Lecture 12: Alignment and Matching

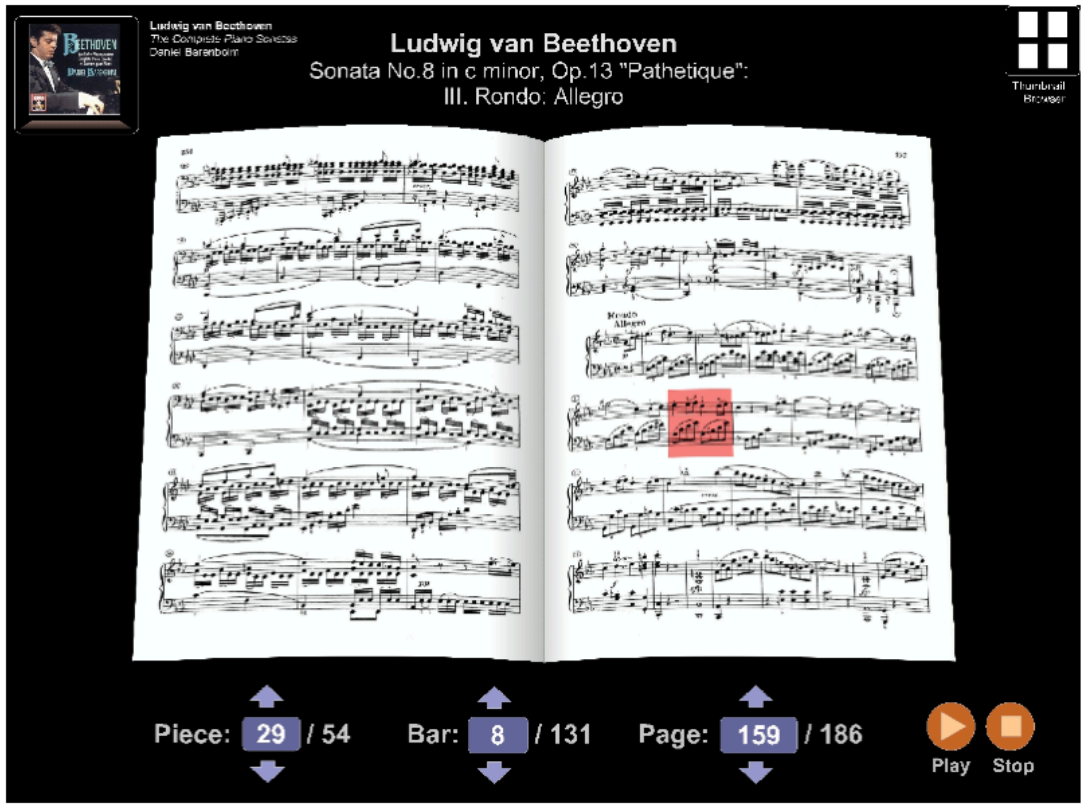
1. Music Alignment
2. Cover Song Detection
3. Echo Nest Analyze

Dan Ellis
Dept. Electrical Engineering, Columbia University
dpwe@ee.columbia.edu http://www.ee.columbia.edu/~dpwe/e4896/

1. Music Alignment

Kurth et al., 2007

- Often have versions of the same music with unmatched time axes
  - different performances
  - performance vs. score



- Various applications for aligning them
  - synchronizing different tracks (with TSM)
  - synchronized score display
  - ground truth transcriptions

The Similarity Matrix

Foote 1999

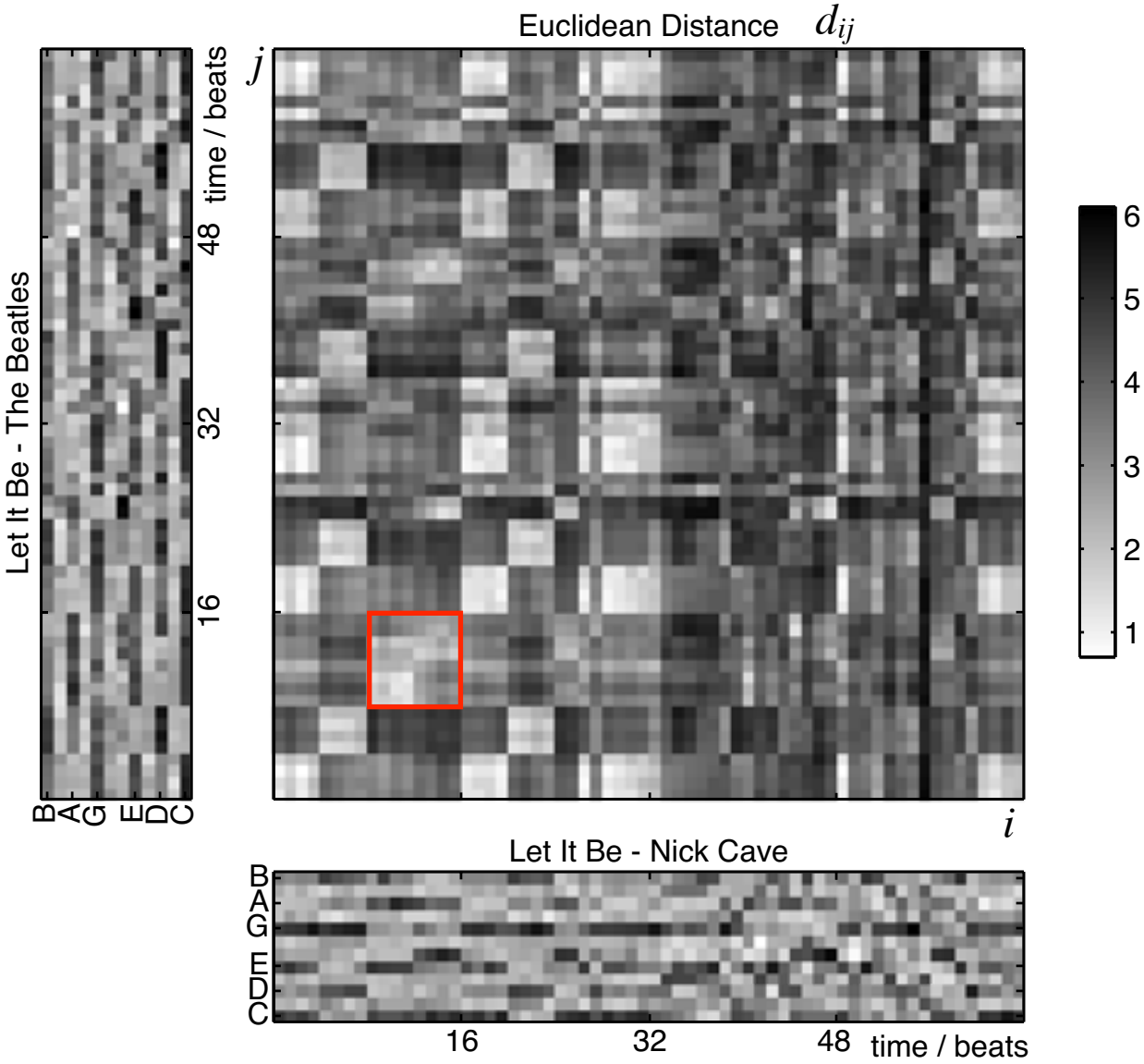
- Point-to-point comparison of sequences

- e.g. Euclidean distance

$$d_{euc}(i,j) = \sum_k |x_i(k) - y_j(k)|^2$$

- or normalized inner product (cosine distance)

$$d_{cos}(i,j) = 1 - \frac{\sum_k x_i(k) y_j(k)}{\sqrt{\sum_k |x_i(k)|^2 \sum_k |y_j(k)|^2}}$$

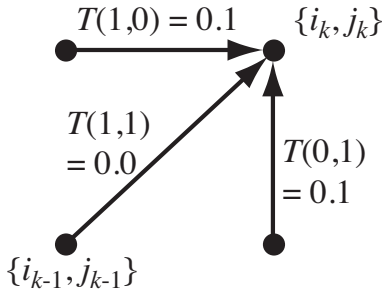


Dynamic Programming

Bellman 1957

- Find best path combining local + transitions
  - works for any kind of similarity matrix

- Allowable transitions

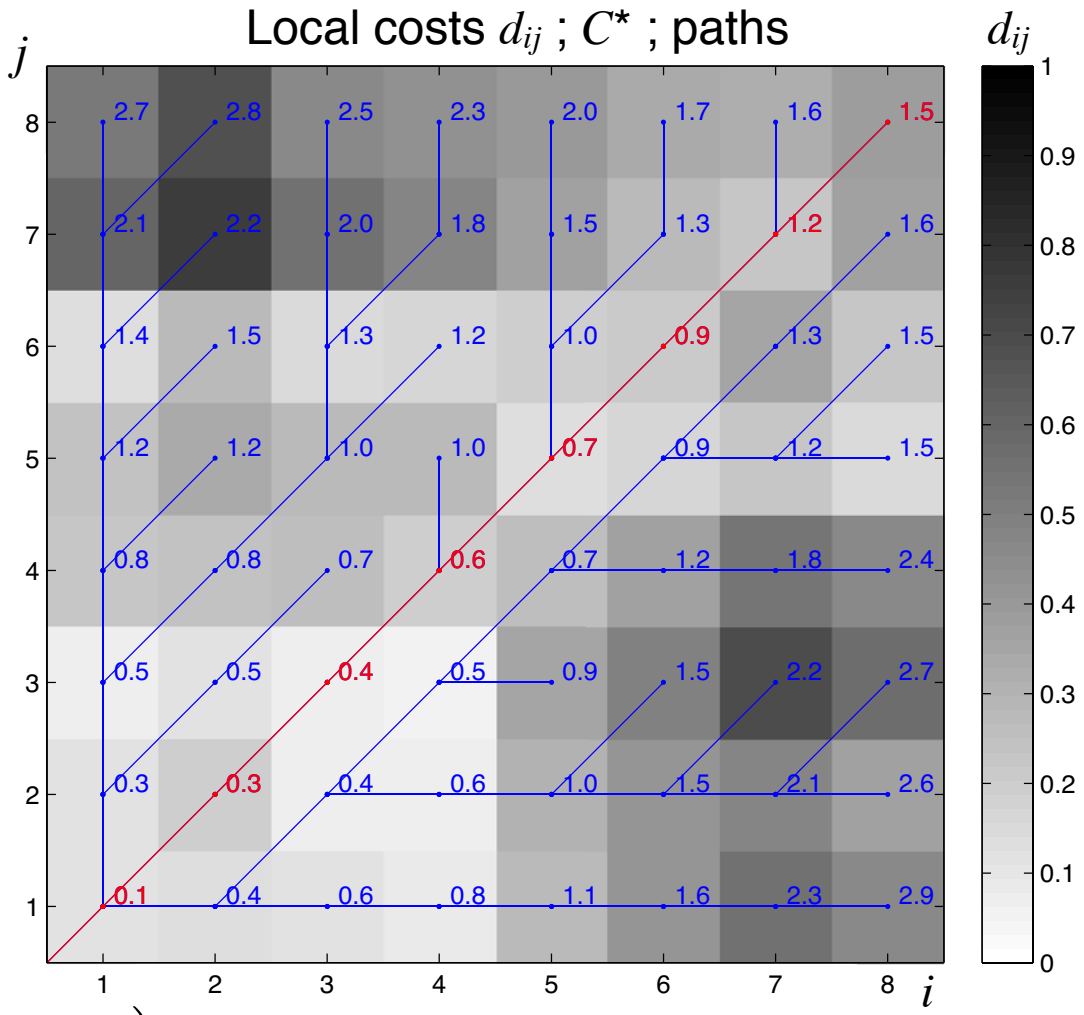


- Finds path $\{i_k, j_k\}$ to minimize cost ...

$$C^*_{i_{max},j_{max}} = \sum_k d(i_k, j_k) + T(i_k - i_{k-1}, j_k - j_{k-1})$$

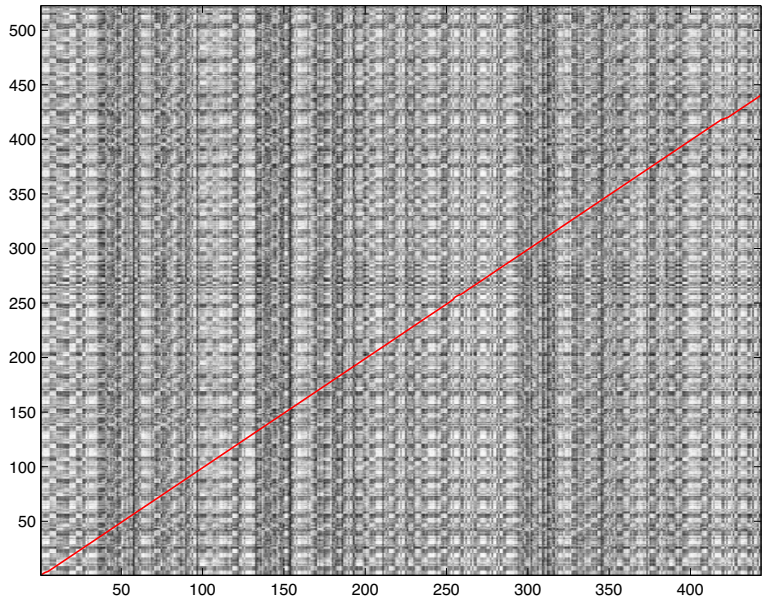
- ... recursively

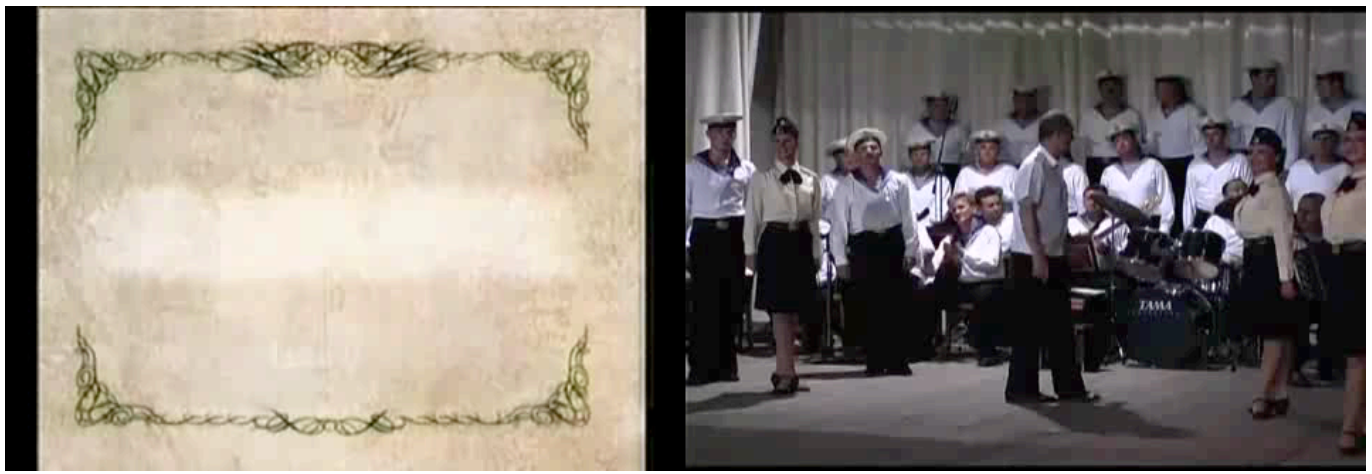
$$C^*_{i,j} = \min_{x,y=\{(1,1),(1,0),(0,1)\}} \left(d(i,j) + T(x,y) + C^*_{i-x,j-y} \right)$$



Audio-to-Audio Alignment

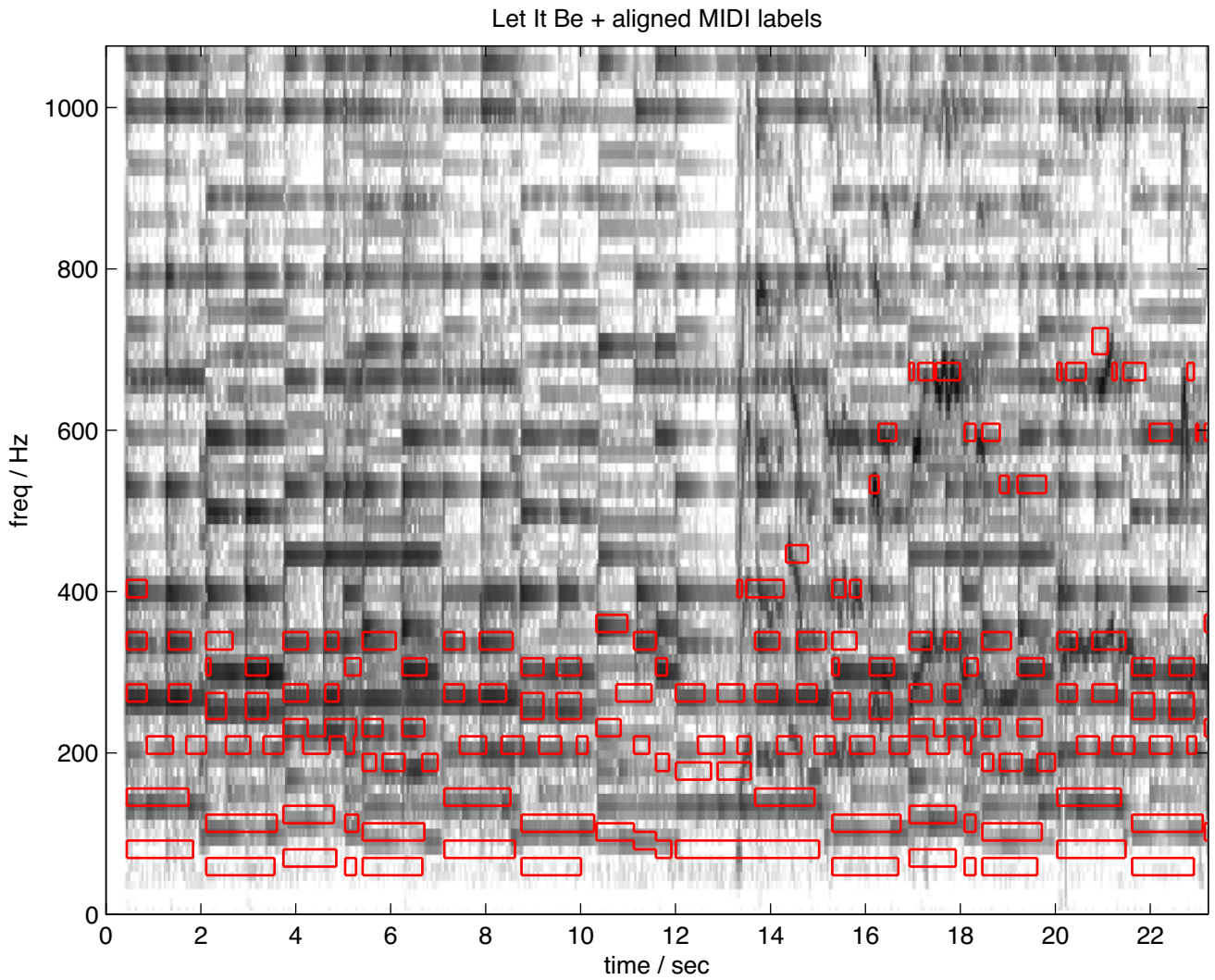
- Dynamic programming to get time mapping
  + phase vocoder time scaling





Audio-Score Alignment

- Aligning a score representation (e.g. MIDI) is a proxy for polyphonic transcription



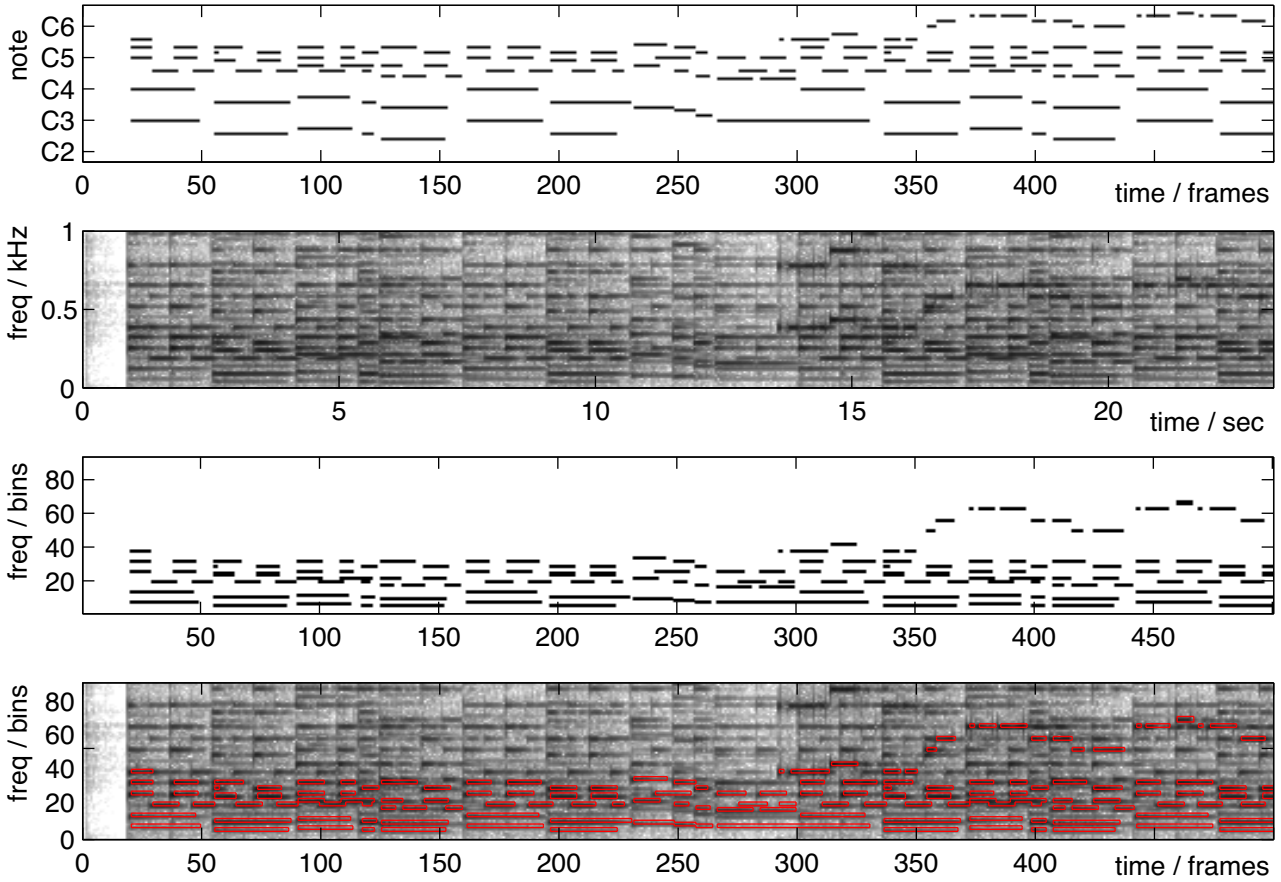
Peak Structure Distance

Orio & Schwartz 2001

- **How do we match spectra to score notes?**
  - synthesize audio from MIDI & compare audio?
  - "Peak Structure distance": is energy where we expect?

$$d_{psd} = 1 - \frac{\sum_k M[k]|X[k]|}{\sum_k |X[k]|}$$

- MIDI "Piano roll"
- Synthesized audio
- Predicted spectrum = mask $M[k]$
- "Peak Structure" = energy blw mask



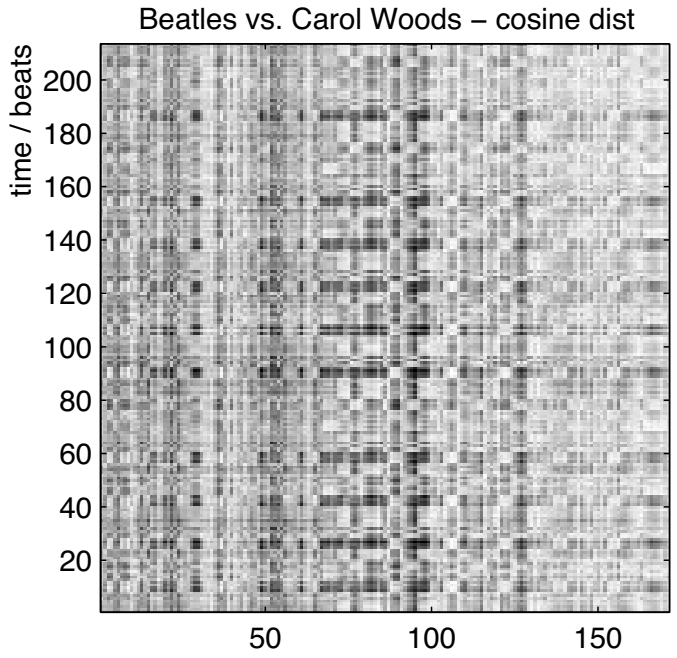
2. Cover Song Detection

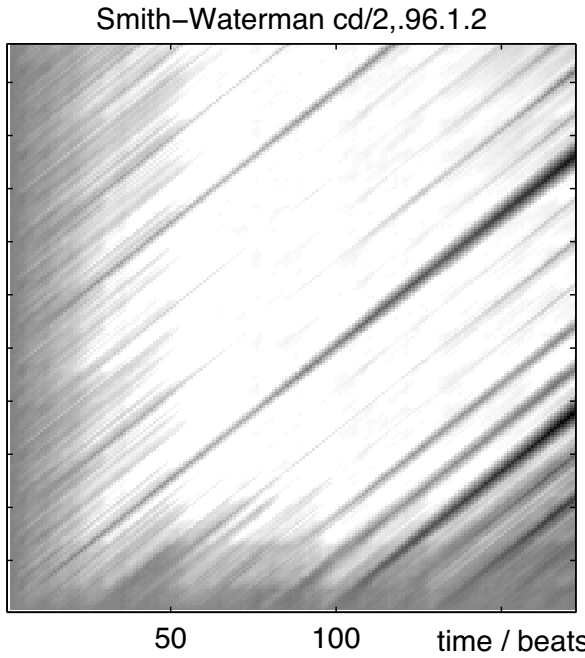
- Musicians are fond of 'cover versions'
  - usually alter melody, harmony, instrumentation, rhythm, style
  - can be hard to spot even for a human!


- Can try to match via alignment
  - .. with some threshold on best alignment cost?

Smith-Waterman Local Alignment

- Cover version may have different form
  - different number, ordering of verse/ chorus/brige
  - want to find any large aligned regions

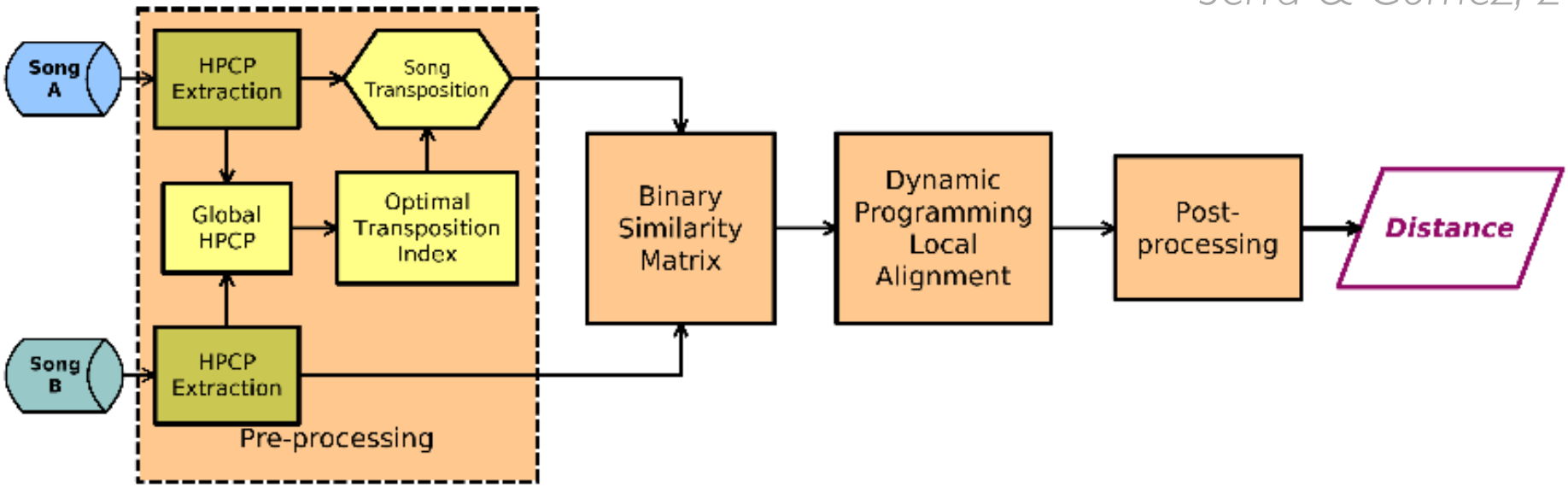




- "Local alignment" measure
  - $S^*_{i,j} = \max_{x,y}\left(\max\{0, s(i,j) - P(x,y) + S^*_{i-x,j-y}\}\right)$
  - want largest score S^*
  - similarity $s(i,j)$ must exceed penalty $P(x,y)$ on avg. (e.g. 0.96 for diagonal, 1.2 for off-diagonal)

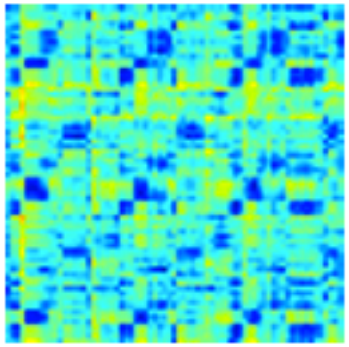
Local Alignment Cover Detection

Serrà & Gòmez, 2008



- Smith-Waterman needs predictable values
  - use binary similarity based on best transposition

Euclidean

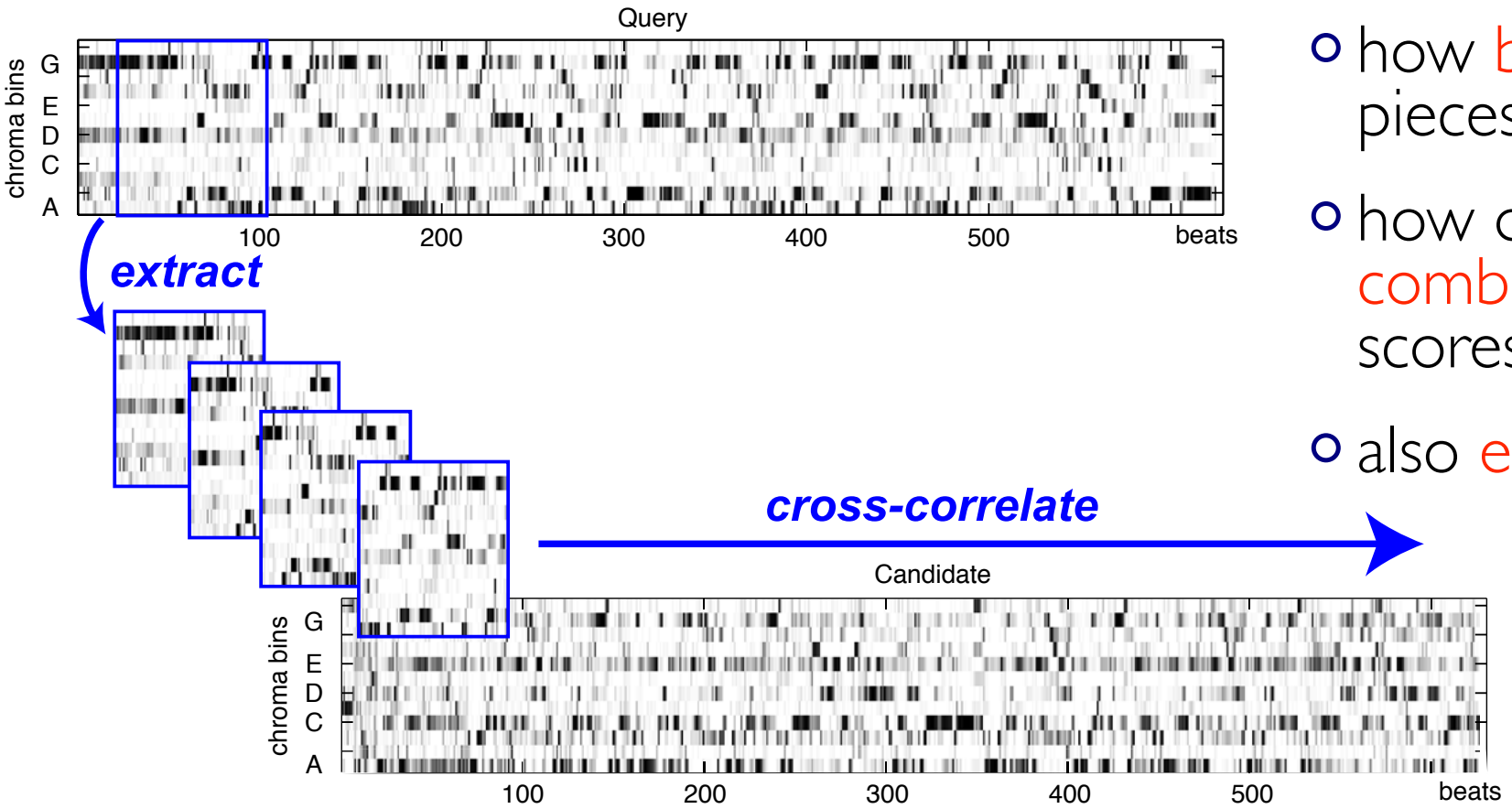


Binary

Non-cover

Cross-correlation Covers System

- DP is good for time-warping, but expensive
  - beat-timing is tempo independent (if it works)
  - simply cross-correlate beat-chroma patches?

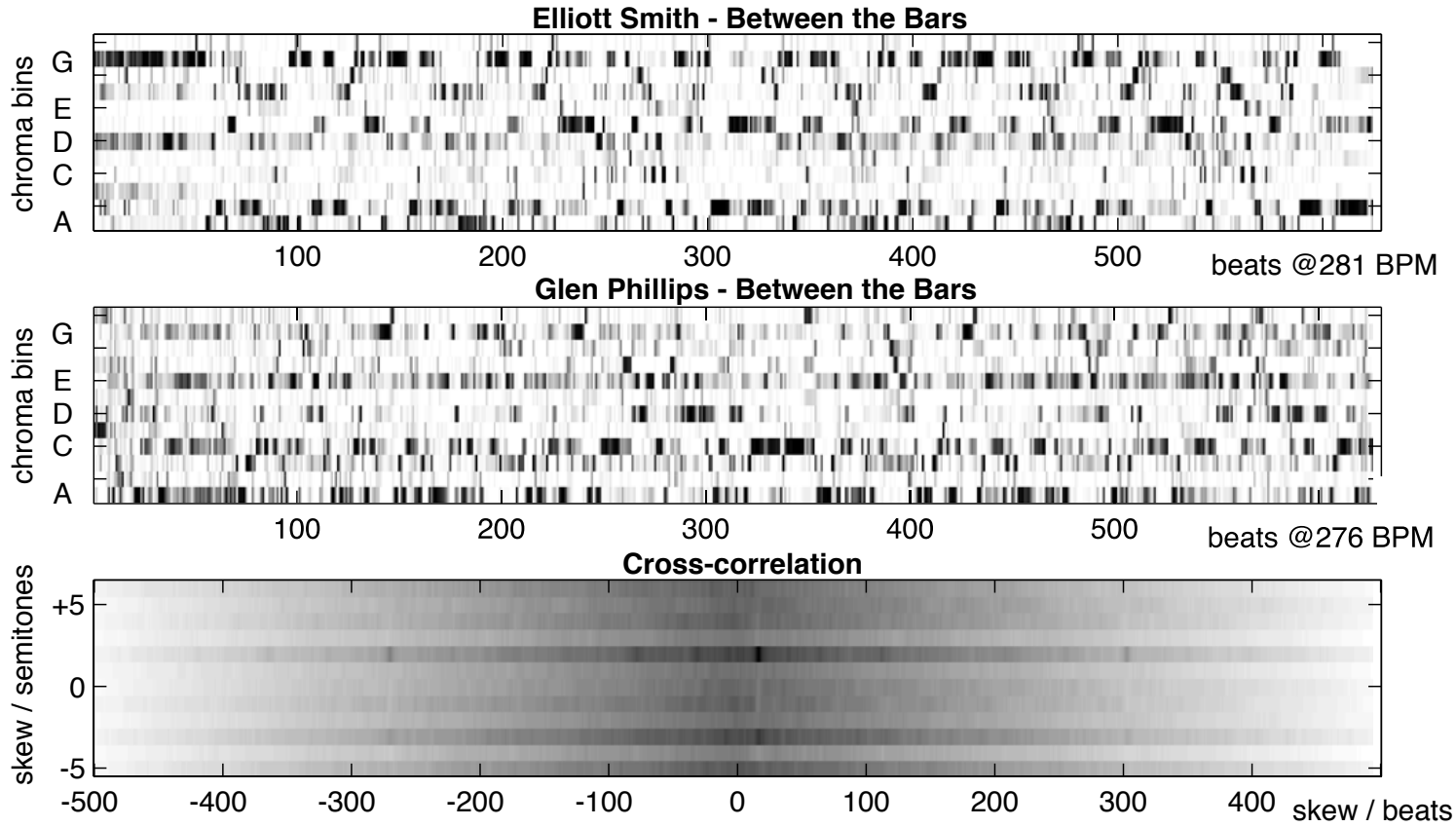


- how big are the pieces?
- how do we combine individual scores?
- also expensive

Global Cross-Correlation

Ellis & Poliner, 2007

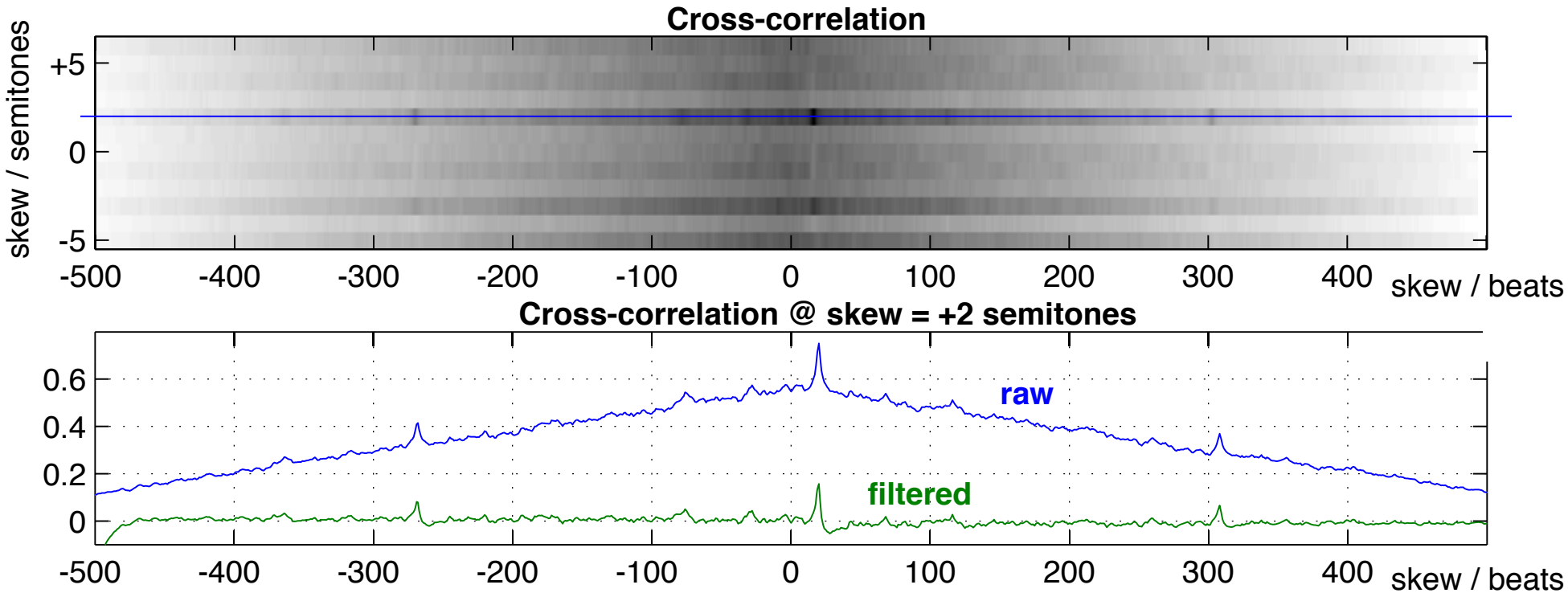
- Cross-correlate *entire* beat-chroma matrices
  - ... at all possible transpositions (circular)
  - implicit combination of match quality and duration



- One good matching fragment is sufficient...?

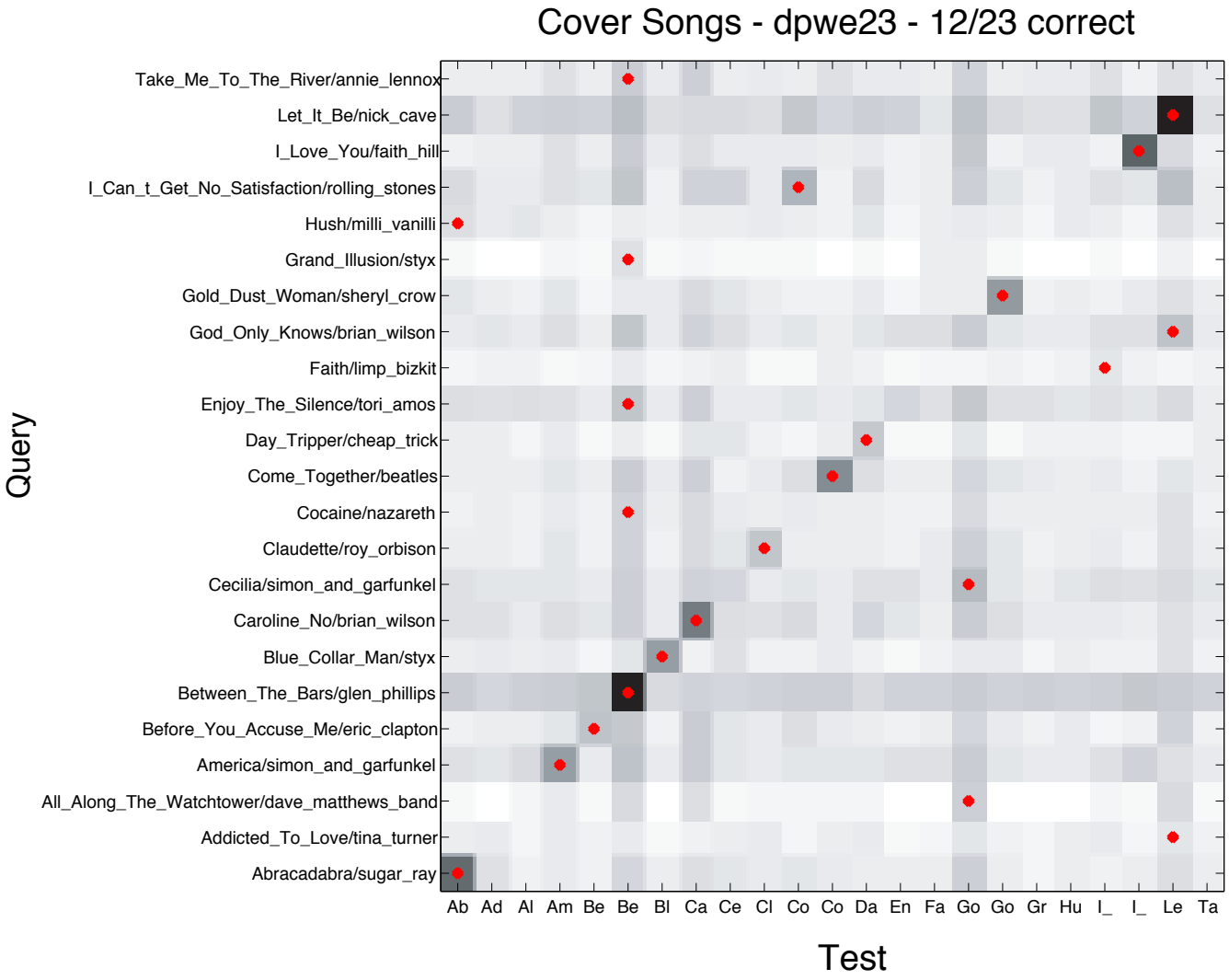
Filtered Cross-Correlation

- Raw correlation not as important as precise local match
  - looking for large contrast at ±1 beat skew
  - i.e. high-pass filter



Cover Song Results

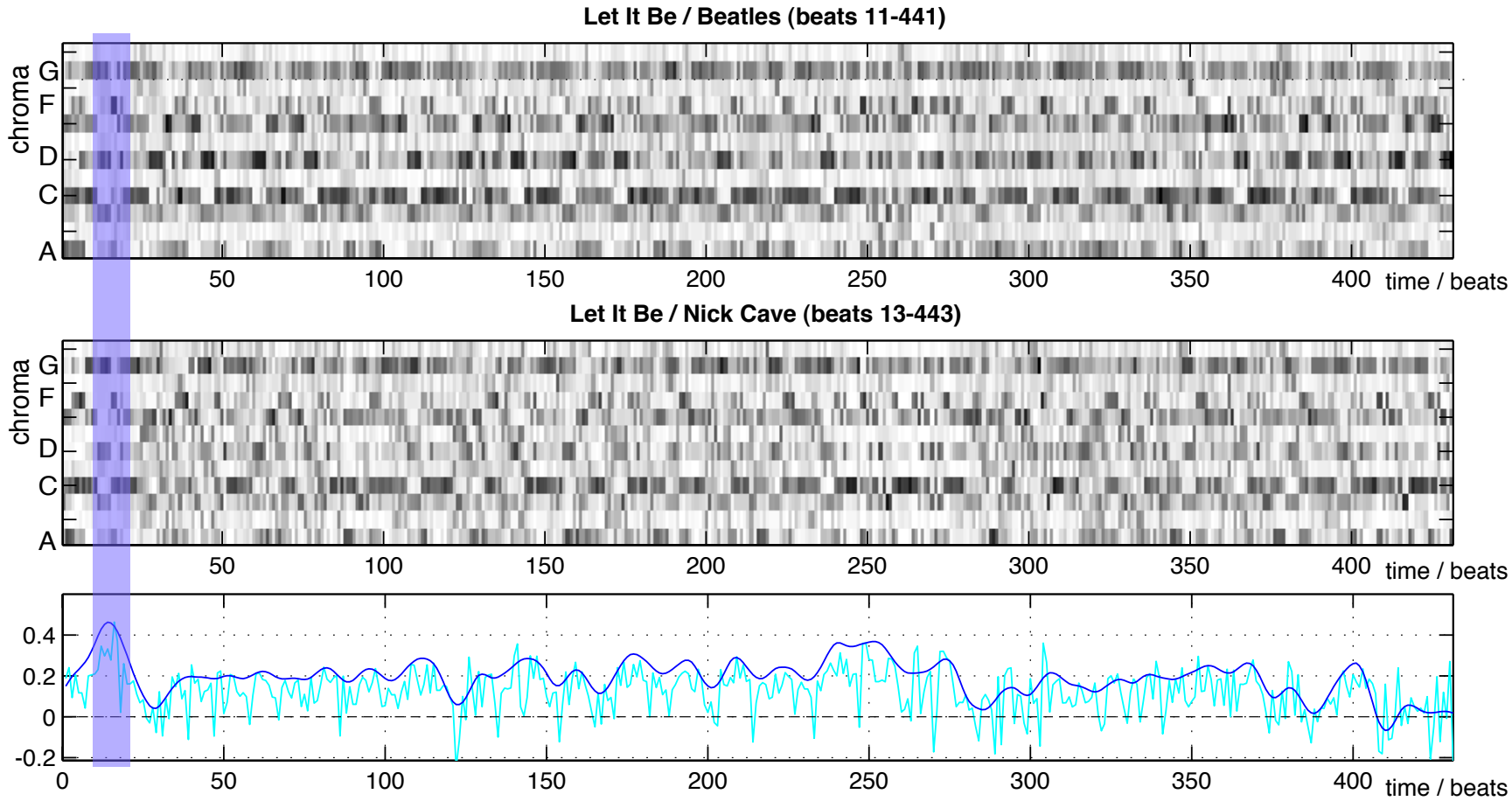
- 23 Covers found in 8700 song 'uspop2002'



- popular 'decoys' – normalization issues

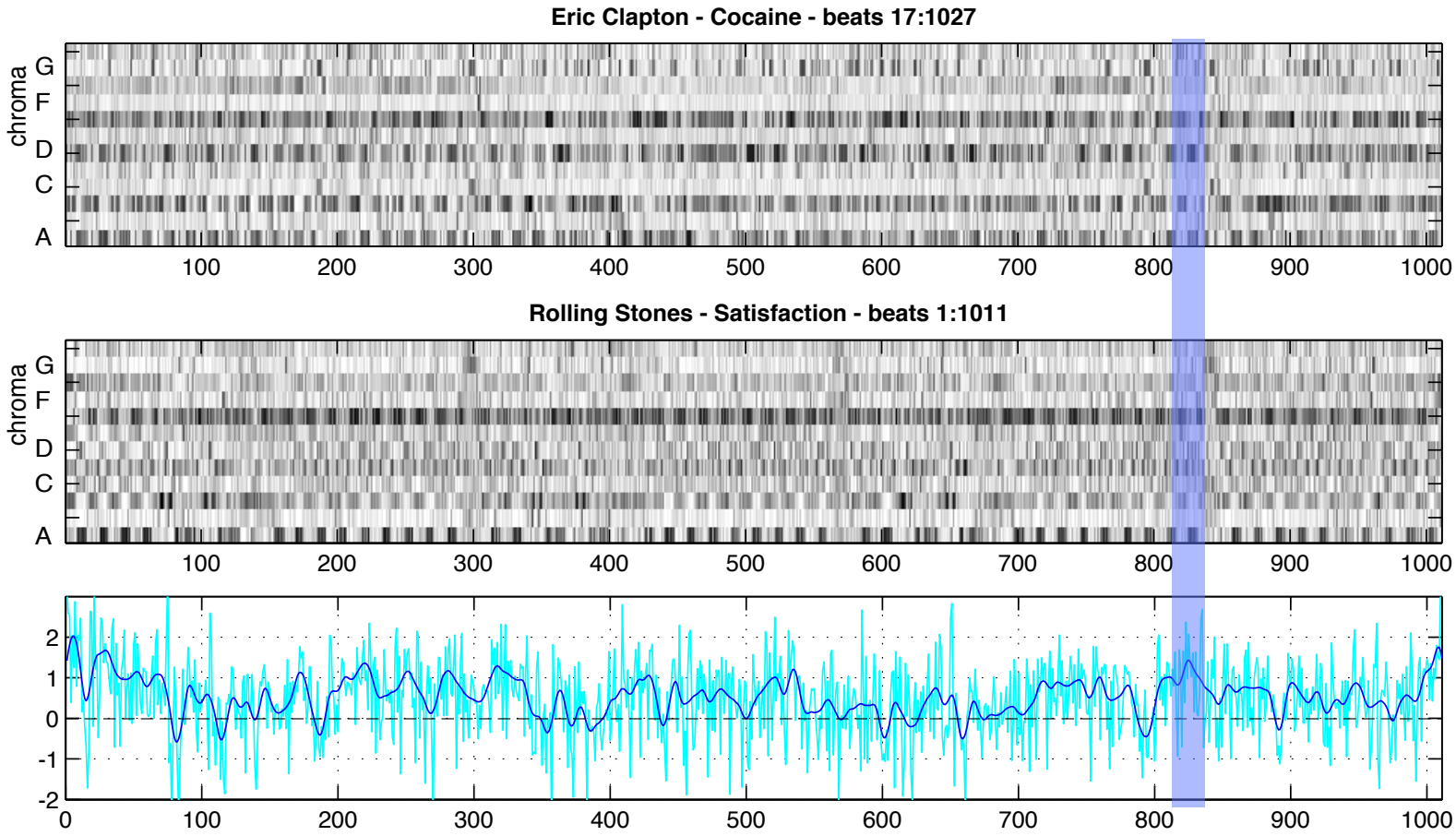
Analyzing Cover Song Correlation

- Look inside global cross-correlation to find matching fragments...
  - $\text{xcorr} = \Sigma_t \Sigma_f \left(C_1(t,f) \cdot C_2(t,f) \right)$ - view along time

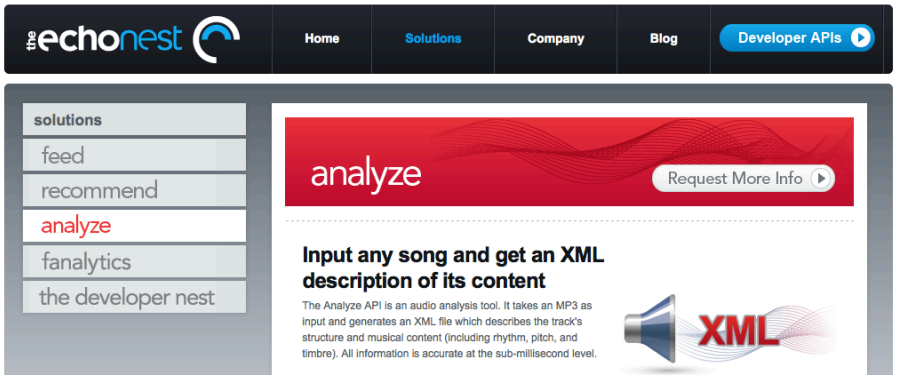


Cover Song False Alarm

- Correlation can be weak
  - "Cocaine" (Clapton) vs. "Satisfaction" (Stones)

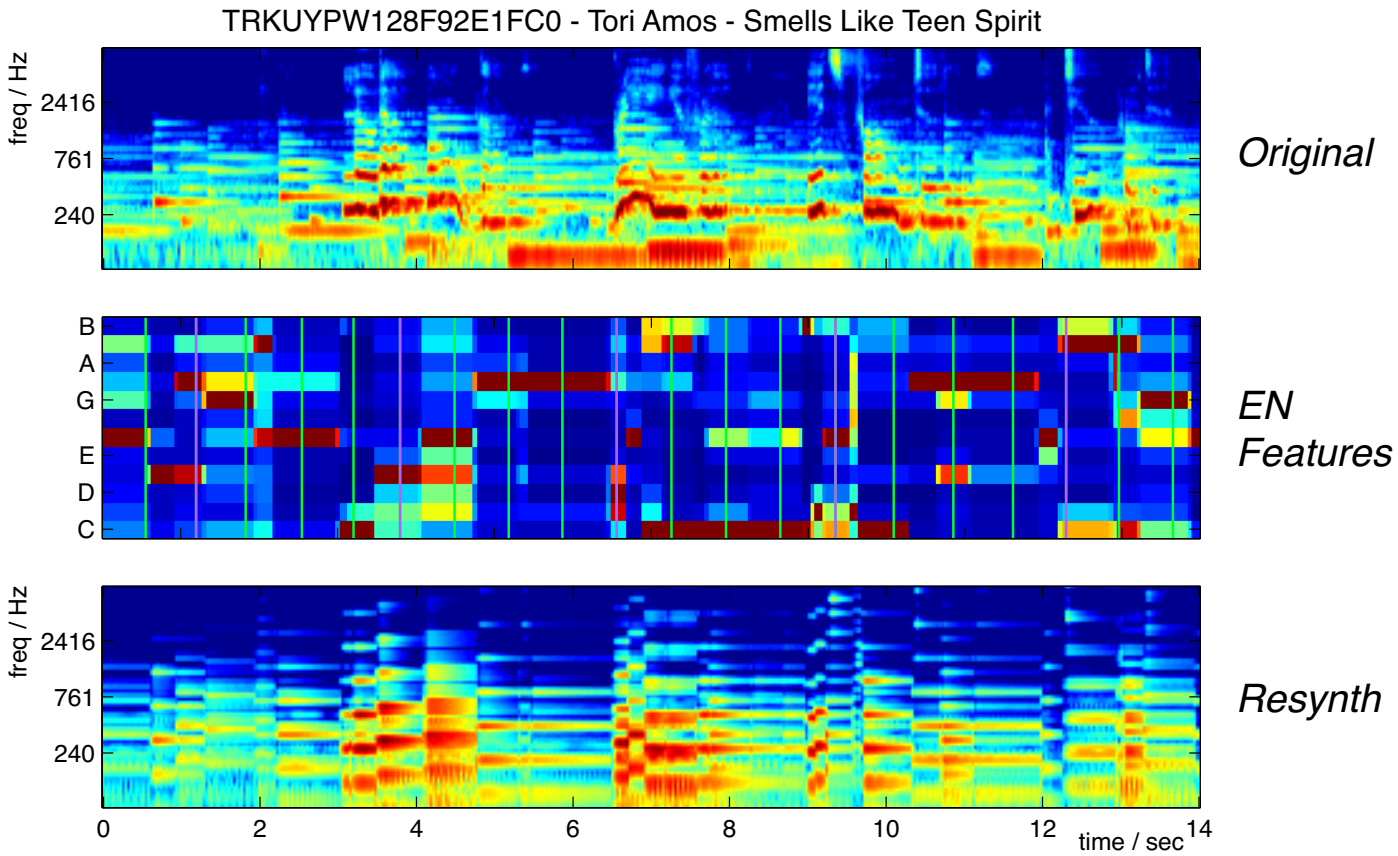


3. Echo Nest Analyze



- Web service to provide beat, chroma, ... analysis (and much more)

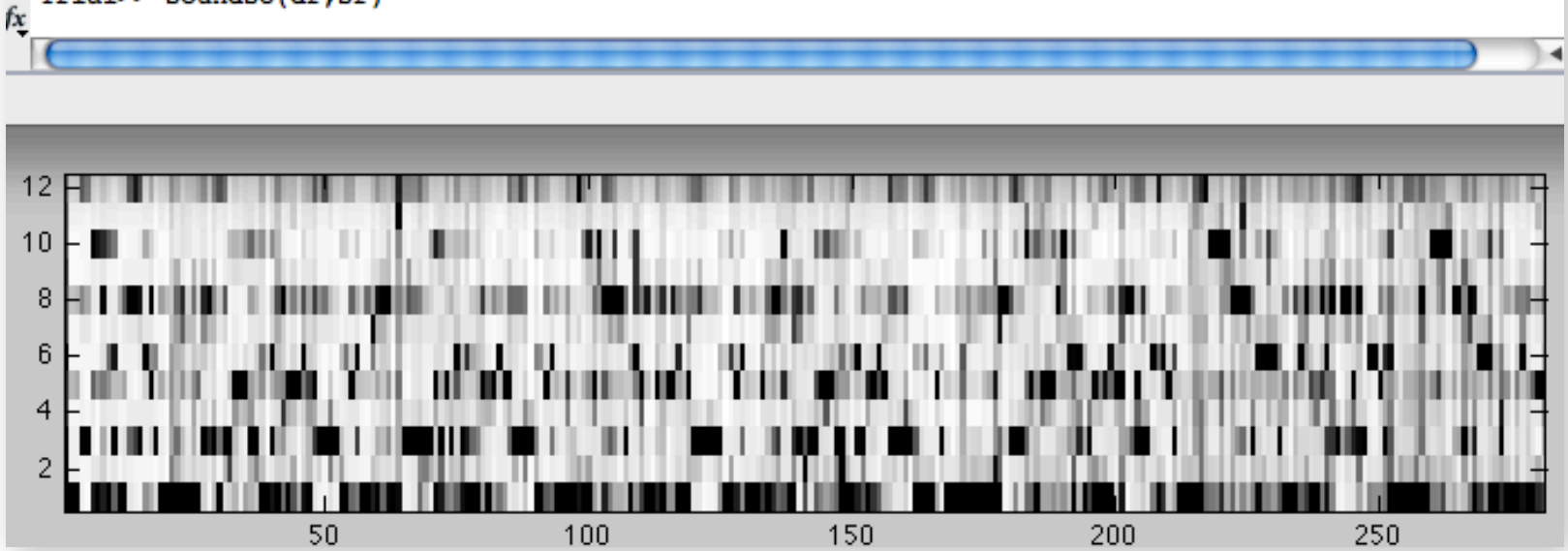
- register for free API key http://developer.echonest.com/account/register/
- upload MP3, get back XML with analysis data



EN Analyze Usage

- Matlab wrapper function

```
Trial>>
Trial>> D = en_analyze('LIB-Beatles-YBPFvp750sc.mp3',1);
uploading mp3
upload complete
track ID=music://id.echonest.com/~/TR/TRCJEDQ127F2594BB5
segment info downloaded
segment info parsed
tatum info downloaded
tatum info parsed
beat info downloaded
beat info parsed
bar info downloaded
bar info parsed
section info downloaded
downloading remaining fields
done
Trial>> imagesc(D.pitches)
Trial>> axis xy
Trial>> dr = synthesize_chroma(D.pitches,D.segment,sr);
Trial>> soundsc(dr,sr)
```



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)

- **EN Analyze features +...**
  - Lyrics,
    Tags,
    Covers,
    Listeners ...

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

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 ch1x
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
TRIHLAW128F429BBF8 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
```

Summary

- Music Alignment
  Dynamic Programming finds correspondence

- Cover Songs
  DP, or cross-correlation for efficiency

- EN Analyze
  Web service to analyze audio

References

- R. Bellman, *Dynamic Programming*, Princeton University Press. 1957.
- D. Ellis and G. Poliner, "Identifying Cover Songs With Chroma Features and Dynamic Programming Beat Tracking," *Proc. ICASSP-07*, Hawai'i, pp. IV-1429-1432, 2007.
- J. Foote, "Visualizing Music and Audio using Self-Similarity," In *Proc. ACM Multimedia*, Orlando, pp. 77-80, 1999.
- Frank Kurth, Meinard Müller, Christian Fremerey, Yoon ha Chang, and Michael Clausen, "Automated synchronization of scanned sheet music with audio recordings," *Proc. 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, pp. 261-266, 2007.
- N. Orio & D. Schwarz, "Alignment of monophonic and polyphonic music to a score," *Proc. Int. Comp. Music Conf.*, Havana, pp.155-158, 2001.
- J. Serrà, E. Gómez, P. Herrera, X. Serra, "Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification," *IEEE Trans. on Audio, Speech and Lang. Proc.*, 16(6), pp. 1138-1151, 2008.