

Mining Large-Scale Music Data Sets

Dan Ellis & Thierry Bertin-Mahieux

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

{dpwe,thierry}@ee.columbia.edu

<http://labrosa.ee.columbia.edu/>

1. Music Audio Representations
2. The Million Song Dataset
3. Finding Similar Items (Cover Songs)
4. Current Work



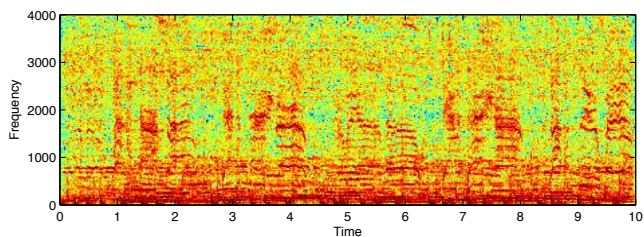
Laboratory for the Recognition and Organization of Speech and Audio



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

The Problem

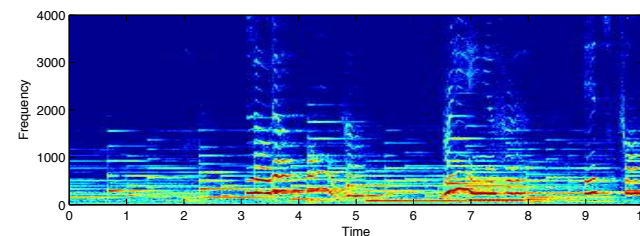
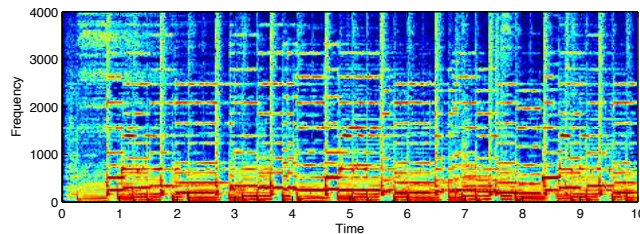
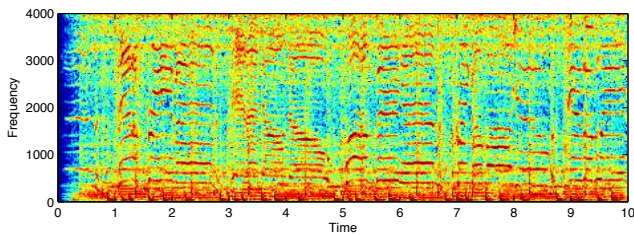
Query track



One Million Songs



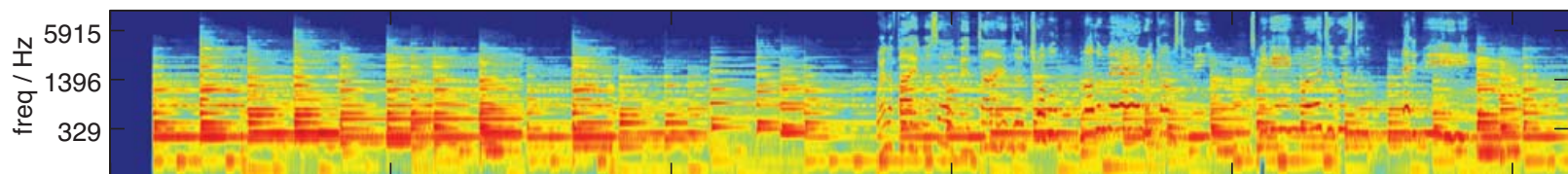
“Similar” tracks



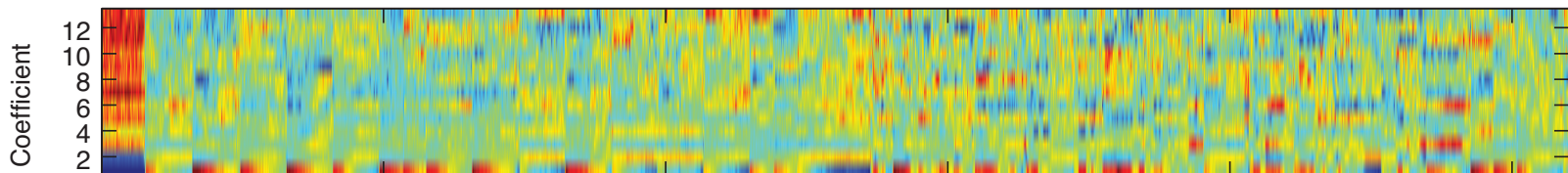
I. Music Audio Representations

- We need a description of the audio that contains “suitable” detail
- Speech Recognition uses **Mel-Frequency Cepstral Coefficients (MFCCs)**

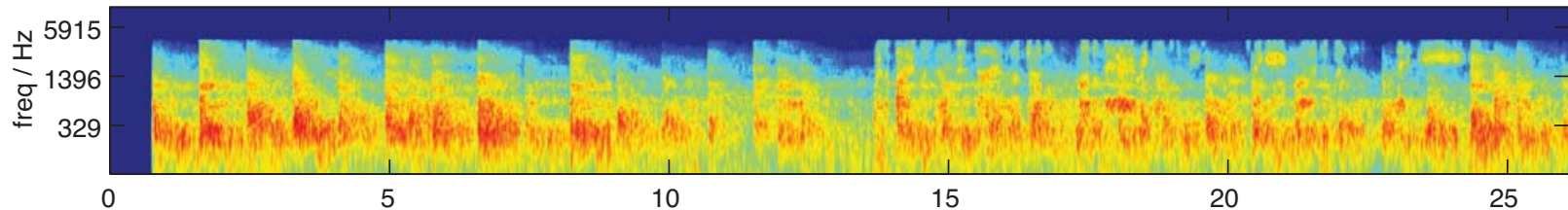
Let It Be (LIB-1) - log-freq specgram



MFCCs

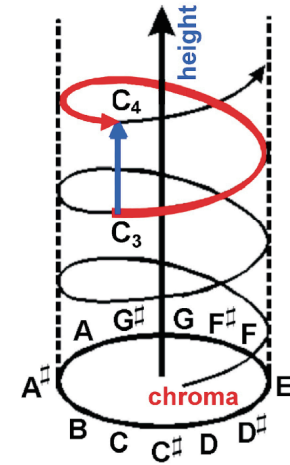


Noise excited MFCC resynthesis (LIB-2)



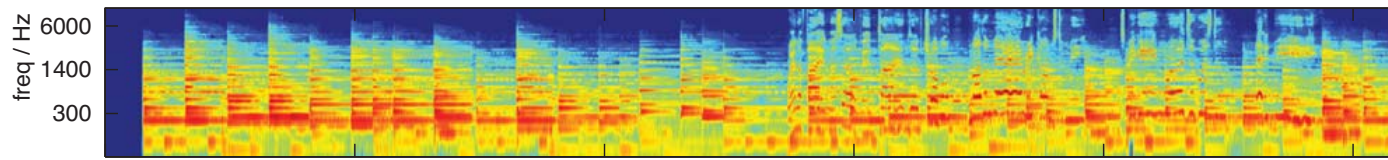
Chroma Features

- We'd like to preserve the notes
 - at least within one octave

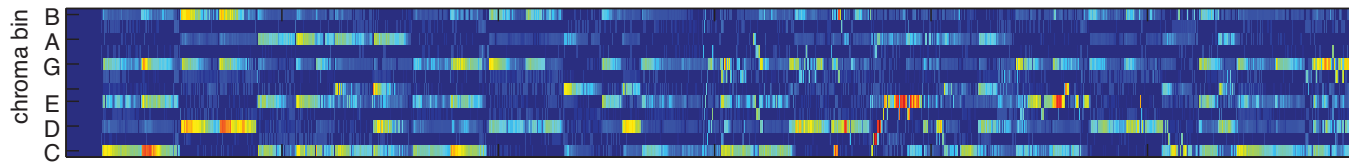


Warren et al. 2003

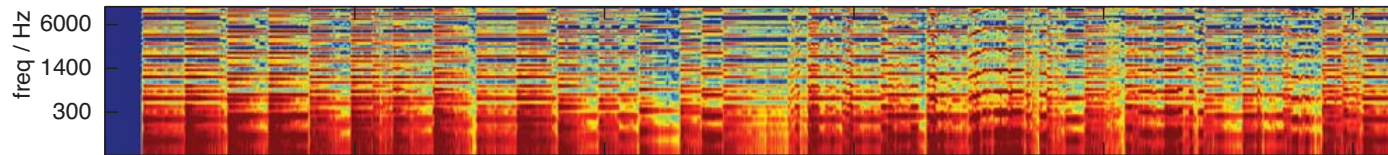
Let It Be - log-freq specgram (LIB-1)



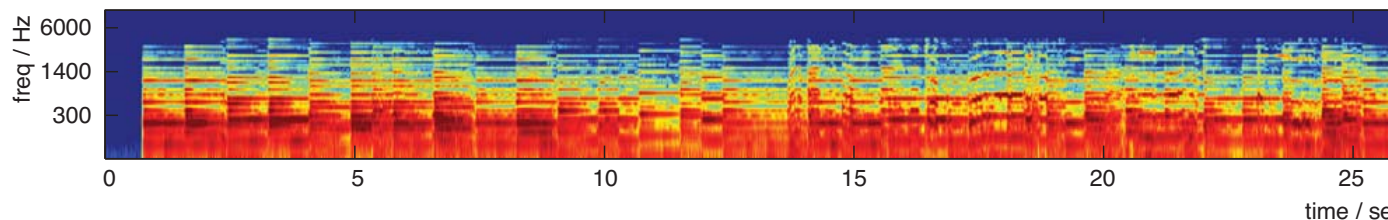
Chroma features



Shepard tone resynthesis of chroma (LIB-3)



MFCC-filtered shepard tones (LIB-4)



2. Million Song Dataset (MSD)

Thierry Bertin-Mahieux

- Commercial-scale dataset available to MIR researchers
 - 1M pop songs
 - 250 GB of features
 - (6 years of listening)



- Many facets:
 - Features,
 - Lyrics,
 - Tags,
 - Covers,
 - Listeners ...

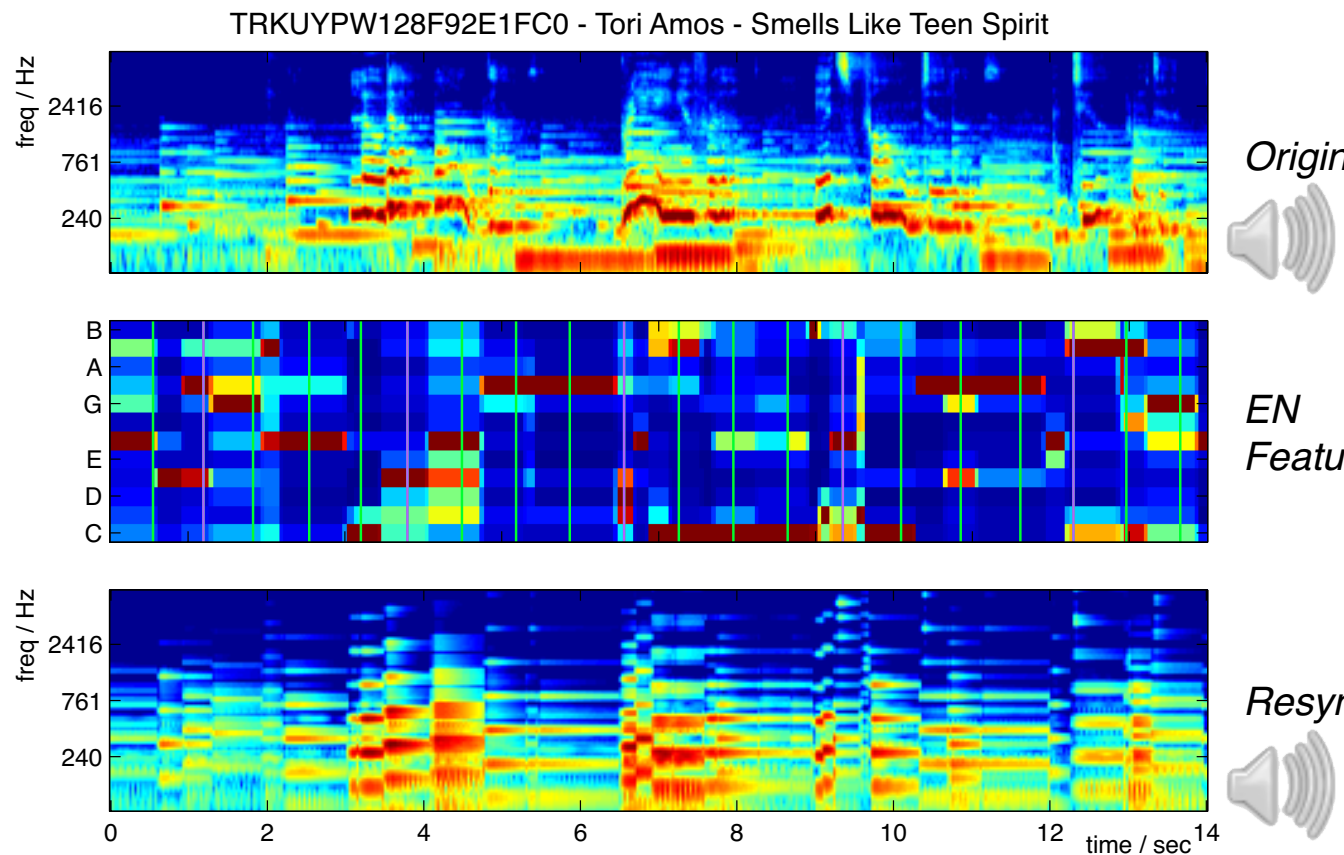
<http://labrosa.ee.columbia.edu/millionsong>

MSD Audio Features

- Use **Echo Nest** “Analyze” features
 - segment audio into variable-length “events”

- represent by 12 chroma + 12 “timbre”

- supports a crude resynthesis:



MSD Metadata

EN Metadata

```
artist: 'Tori Amos'  
release: 'LIVE AT MONTREUX'  
title: 'Smells Like Teen Spirit'  
id: 'TRKUYPW128F92E1FC0'  
key: 5  
mode: 0  
loudness: -16.6780  
tempo: 87.2330  
time_signature: 4  
duration: 216.4502  
sample_rate: 22050  
audio_md5: '8'  
7digitalid: 5764727  
familiarity: 0.8500  
year: 1992
```

Last.fm Tags

100.0 – cover	5.0 – cover songs
57.0 – covers	4.0 – soft rock
43.0 – female vocalists	4.0 – nirvana cover
42.0 – piano	4.0 – Mellow
34.0 – alternative	4.0 – alternative rock
14.0 – singer-songwriter	3.0 – chick rock
11.0 – acoustic	3.0 – Ballad
8.0 – tori amos	3.0 – Awesome Covers
7.0 – beautiful	2.0 – melancholic
6.0 – rock	2.0 – k00l chlX
6.0 – pop	2.0 – indie
6.0 – Nirvana	2.0 – female vocalistist
6.0 – female vocalist	2.0 – female
6.0 – 90s	2.0 – cover song
5.0 – out of genre covers	2.0 – american

SHS Covers

```
%5489,4468, Smells Like Teen Spirit  
TRTUOVJ128E078EE10 Nirvana  
TRFZJOZ128F4263BE3 Weird Al Yankovic  
TRJHCKN12903CDD274 Pleasure Beach  
TRELTOJ128F42748B7 The Flying Pickets  
TRJKBXL128F92F994D Rhythms Del Mundo feat. Shanade  
TRIHLOW128F429BBF8 The Bad Plus  
TRKUYPW128F92E1FC0 Tori Amos
```

MxM Lyric Bag-of-Words

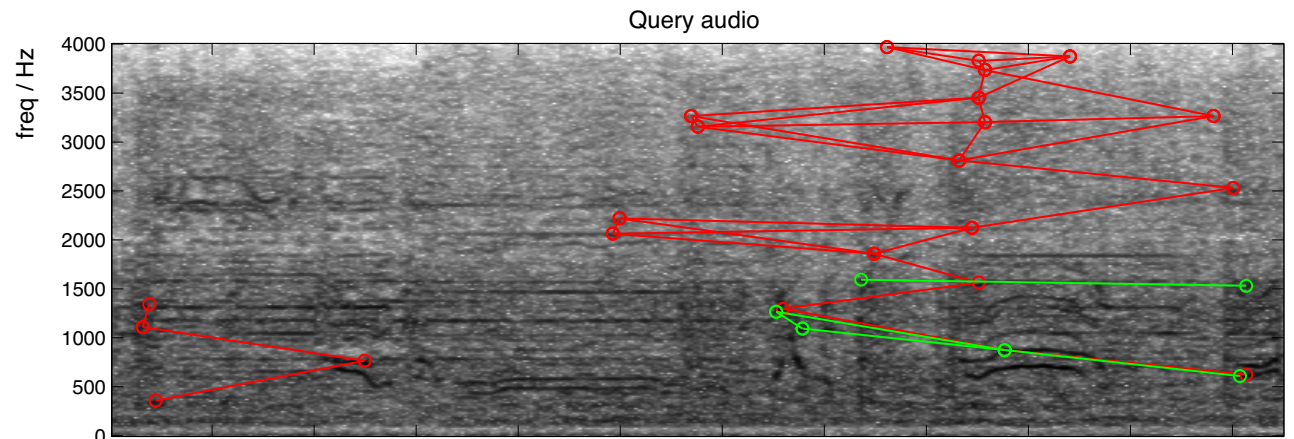
12 hello	6 here	3 is
11 i	6 us	3 with
10 a	6 entertain	3 oh
9 and	4 the	3 out
7 it	4 feel	3 an
6 are	4 yeah	3 light
6 we	3 to	3 less
6 now	3 my	3 danger

3. Finding Similar Items

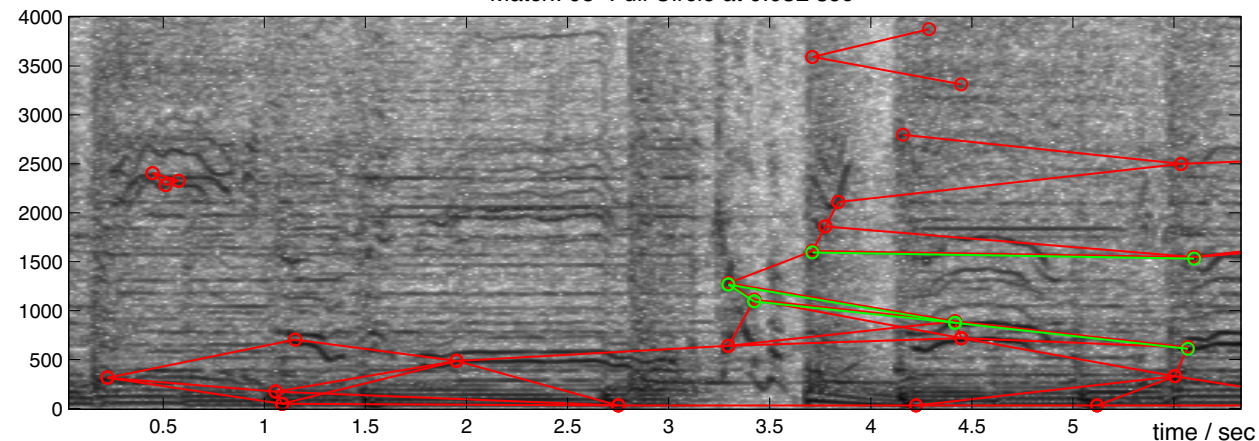
Avery Wang '03

- If it's really the same, we can use **fingerprinting** (e.g. Shazam):

- find local “landmarks”
- quantize by pairs
- inverted index



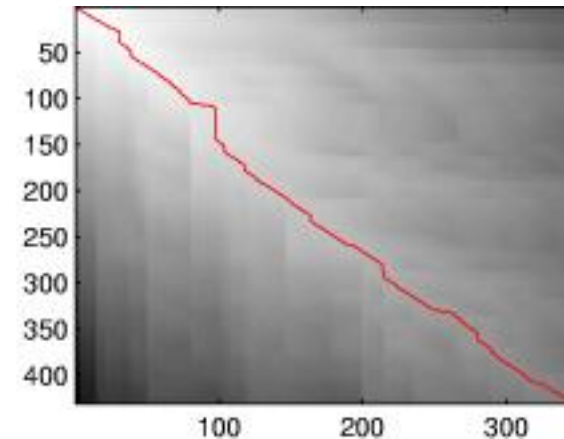
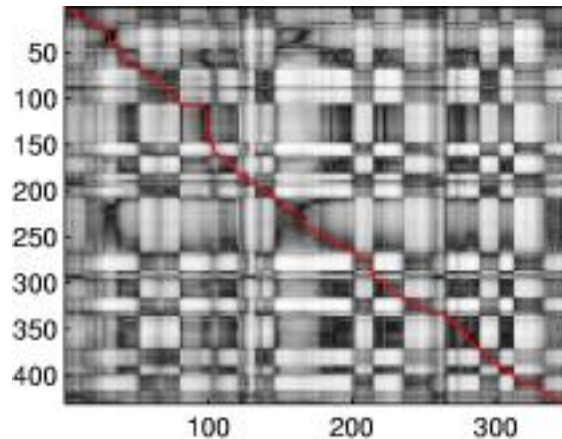
Match: 05-Full Circle at 0.032 sec



Dynamic Programming Alignment

Serra, Gomez et al. '08

- We can match the chroma representations by **DP alignment**

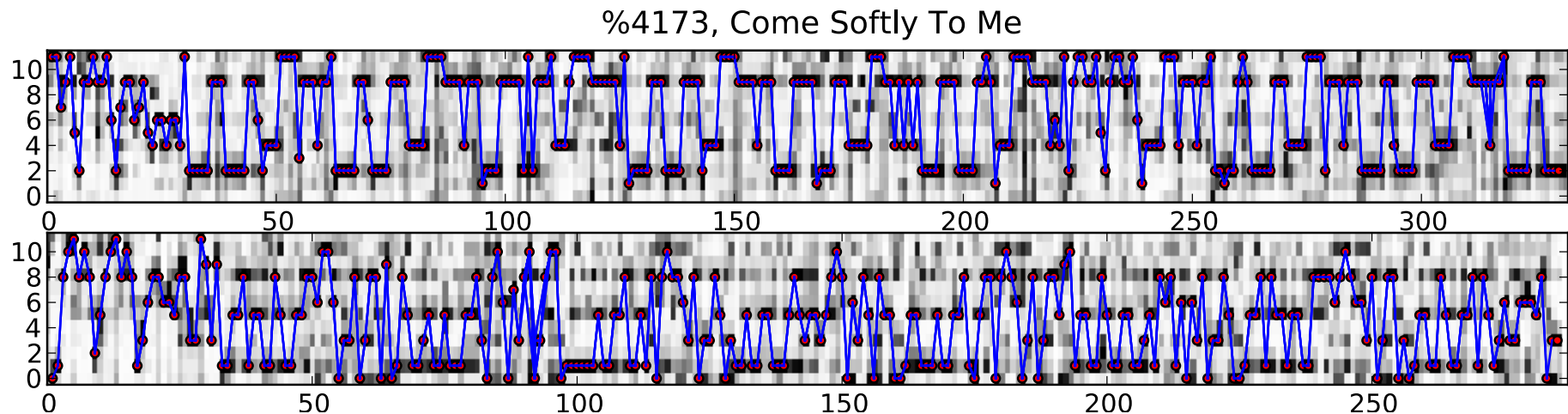


- robust to timing changes
- quite efficient
- **best performer** in MIREX Cover Song evaluations

Large-Scale Cover Matching

Bertin-Mahieux & Ellis '11

- How can we find covers in **1M songs?**
 - @ 1 sec / comparison, one search = 11.5 CPU-days
 - full N^2 mining = 16,000 CPU-years
- **Hashing?**
 - landmarks from chroma patches (like Shazam)

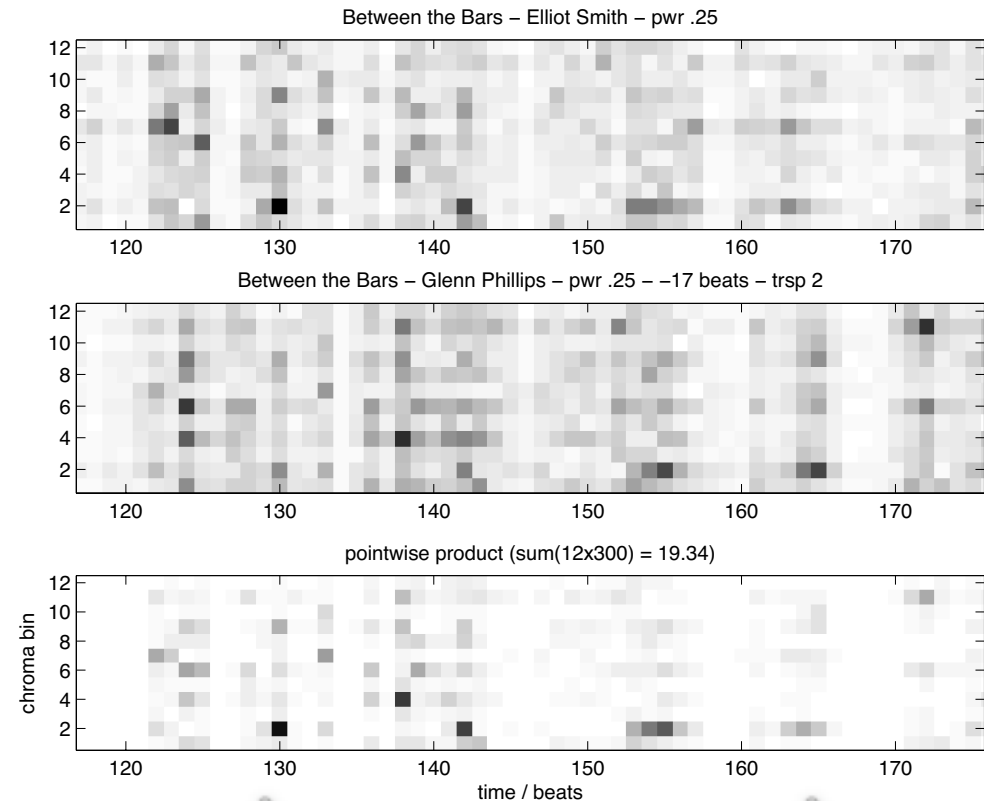


- represent item by distribution of contour fragments
- one stage in a filtering hierarchy...

Euclidean Cover Matching

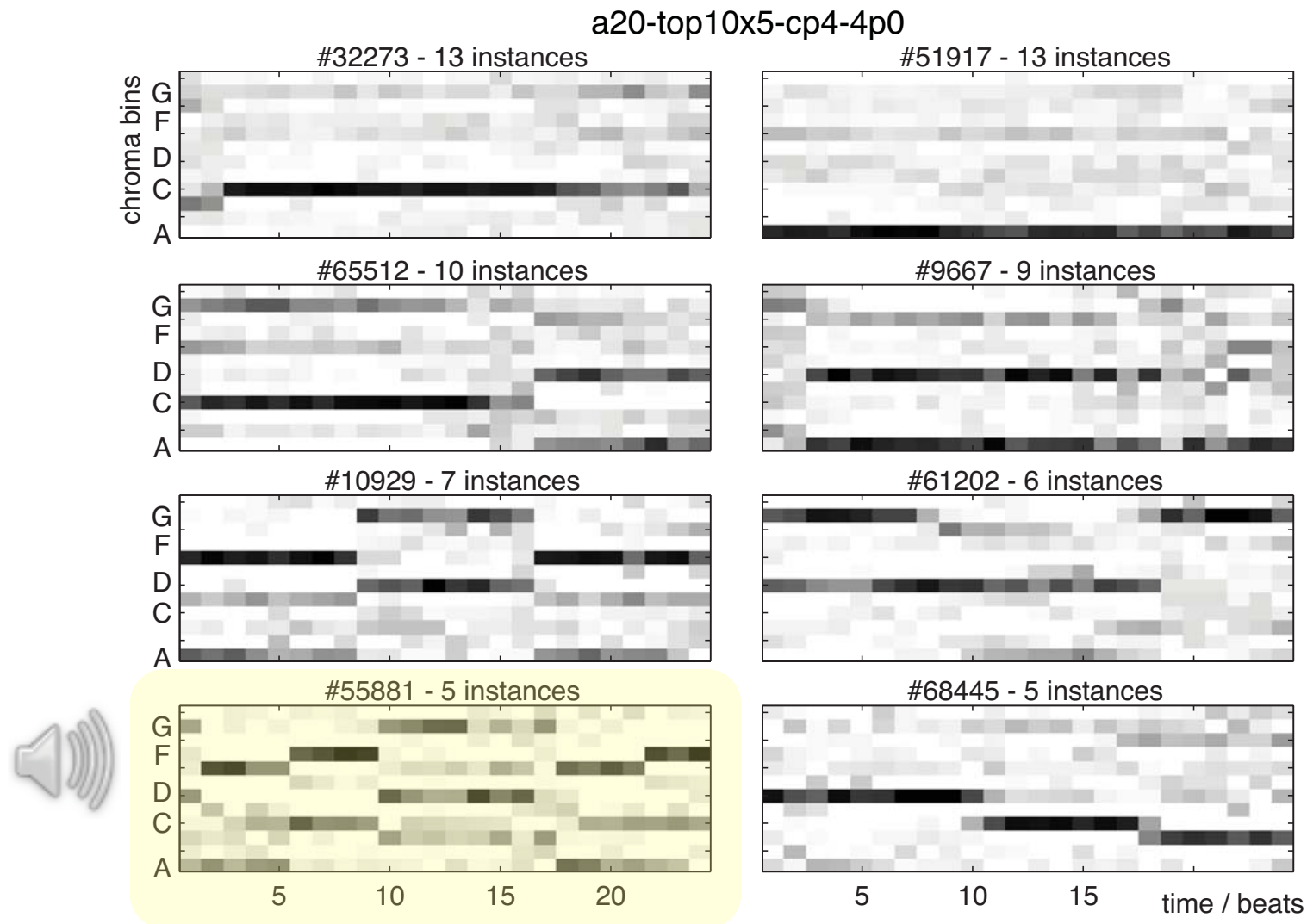
Ellis & Poliner '08
Bertin-Mahieux & Ellis '12

- Reduce cover matching to **nearest-neighbor search** in fixed-size beat-chroma patches
- Right “**distance**”?
 - principal components
 - learned weightings
- **Alignment/segmentation?**
 - music segmentation
 - multiple probes
 - **framing-insensitive** representation



4. Other Work: Pattern Mining

- **Cluster** beat-synchronous chroma **patches**



Other Work: MSD Challenge

with Brian McFee

- Using “Taste Profile” data
 - User-Song-playcount

b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAPDEY12A81C210A9	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2
b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOCNMUH12A6D4F6E6D	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODDNQT12A6D4F5F7E	5

- Challenge:
 - Rank IM songs to **complete half a playlist**
 - using audio, metadata, lyrics, whatever
- Kaggle.com based contest
 - self-service evaluation, leader board
- Results at MIREX, AdMIRe

Summary

- Finding Musical Similarity at Large Scale
- Beat Chroma Representation
- Million Song Dataset
- MSD Challenge