Lecture 11: Chroma and Chords

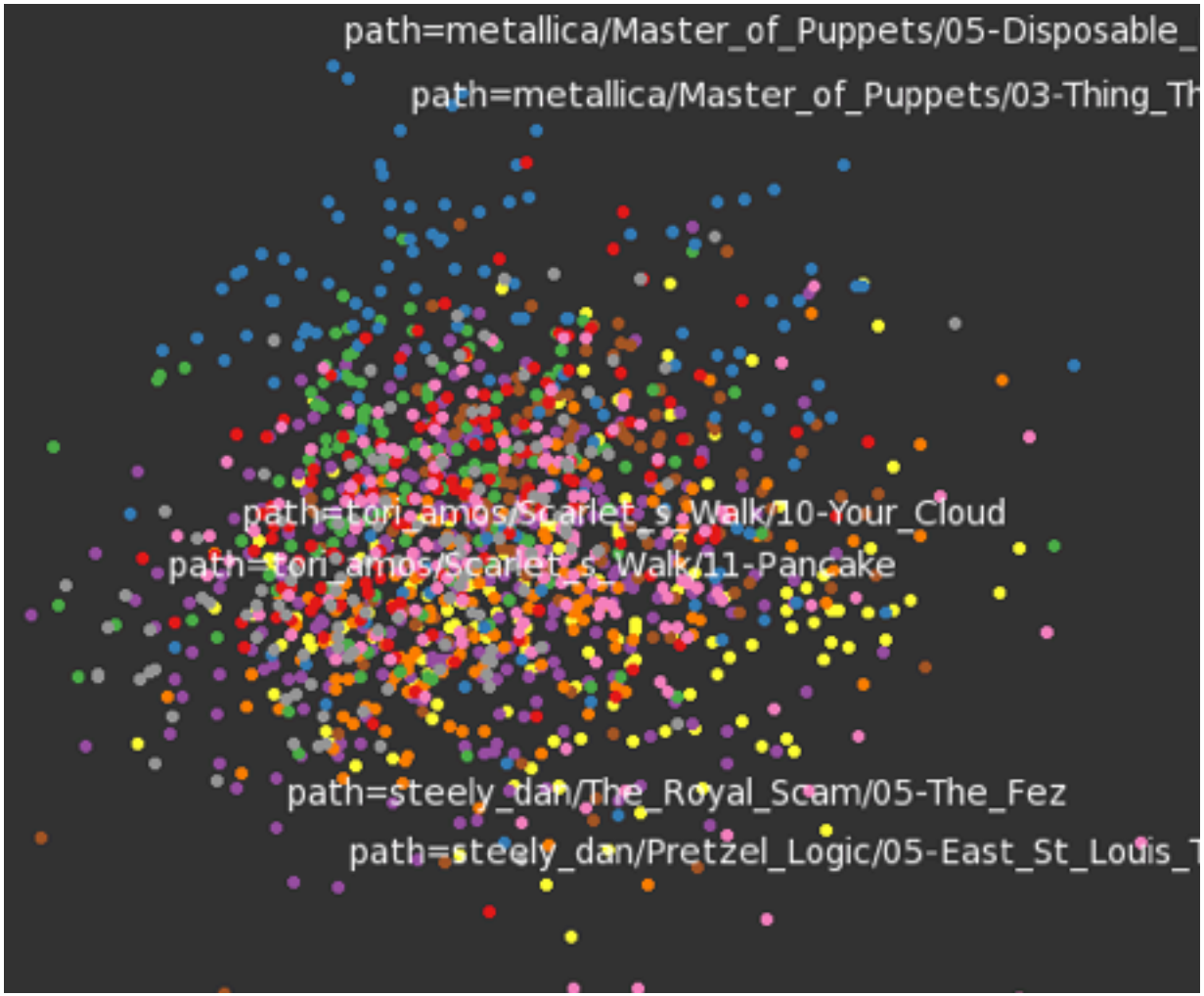
1. Features for Music Audio
2. Chroma Features
3. Chord Recognition

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

1. Features for Music Audio

- Challenges of large music databases
  - how to find "what we want"...

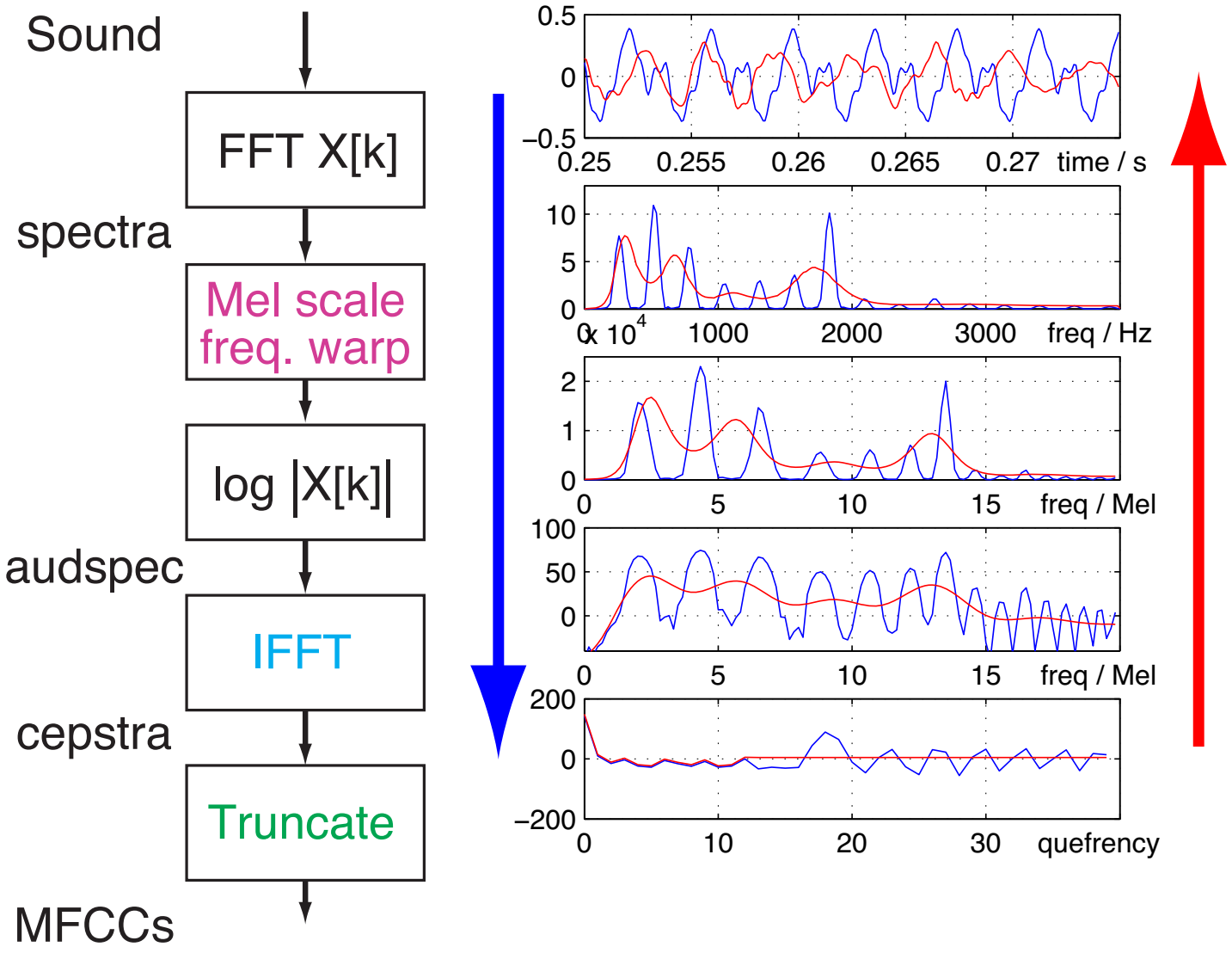
- Euclidean metaphor
  - music tracks as points in space
- What are the dimensions?
  - "sound" - timbre, instruments → MFCC
  - melody, chords → Chroma
  - rhythm, tempo → Rhythmic bases



MFCCs

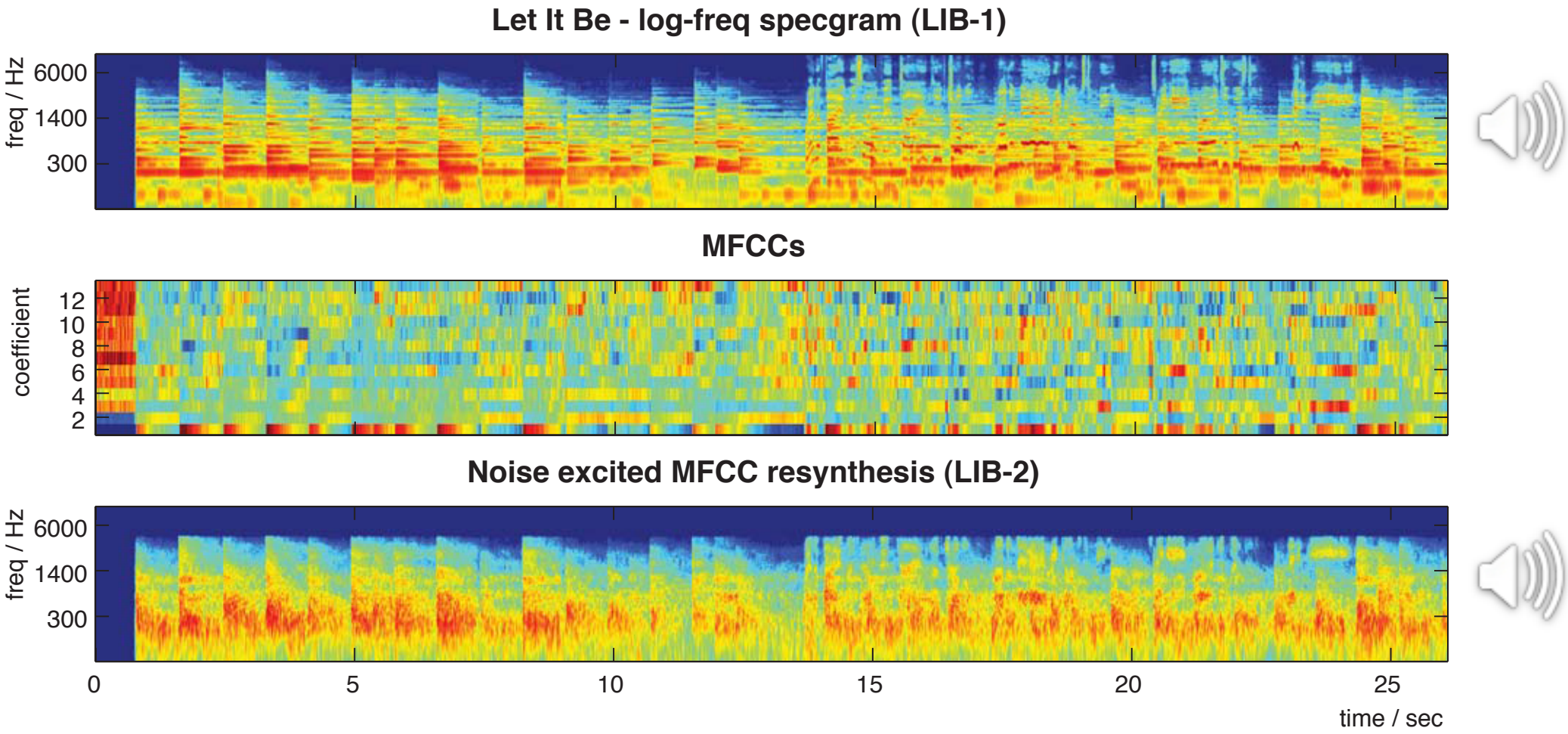
Logan 2000

- The standard feature for speech recognition



MFCC Example

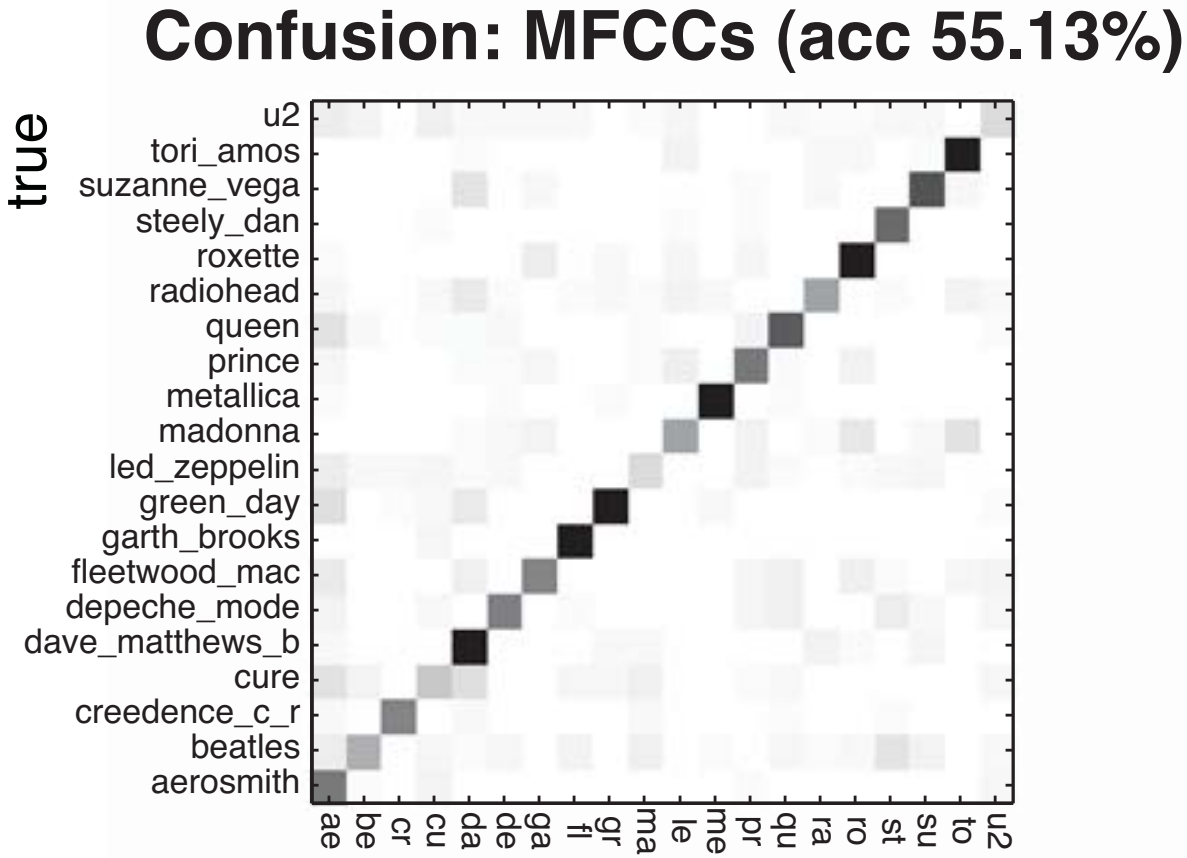
- Resynthesize by imposing spectrum on noise
  - MFCCs capture instruments, not notes



MFCC Artist Classification

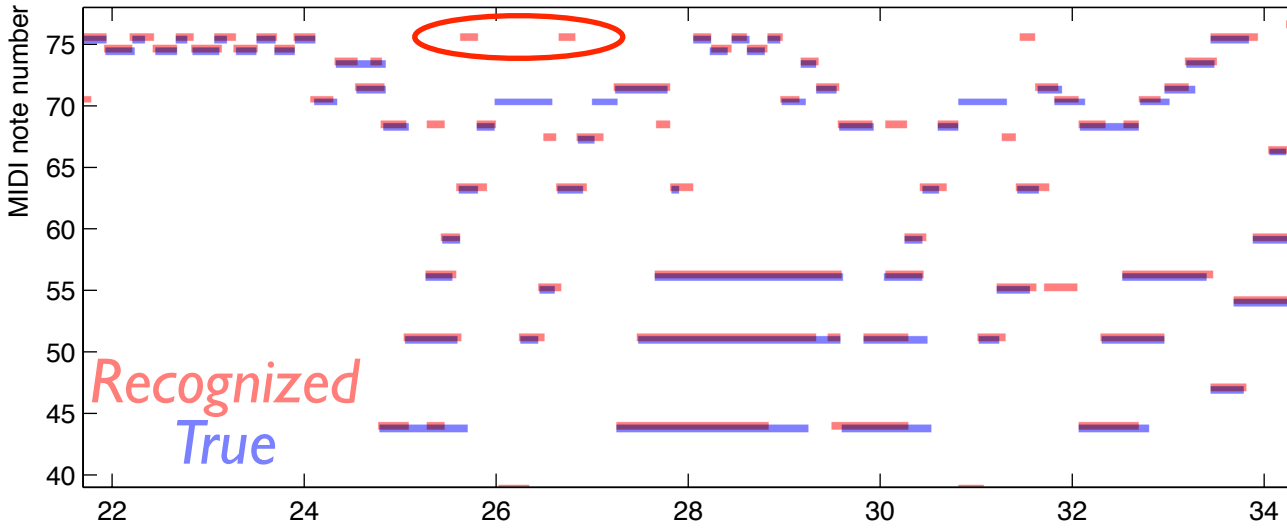
Ellis 2007

- **20 Artists x 6 albums each**
  - train models on 5 albums, classify tracks from last
- **Model as MFCC mean + covariance per artist**
  - "single Gaussian" model
  - 20 (mean) + 10 x 19 (covariance) parameters
  - 55% correct (guessing ~5%)



2. Chroma Features

- What about modeling tonal content (notes)?
  - melody spotting
  - chord recognition
  - cover songs...
- MFCCs exclude tonal content
- Polyphonic transcription is too hard
  - e.g. sinusoidal tracking: confused by harmonics

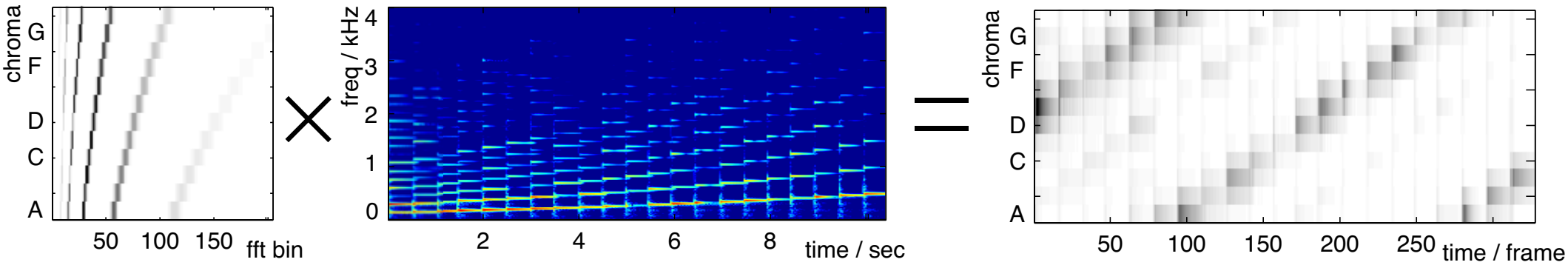


- Chroma features as solution...

Chroma Features

Fujishima 1999

- Idea: Project all energy onto 12 semitones regardless of octave
  - maintains main "musical" distinction
  - invariant to musical equivalence
  - no need to worry about harmonics?

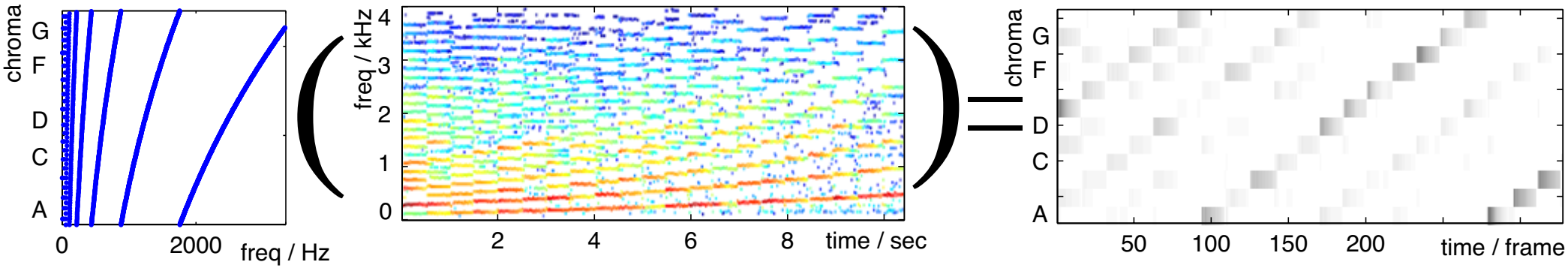


$$C(b) = \sum_{k=0}^{N_M} B(12\log_2(k/k_0) - b)W(k)|X[k]|$$

- $W(k)$ is weighting, $B(b)$ selects every ~ mod12

Better Chroma

- Problems:
  - blurring of bins close to edges
  - limitation of FFT bin resolution
- Solutions:
  - peak picking - only keep energy at center of peaks



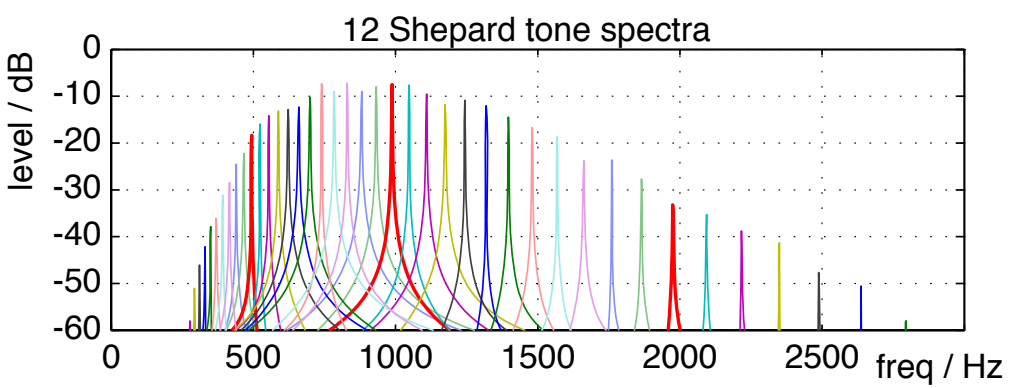
  - Instantaneous Frequency - high-resolution estimates
  - adapt tuning center based on histogram of pitches

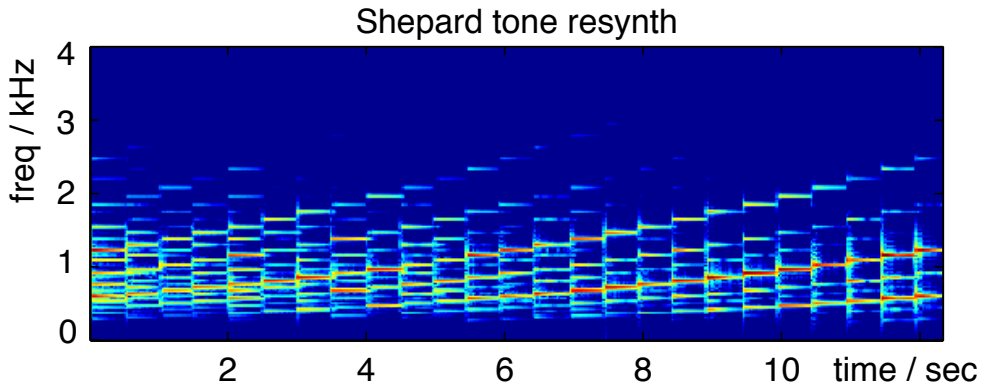
Chroma Resynthesis

Ellis & Poliner 2007

- Chroma describes the notes in an octave
  - ... but not the octave
- Can resynthesize by presenting all octaves
  - ... with a smooth envelope
  - "Shepard tones" - octave is ambiguous

$$y_b(t) = \sum_{o=1}^{M} W(o + \frac{b}{12}) \cos 2^{o+\frac{b}{12}} w_0 t$$

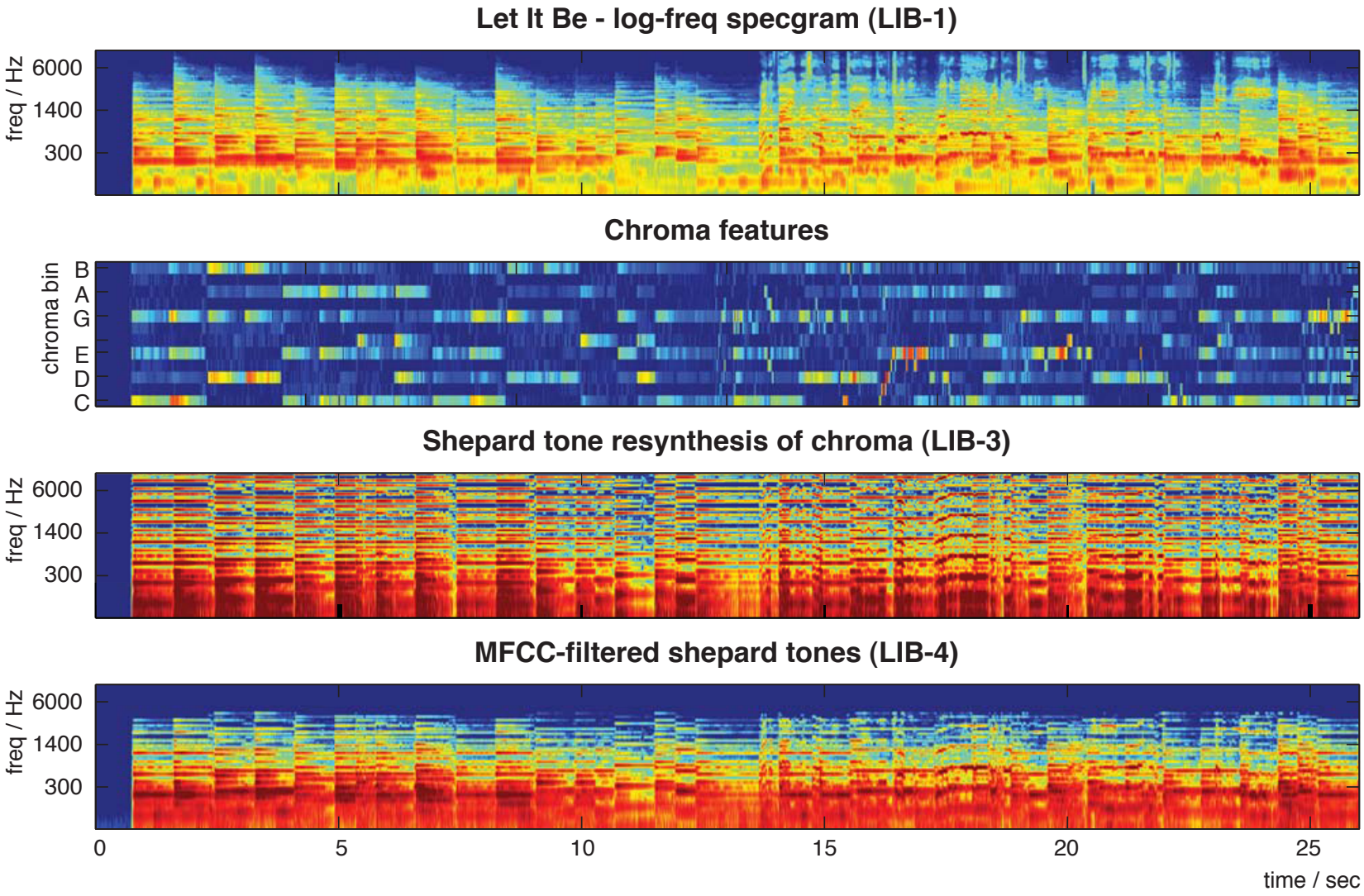




- endless sequence illusion

Chroma Example

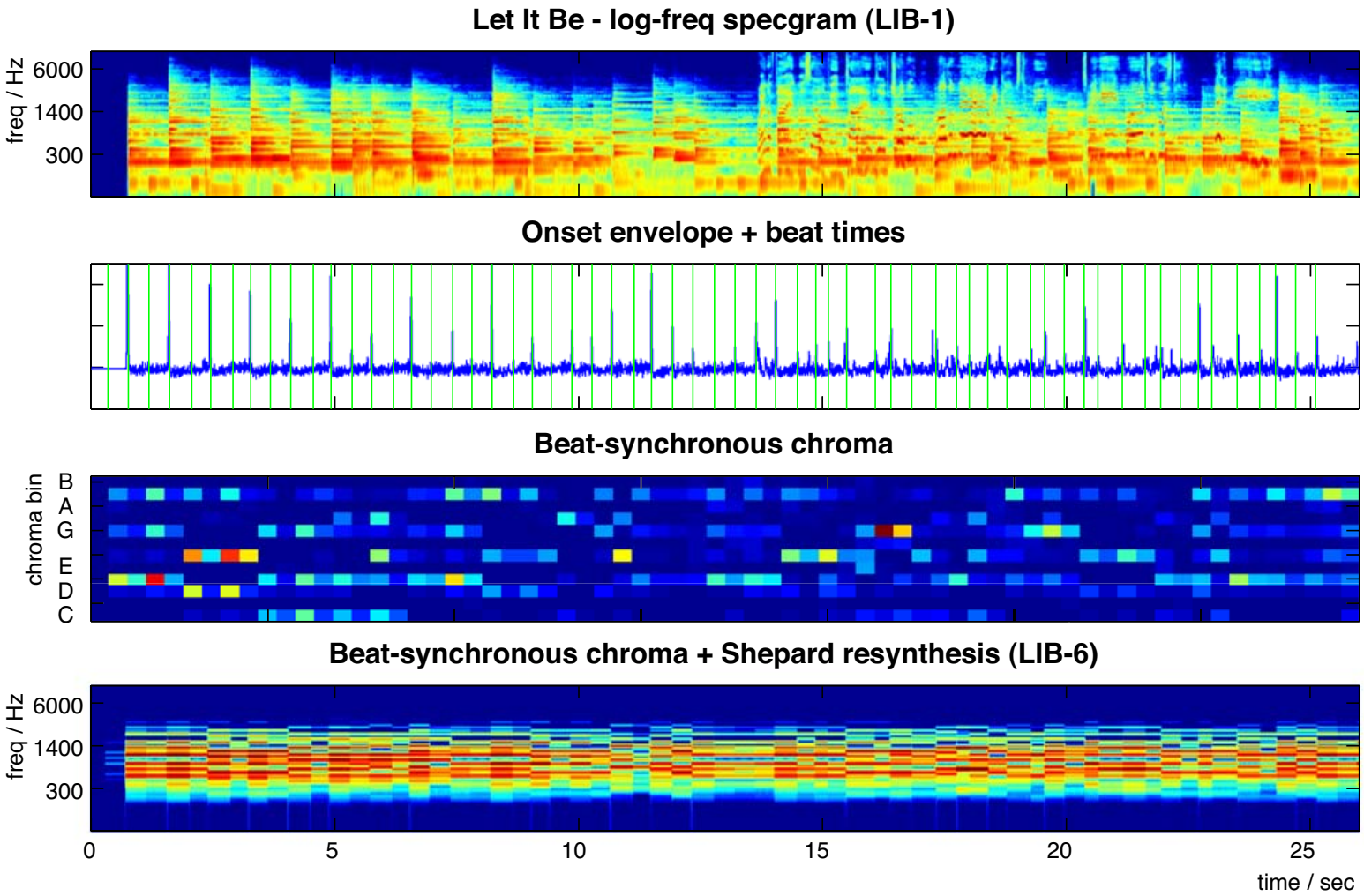
- Simple Shepard tone resynthesis
  - can also reimpose broad spectrum from MFCCs



Beat-Synchronous Chroma

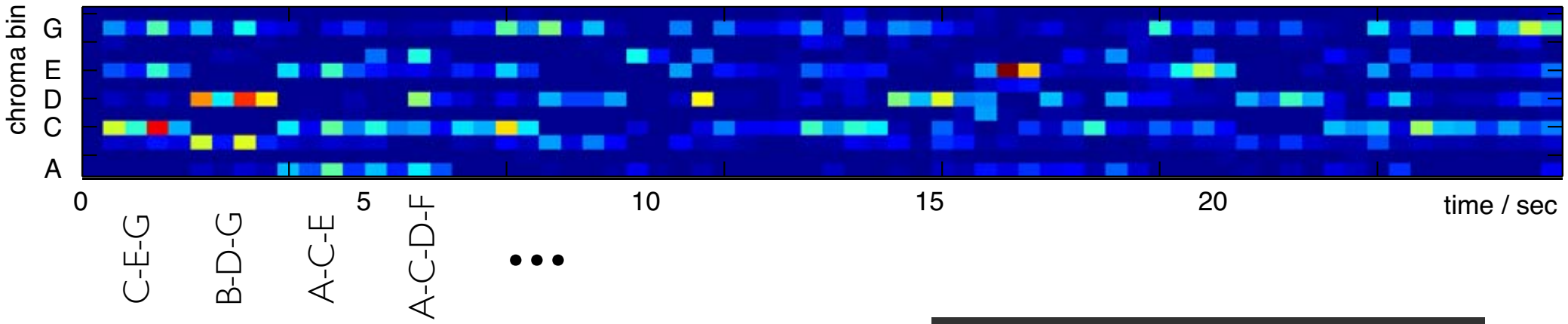
Bartsch & Wakefield 2001

- Drastically reduce data size by recording one chroma frame per beat



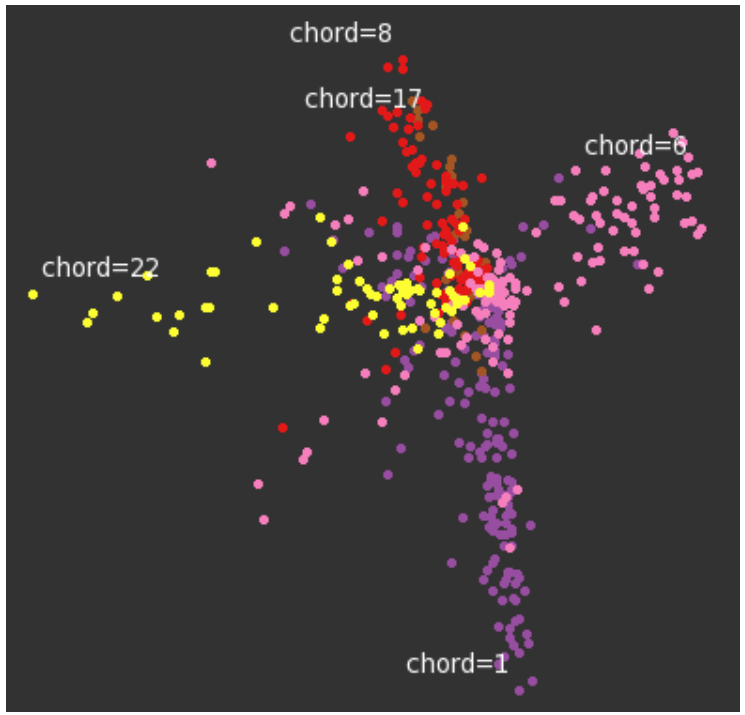
3. Chord Recognition

- Beat synchronous chroma look like chords



  - can we transcribe them?

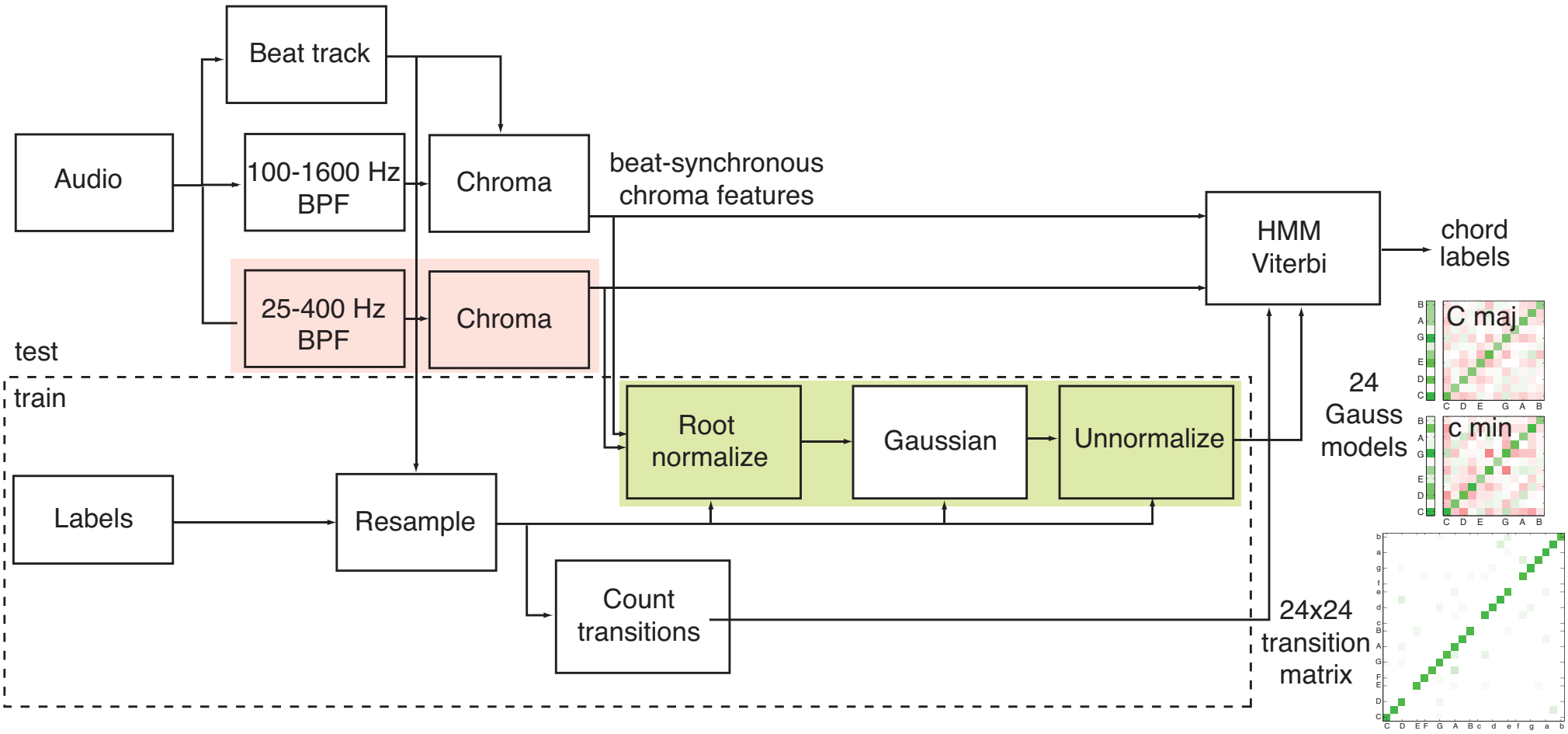
- Two approaches
  - manual templates (prior knowledge)
  - learned models (from training data)



Chord Recognition System

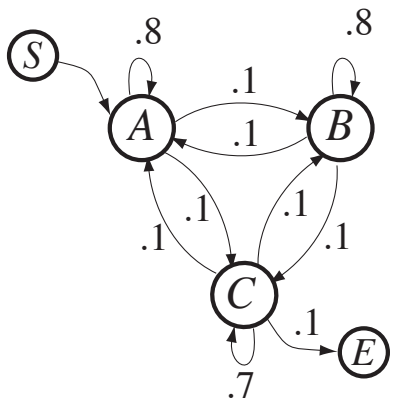
Sheh & Ellis 2003

- Analogous to speech recognition
  - Gaussian models of features for each chord
  - Hidden Markov Models for chord transitions



HMMs

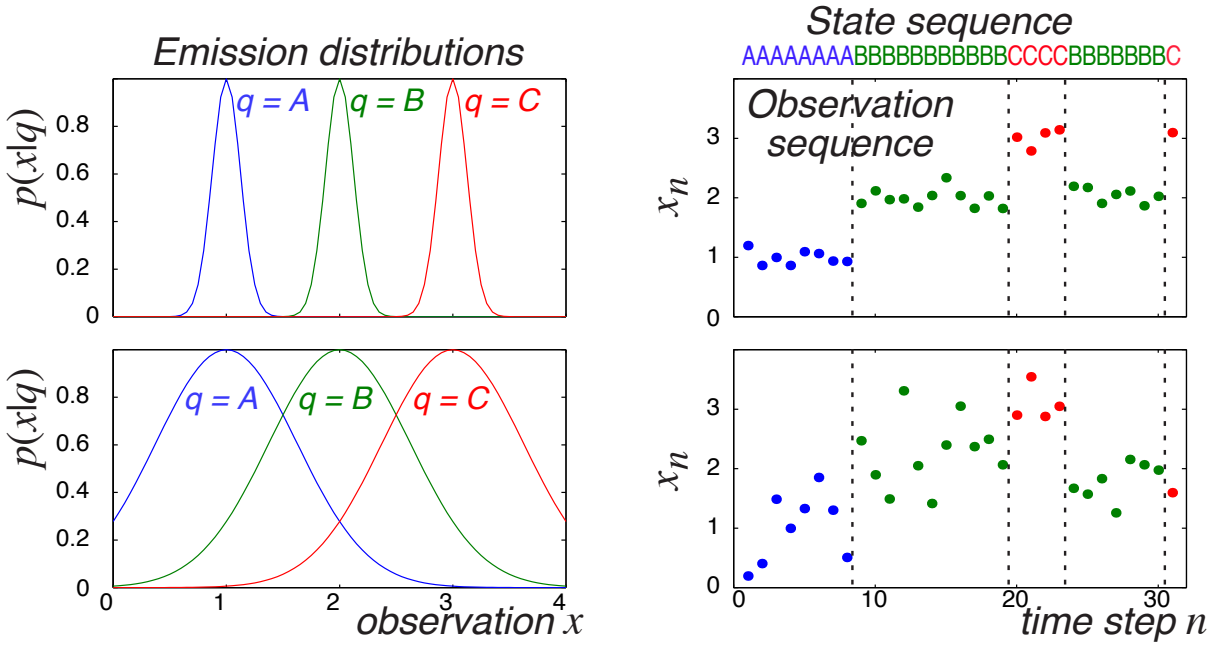
- Hidden Markov Models are good for inferring hidden states
  - underlying Markov "generative model"
  - each state has emission distribution



$p(q_{n+1}\|q_n)$		q_{n+1} S	A	B	C	E
q_n	S	0	1	0	0	0
	A	0	.8	.1	.1	0
	B	0	.1	.8	.1	0
	C	0	.1	.1	.7	.1
	E	0	0	0	0	1

S A A A A A A A A B B B B B B B B B C C C C B B B B B B C E

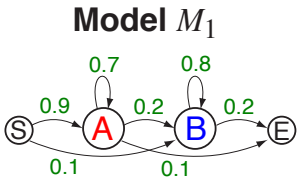
  - observations tell us *something* about state...
  - infer smoothed state sequence



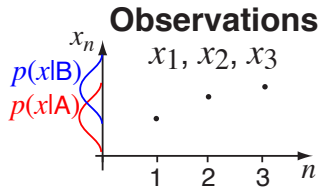
HMM Inference

- HMM defines emission distribution $p(x|q)$ and transition probabilities $p(q_n|q_{n-1})$
- Likelihood of observed given state sequence:

$$p(\{x_n\}|\{q_n\}) = \prod_n p(x_n|q_n)p(q_n|q_{n-1})$$

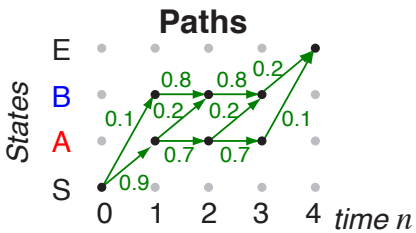


	S	A	B	E
S	•	0.9	0.1	•
A	•	0.7	0.2	0.1
B	•	•	0.8	0.2
E	•	•	•	1



Observation likelihoods

$p(x\|q)$	x_1	x_2	x_3
A	2.5	0.2	0.1
B	0.1	2.2	2.3



All possible 3-emission paths Q_k from S to E

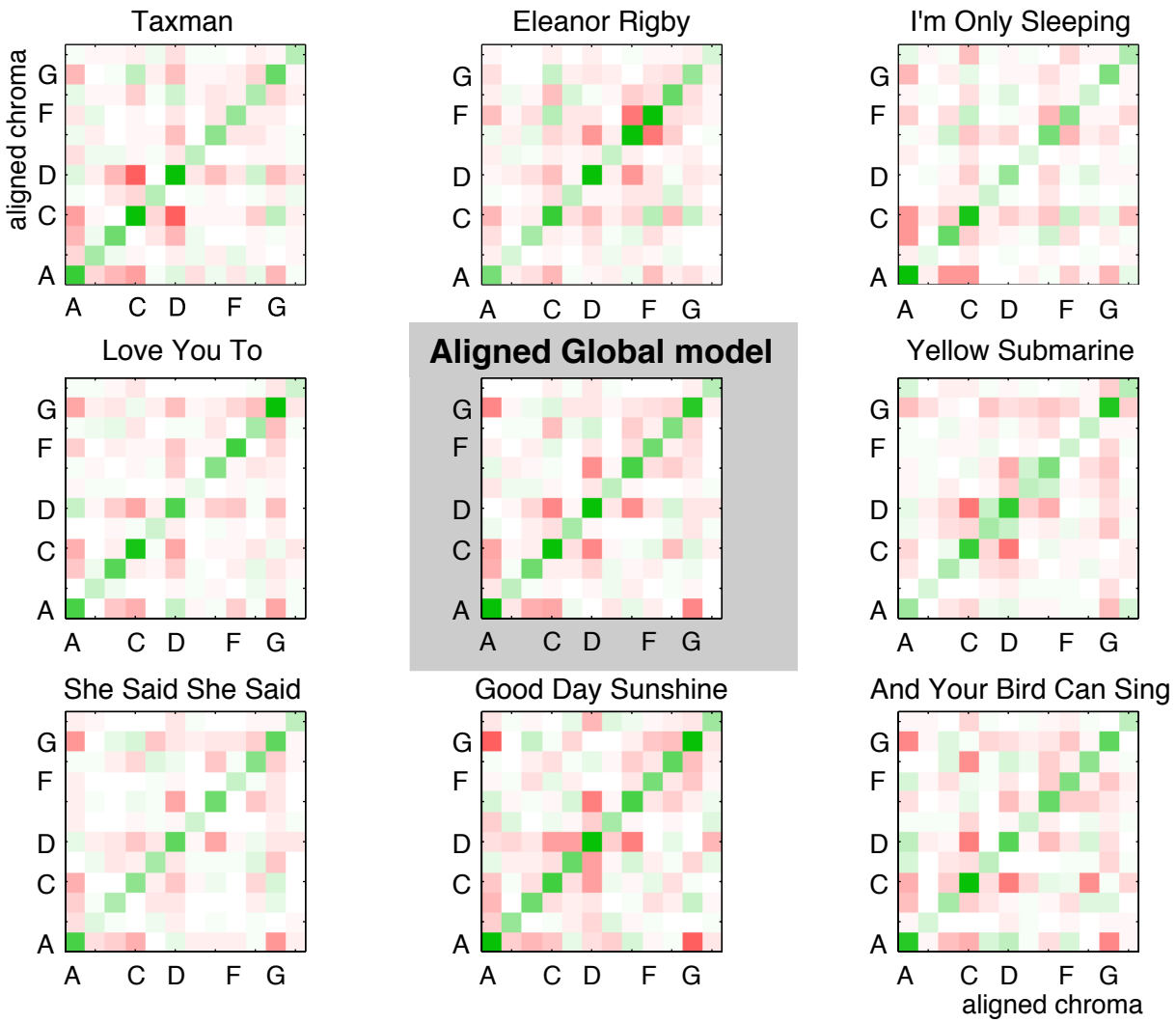
q_0	q_1	q_2	q_3	q_4	$p(Q \mid M) = \prod_n p(q_n\|q_{n-1})$	$p(X \mid Q,M) = \prod_n p(x_n\|q_n)$	$p(X,Q \mid M)$
S	A	A	A	E	.9 x .7 x .7 x .1 = **0.0441**	2.5 x 0.2 x 0.1 = 0.05	0.0022
S	A	A	B	E	.9 x .7 x .2 x .2 = 0.0252	2.5 x 0.2 x 2.3 = 1.15	0.0290
S	A	B	B	E	.9 x .2 x .8 x .2 = 0.0288	2.5 x 2.2 x 2.3 = 12.65	**0.3643**
S	B	B	B	E	.1 x .8 x .8 x .2 = 0.0128	0.1 x 2.2 x 2.3 = 0.506	0.0065
					Σ = 0.1109		$\Sigma = p(X \mid M) =$ **0.4020**

- By dynamic programming, we can also identify the *best* state sequence given just the observations

Key Normalization

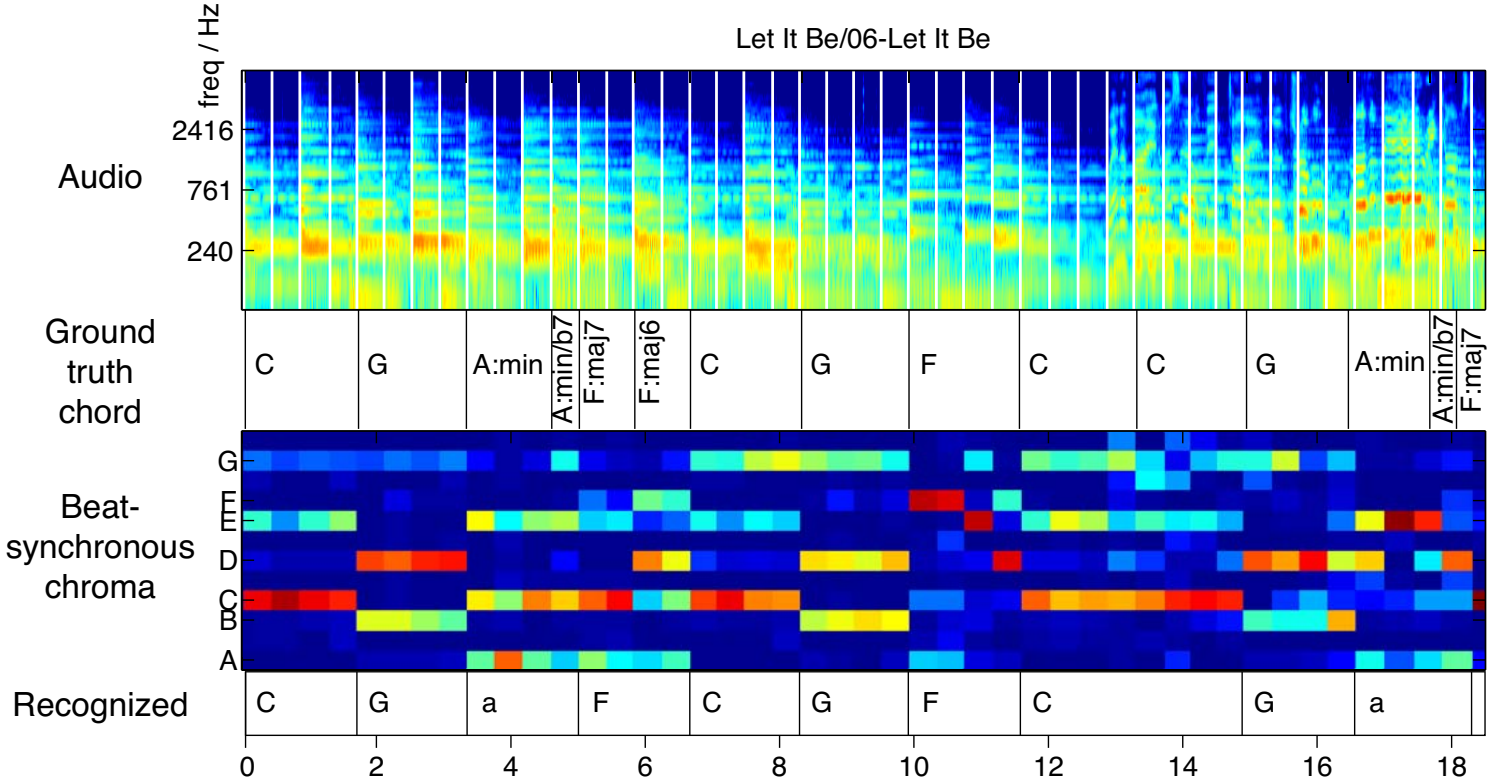
- Chord transitions depend on key of piece
  - dominant, relative minor, etc...

- Chord transition probabilities should be key-relative
  - estimate main key of piece
  - rotate all chroma features
  - learn models



Chord Recognition

- Often works:



- But only about 60% of the time

	12 chroma	+bass
indep. models	0.539	0.552
pooled models	0.556	0.578

Summary

- Music Audio Features
  capture information useful for classification

- Chroma Features
  12 bins to robustly summarize notes

- Chord Recognition
  Sometimes easy, sometimes subtle

References

- B. Logan, "Mel frequency cepstral coefficients for music modeling," in *Proc. Int. Symp. Music Inf. Retrieval ISMIR*, Plymouth, September 2000.
- D. Ellis, "Classifying Music Audio with Timbral and Chroma Features," in *Proc. Int. Symp. Music Inf. Retrieval ISMIR-07*, pp. 339-340, Vienna, October 2007.
- T. Fujishima, "Realtime chord recognition of musical sound: A system using common lisp music," In *Proc. Int. Comp. Music Conf.*, pp. 464–467, Beijing, 1999.
- D. Ellis and G. Poliner, "Identifying Cover Songs With Chroma Features and Dynamic Programming Beat Tracking," *Proc. ICASSP-07*, pp. IV-1429-1432, Hawai'i, April 2007.
- M. A. Bartsch and G. H. Wakefield, "To catch a chorus: Using chroma-based representations for audio thumbnailing," in *Proc. IEEE WASPAA*, Mohonk, October 2001.
- A. Sheh and D. Ellis, "Chord Segmentation and Recognition using EM-Trained Hidden Markov Models," *Int. Symp. Music Inf. Retrieval ISMIR-03*, pp. 185-191, Baltimore, October 2003.