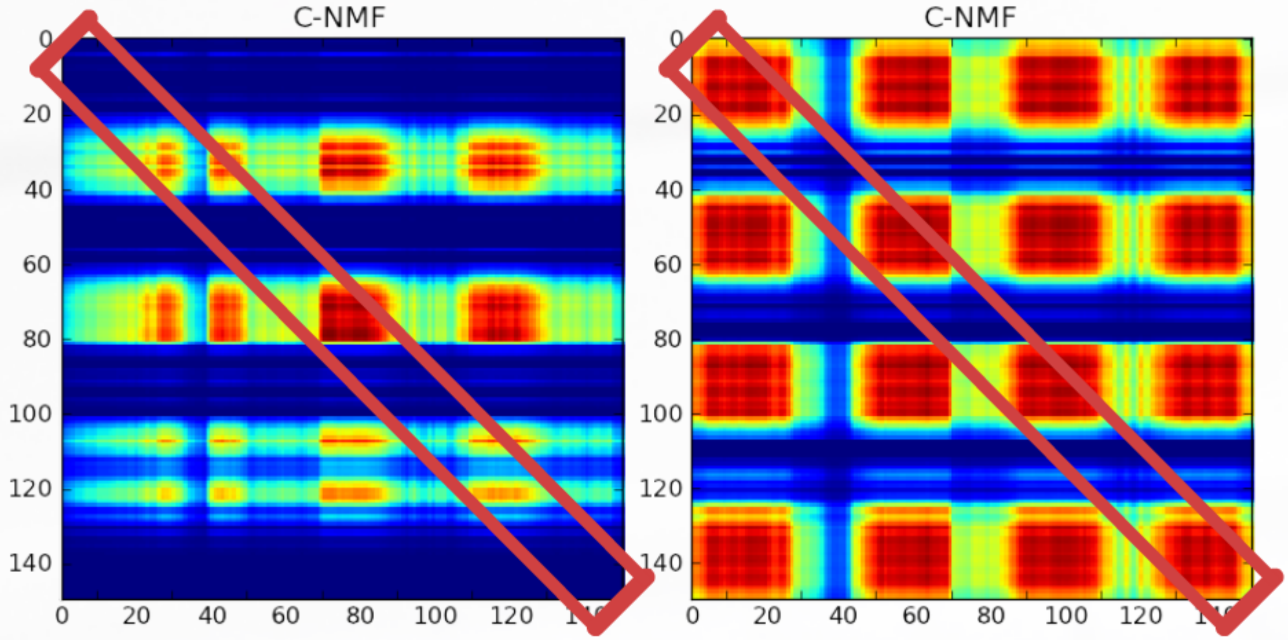
Finding Boundaries

- Run k-means on the decomposition matrices
- Set number of clusters to k=2
- Average the resulting boundaries of each matrix



Clustering Sections

- Run k-means on the feature space formed by the diagonals of the decomposition matrices (Kaiser 2010)
- Make use of previously found boundaries
- The number of possible sections is fixed (K) when running k-means



Results

- Datasets:
  - The Beatles dataset
    - 176 annotated songs
    - Typically used for this task
    - Overfitting The Beatles?
  - SALAMI dataset (Smith 2011)
    - Subset of 253 songs freely available on-line
    - Two different annotations per song

- Methods:
  - C-NMF
  - NMF
  - SI-PLCA (Weiss 2010)

- Parameters:
  - r = 2
  - K = 4

Results on The Beatles

TUT Beatles Dataset								
	Clustering					Boundaries		
Method	F	P	R	S_o	S_u	F	P	R
C-NMF	**59.3**	48.9	83.2	49.8	47.8	57.3	54.9	64.6
NMF	56.6	48.8	77.7	43.7	49.6	**58.9**	54.7	67.7
SI-PLCA	55.8	46.3	80.7	41.0	50.6	23.2	50.9	17.2
Kaiser	60.8	61.5	64.6	–	–	50.0	46.5	52.2

F: F-measure
P: Precision
R: Recall
So: Over-Segmentation entropy
Su: Under-Segmentation entropy

Results on SALAMI

SALAMI (Internet Archive) Dataset								
	Clustering					Boundaries		
Method	F	P	R	S_o	S_u	F	P	R
C-NMF	**53.1**	**44.0**	81.0	50.6	**44.3**	45.1	43.0	52.3
NMF	51.5	42.8	77.6	37.9	45.6	**48.8**	44.0	62.7
SI-PLCA	51.3	55.8	52.1	44.2	51.4	24.8	45.1	18.4

F: F-measure
P: Precision
R: Recall
So: Over-Segmentation entropy
Su: Under-Segmentation entropy

Results on The Beatles

TUT Beatles Dataset								
	Clustering					Boundaries		
Method	F	P	R	S_o	S_u	F	P	R
C-NMF	**59.3**	48.9	83.2	49.8	47.8	57.3	54.9	64.6
NMF	56.6	48.8	77.7	43.7	49.6	**58.9**	54.7	67.7
SI-PLCA	55.8	46.3	80.7	41.0	50.6	23.2	50.9	17.2
Kaiser	60.8	61.5	64.6	–	–	50.0	46.5	52.2

F: F-measure

P: Precision

R: Recall

So: Over-Segmentation entropy

Su: Under-Segmentation entropy

Results on SALAMI

SALAMI (Internet Archive) Dataset								
	Clustering					Boundaries		
Method	F	P	R	S_o	S_u	F	P	R
C-NMF	**53.1**	44.0	81.0	50.6	44.3	45.1	43.0	52.3
NMF	51.5	42.8	77.6	37.9	45.6	**48.8**	44.0	62.7
SI-PLCA	51.3	55.8	52.1	44.2	51.4	24.8	45.1	18.4

F: F-measure
P: Precision
R: Recall
So: Over-Segmentation entropy
Su: Under-Segmentation entropy

Conclusions and Future Work

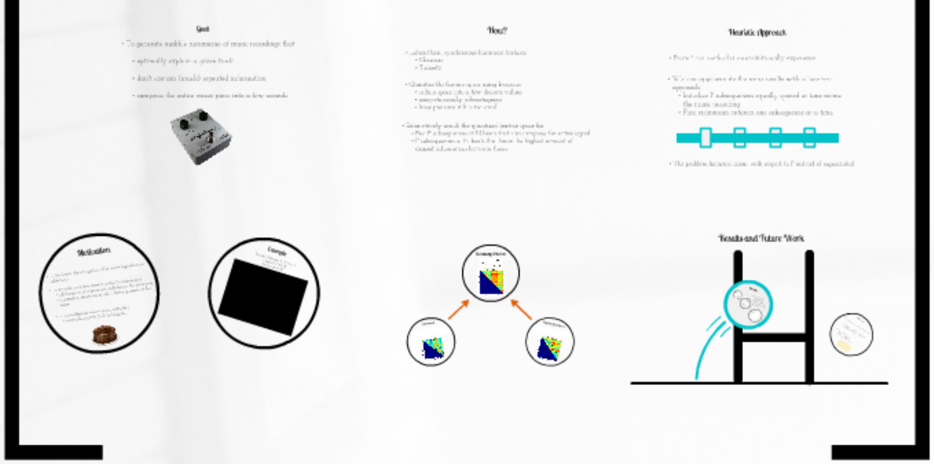
- C-NMF is more consistent than NMF in the same number of iterations

- C-NMF is the best option for clustering sections using matrix factorization
  - better results
  - faster

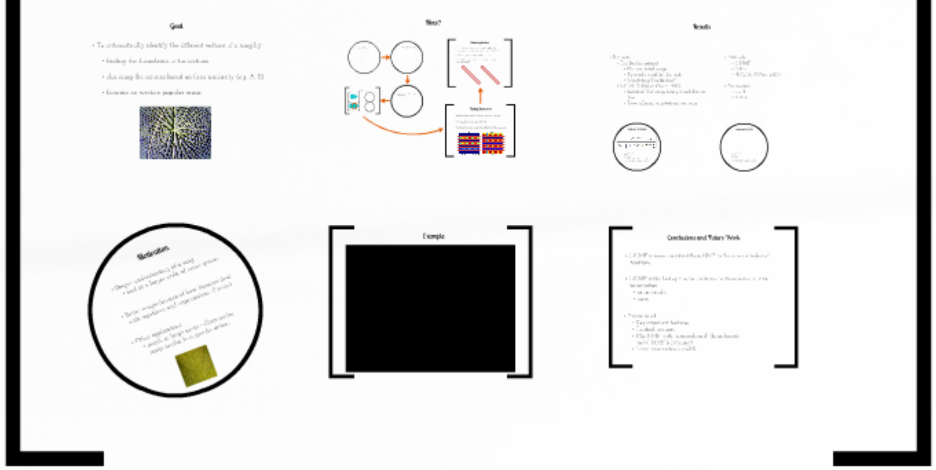
- Future work:
  - Key-invariant features
  - Timbral features
  - Mix NMF with "checkerboard" (boundaries) and C-NMF (clustering)
  - Learn parameters r and K

Music Structure Analysis

Music Summaries



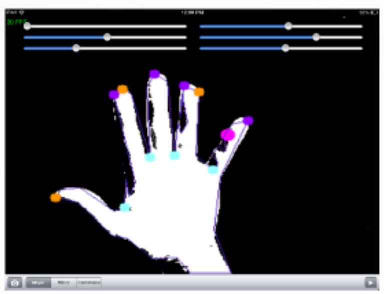
Matrix Factorization for Music Structure

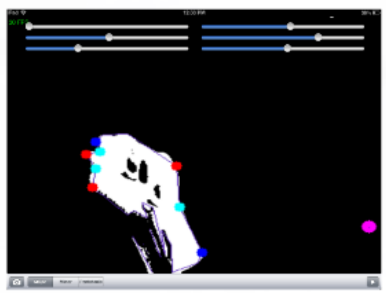


New Musical Interfaces

AirSynth

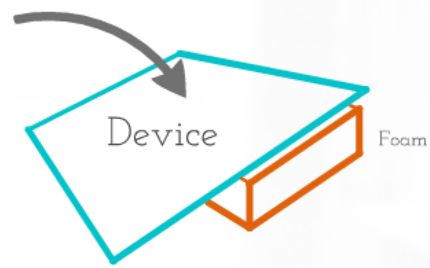
- Detect hand with camera
  - Color coded
  - Convex-Hull algorithm
- Detect number of fingers
- Map movement to different sound synthesis parameters





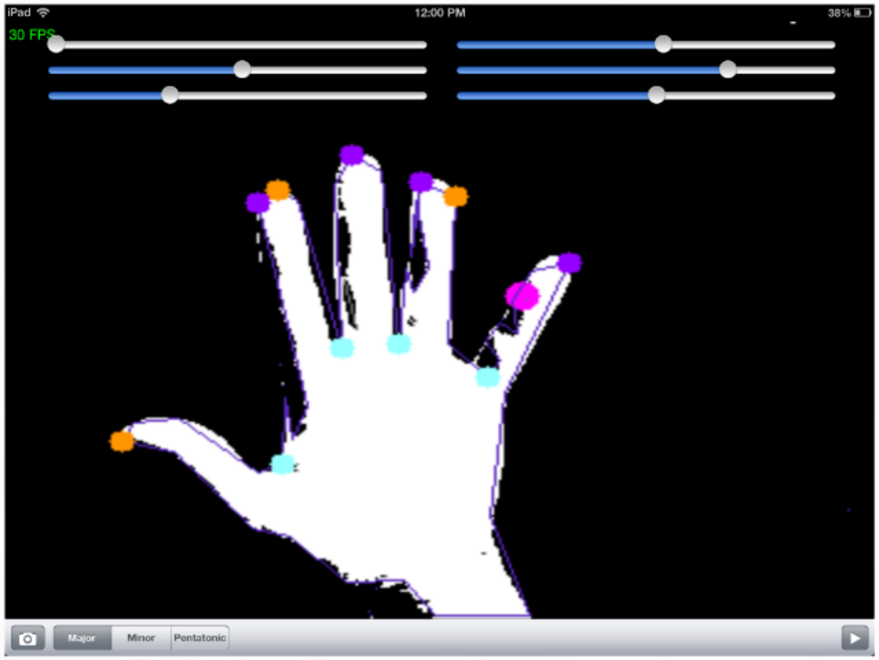
Foam Force Feedback

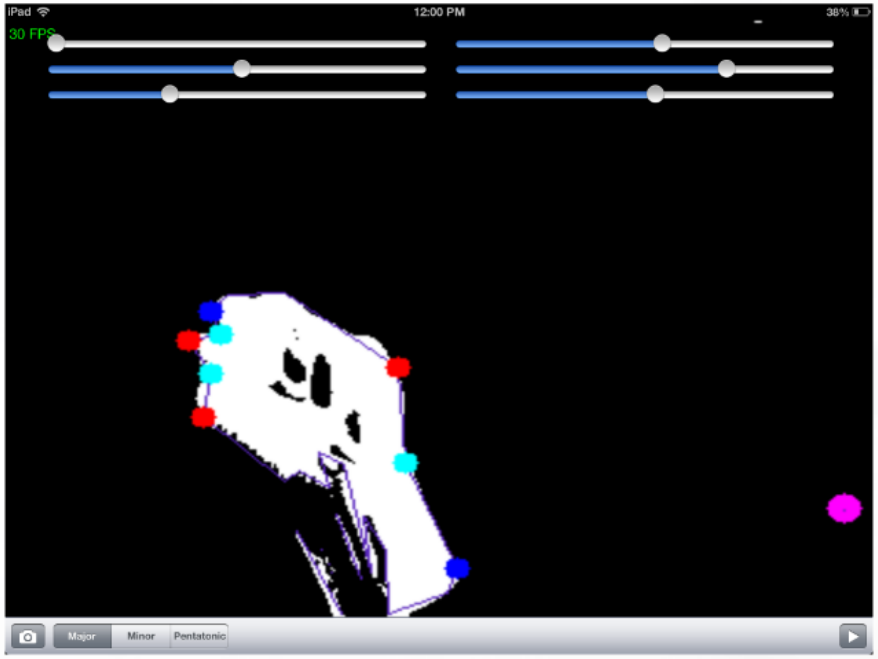
- Add small piece of foam under device
- Use accelerometer to track distance
- Distance will mark the amount of pressure



AirSynth

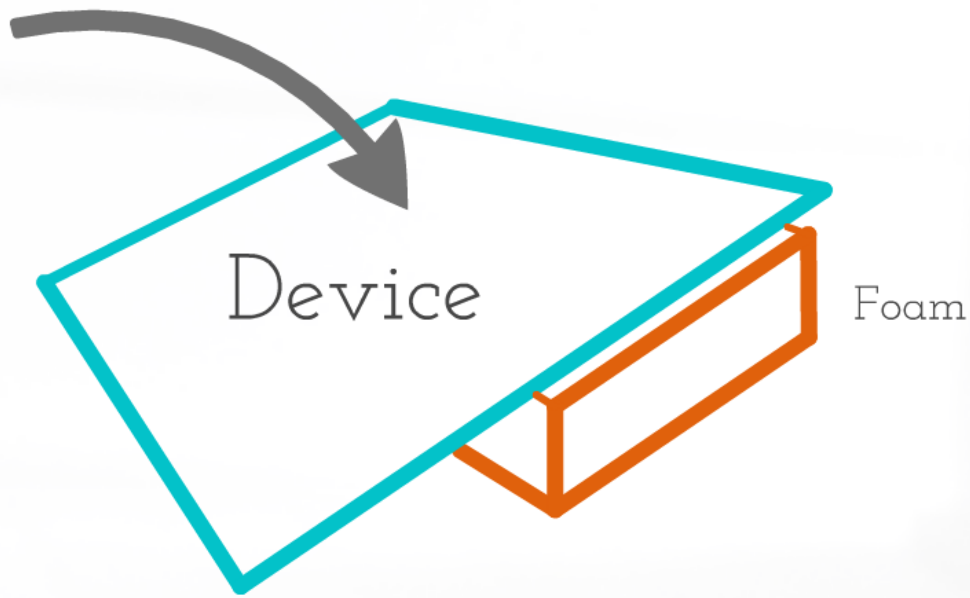
- Detect hand with camera
  - Color coded
  - Convex-Hull algorithm
- Detect number of fingers
- Map movement to different sound synthesis parameters



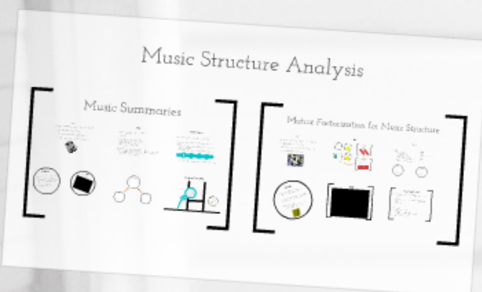


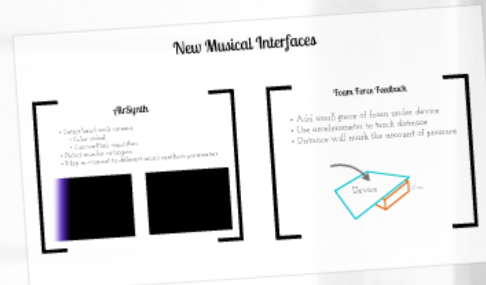
Foam Force Feedback

- Add small piece of foam under device
- Use accelerometer to track distance
- Distance will mark the amount of pressure



Music Structure Analysis and New Musical Interfaces





Oriol Nieto
Music and Audio Research Lab
New York University
Jan 10th 2013

MARL
NYU Music and Audio Research Laboratory

the echonest