

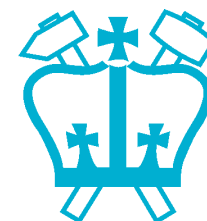
Searching for Similar Phrases in Music Audio

Dan Ellis

Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Engineering, Columbia University, NY USA

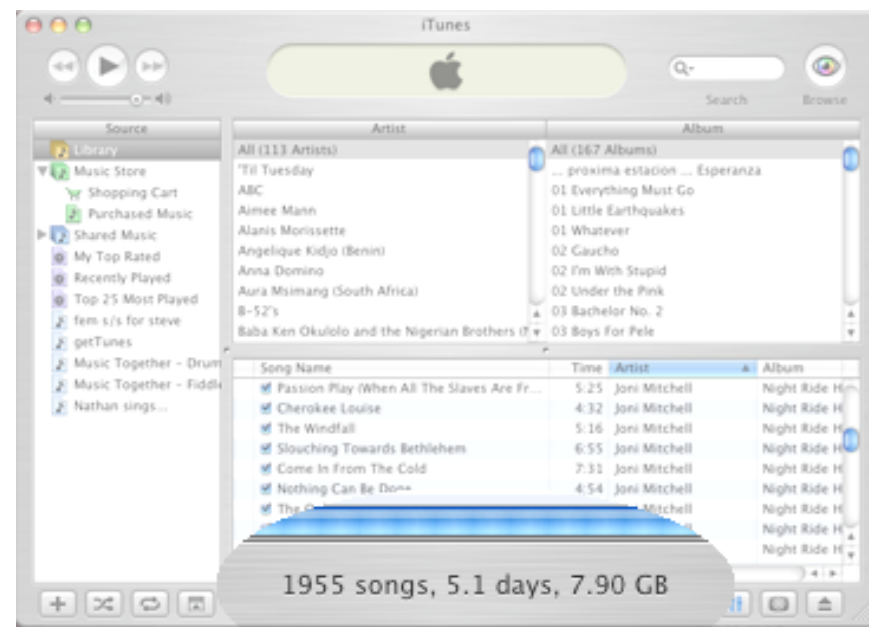
<http://labrosa.ee.columbia.edu/>

1. Motivation: Similar Phrases
2. Phrase Matching System
3. Experiments
4. Conclusions & Future

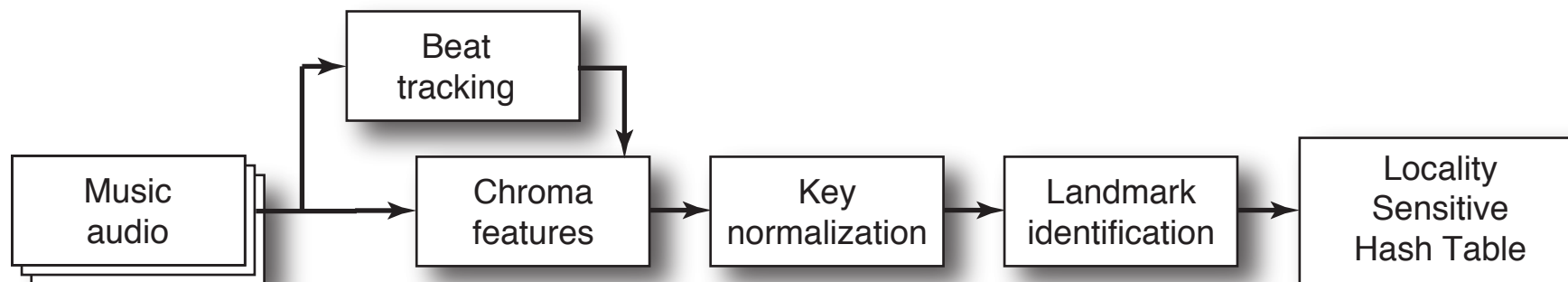


I. Motivation: Similar Phrases

- Idea: Music is a sequence of **reused** pieces
 - e.g. melodic runs, chord sequences, ...
- Can we identify them in **large music databases?**
 - ... which we have
 - i.e. machine learning
- **Applications**
 - classification and matching of pieces
 - compressed representation
 - **data-driven musicology**



Common Phrase Discovery

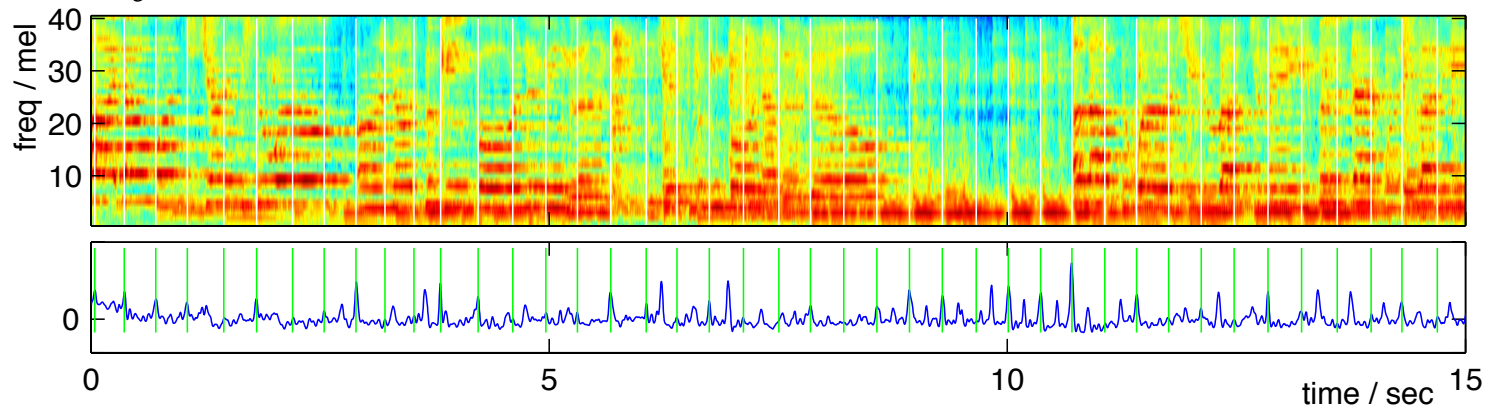


- Chop up music into short descriptions of musical content
 - 24-beat beat-chroma matrices?
- Choose a few that appear to be “starts”
- Put into LSH table (similar items fall in same bin)
- Find the bins with most entries

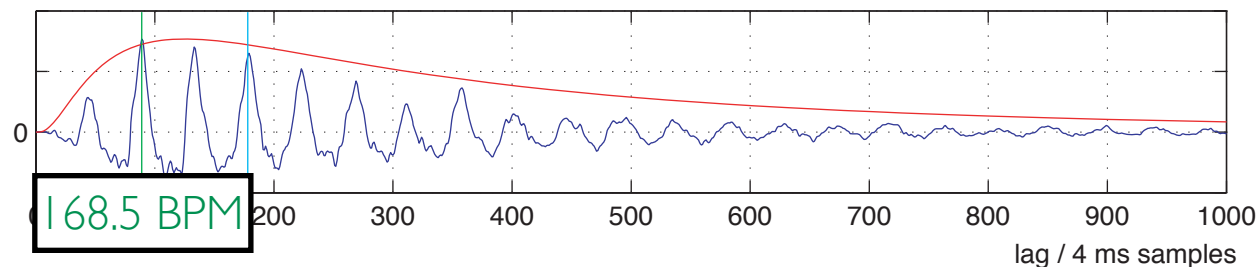
2. Phrase Matching: Beat Tracking

Ellis '06,'07

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



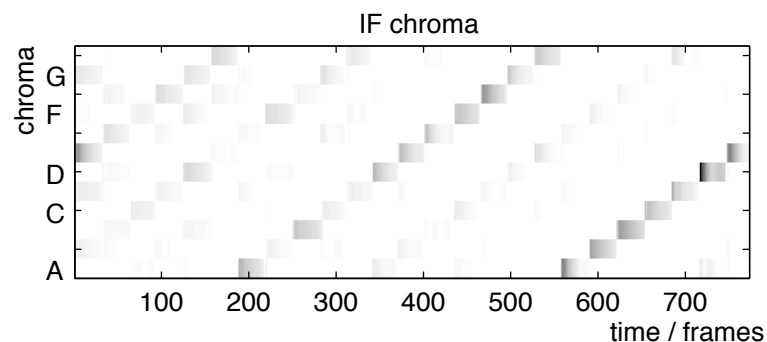
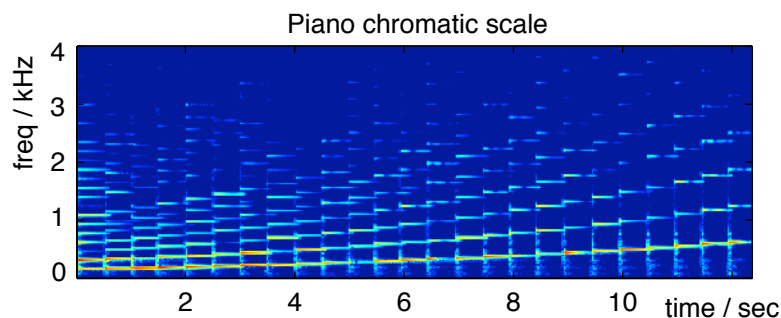
- Autocorr. + window → global tempo estimate



Chroma Features

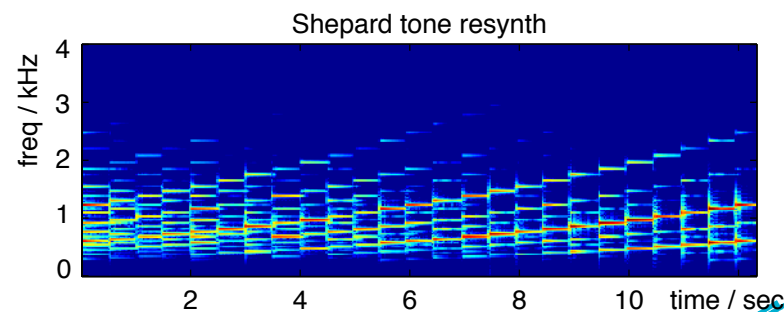
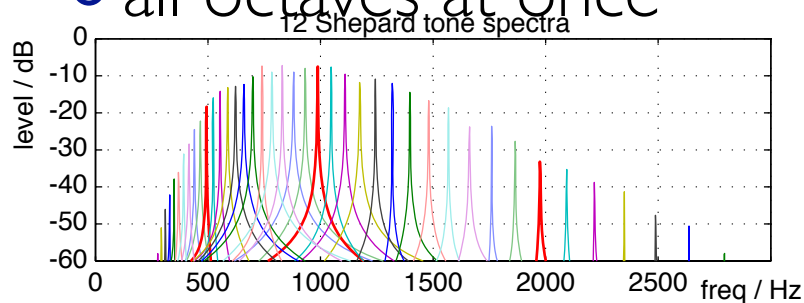
- Chroma features convert spectral energy into musical weights in a **canonical octave**
 - i.e. 12 semitone bins

Piano scale



- Can resynthesize as “Shepard Tones”

- all octaves at once

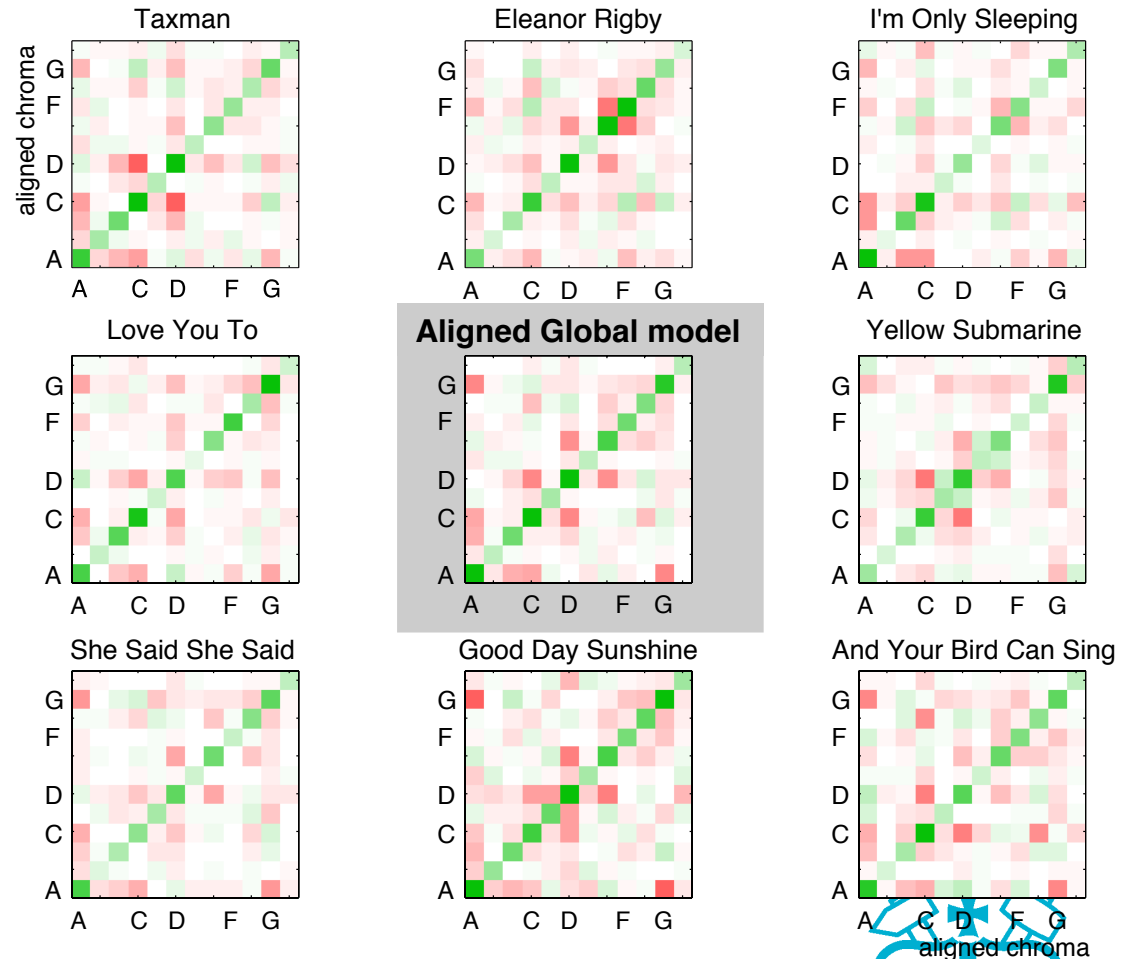


Key Estimation

Ellis ICASSP '07

- Covariance of chroma reflects **key**
- Normalize by **transposing** for best fit

- single Gaussian model of one piece
- find ML rotation of other pieces
- model **all** transposed pieces
- iterate until **convergence**



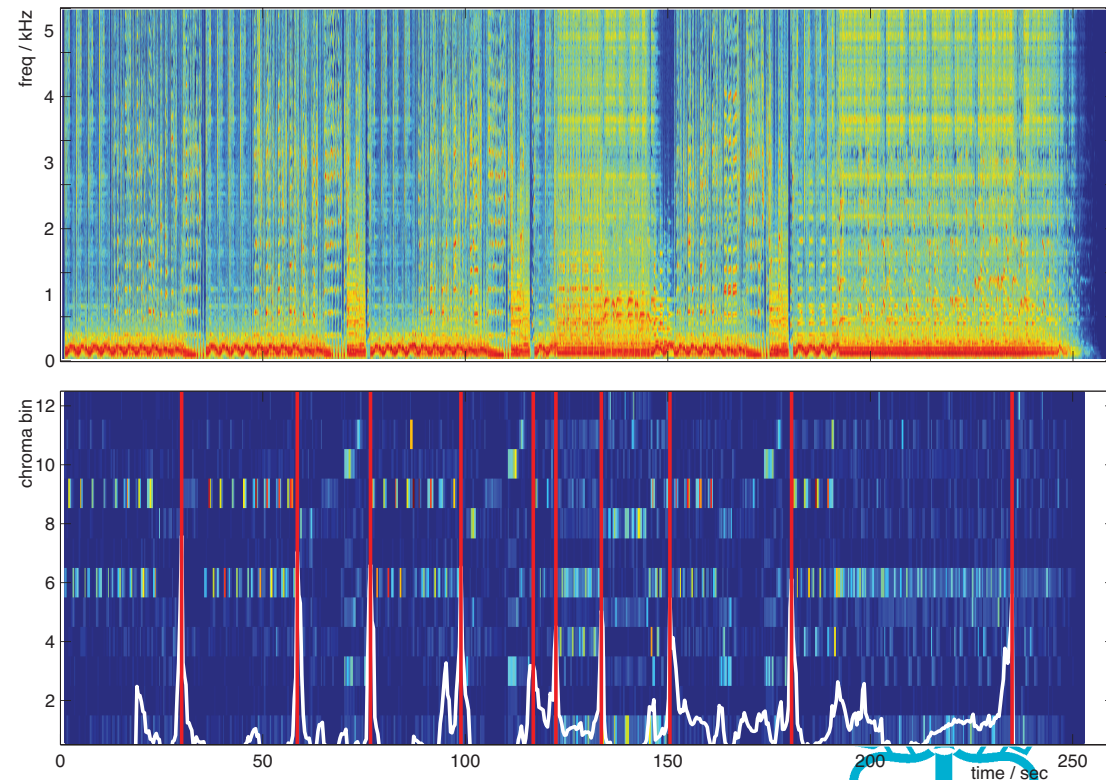
Landmark Location

- Looking for “beginnings” of phrases
 - e.g. abrupt change in harmony, instruments, etc.
 - use likelihood ratio test:
weighted windows either side of boundary vs. all

- Choose top 10 locally-normalized peaks

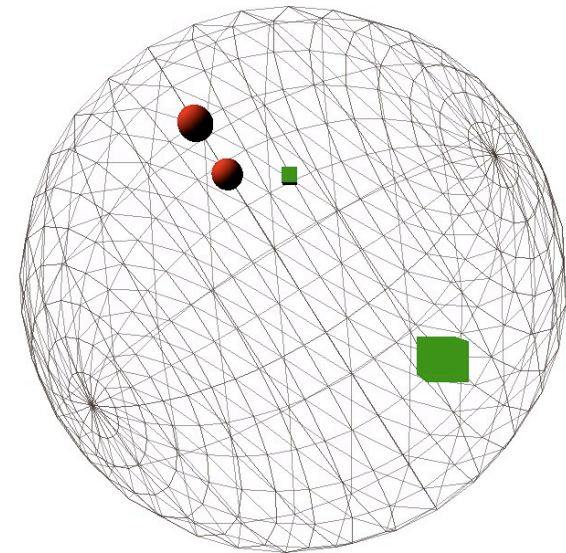
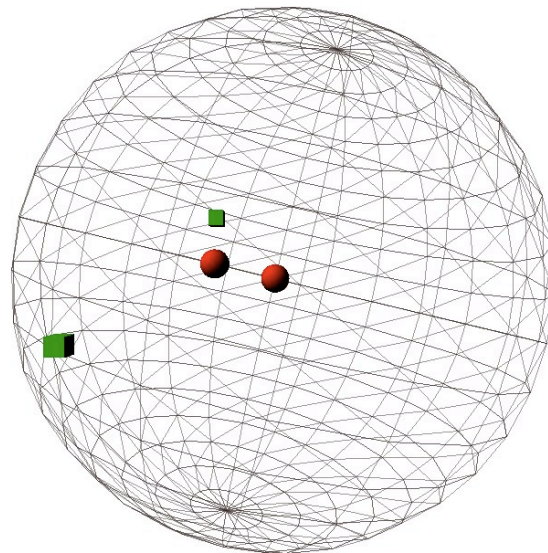
- .. to control data size

Come Together - Spectrogram, Beat-sync chromogram, and top 10 segment points



Locality Sensitive Hashes

- Goal: Quantize **high-dimensional** data so 'similar' items fall into same bin
 - .. for fast and scalable nearest-neighbor search
- Idea: Multiple **random** scalar projections
 - each one will tend to keep neighbors nearby
 - items close together in all projections are probably neighbors



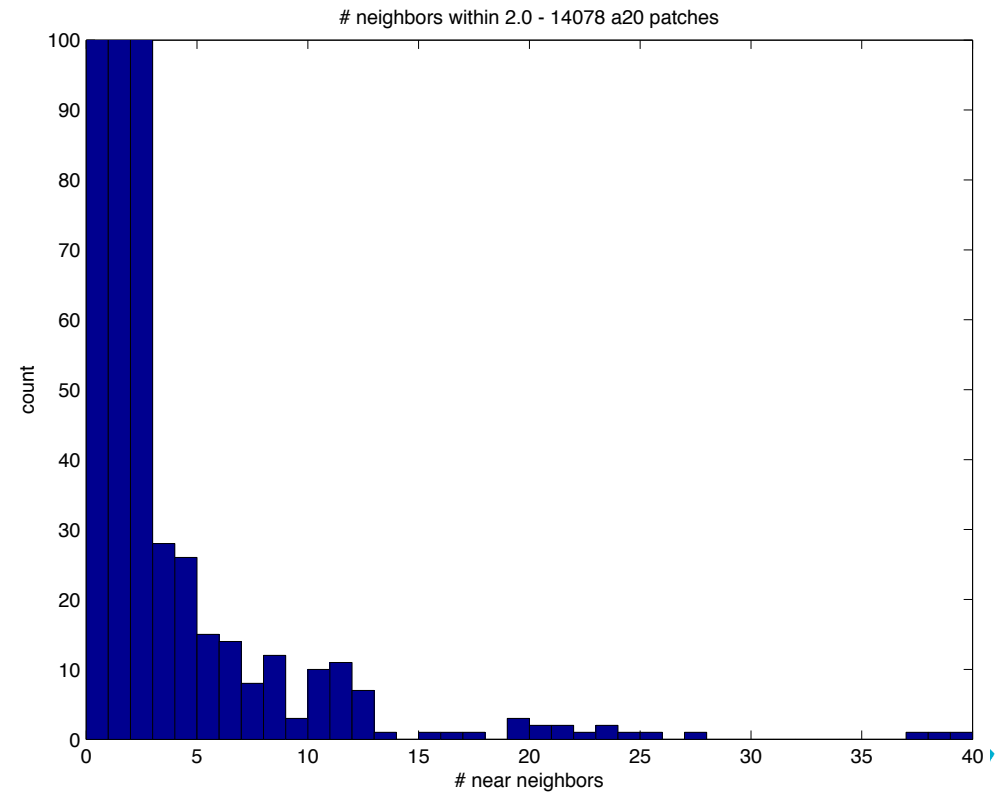
3. Experiments

- Data

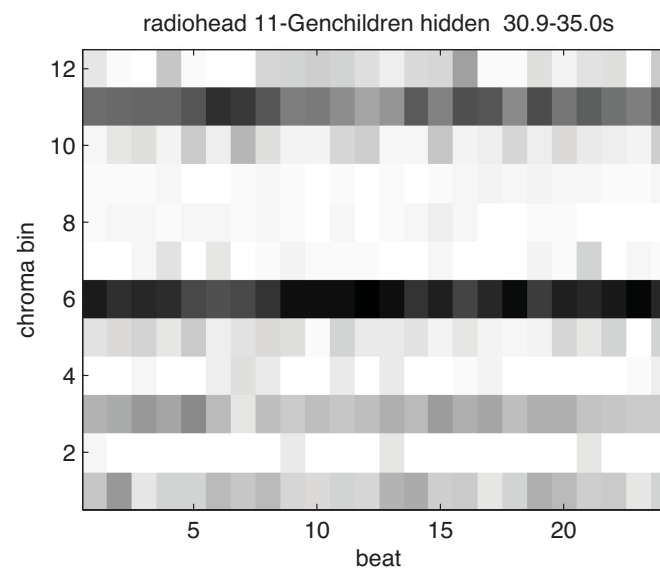
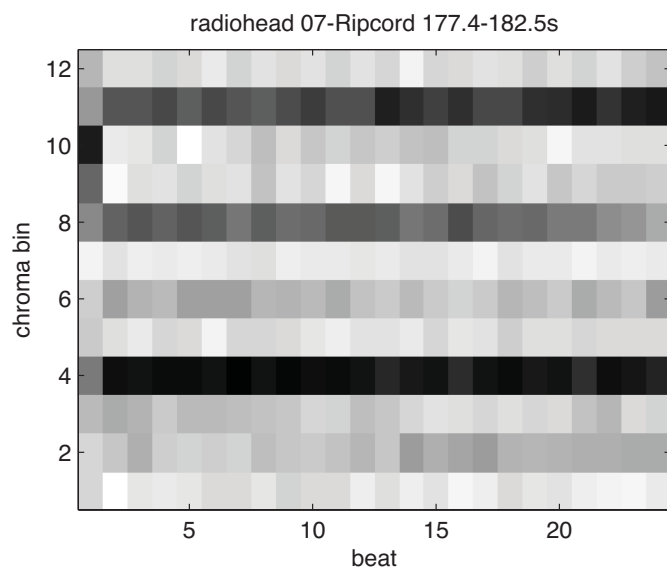
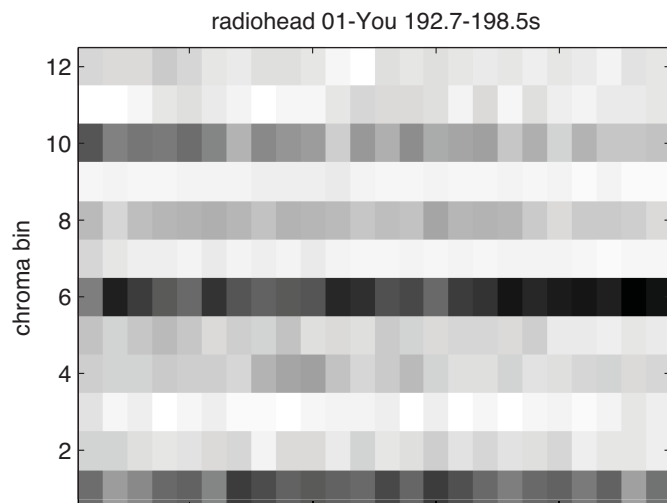
- “**artist20**” - 20 artist x 6 albums = 1413 tracks
- (up to) 10 landmarks/track = 14,078 patches
- each patch = 12 chroma bins x 24 beats (288 dims)

- Performance

- feature calculation:
 - ~ 60 min
- LSH 14k NNs:
 - ~ 30 sec
- 51 patches have >10 NNs within $r = 2.0$



Results - artist20



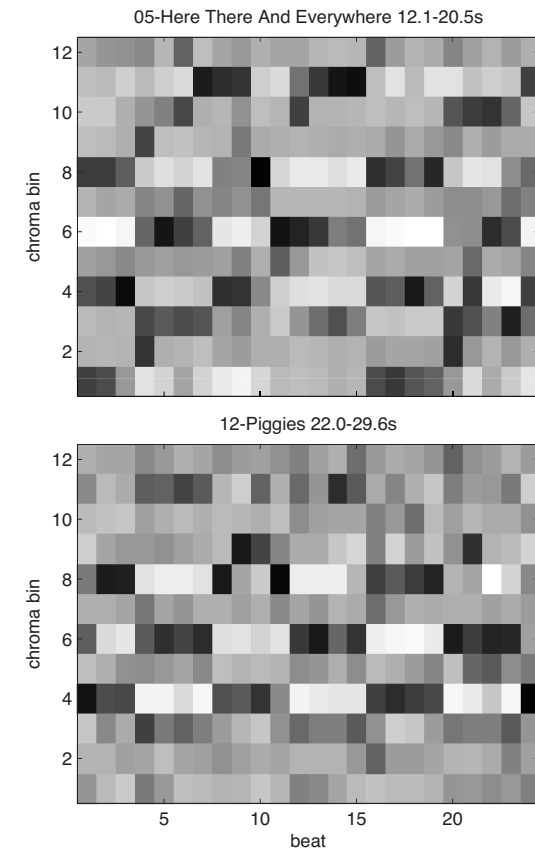
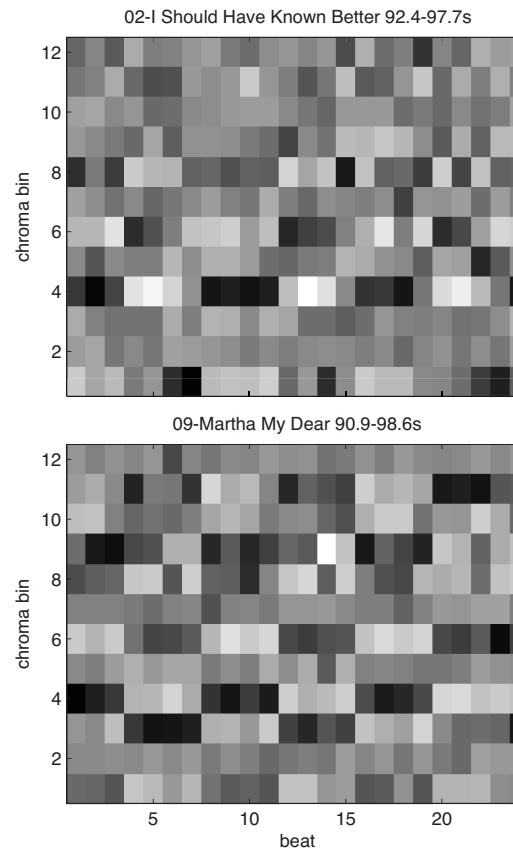
○ mainly sustained notes

Similar Phrases in Music - Ellis

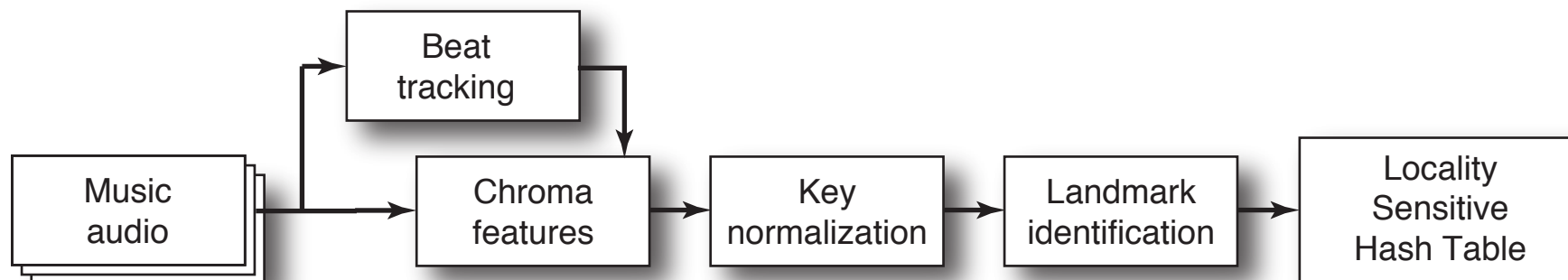
2007-12-18 p. 10/12

Results - Beatles

- Only the 86 Beatles tracks
- **All** beat offsets = 41,705 patches
 - LSH takes 300 sec - approx $N \log N$ in patches?
- **High-pass** along time
 - to avoid sustained notes
- **Song filter**
 - remove hits in same track



Summary / Conclusions



- **Lots of data**
 - find motifs by counting near neighbors
- **Common patterns**
 - e.g. melodic/harmonic-beat sequences
- **Future**
 - different features and/or pre-emphasis
 - better landmark points
 - complete dictionary