Beat-Synchronous Chroma Representations for Music Analysis

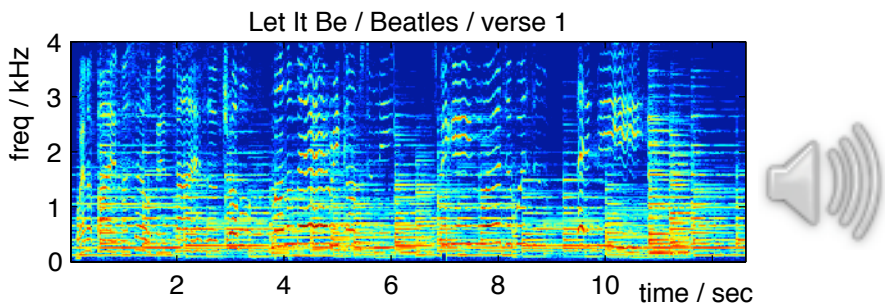
Dan Ellis
Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Eng., Columbia Univ., NY USA

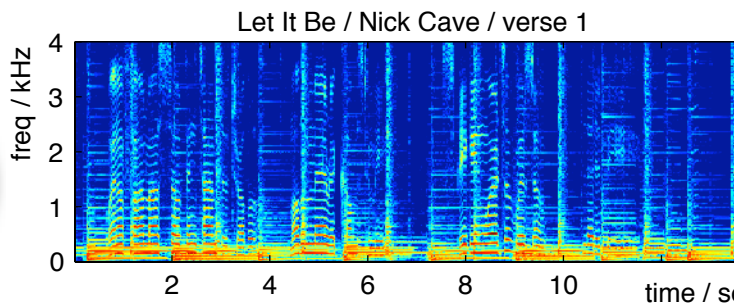
dpwe@ee.columbia.edu http://labrosa.ee.columbia.edu/

1. Chroma Features
2. Beat Tracking
3. Matching Cover Songs
4. Artist Identification

Beyond MFCCs...

- MFCCs have been useful in Audio Music IR
  - "timbral similarity"
  - artist ID, segmentation, thumbnailing, singing ...
- Separate tradition of Symbolic MIR
  - melody matching, chord detection, meter analysis
- It's time to bring them together
  - ... with robust audio mid-level representations
  - ... that capture tonal (melodic-harmonic) content



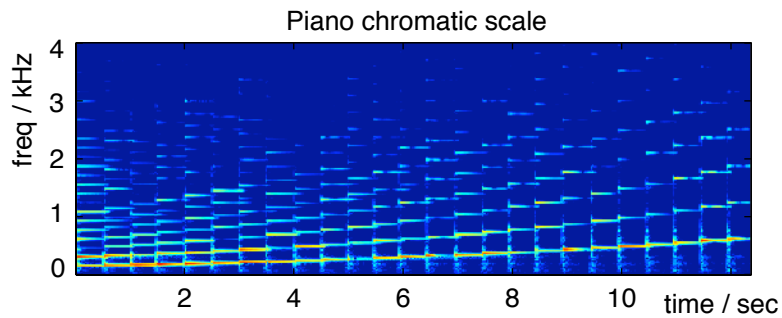


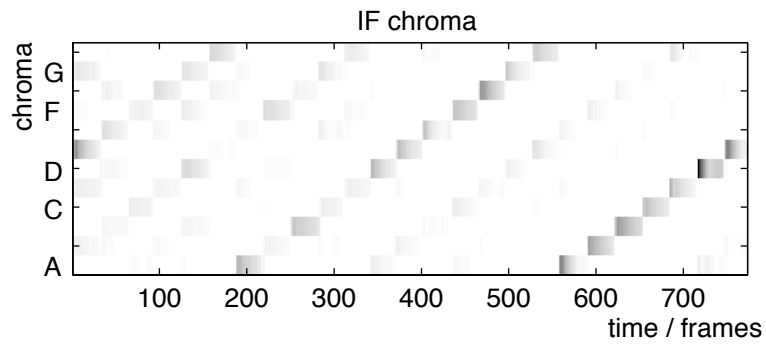
- = beat-synchronous chroma features

1. Chroma Features

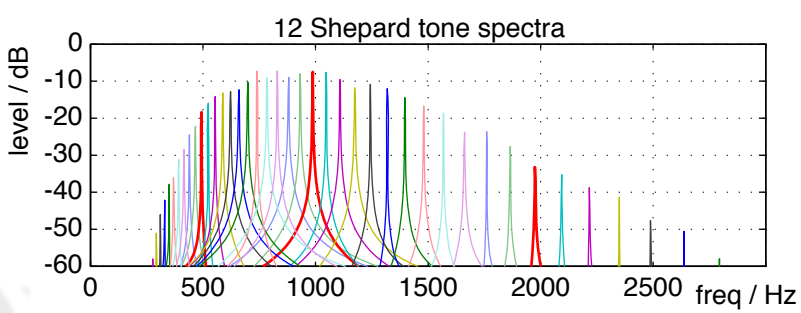
- Chroma features map spectral energy into one canonical octave
  - i.e. 12 semitone bins

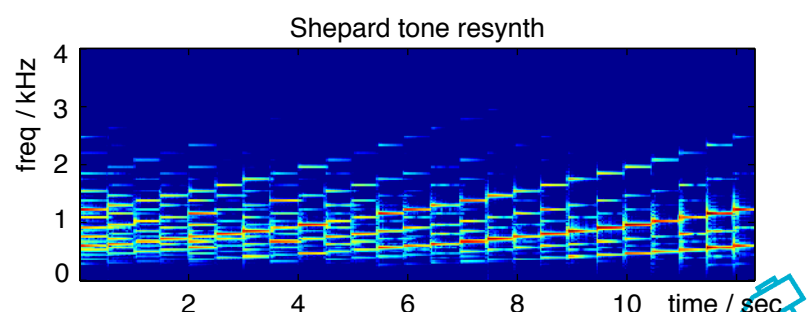
Piano scale





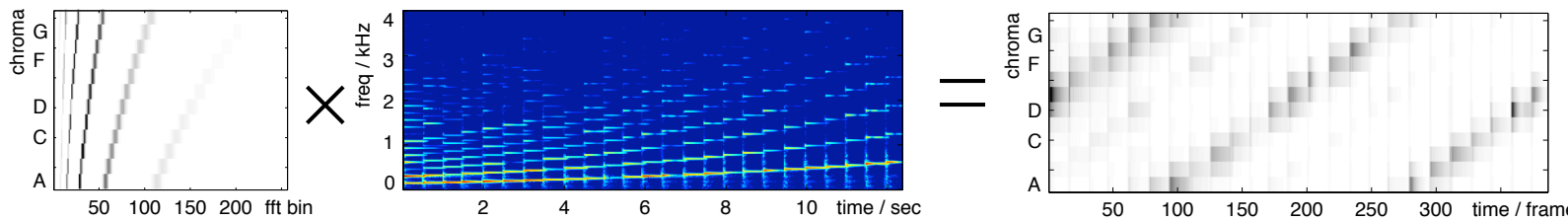
- Can resynthesize as "Shepard Tones"
  - all octaves at once



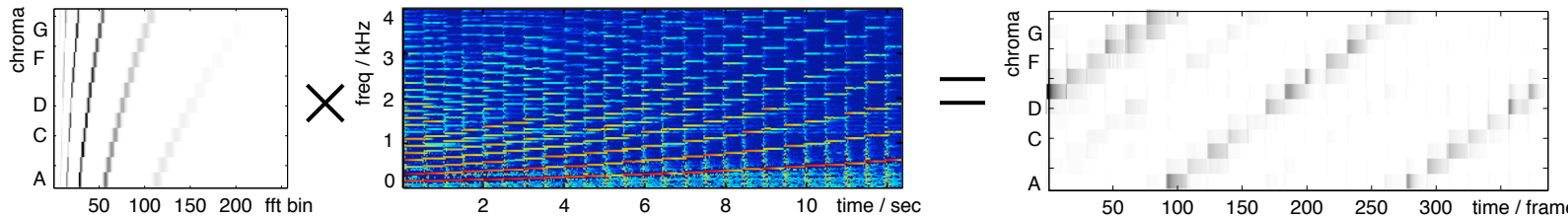


Calculating Chroma Features

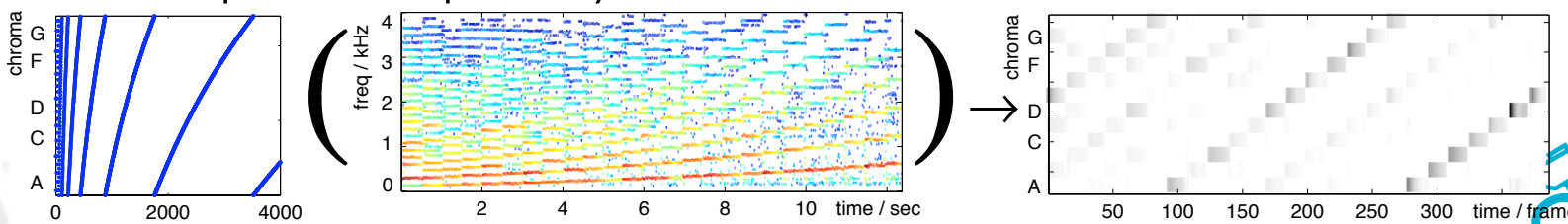
- **Method 1: Map every STFT bin**
  - blurs non-tonal energy



- **Method 2: Map only STFT peaks**
  - still blurry at low frequencies

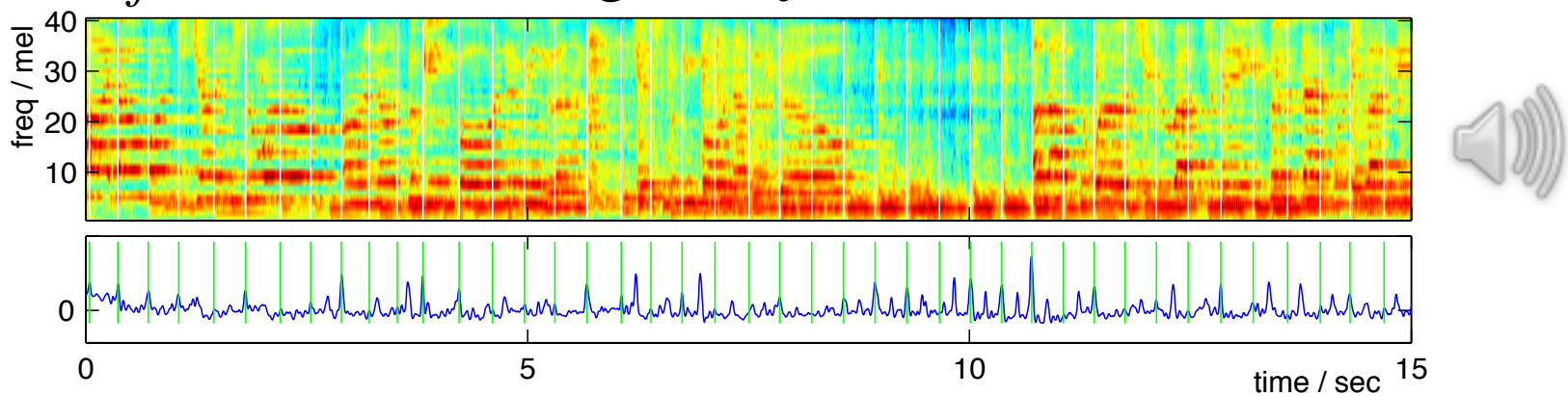


- **Method 3: Instantaneous Frequency $\delta\theta/\delta t$**
  - escapes frequency resolution limit

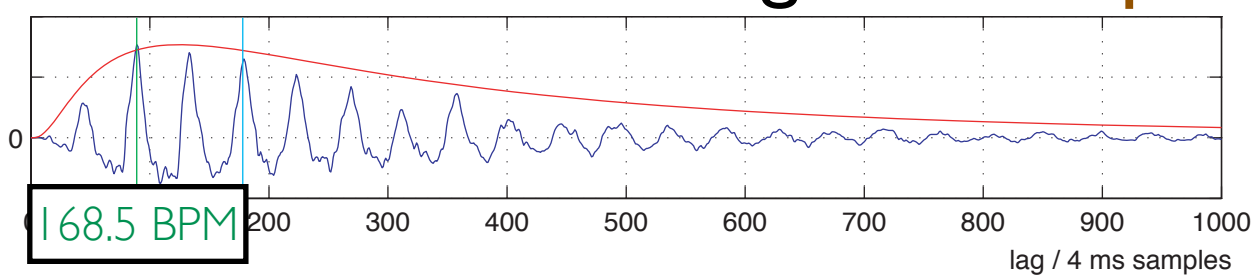


2. Beat Tracking (1)

- Goal: One feature vector per 'beat' (tatum)
  - for tempo normalization, efficiency
- "Onset Strength Envelope"
  - $\text{sum}_f(\max(0, \text{diff}_t(\log |X(t,f)|)))$

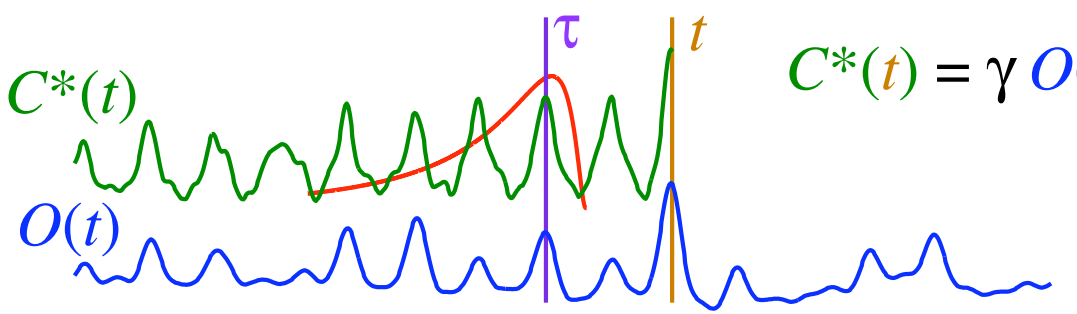


- Autocorr. + window → global tempo estimate



Beat Tracking (2)

- Dynamic Programming finds beat times $\{t_i\}$
  - optimizes $\Sigma_i O(t_i) + \alpha \Sigma_i W\big((t_{i+1} - t_i - \tau_p)/\beta\big)$
  - where $O(t)$ is onset strength envelope (local score)
    $W(t)$ is a log-Gaussian window (transition cost)
    τ_p is the default beat period per measured tempo
  - incrementally find best predecessor at every time
  - backtrace from largest final score to get beats

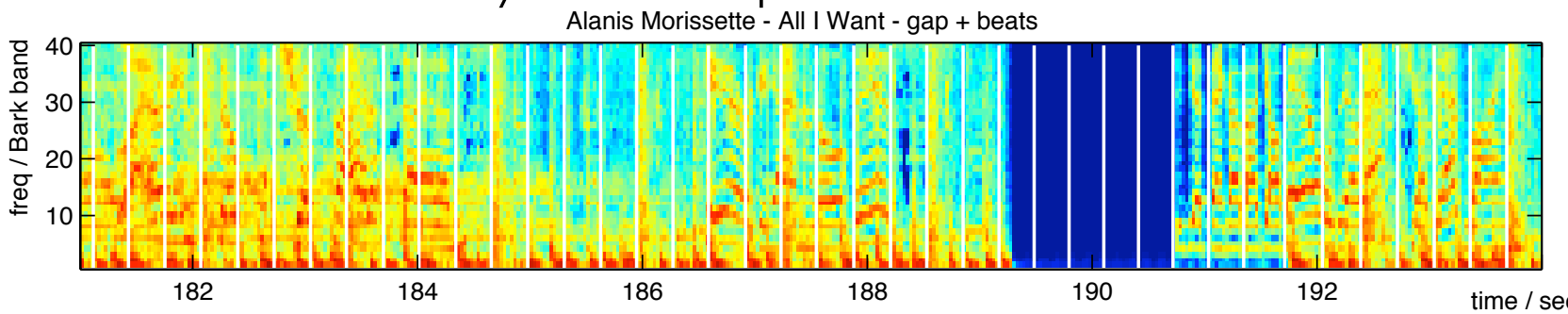


$$C^*(t) = \gamma\, O(t) + (1-\gamma)\max_{\tau}\{W((\tau - \tau_p)/\beta)C^*(\tau)\}$$

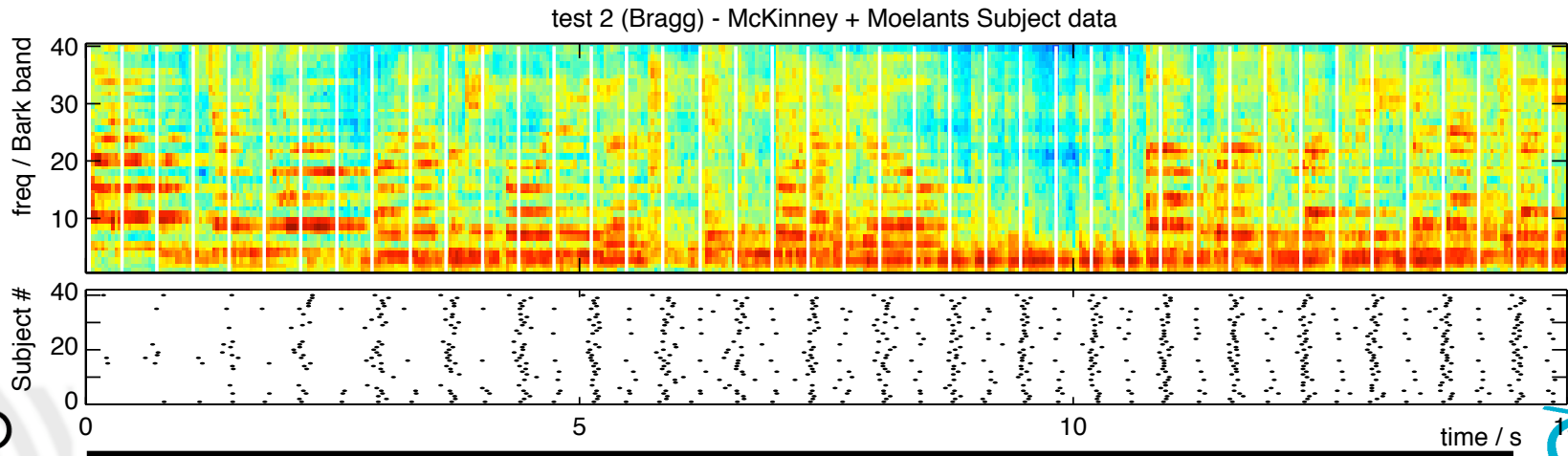
$$P(t) = \operatorname{argmax}_{\tau}\{W((\tau - \tau_p)/\beta)C^*(\tau)\}$$

Beat Tracking Results

- DP will bridge gaps (non-causal)
  - there is always a best path ...

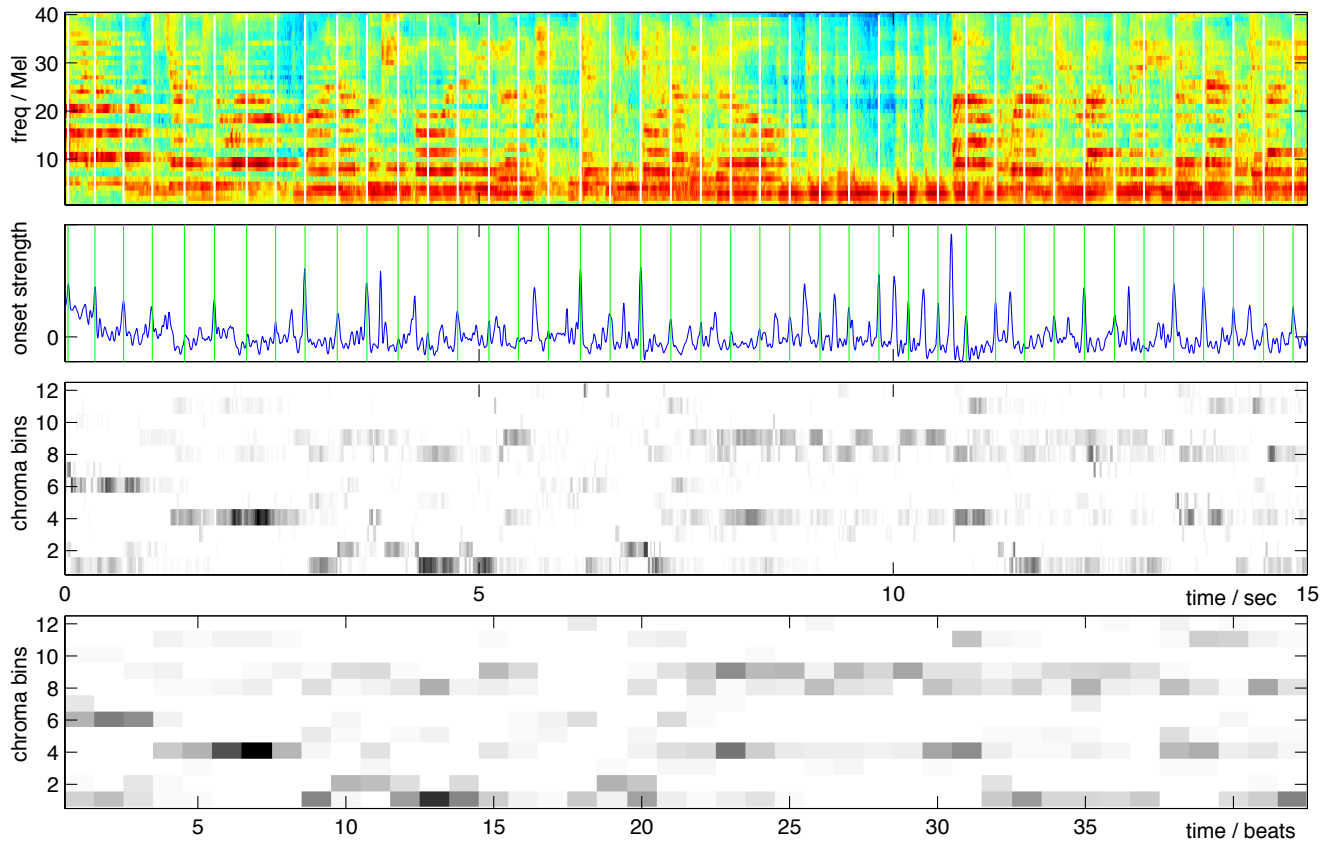


- 2nd place in MIREX 2006 Beat Tracking
  - compared to McKinney & Moelants human data



Beat-Synchronous Chroma Features

- Beat + chroma features / 30ms frames
  → average chroma within each beat
  - compact; sufficient?

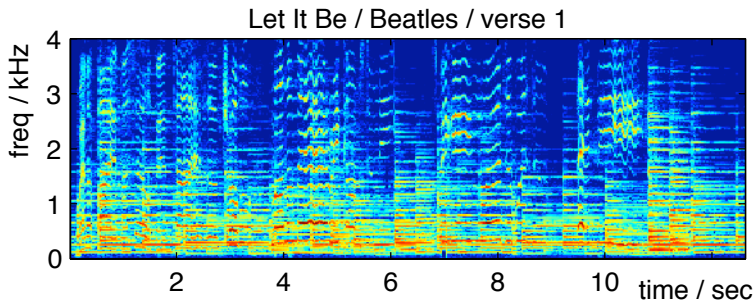


3. Cover Song Detection

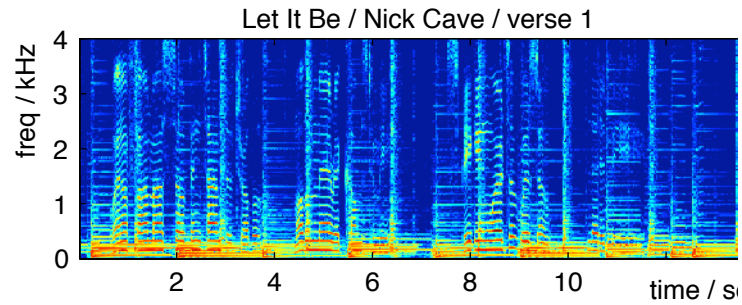
with Graham Poliner

- "Cover Songs" = reinterpretation of a piece
  - different instrumentation, character
  - no match with "timbral" features

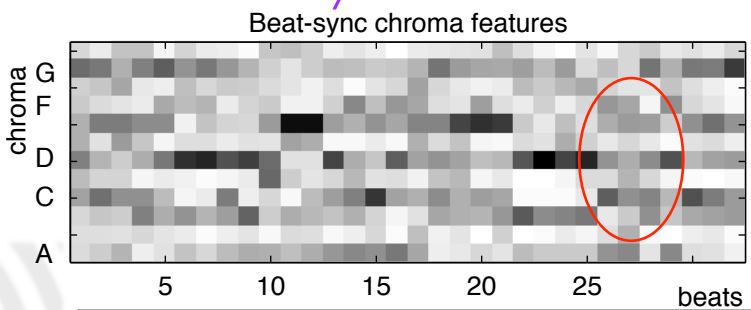
Let It Be - The Beatles

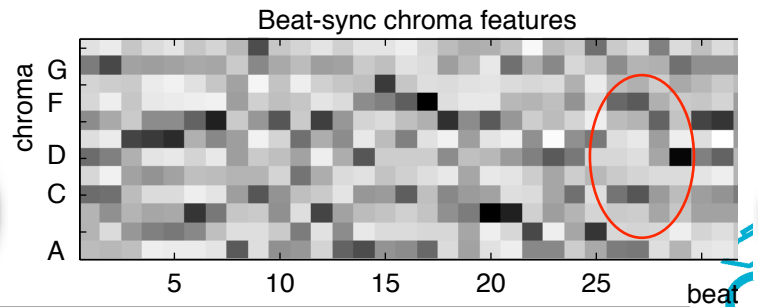


Let It Be - Nick Cave



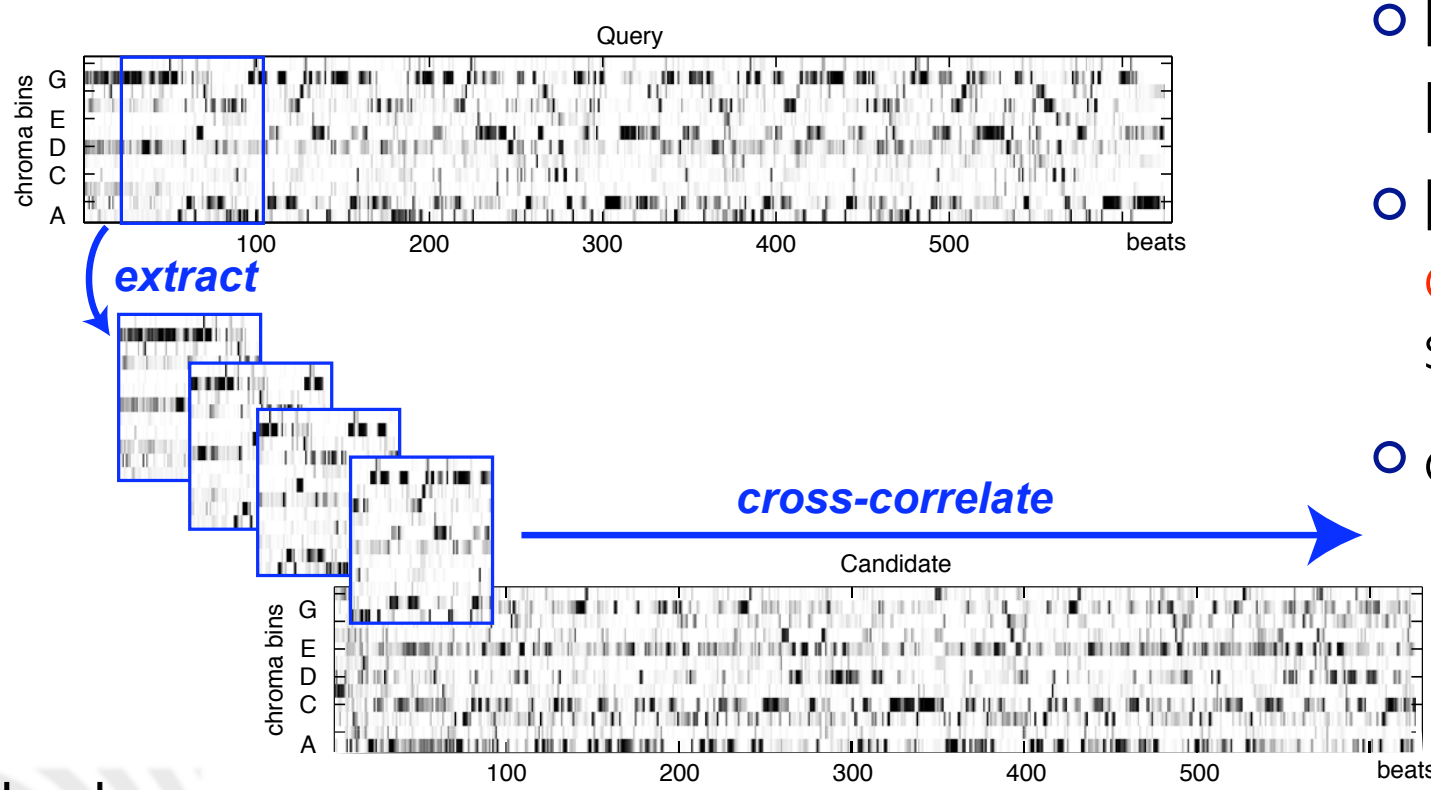
- Need a different representation!
  - beat-synchronous chroma features





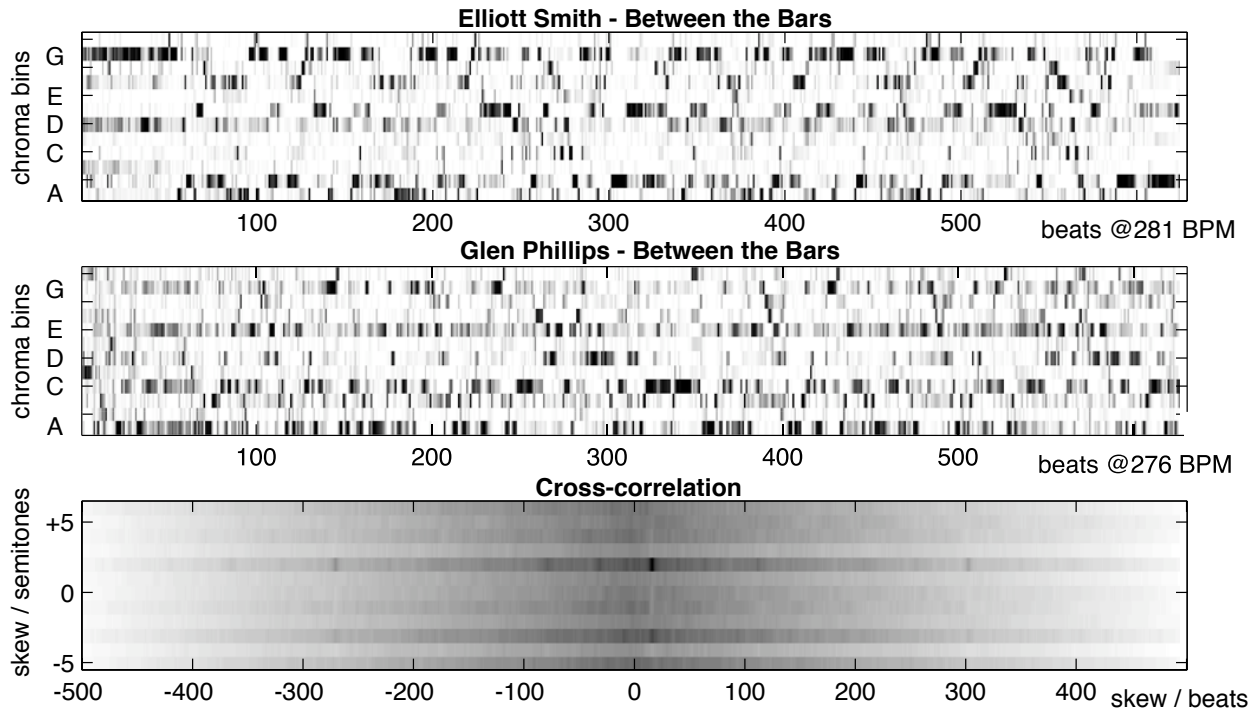
Matching (1): Little Fragments

- Cover versions may change song structure
  - multiple local matches at different alignments
- Match query and target as many small pieces?
  - how big are the pieces?
  - how do we combine individual scores?
  - do we have all day?



Matching (2): Global Correlation

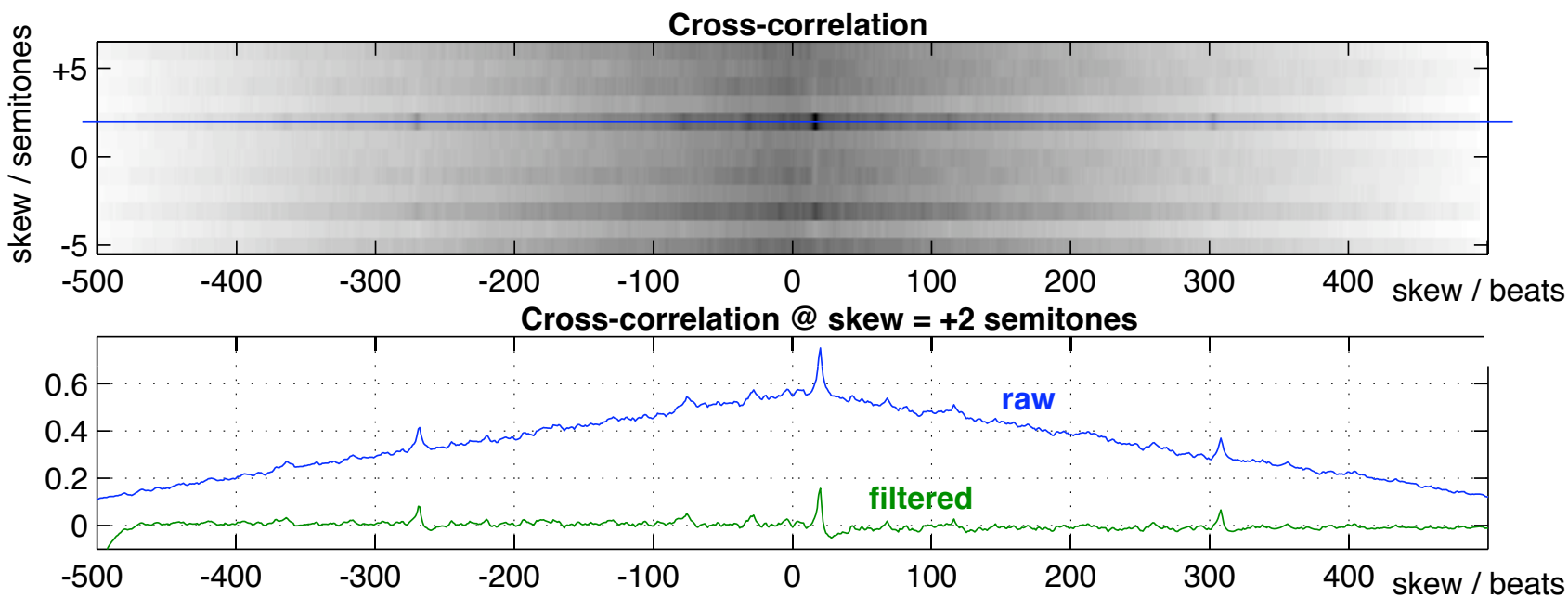
- Cross-correlate *entire* beat-chroma matrices
  - ... at all possible transpositions
  - implicit combination of match quality and duration



- One good matching fragment is sufficient...?

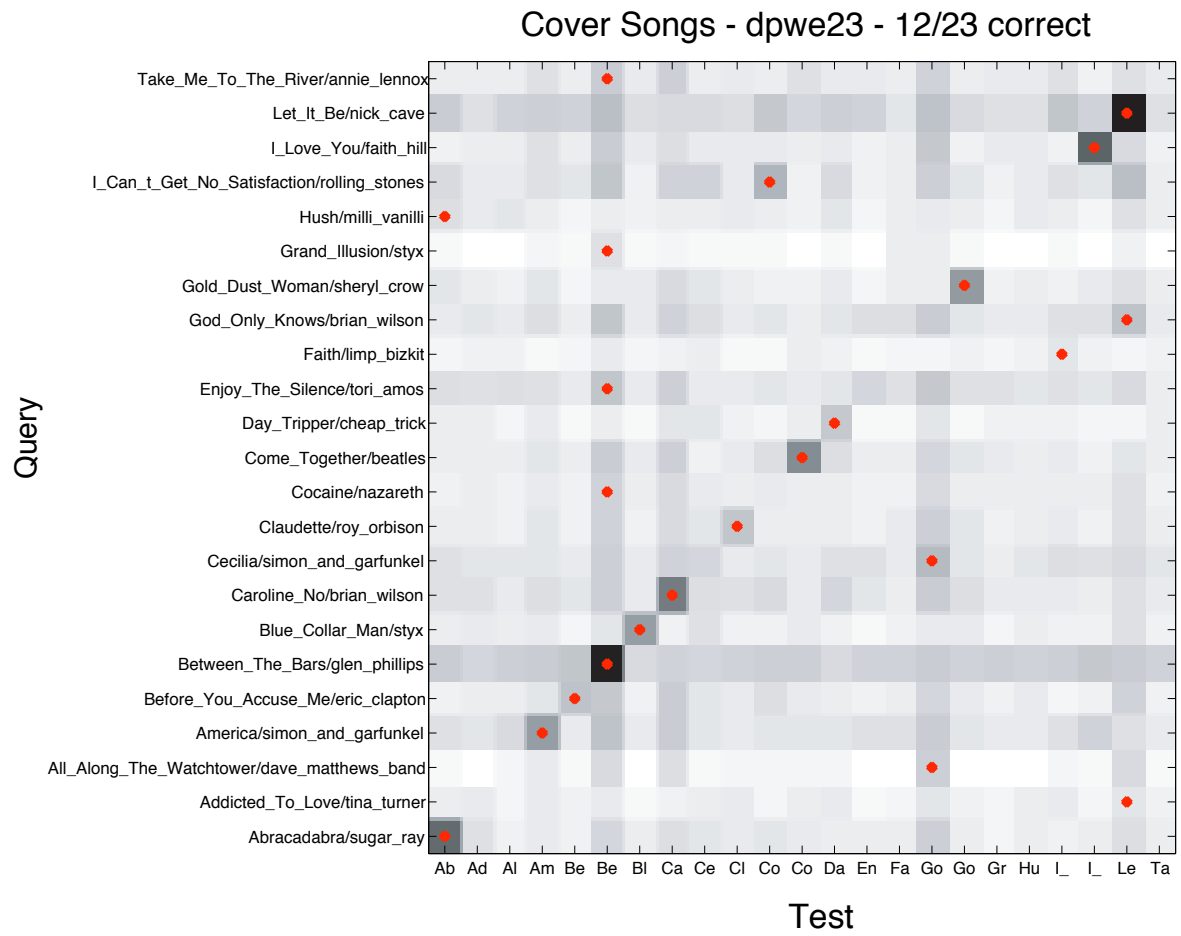
Filtered Cross-Correlation

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



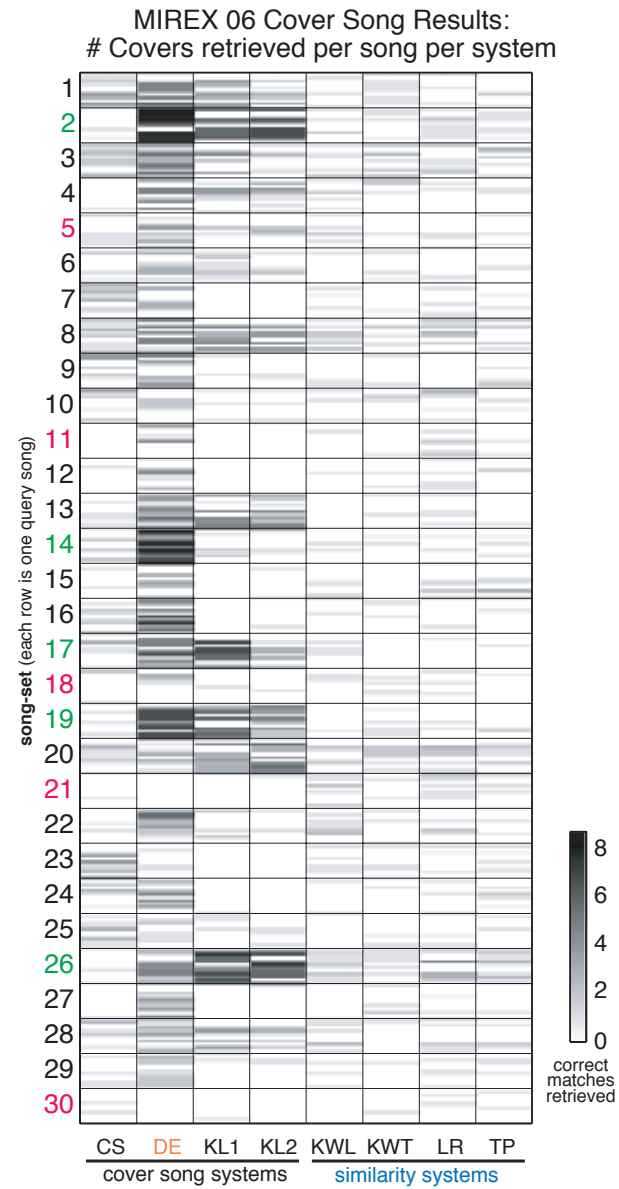
Results (1): Ellis 23 set

- 23 pairs of cover songs from uspop2002 +...
  - one correct match per query



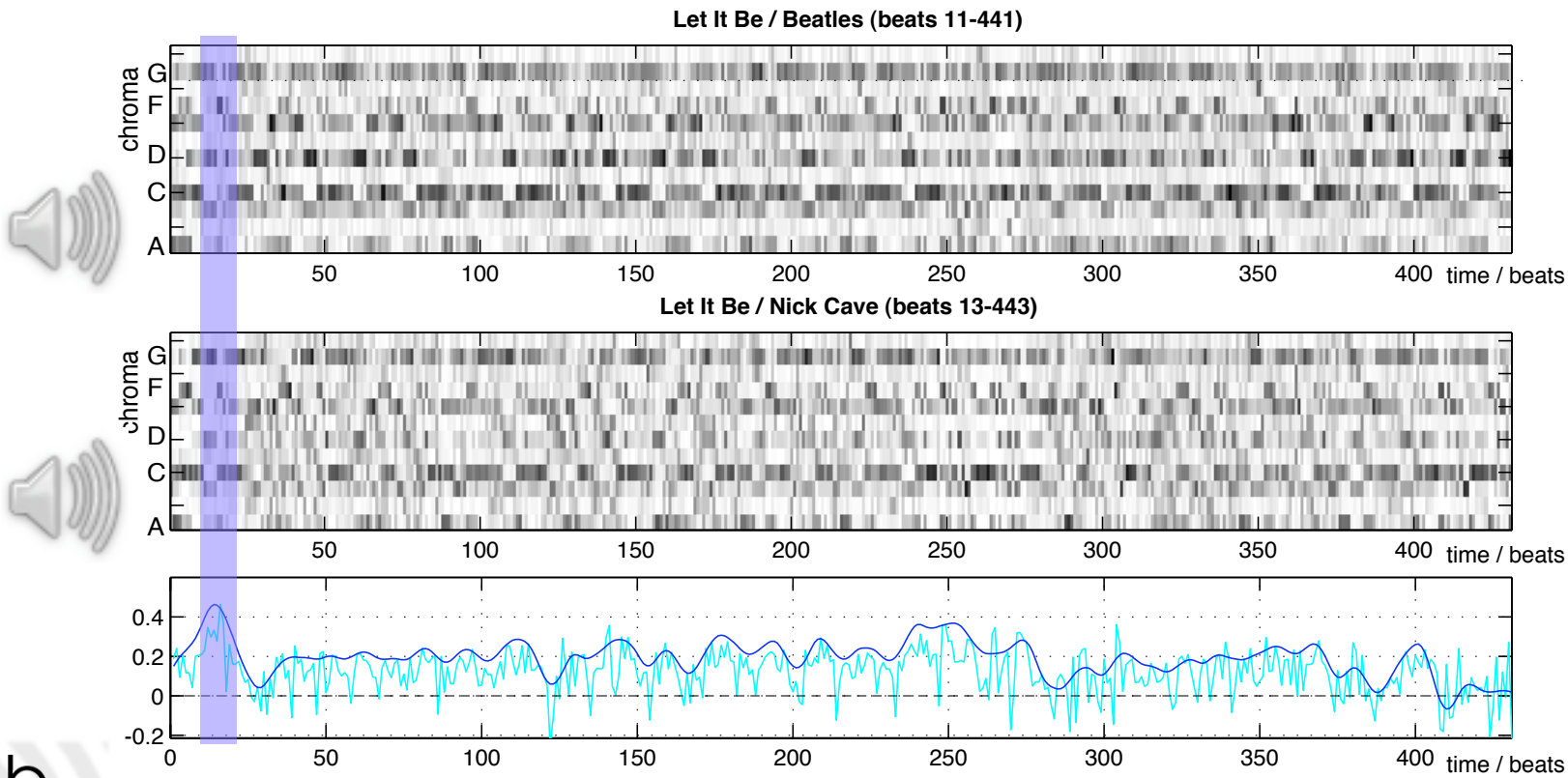
Results (2): MIREX 06

- **Cover song contest**
  - 30 songs x 11 versions of each (!)
  - (data has not been disclosed)
  - # true covers in top 10
  - 8 systems compared (4 cover song + 4 similarity)
- **Found 761/3300 = 23% recall**
  - next best: 11%
  - guess: 3%



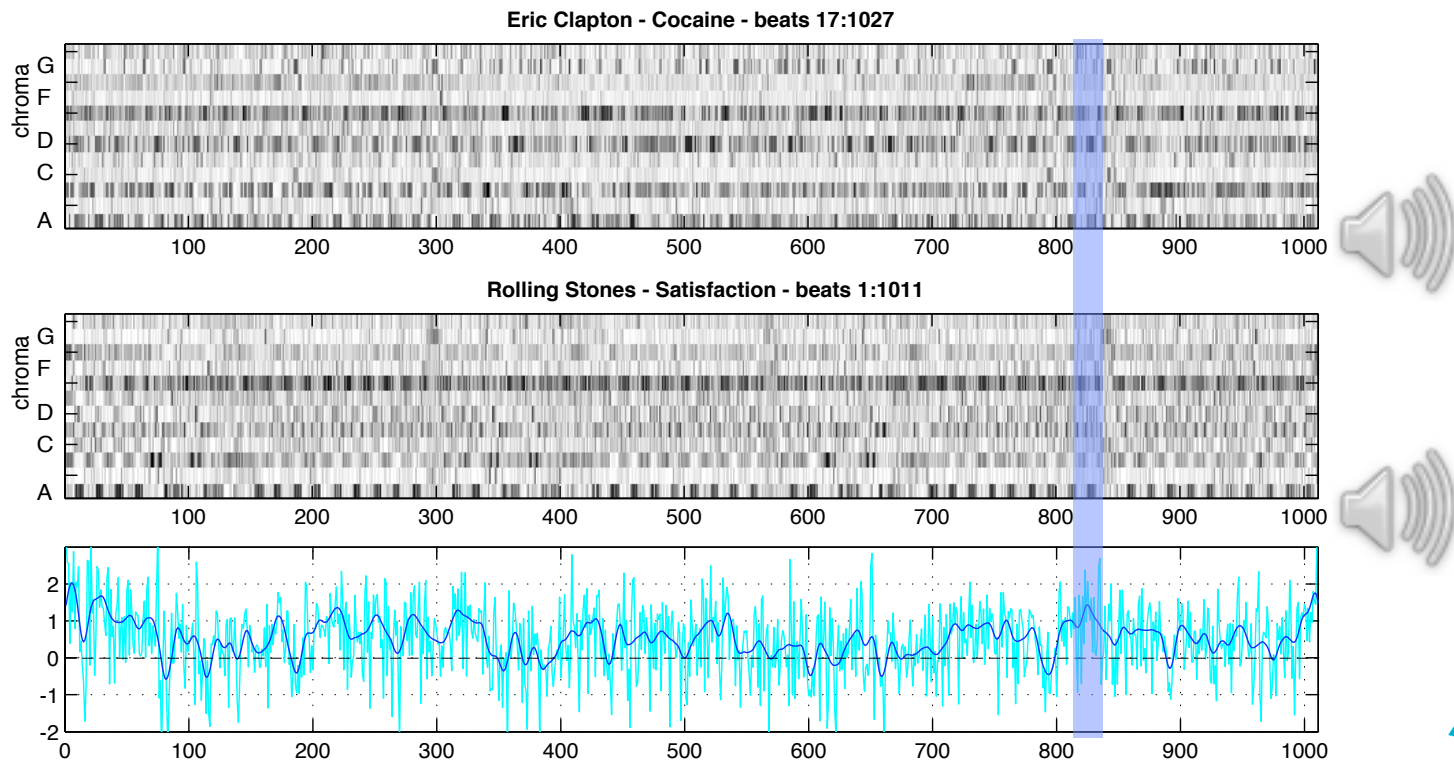
Where are the matches?

- 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



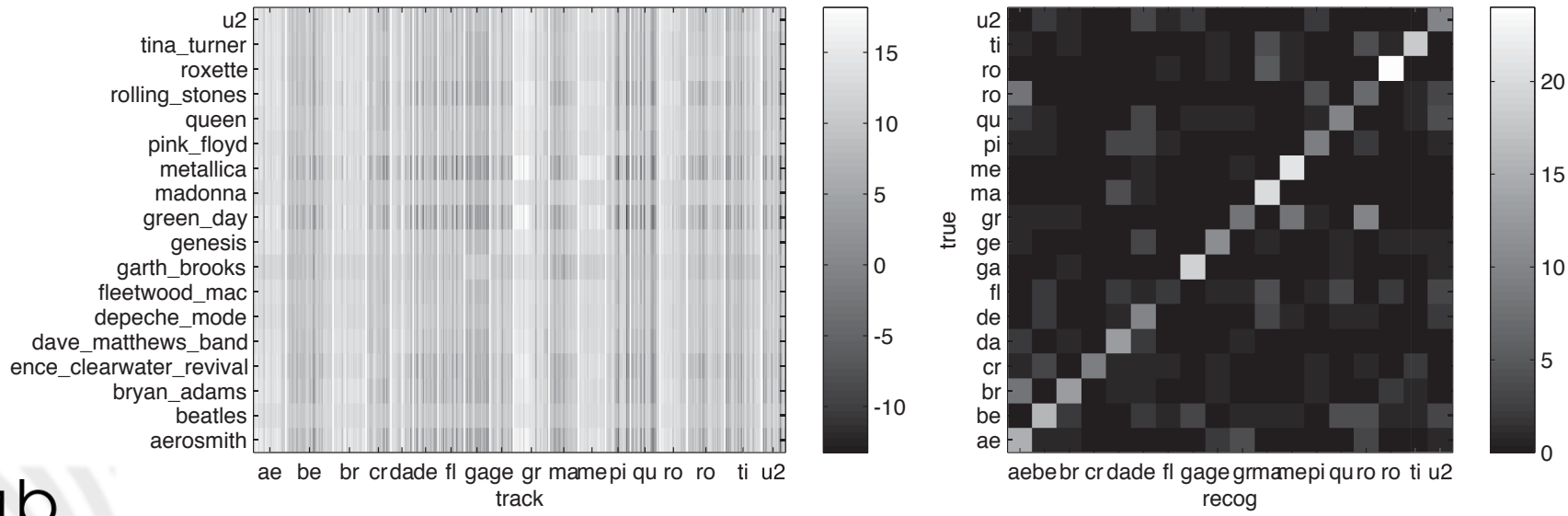
What are the mistakes?

- **False reject - missed true match**
  - cover version is too different, beat tracking wrong ...
- **False alarm - invalid match**
  - "Cocaine" (Clapton) vs. "Satisfaction" (Stones)



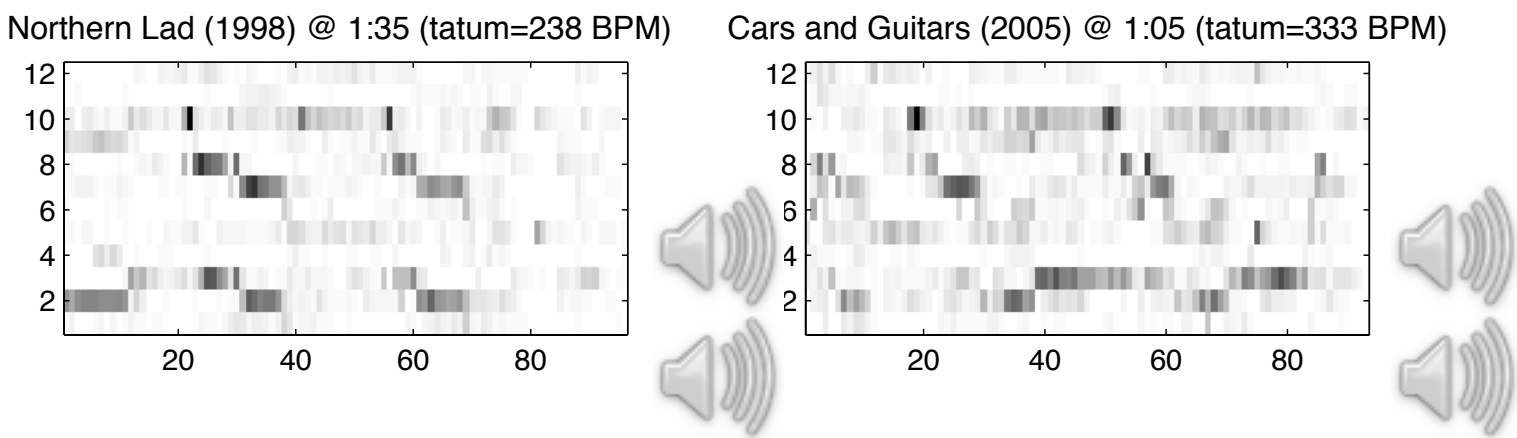
4. Artist Identification (AID)

- **Baseline system: "Bag of (timbral) frames"**
  - MFCC frames, model as Gaussian or GMM
  - distance by likelihood or KL
- **Dataset: [Mandel et al. 2006]**
  - 18 artists x 5 or 6 albums each
  - 18x3 albums for training, 18x2 for test, 10x1 dev



Beat Chroma Features for AID?

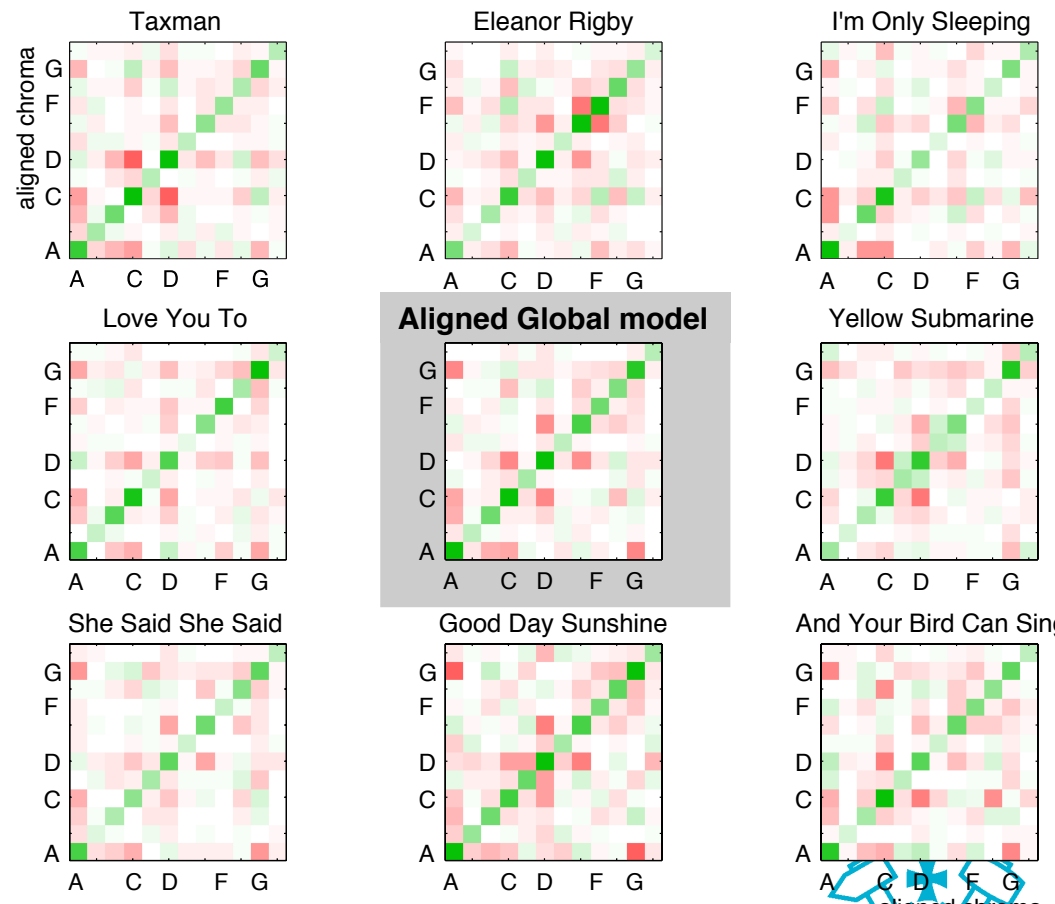
- Artists may use tonality in particular ways...
  - density, variety
  - particular chords
  - (influence of instruments on chroma features)



- Try bag-of-frames on beat-chroma rep'n
  - use several consecutive beats?
  - key-normalization of each piece?

Key Normalization

- Could try matching at all possible rotations..
- .. or just transpose every piece initially
  - single Gaussian model of one piece
  - find ML rotation of other pieces
  - model all transposed pieces
  - iterate until convergence



Timbre+Chroma AID

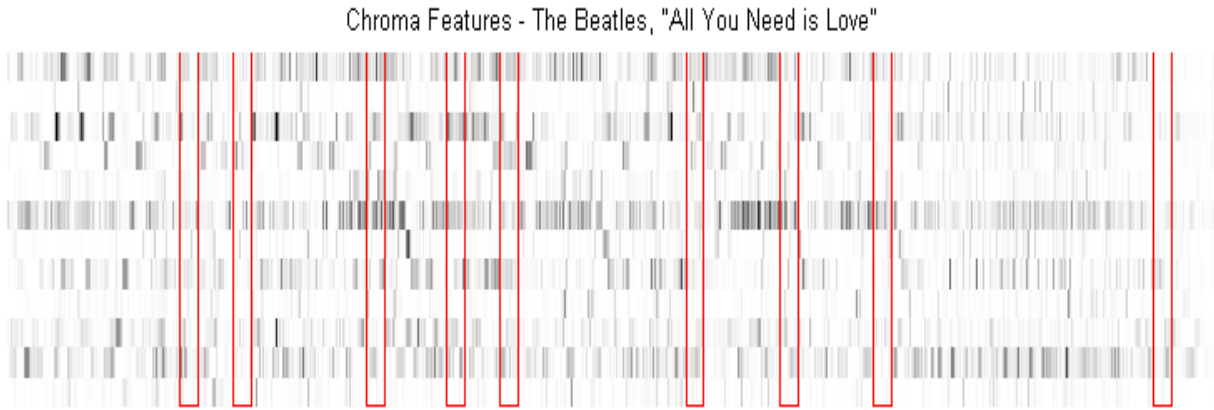
- Preliminary Mandel18 Artist ID accuracy:

Feature	Model	T win	Acc	Exec. time
MFCC20	FullCov	1	**48%**	212 s
MFCC20	64 GMM	1	33%	1952 s
Chroma	FullCov	1	15%	46 s
Chroma	FullCov	4	14%	117 s
Chroma	64GMM	1	24%	850 s
Chroma	64GMM	4	15%	2242 s
ChromaKN	FullCov	1	17%	110 s
ChromaKN	FullCov	4	14%	258 s
ChromaKN	64GMM	1	**25%**	2533 s
ChromaKN	64GMM	4	16%	5803 s
MFCC + Chroma fusion			**52%**	

Artist Fragments

with Courtenay Cotton

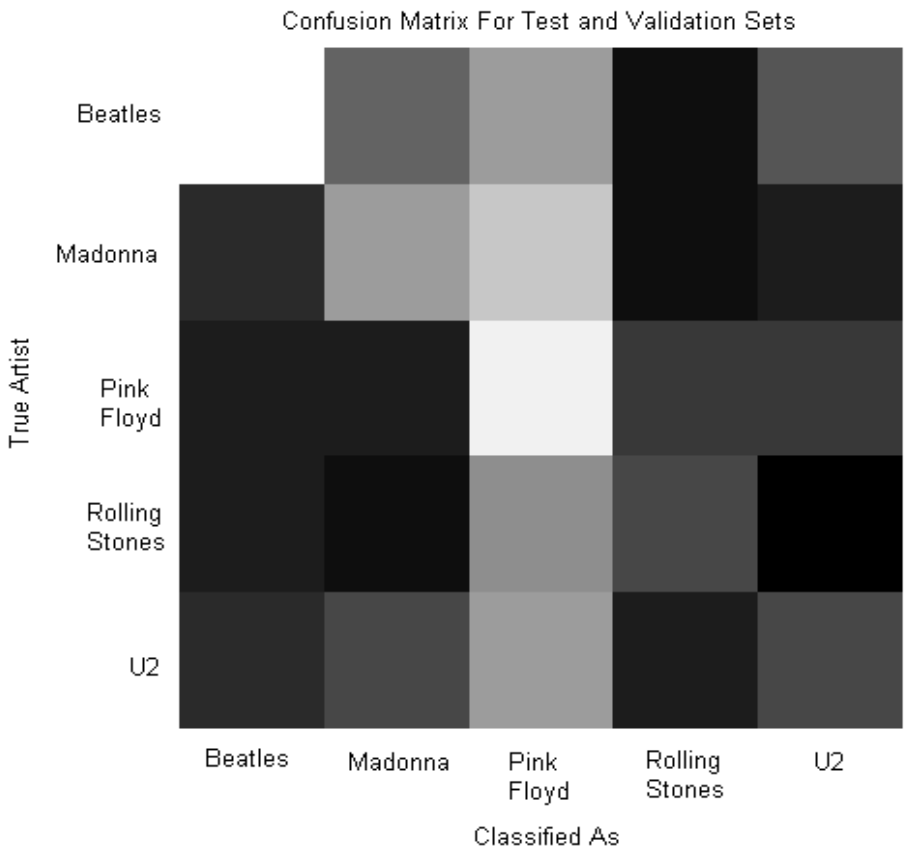
- **Idea: Find the most discriminant beat-chroma fragments per artist**
  - k-means cluster 16 beat fragments within piece



  - keep fragments largest ratio (avg. similarity to same artist)/(avg. sim. to others)
  - classify test pieces by ID of best-scoring fragment

Artist Fragment Results

- Preliminary, 5 way artist ID, ~32% correct



- need to search more fragments
- way to choose phrase beginnings?
- a basis set for all tonal content?

Conclusions and Future Work

- Beat-synchronous chroma features
  are successful for matching cover songs
  - captures melody-harmony, not instruments
- Further uses:
  Beat-chroma fragments
  as musical building blocks
  - e.g. VQ over large body of music
  - find recurrent motifs
  - artist identification?

- Code available!
  Google "matlab chroma features"