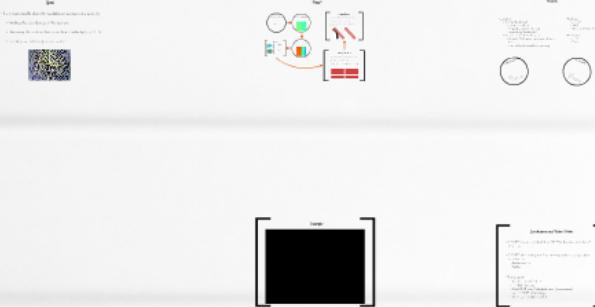


Music Structure Analysis by Matrix Factorization



Oriol Nieto & Tristan Jehan

Music and Audio Research Lab
New York University
Jan 26th 2013

 MARL
NYU Music and Audio Research Laboratory

the  echo nest

Music Structure Analysis by Matrix Factorization



Oriol Nieto & Tristan Jehan

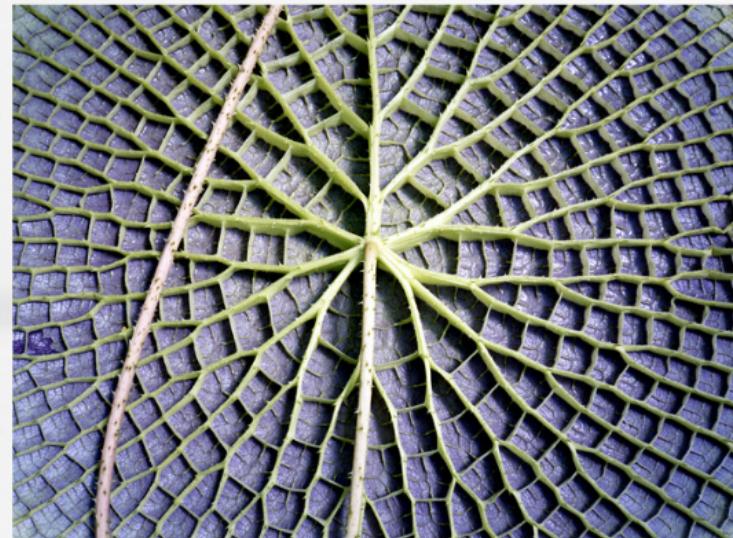
Music and Audio Research Lab
New York University
Jan 26th 2013

 MARL
NYU Music and Audio Research Laboratory

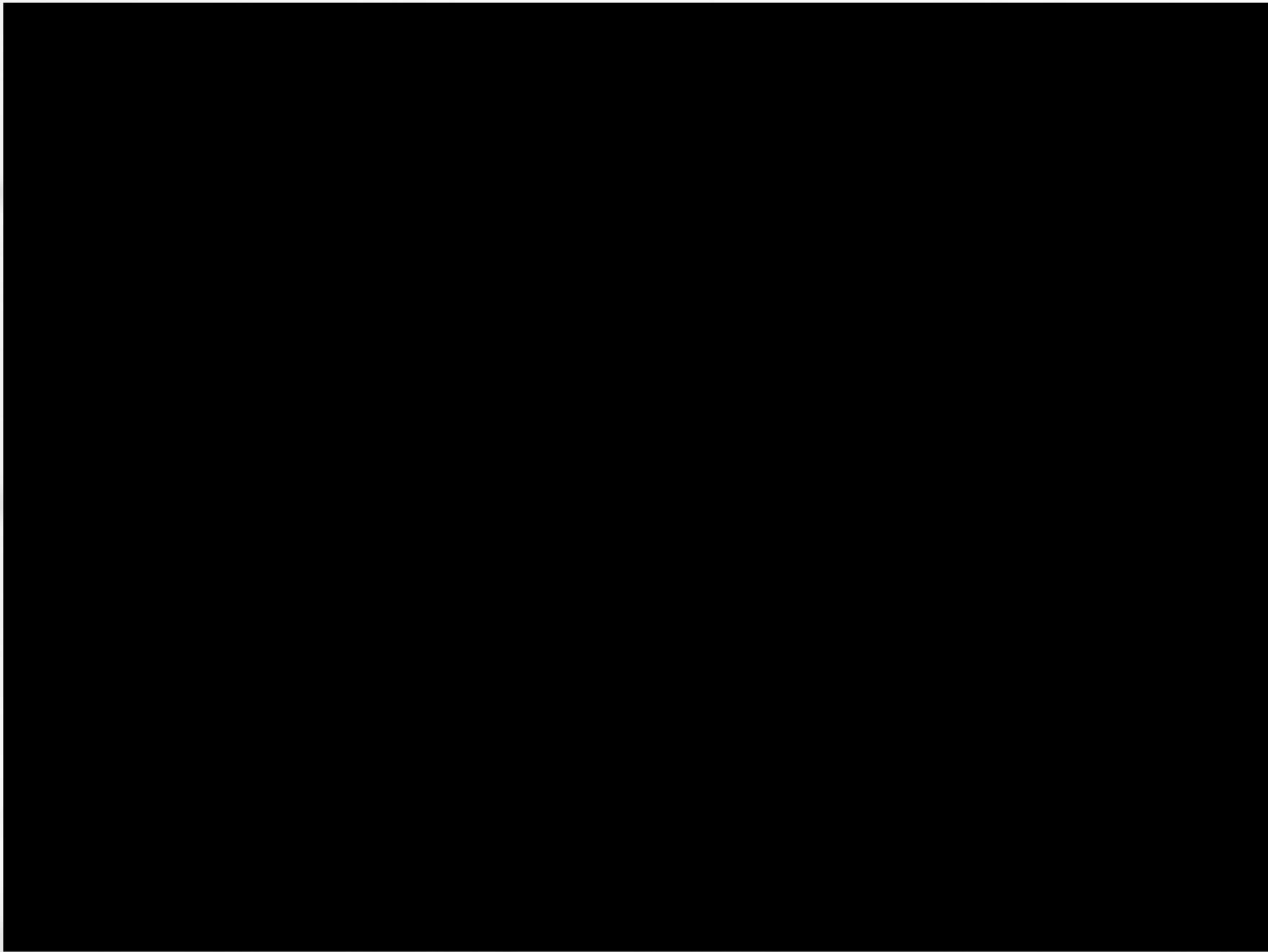
 theechoonest

Goal

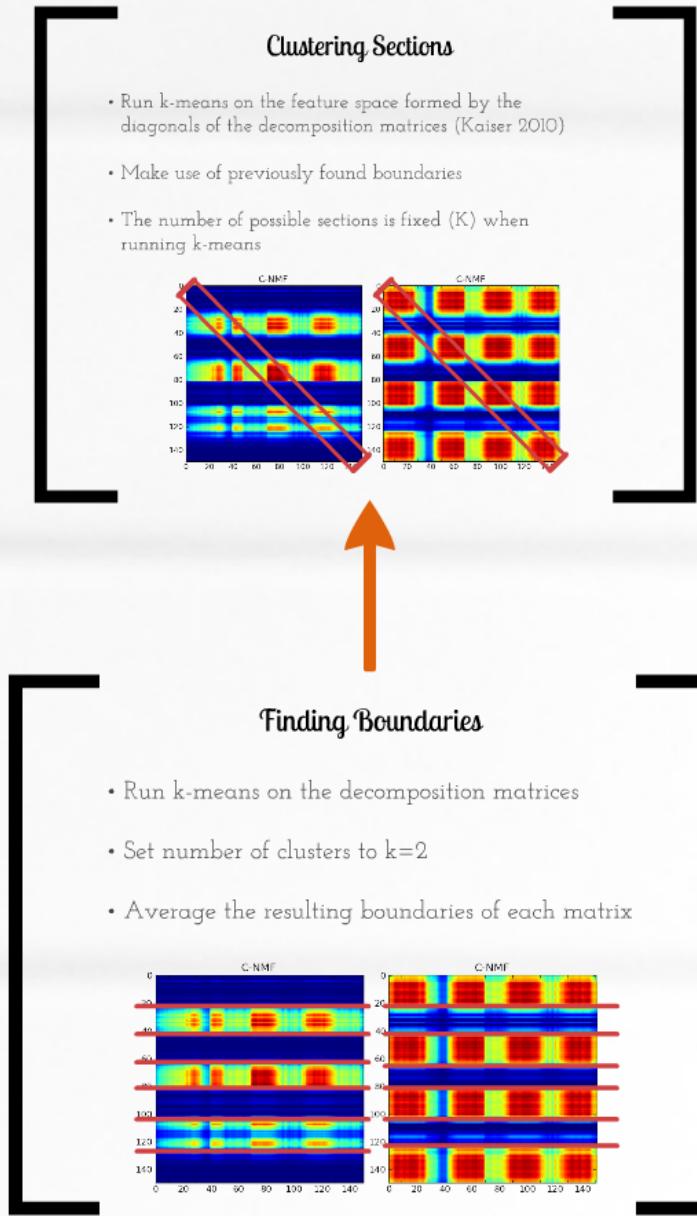
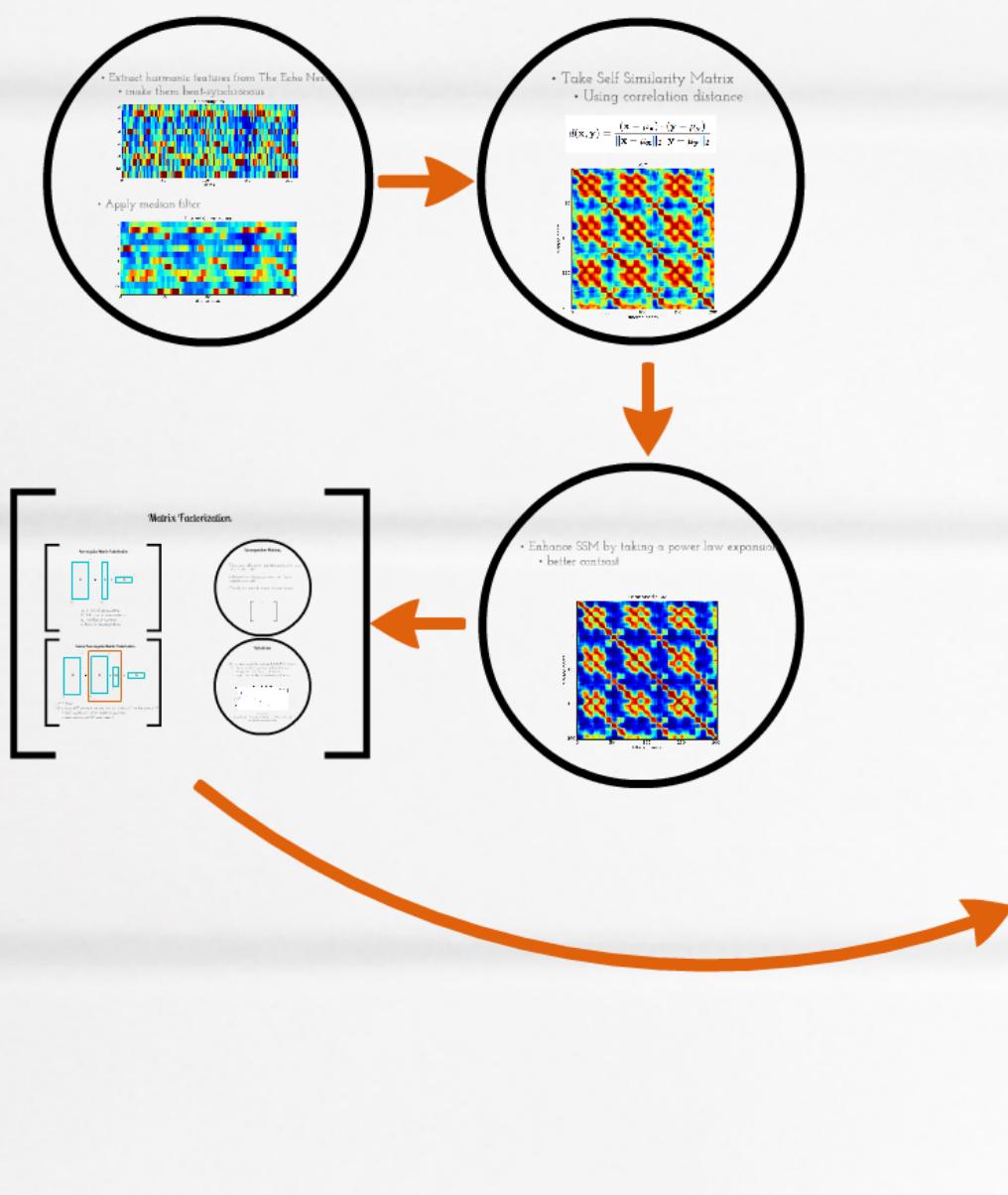
- To automatically identify the different sections of a song by
 - finding the boundaries of the sections
 - clustering the sections based on their similarity (e.g. A, B)
 - focusing on western popular music



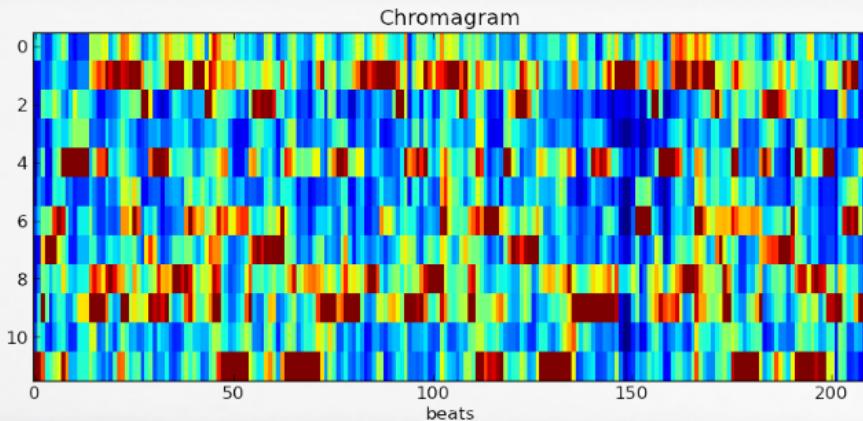
Example



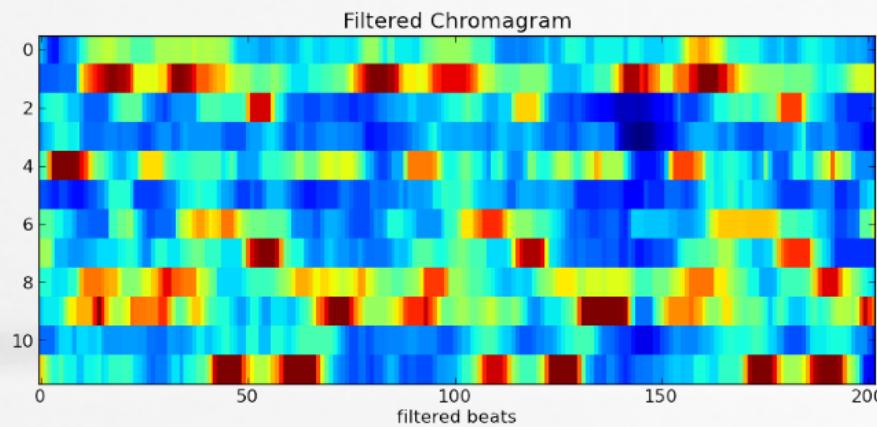
How?



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

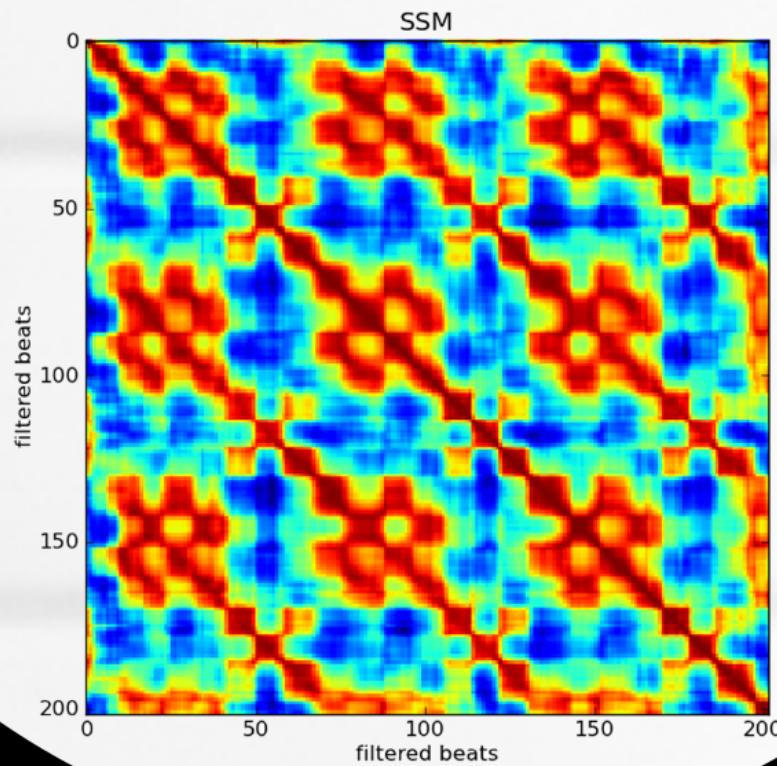


- Apply median filter

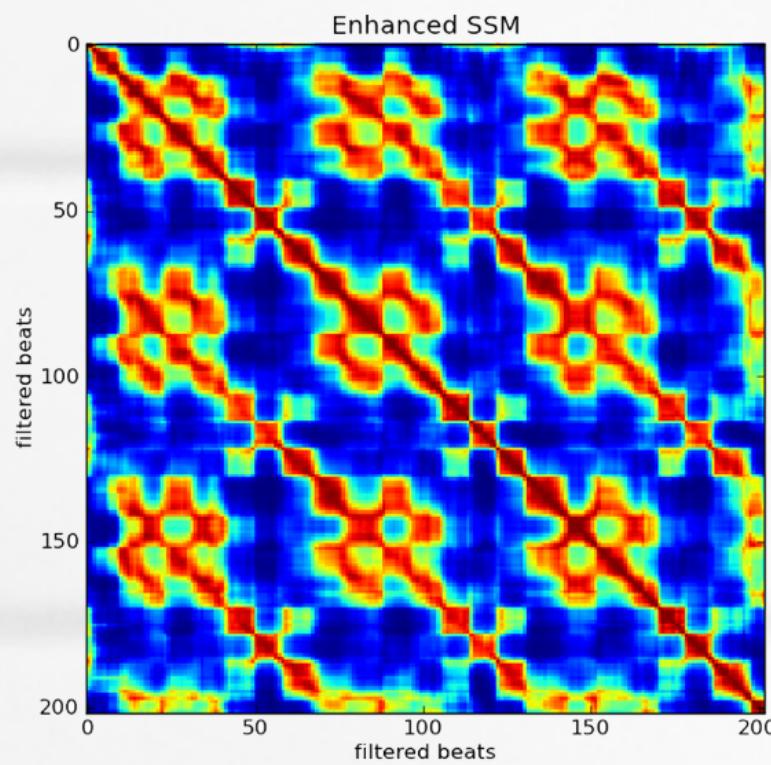


- Take Self Similarity Matrix
 - Using correlation distance

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



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



Matrix Factorization

Non-negative Matrix Factorization

$$\begin{matrix} 1 & p \\ X & \end{matrix} \approx \begin{matrix} 1 & r \\ F & \end{matrix} \times \begin{matrix} 1 & p \\ G & \end{matrix}$$

N N r

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

Convex Non-negative Matrix Factorization

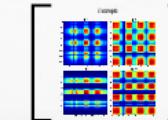
$$\begin{matrix} 1 & p \\ X & \end{matrix} \approx \begin{matrix} 1 & p \\ X & \end{matrix} \times \begin{matrix} 1 & r \\ W & \end{matrix} \times \begin{matrix} 1 & p \\ G & \end{matrix}$$

N N r

- $F = XW$
- 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

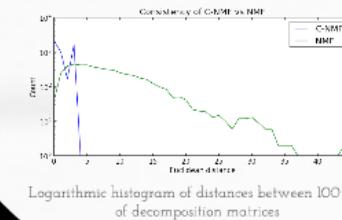
Decomposition Matrices

- 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

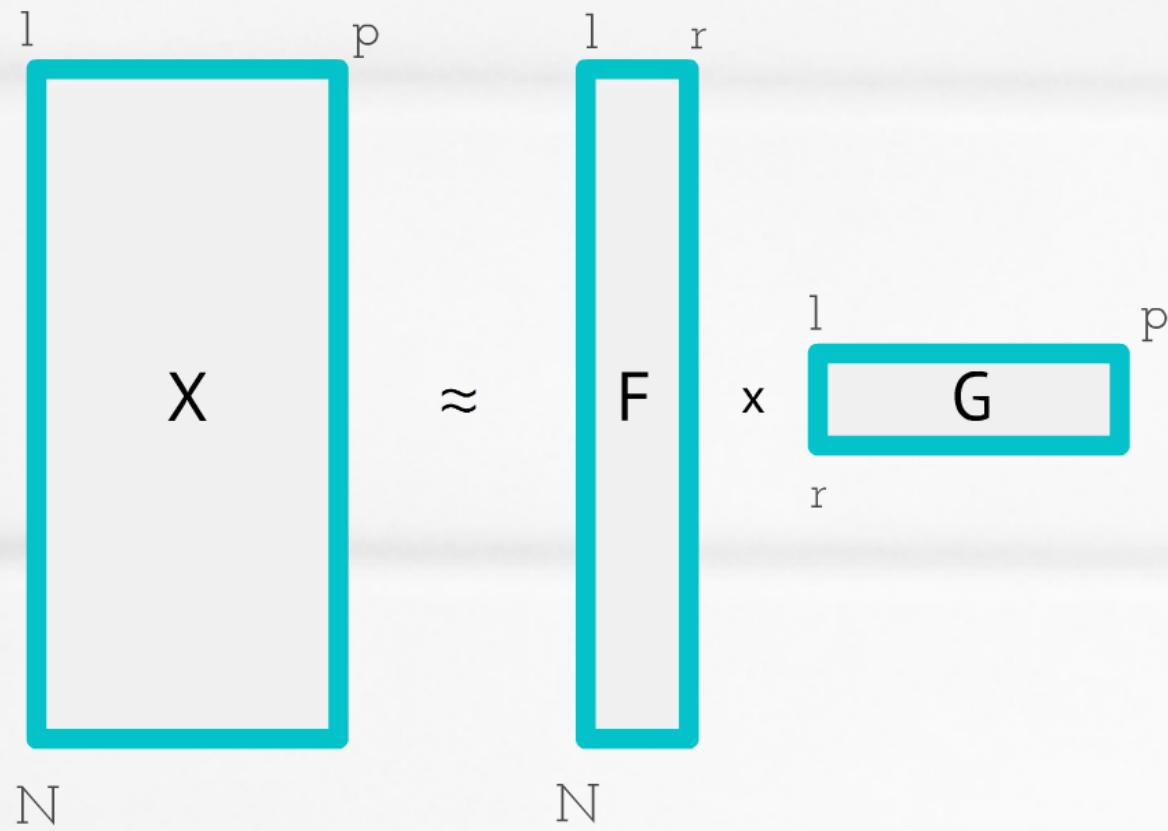


Robustness

- 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



Non-negative Matrix Factorization



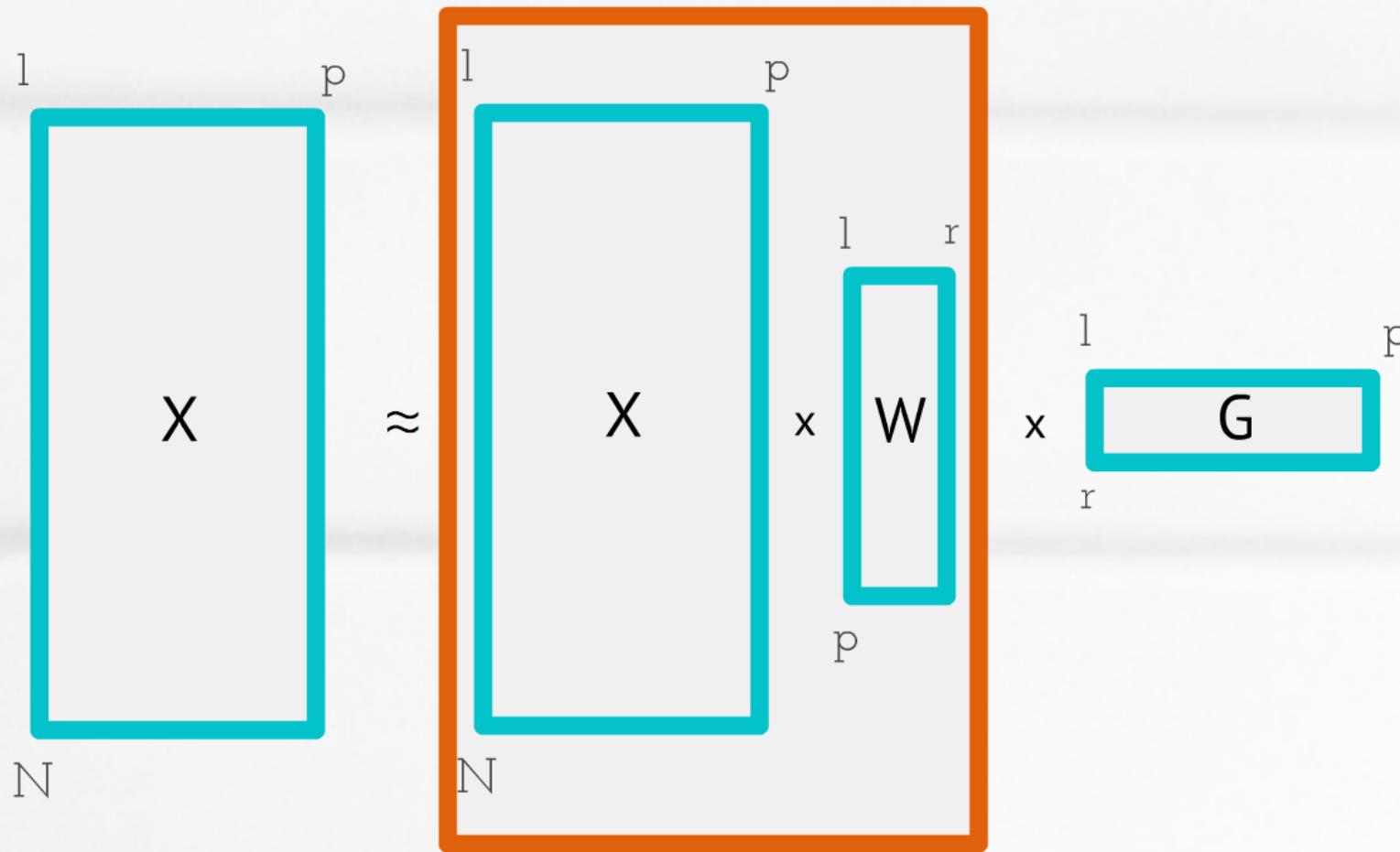
X , F , and G are positive

N : Number of observations

p : Number of features

r : Rank of decomposition

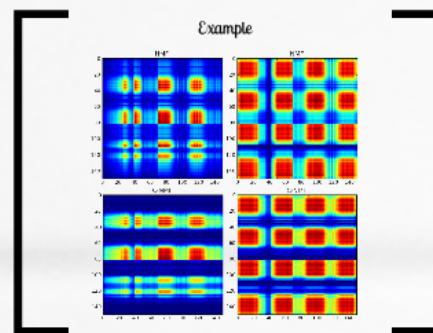
Convex Non-negative Matrix Factorization



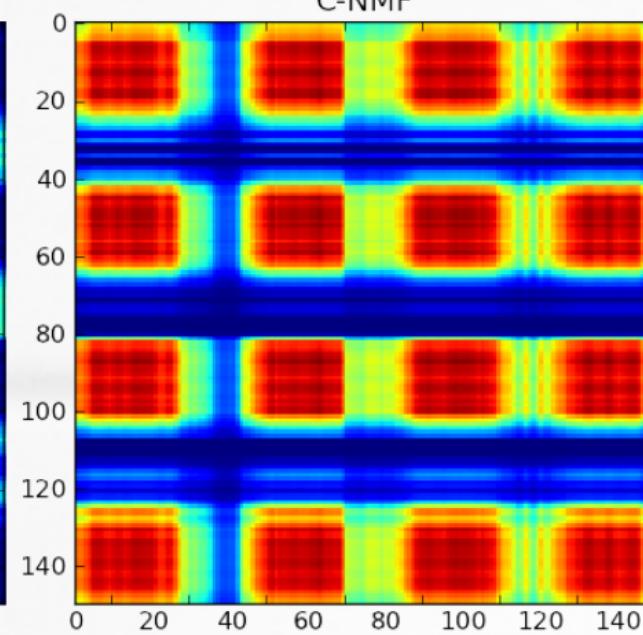
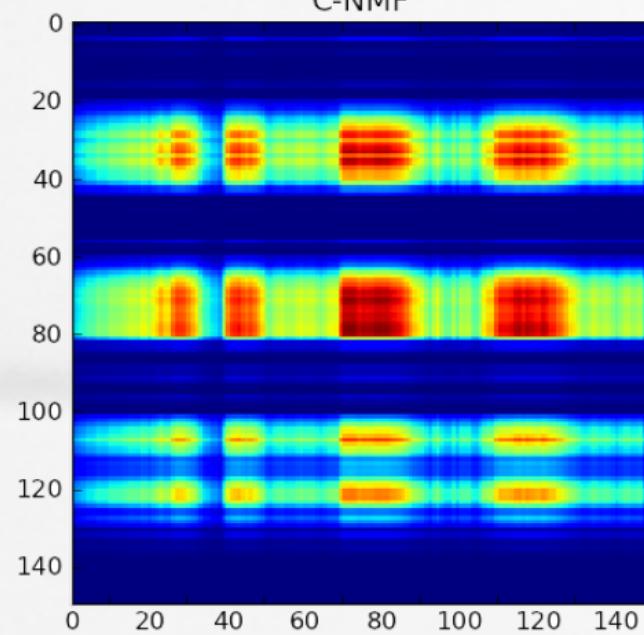
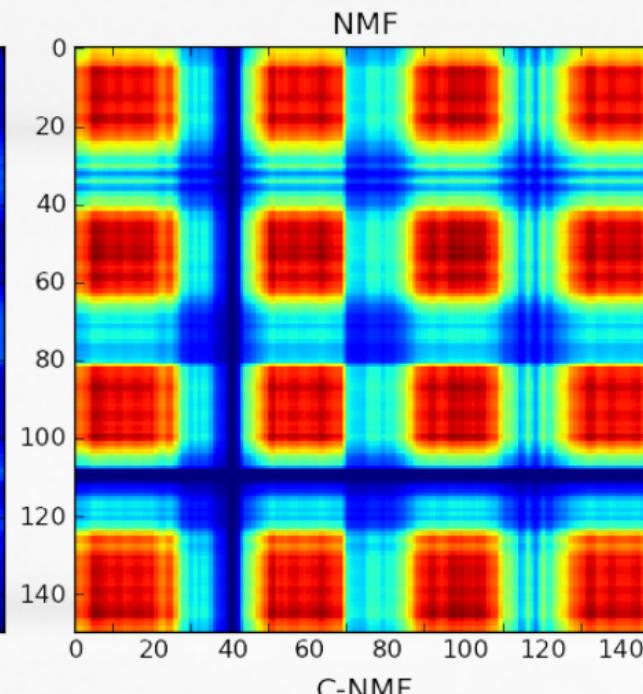
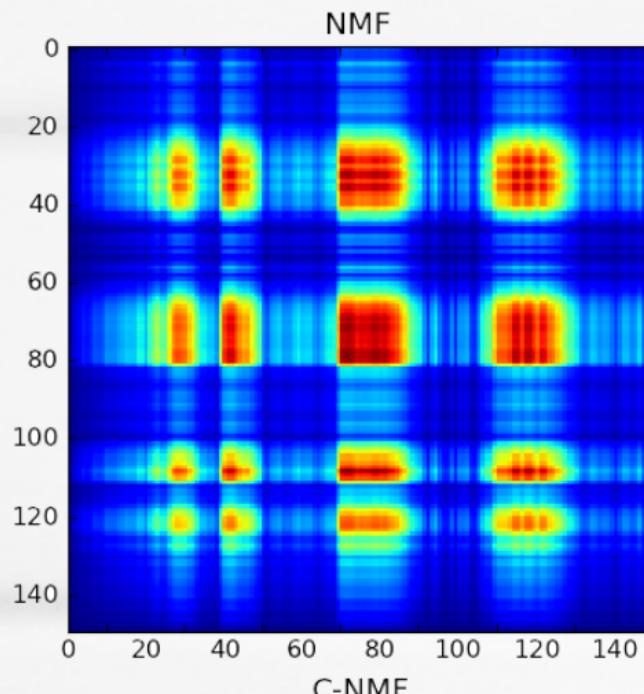
- $F = XW$
- Columns of F become convex combinations of the features of X
 - Each coefficient of W must be positive
 - Each column of W must sum 1

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

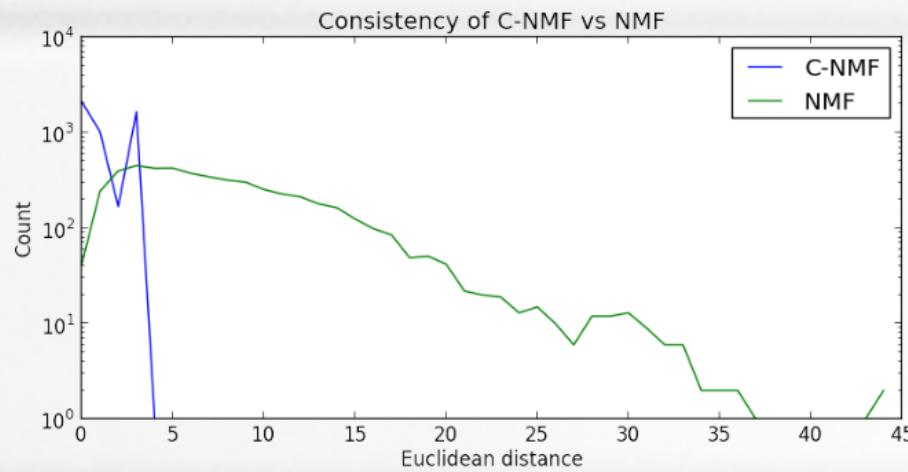


Example



Robustness

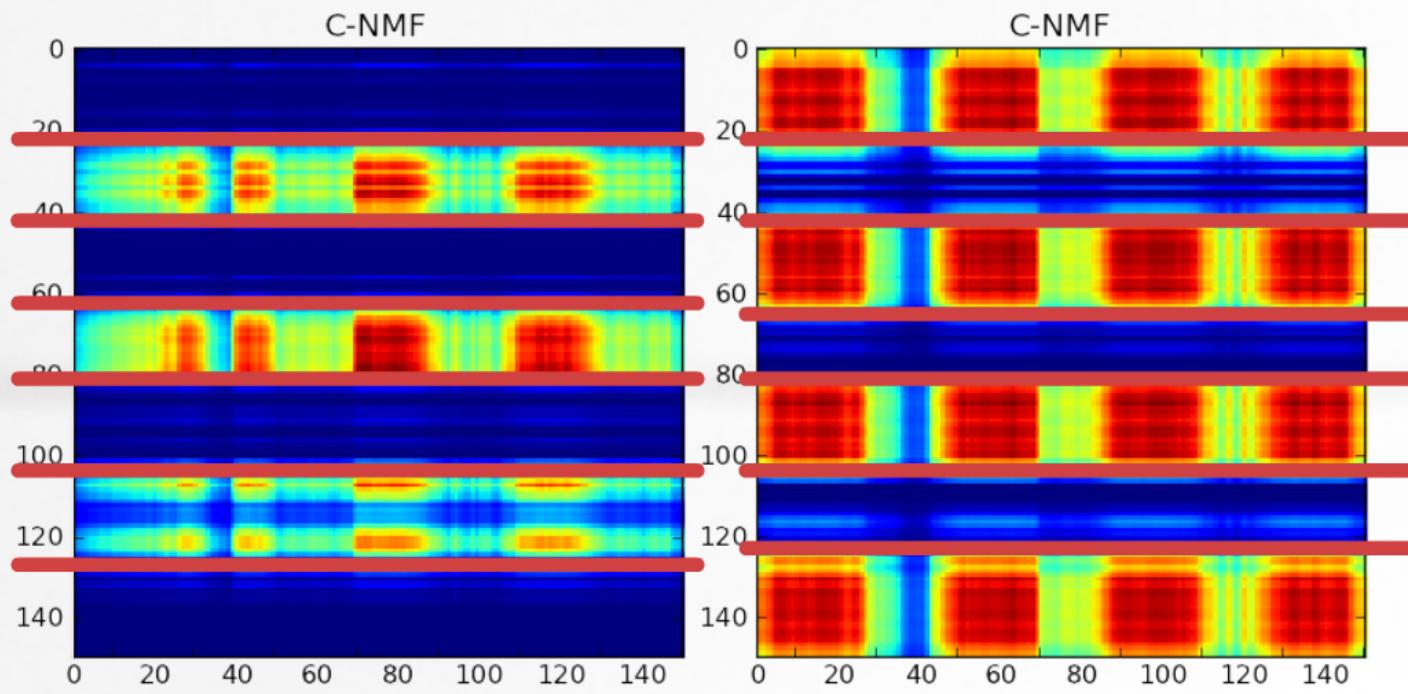
- By adding a convex constrain to NMF we tend to find more consistent solutions in less iterations
 - Less prone to fall into local minima
 - Lower the number of iterations -> Faster



Logarithmic histogram of distances between 100 sets
of decomposition matrices

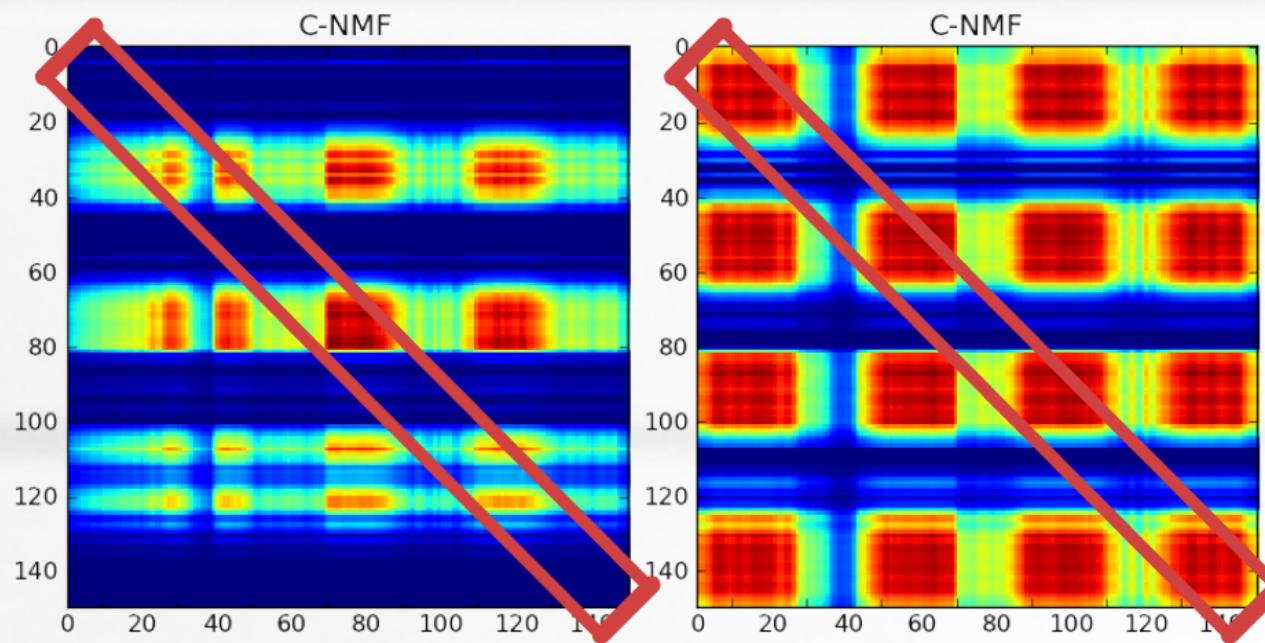
Finding Boundaries

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



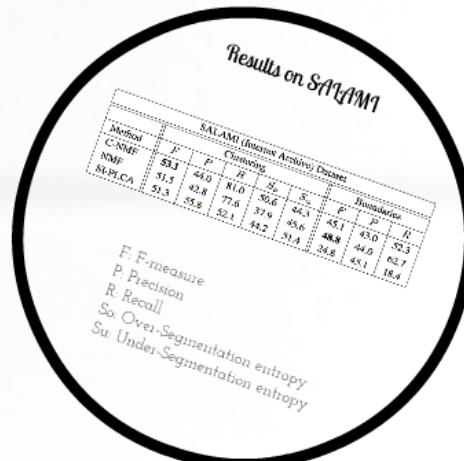
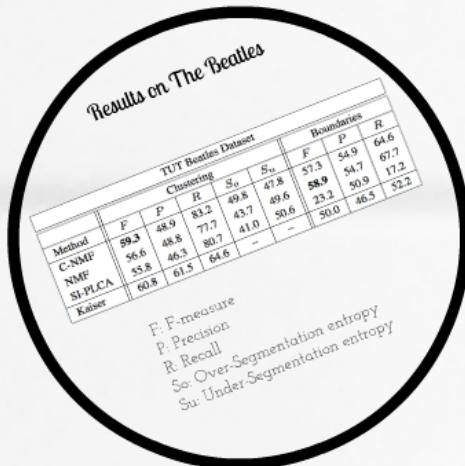
Clustering Sections

- 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 online
 - 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								
Method	Clustering					Boundaries		
	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

S_o : Over-Segmentation entropy

S_u : Under-Segmentation entropy

Results on SALAMI

SALAMI (Internet Archive) Dataset								
Method	Clustering					Boundaries		
	<i>F</i>	<i>P</i>	<i>R</i>	S_o	S_u	<i>F</i>	<i>P</i>	<i>R</i>
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

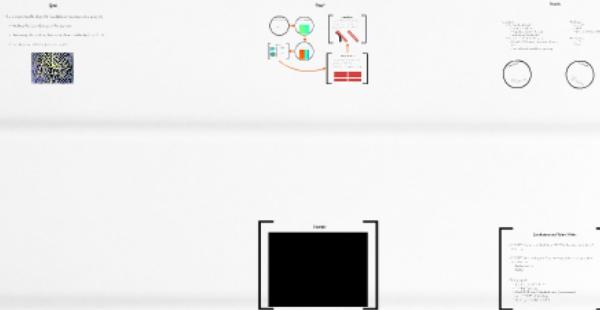
S_o : Over-Segmentation entropy

S_u : 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 by Matrix Factorization



Oriol Nieto & Tristan Jehan

Music and Audio Research Lab

New York University

Jan 26th 2013

