The State of Music at LabROSA

Dan Ellis
Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Eng., Columbia Univ., NY USA

dpwe@ee.columbia.edu http://labrosa.ee.columbia.edu/

1. The State of LabROSA
2. Music Projects
3. The Big Picture

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

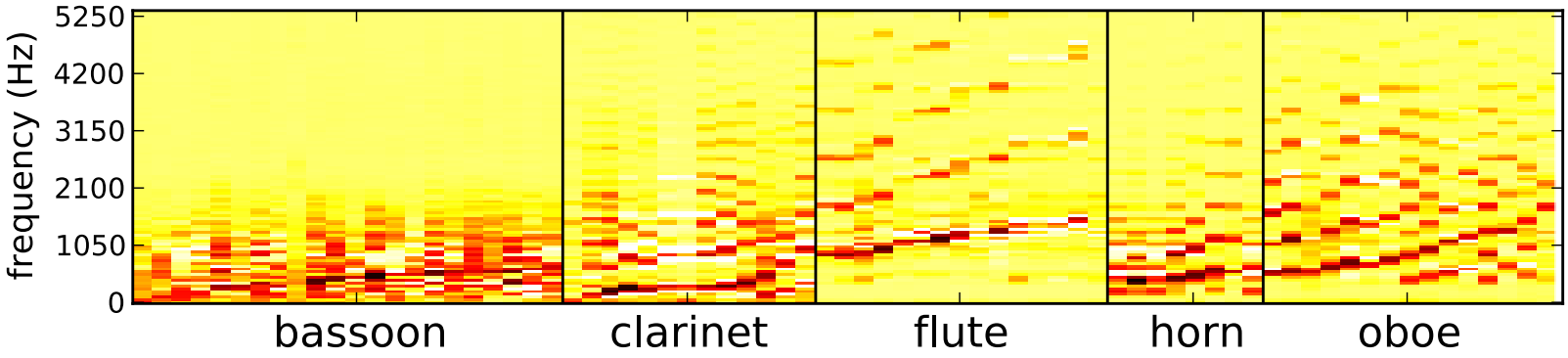
Columbia University
In the City of New York

Beta Process NMF

Liang, Hoffman

- Automatically choose how many components to use

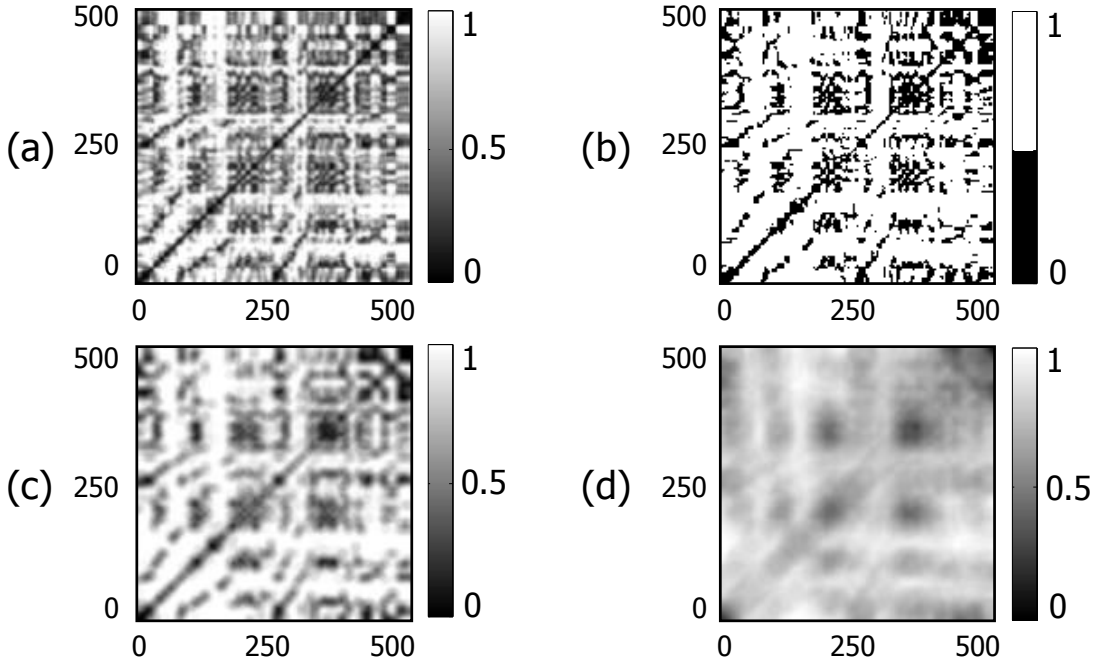
$$\mathbf{X} = \mathbf{D}(\mathbf{S} \odot \mathbf{Z}) + \mathbf{E}$$



Structure Similarity

Silva, Papadopoulos

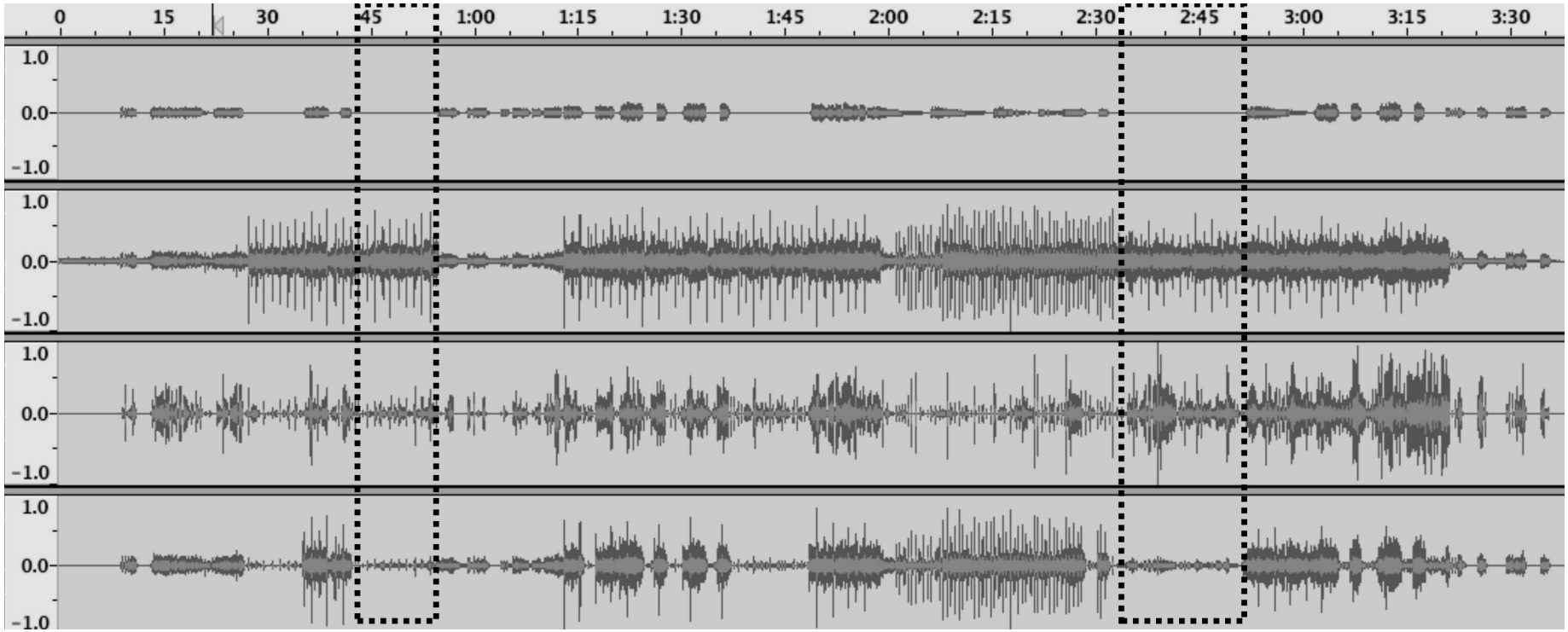
- CK-1 image similarity uses MPEG Video compression
  - can exploit shifted parts of image
- Match pieces based on structure recurrence plots (Bello'11)



Block Structure RPCA

Papadopoulos

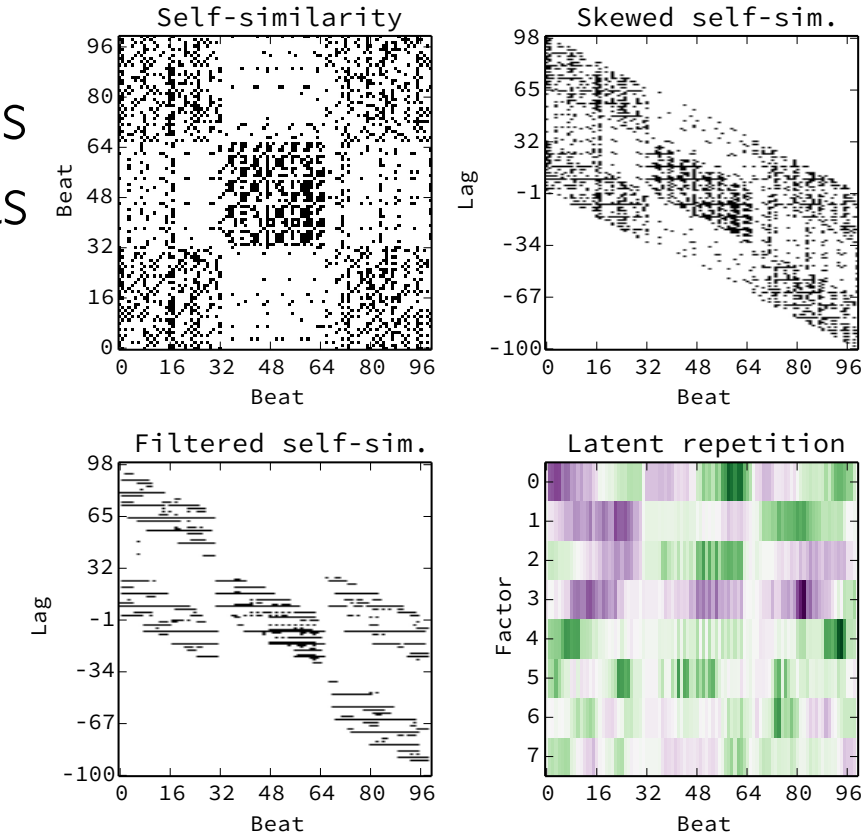
- RPCA separates vocals and background based on low rank optimization
  - single trade-off parameter
  - adjust based on higher-level musical features?



Ordinal LDA Segmentation

McFee

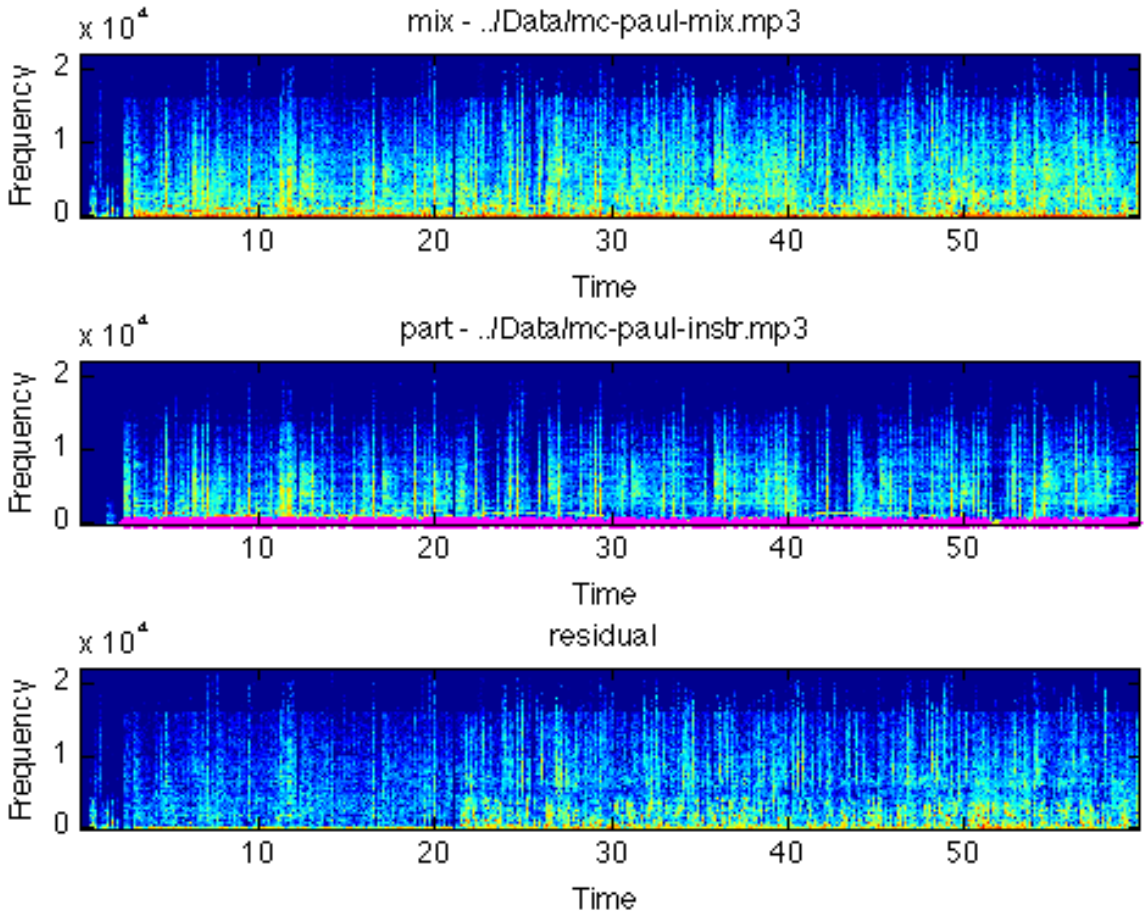
- Low-rank decomposition of skewed self-similarity to identify repeats
- Learned weighting of multiple factors to segment
  - Linear Discriminant Analysis between adjacent segments



"Remixavier"

Raffel

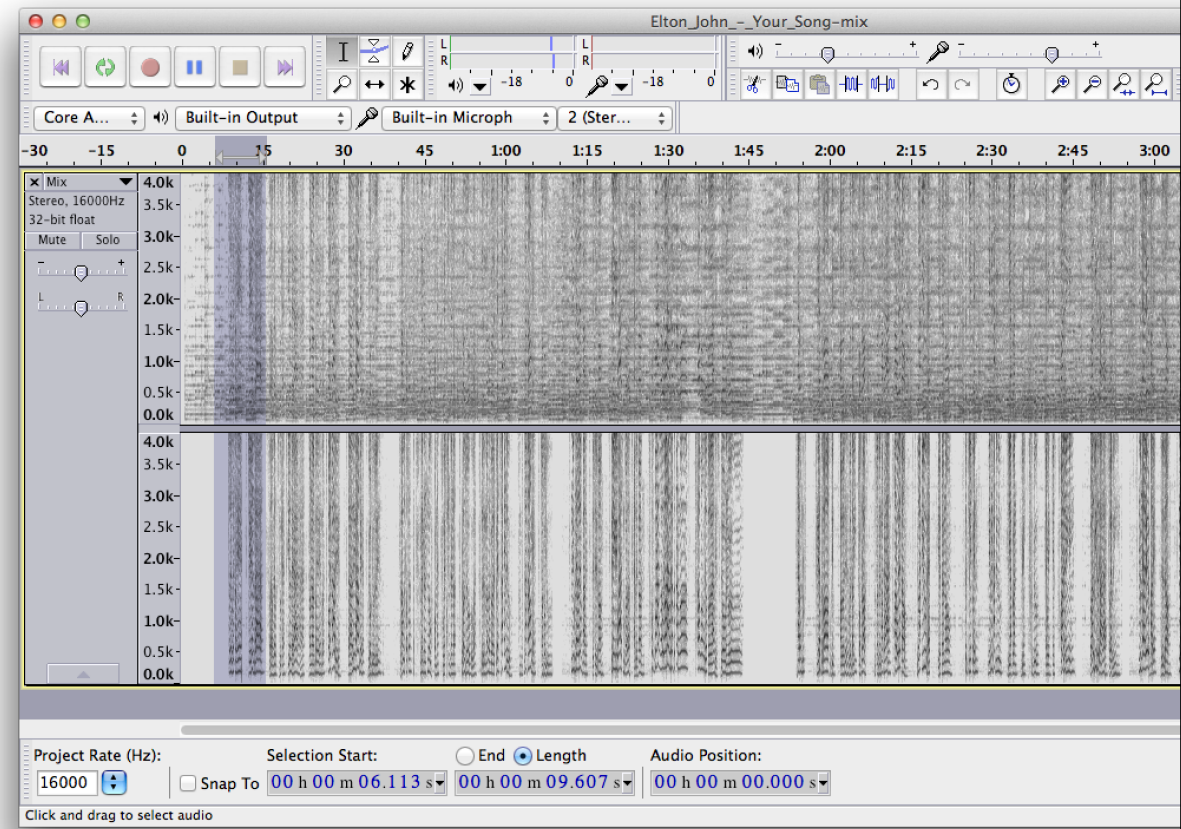
- **Optimal align-and-cancel of mix and acapella**
  - timing and channel may differ



Singing ASR

McVicar

- Speech recognition adapted to singing
  - needs aligned data
- Extensive work to match up scraped "acapellas" and full mix
  - including jumps!



Million Song Dataset

Bertin-Mahieux
McFee

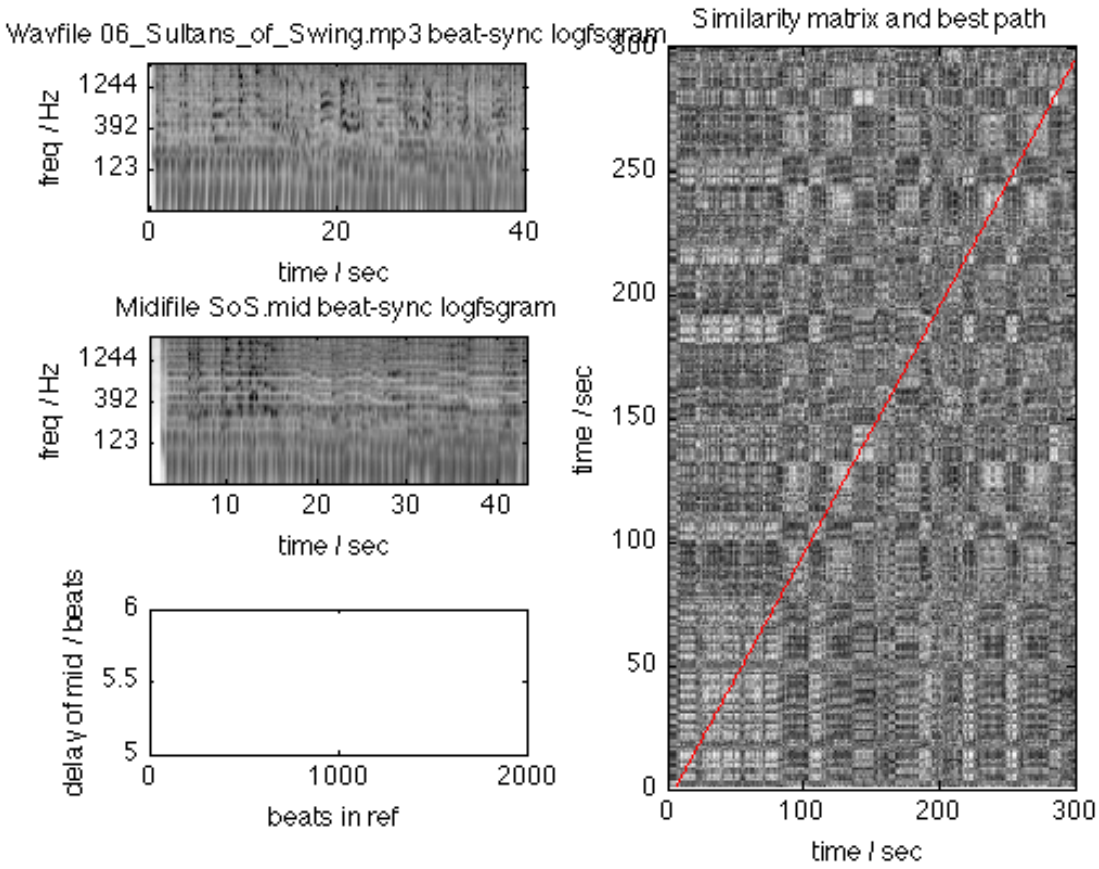
- **Many Facets**
  - Echo Nest audio features + metadata
  - Echo Nest "taste profile" user-song-listen count
  - Second Hand Song covers
  - musiXmatch lyric BoW
  - last.fm tags

- **Now with audio?**
  - resolving artist / album / track / duration against what.cd

MIDI-to-MSD

Raffel Shi

- Aligned MIDI to Audio is a nice transcription



- Can we find matches in large databases?

Summary

- **Basic techniques**
  - beat tracking, segmentation, chord recognition, transcription
- **More data**
  - audio
  - alignments
  - aligned transcriptions
- **Sharing code and data**