

Lecture 15: Research at LabROSA

1. Sources, Mixtures, & Perception
2. Spatial Filtering
3. Time-Frequency Masking
4. Model-Based Separation

Dan Ellis

Dept. Electrical Engineering, Columbia University

dpwe@ee.columbia.edu <http://www.ee.columbia.edu/~dpwe/e4896/>

Sparse + Low-Rank + NMF

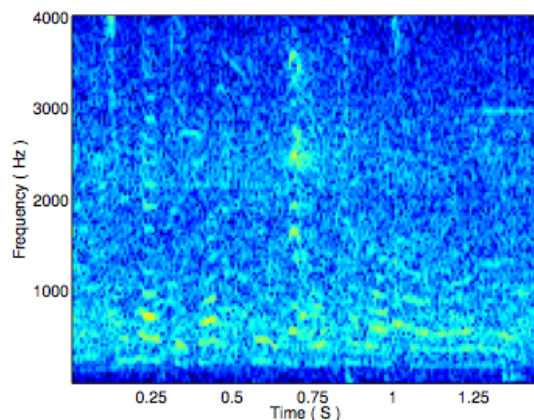
Zhuo Chen

- Optimization to decompose spectrogram:

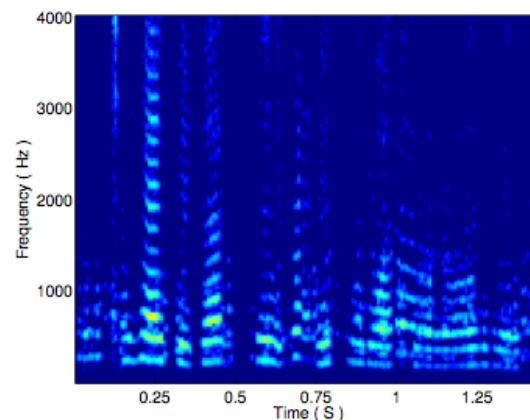
- minimize

- s.t.
$$|\mathbf{S}|_1 + |\mathbf{L}|_* + D_{KL}(\mathbf{Y} - \mathbf{S} - \mathbf{L} || \mathbf{H} \cdot \mathbf{W})$$
$$\mathbf{Y} = \mathbf{S} + \mathbf{L} + \mathbf{H} \cdot \mathbf{W}$$

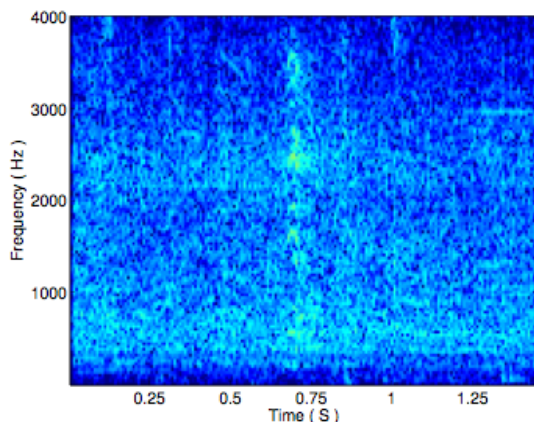
Y



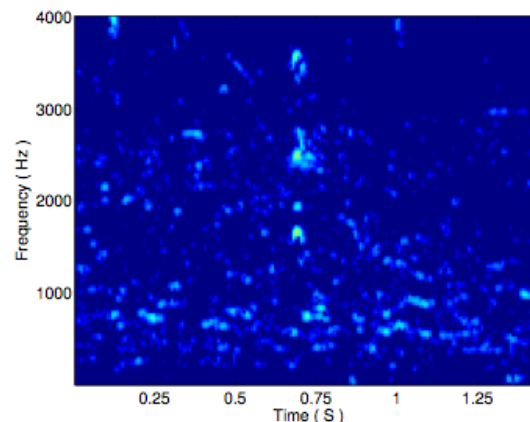
H·W



L



S

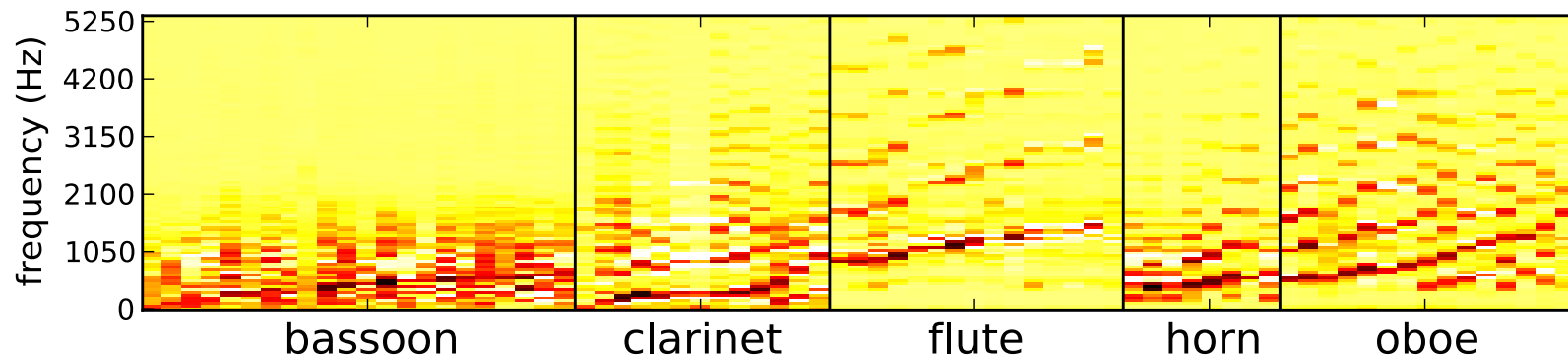


Beta Process NMF

Liang, Hoffman

- Automatically choose how many components to use

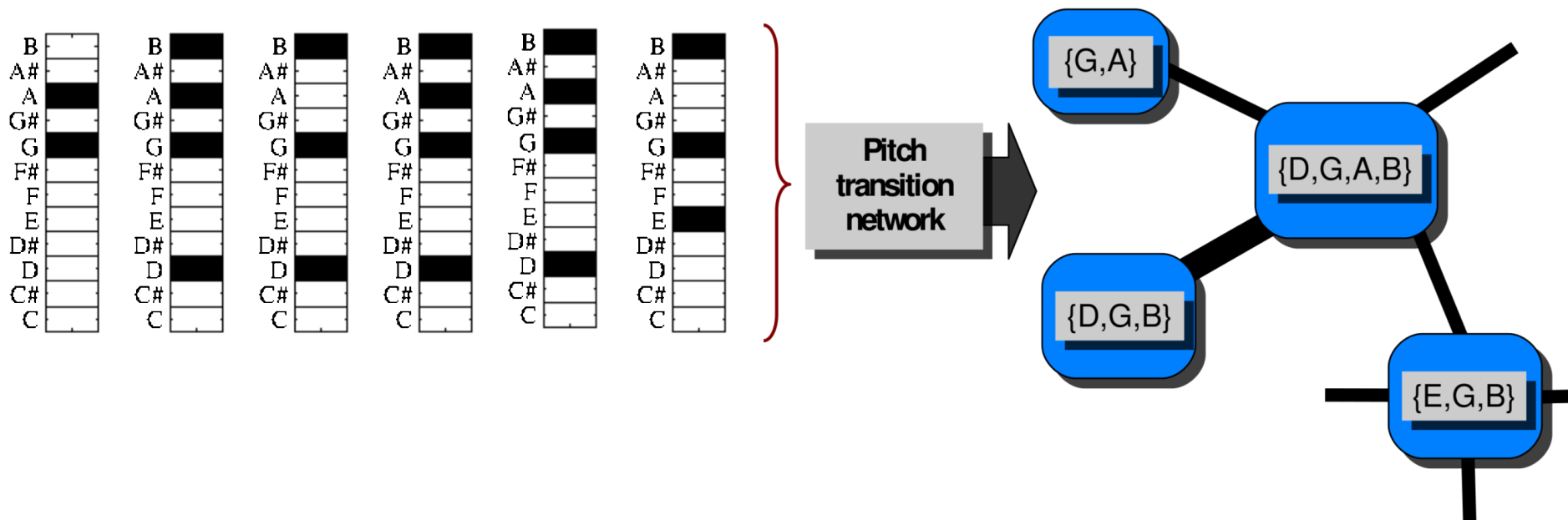
$$\mathbf{X} = \mathbf{D}(\mathbf{S} \odot \mathbf{Z}) + \mathbf{E}$$



Music Complexity

Colin Raffel

- How can we capture musical patterns in the Million Song Dataset?
- Network analysis of quantized simultaneities
 - after Serrà et al. 2012

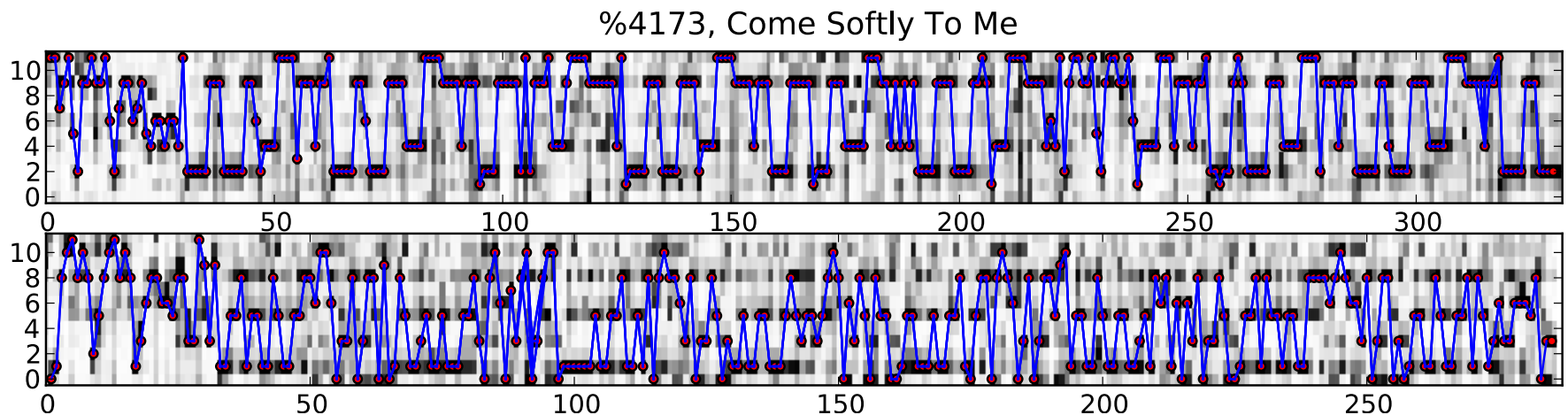


from Serrà, Corral, Boguña, Haro, & Arcos, 2012

Large-Scale Cover Recognition I

Thierry Bertin-Mahieux

- How can we find covers in IM songs?
 - @ 1 sec / comparison, one search = 11.5 CPU-days
 - full N^2 mining = 16,000 CPU-years
- Need a hashing technique
 - landmark-based description of chroma patches

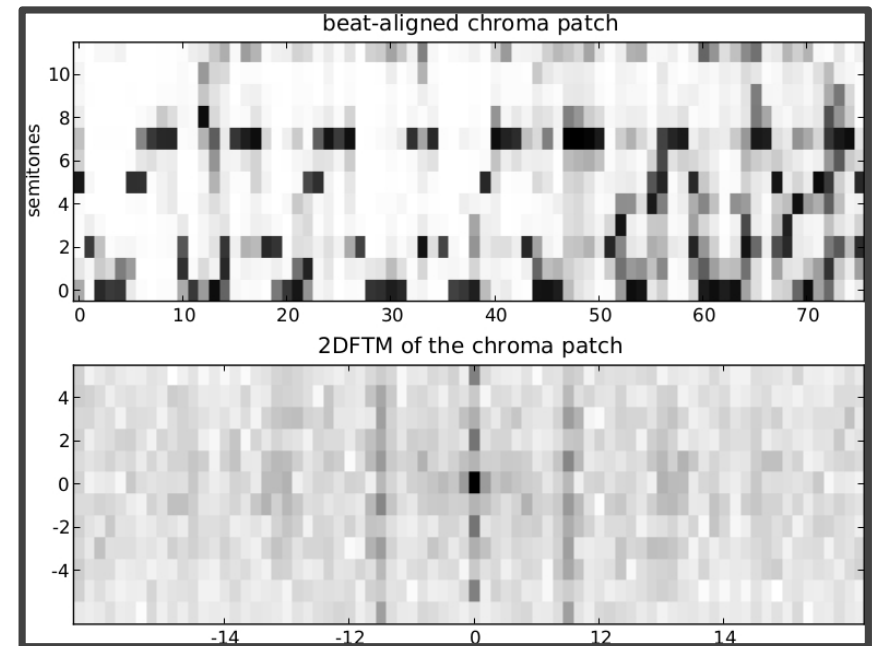


- Euclidean space projection?

Large-Scale Cover Recognition 2

Thierry Bertin-Mahieux

- 2D Fourier Transform Magnitude (2DFTM)
 - fixed-size feature to capture “essence” of chromagram:

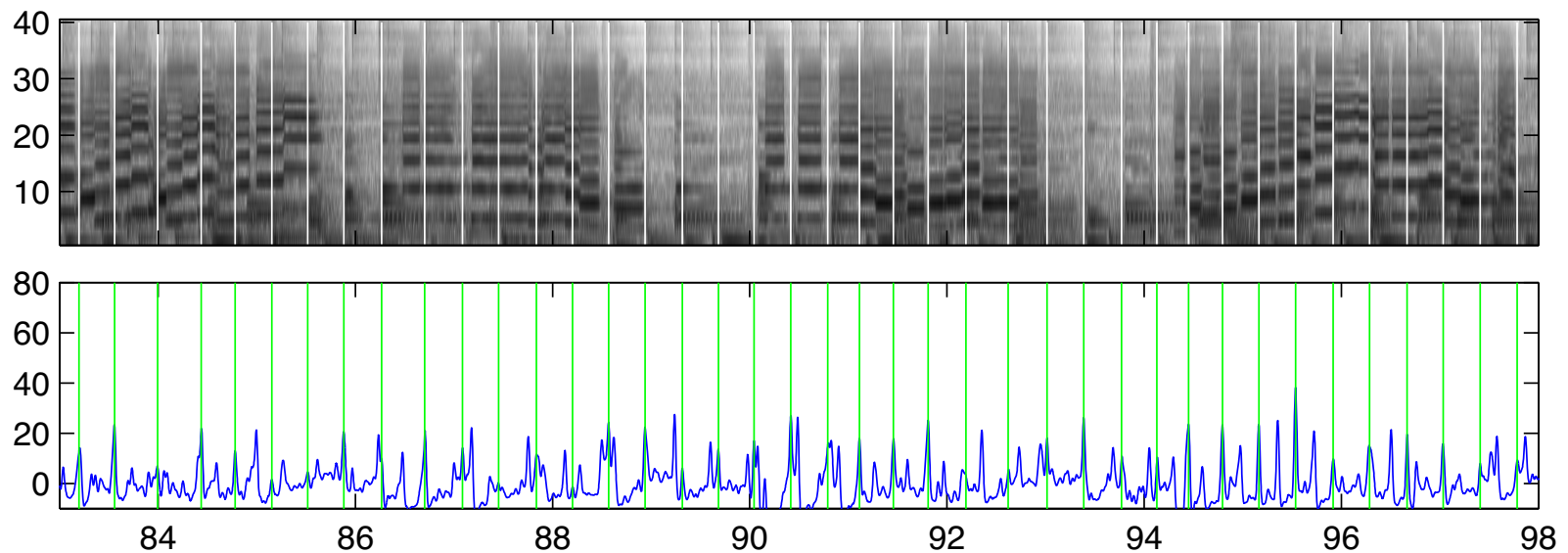


- First results on finding covers in IM songs

	Average rank	meanAP
random	500,000	0.000
jumpcodes 2	308,369	0.002
2DFTM (50 PC)	137,117	0.020

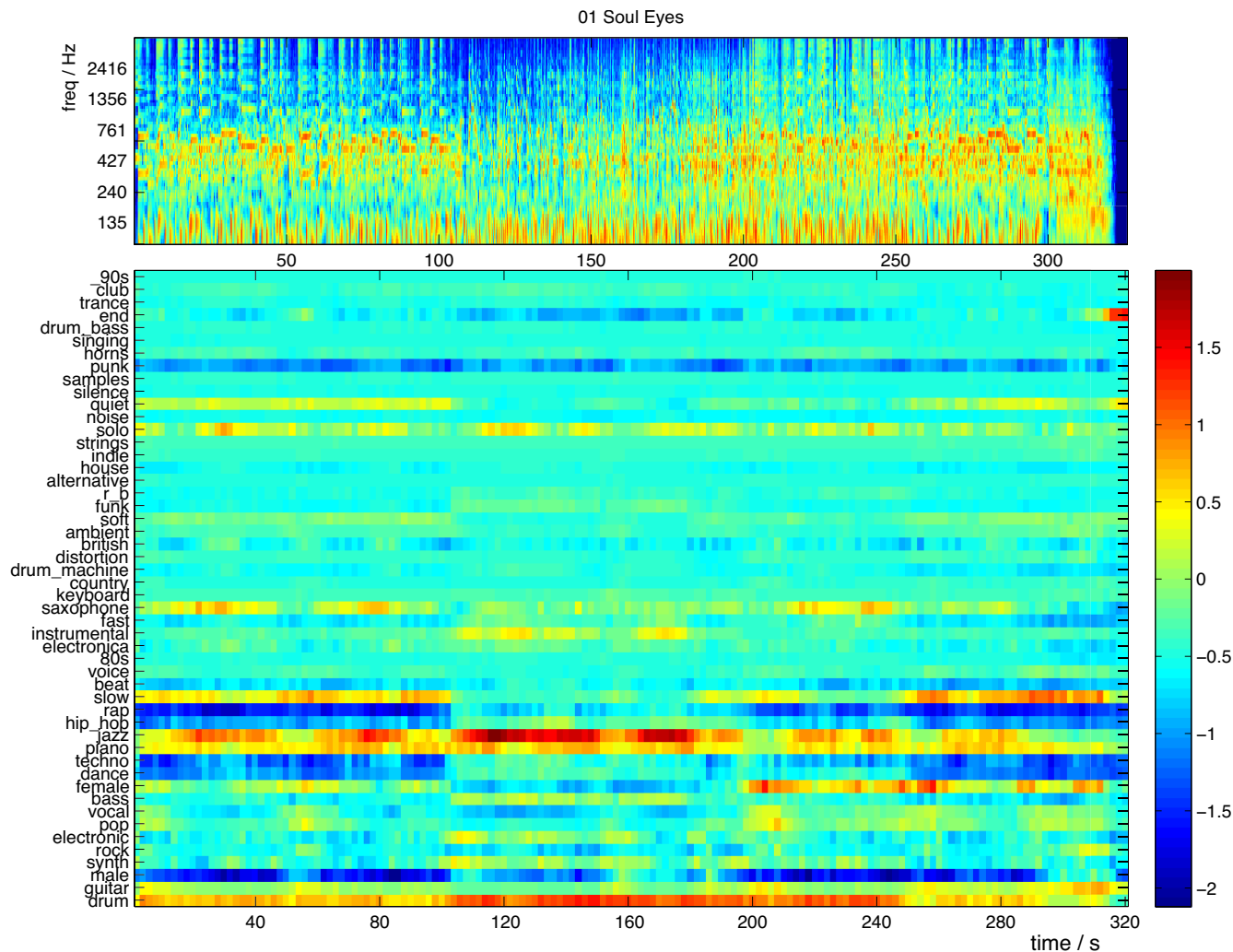
Jazz Discography Project

- How can MIR help organize jazz collections?
 - our tools are quite genre-specific
 - e.g. beat tracker is fine for pop, useless for Jazz



Local Tagging

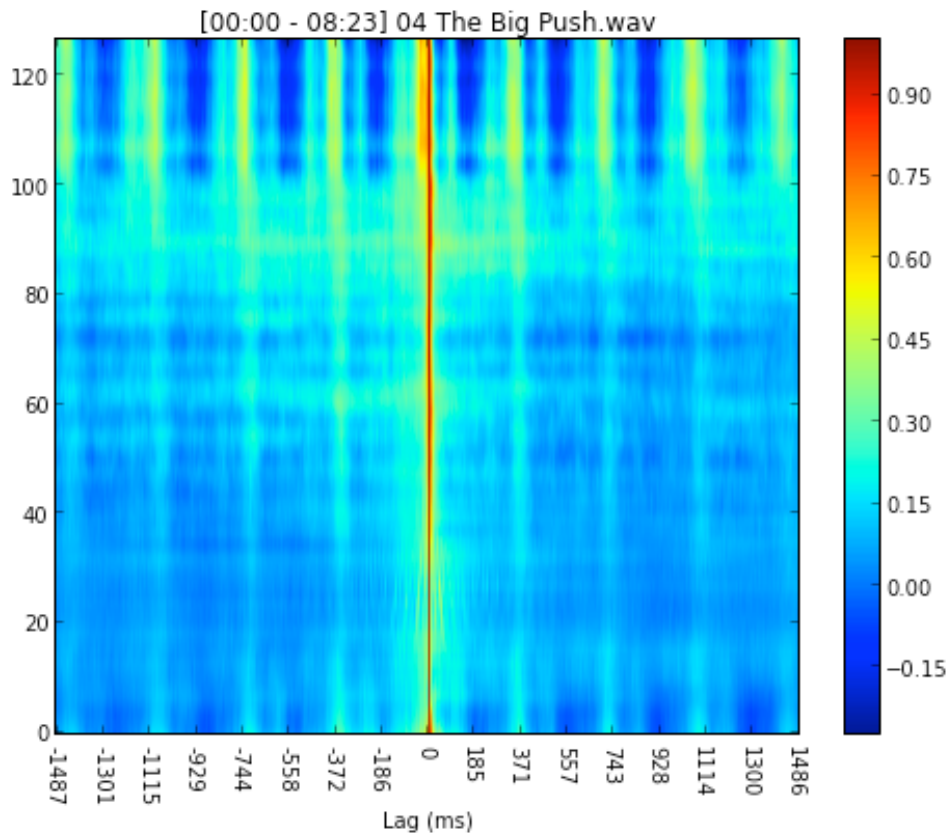
- MFCC-statistics classifiers on 5 sec windows trained from MajorMiner data



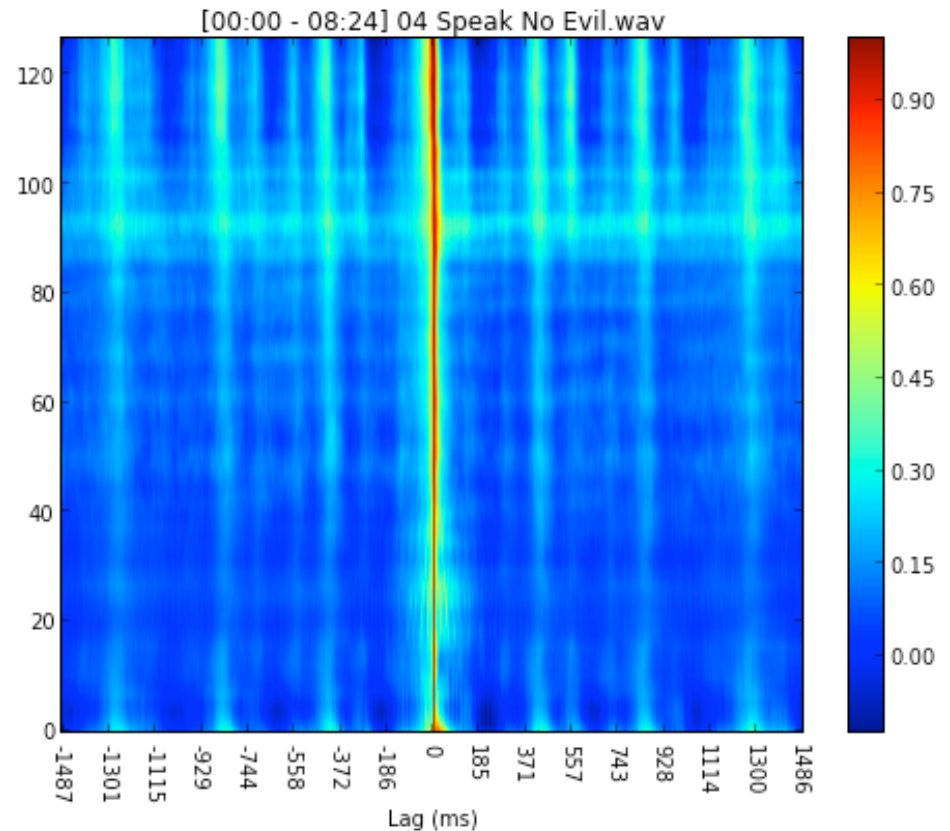
Onset Correlation

Brian McFee

- “Ahead of” or “behind” the beat?



Tony Williams



Elvin Jones

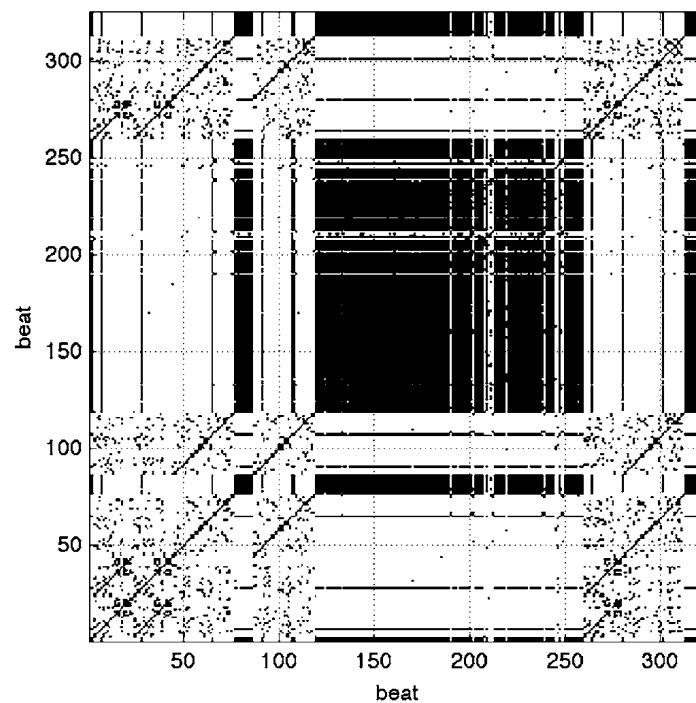
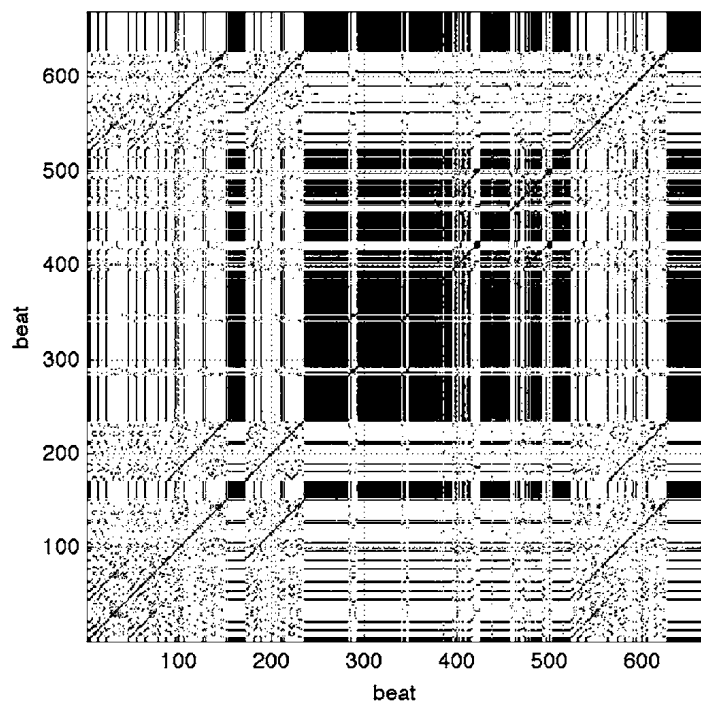


Structural Similarity

Diego Silva

Helene Papadopoulou

- Self-similarity shows repeating structure in music
- Can we find similar pieces by finding similar structures?

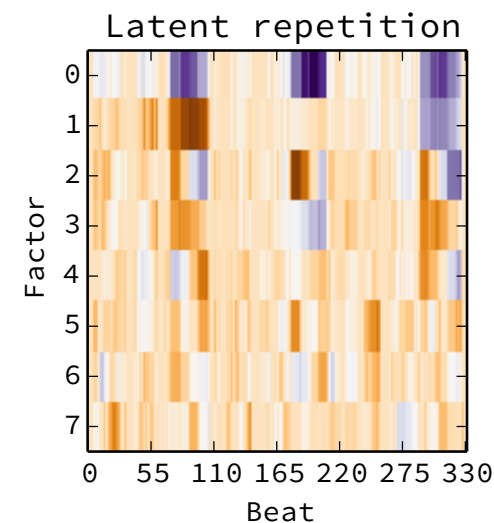
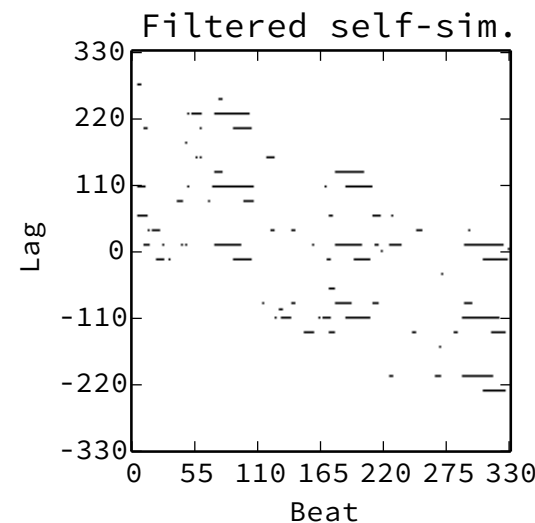
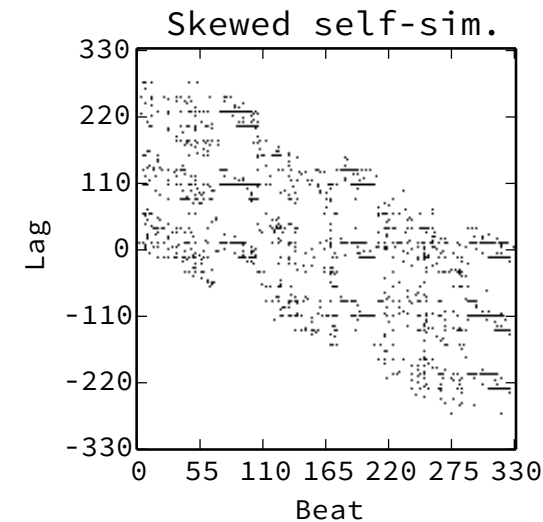
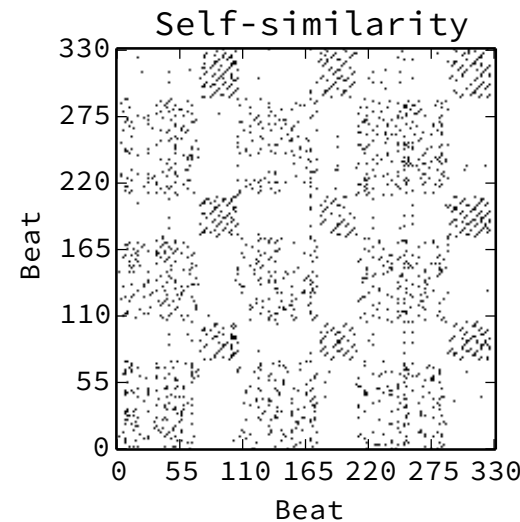


from Bello 2011

Ordinal LDA Segmentation

McFee

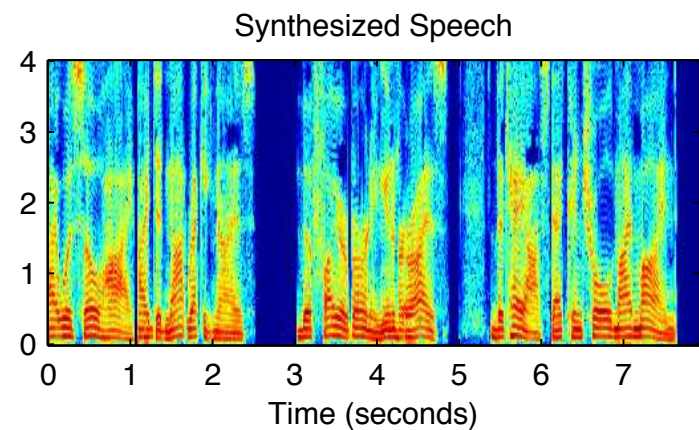
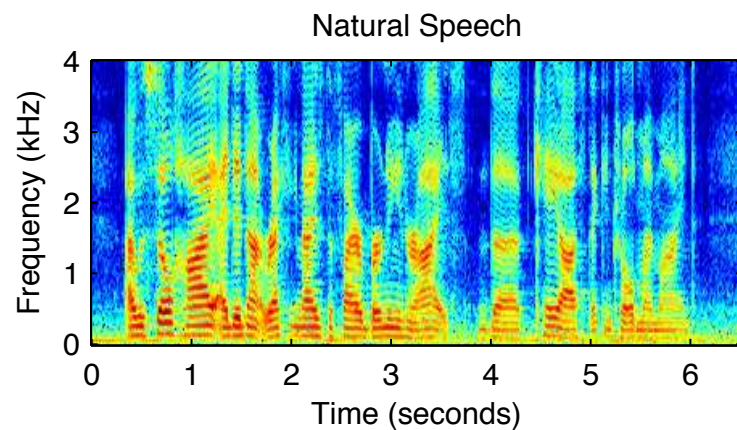
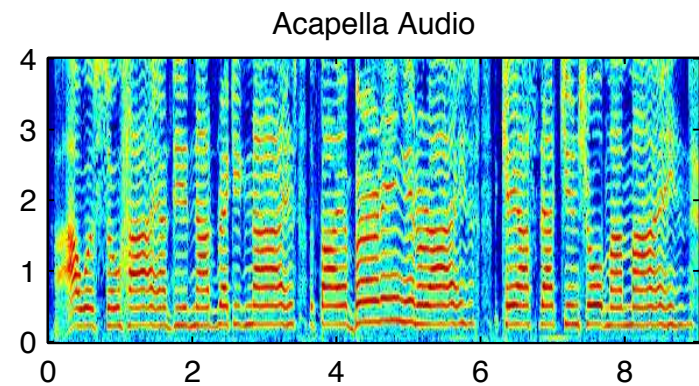
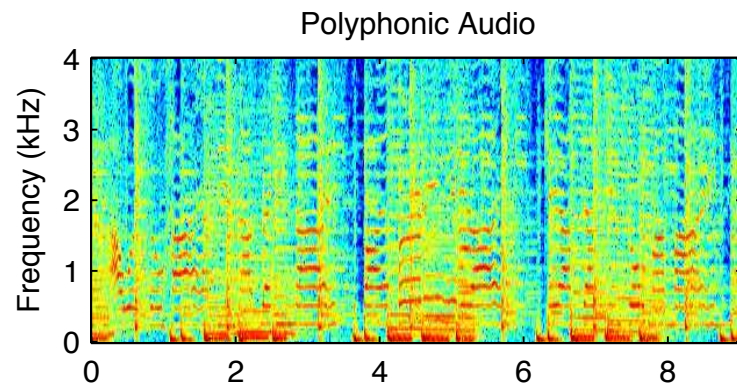
- Low-rank decomposition of skewed self-similarity to identify repeats
- Learned weighting of multiple factors to segment
 - Linear Discriminant Analysis between adjacent segments



Lyric Recognition

Matt McVicar

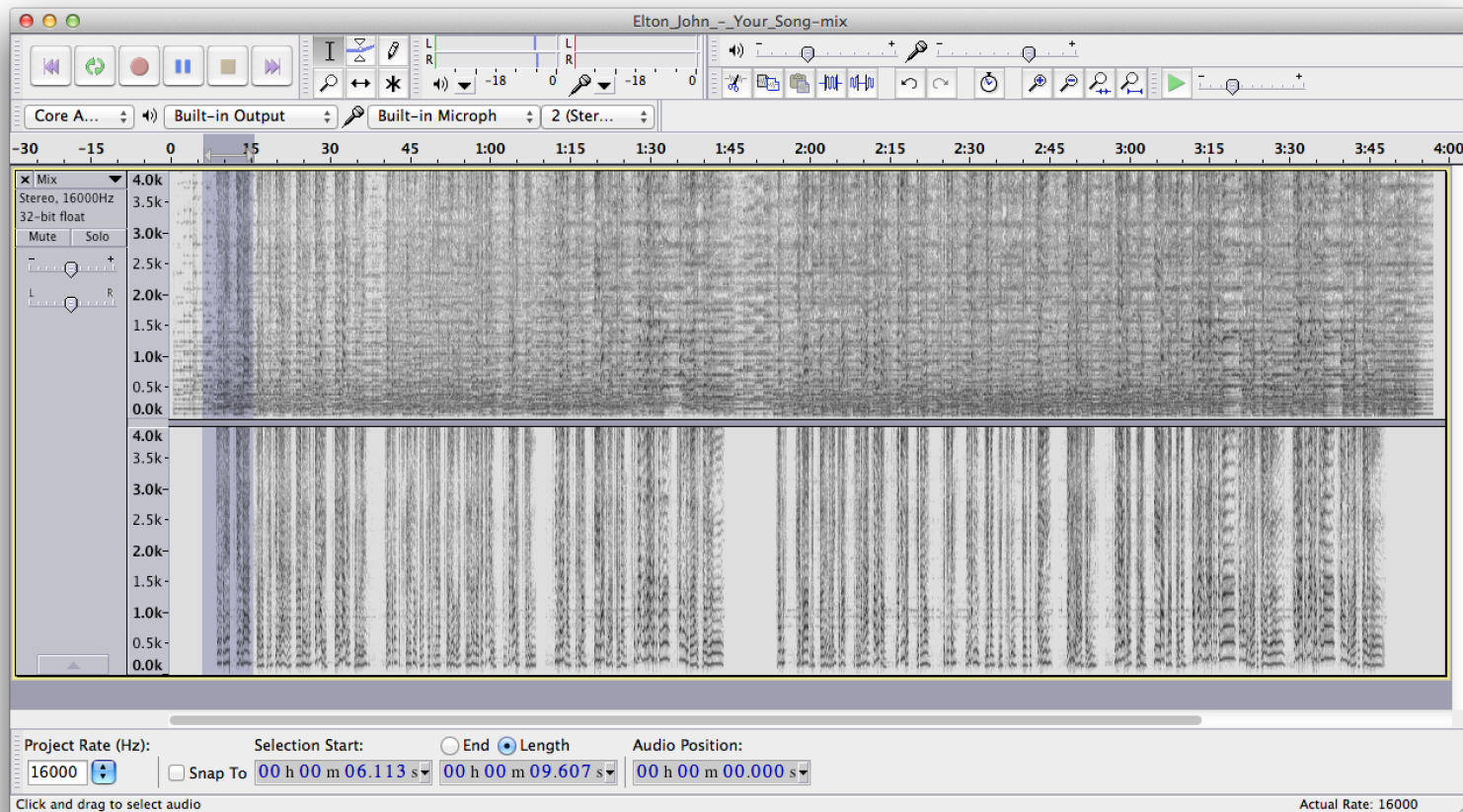
- **Speech Recognition for Songs**
 - lots of interference
 - atypical speech



Singing ASR

McVicar

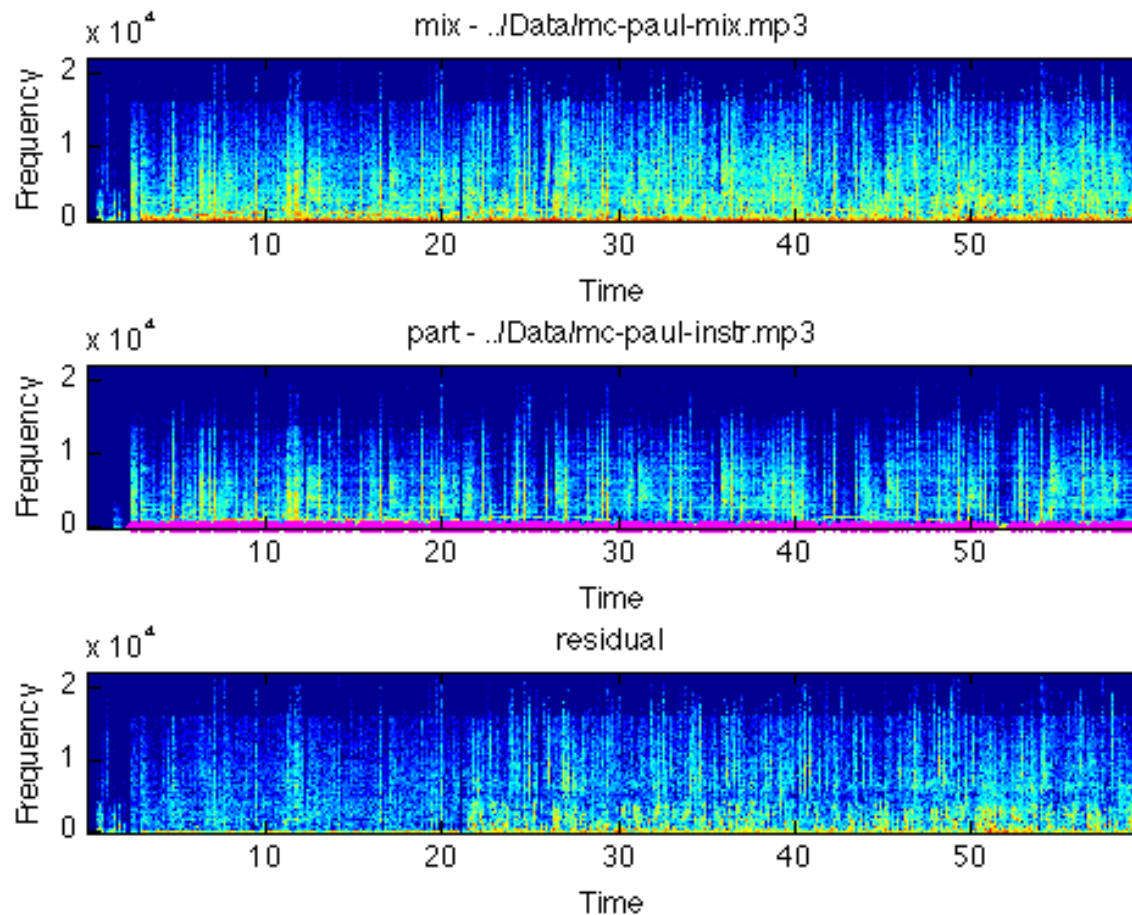
- Speech recognition adapted to singing
 - needs aligned data
- Align scraped “acapellas” and full mix
 - including jumps!



“Remixavier”

Raffel

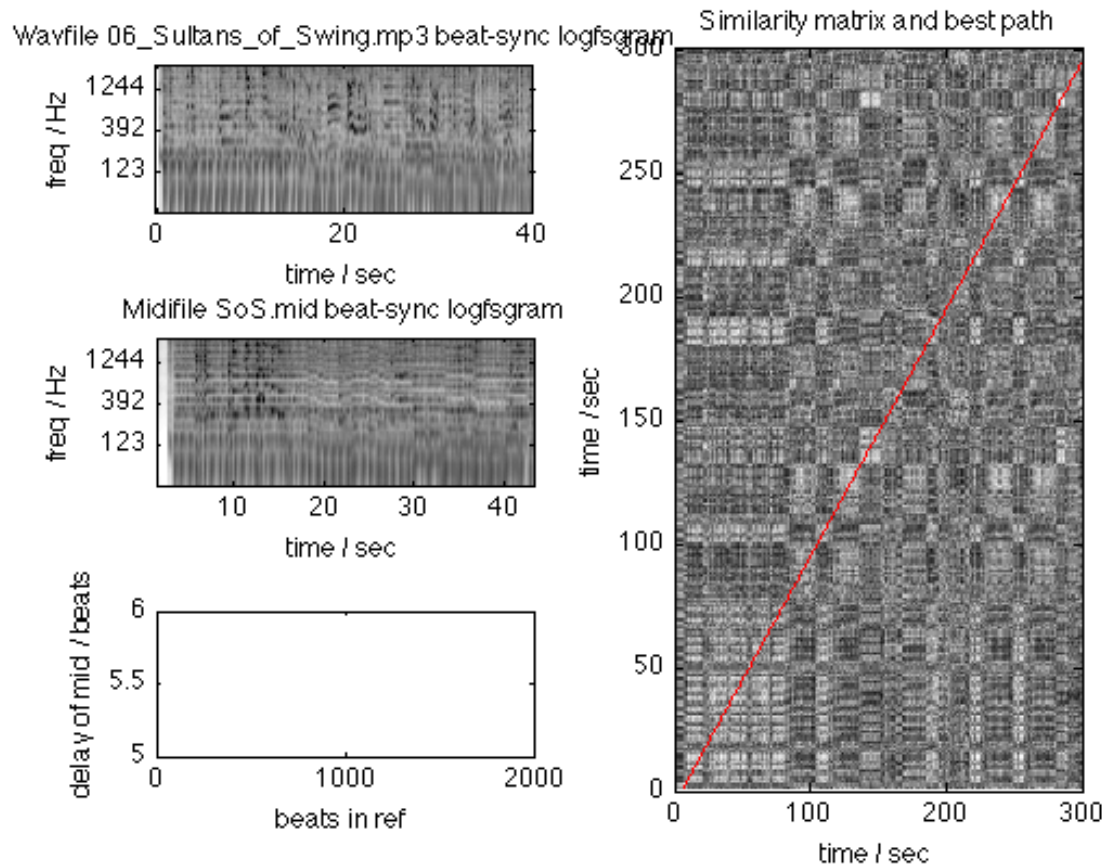
- Optimal align-and-cancel of mix and acapella
 - timing and channel may differ



MIDI-to-MSD

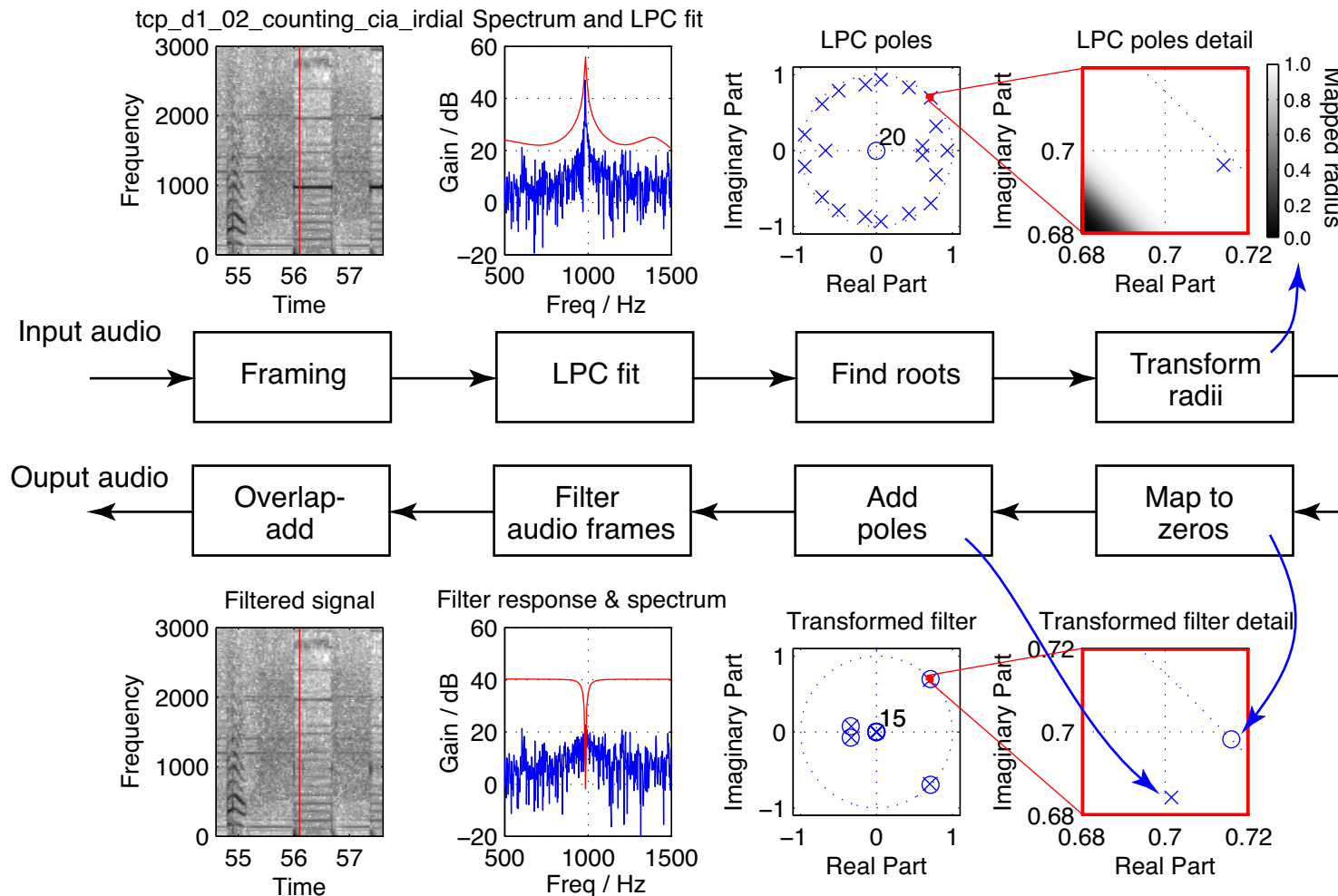
Raffel
Chi

- Aligned MIDI to Audio is a nice transcription



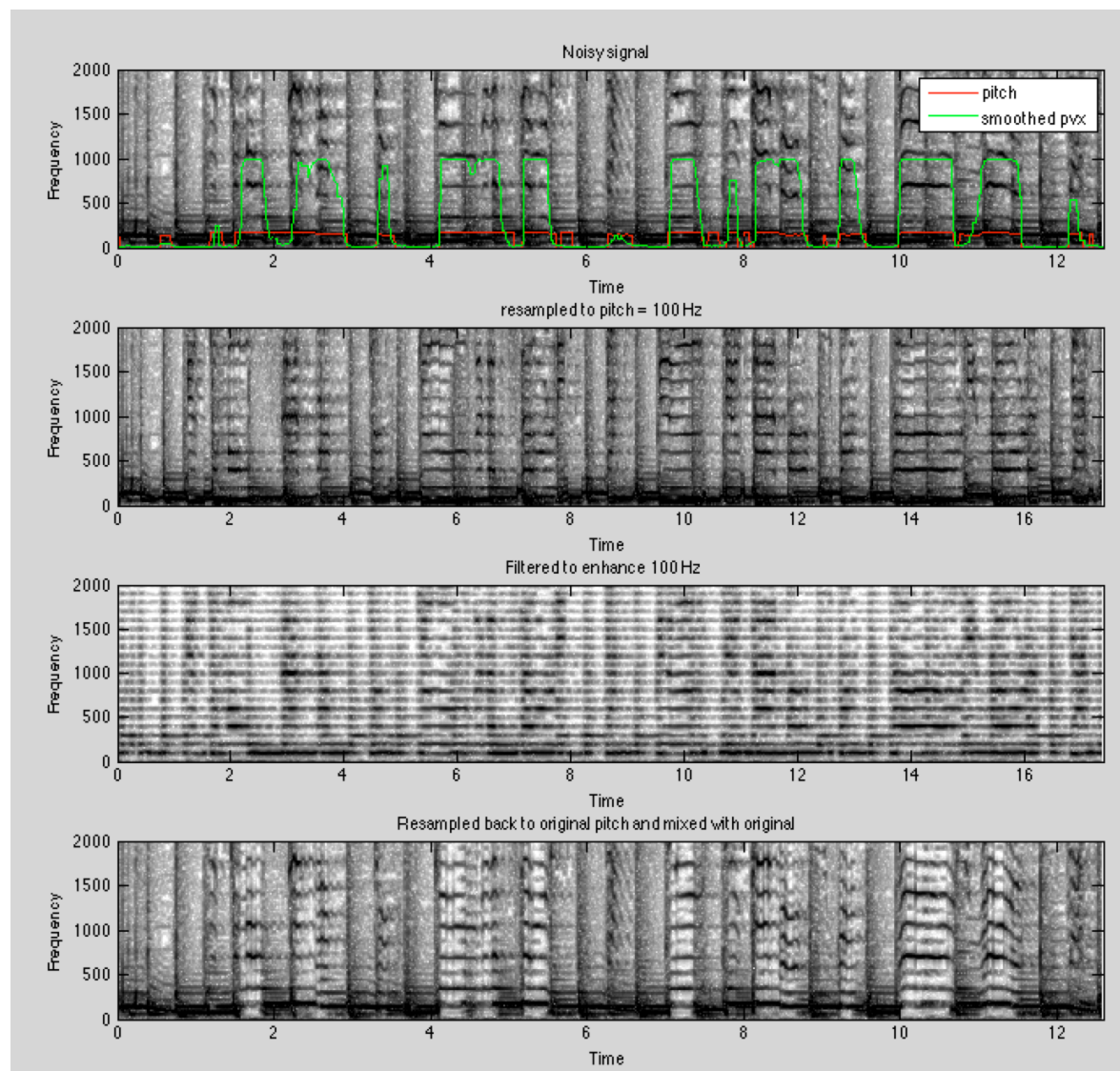
De-DTMF

- Problem:
 - Stationary tones confuse speech detector
 - Adaptively filter sinusoids with steady amplitude



Pitch-based Filtering

- Resample to flatten pitch, then filter



Summary

- **Signal Separation**
 - NMF, RPCA, cancellation, filtering

- **Music Information**
 - Beat tracking, segmentation
 - Large datasets
 - Indexing & retrieval

- **Speech**
 - Lyric recognition
 - Speech detection & enhancement

References

[Bello 2011] J P Bello, “Measuring structural similarity in music”, IEEE Tr.Audio, Speech, & Lang., 19(7): 2013-2025, 2011.

[Serra et al. 2012] J Serrà, A Corral, M Boguña, M. Haro, & J. Arcos, “Measuring the evolution of contemporary western popular music”, Scientific Reports, 2:521, 2012.