

# Data Driven Music Understanding

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: What is Music?
2. Eigenrhythms
3. Melodic-Harmonic Fragments
4. Example Applications

# LabROSA : Machine Listening

- Extracting **useful information** from sound
  - ... like (we) animals do

Task	Describe	Automatic Narration	Emotion	Music Recommendation
	Classify	Environment Awareness	ASR	Music Transcription
	Detect	“Sound Intelligence”	VAD	Speech/Music
		Environmental Sound	Speech	Music <i>Domain</i>

# I. Motivation: What is music?

- What does music **evoke** in a listener's **mind**?
- Which are the **things** that we call “**music**”?



# Oodles of Music

*Bertin-Mahieux et al. '09*

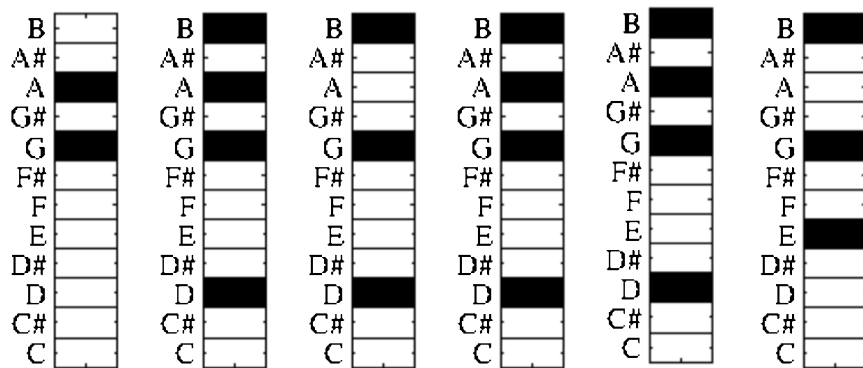


- What can you do with a million tracks?

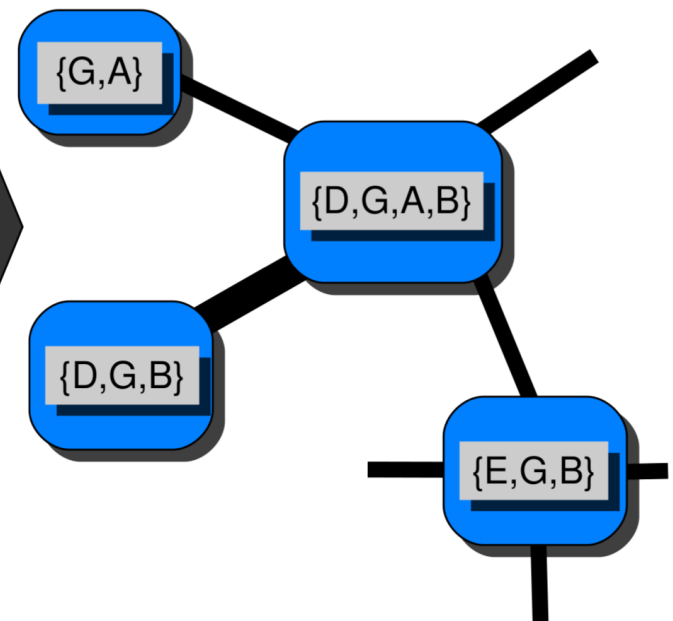
# Re-use in Music

Serrà et al. 2012

- What are the most **popular chord progressions** in pop music?



Pitch transition network



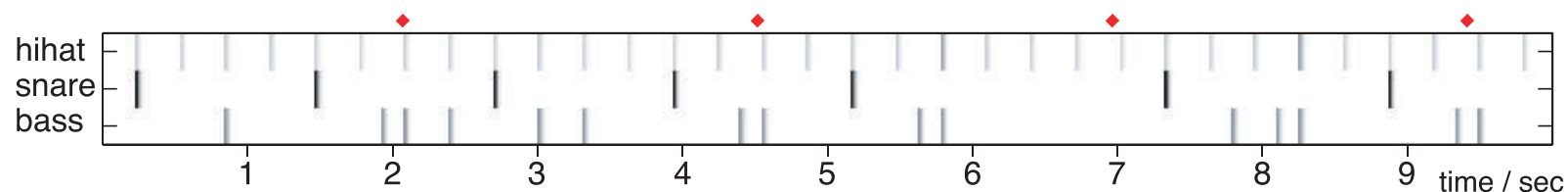
# Potential Applications

- Compression
- Classification
- Manipulation

## 2. Eigenrhythms: Drum Track Structure

*Ellis & Arroyo '04*

- To first order,  
all pop music has the **same beat**:

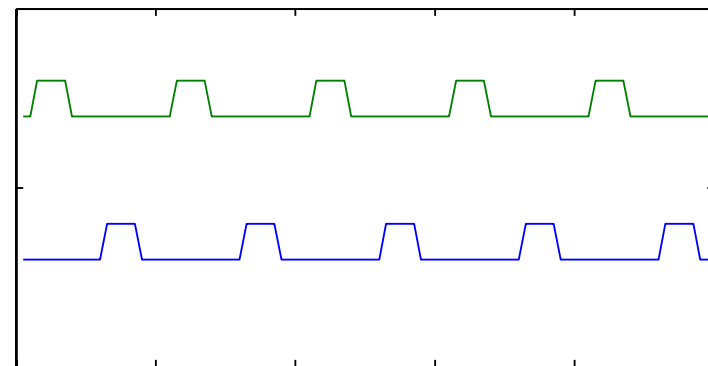
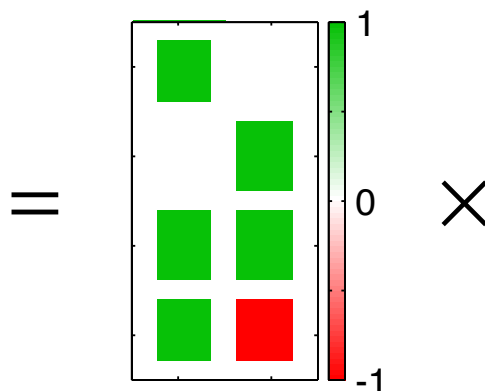
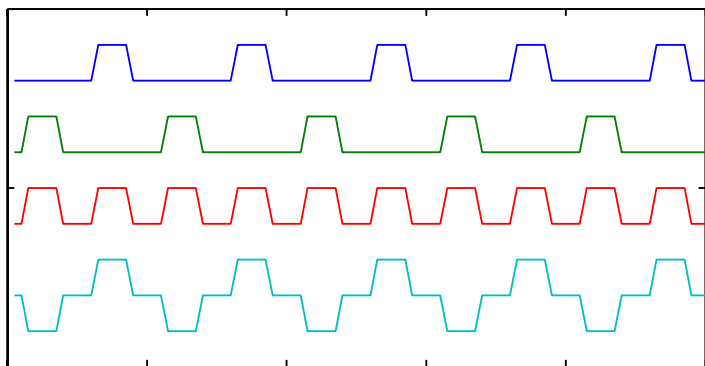


- Can we **learn** this from **examples**?

# Basis Sets

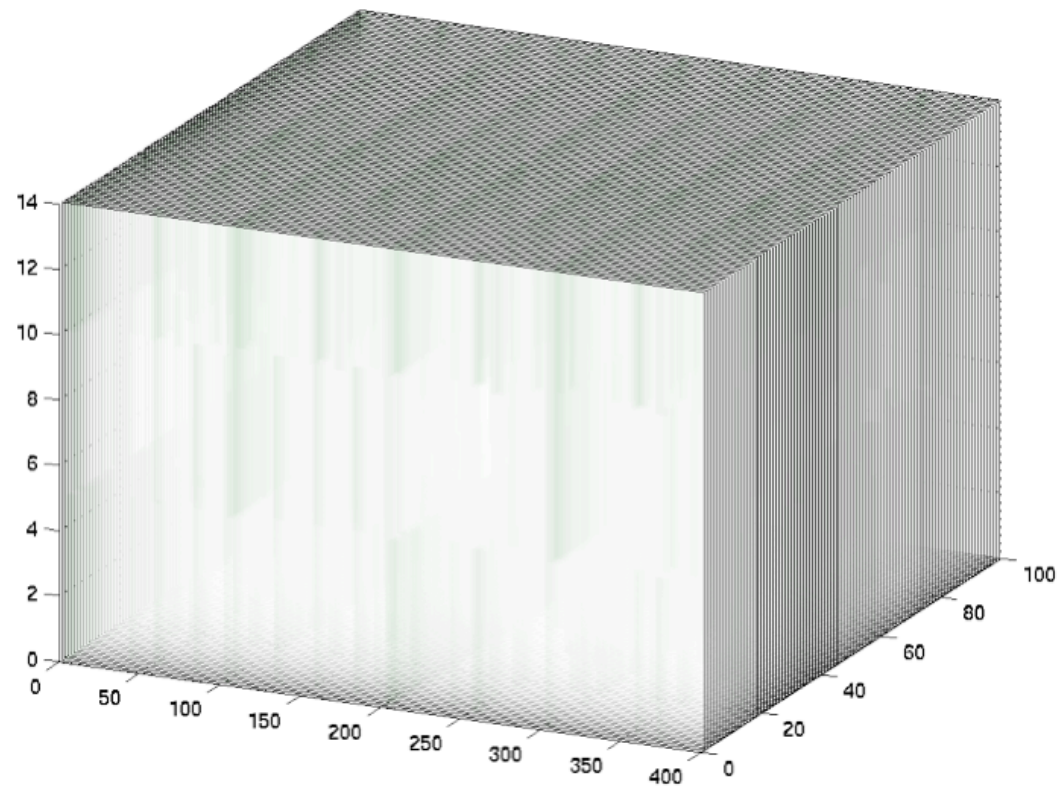
- Combine a few basic patterns to make a larger dataset

$$\text{data } X = \text{weights } W \times \text{patterns } H$$



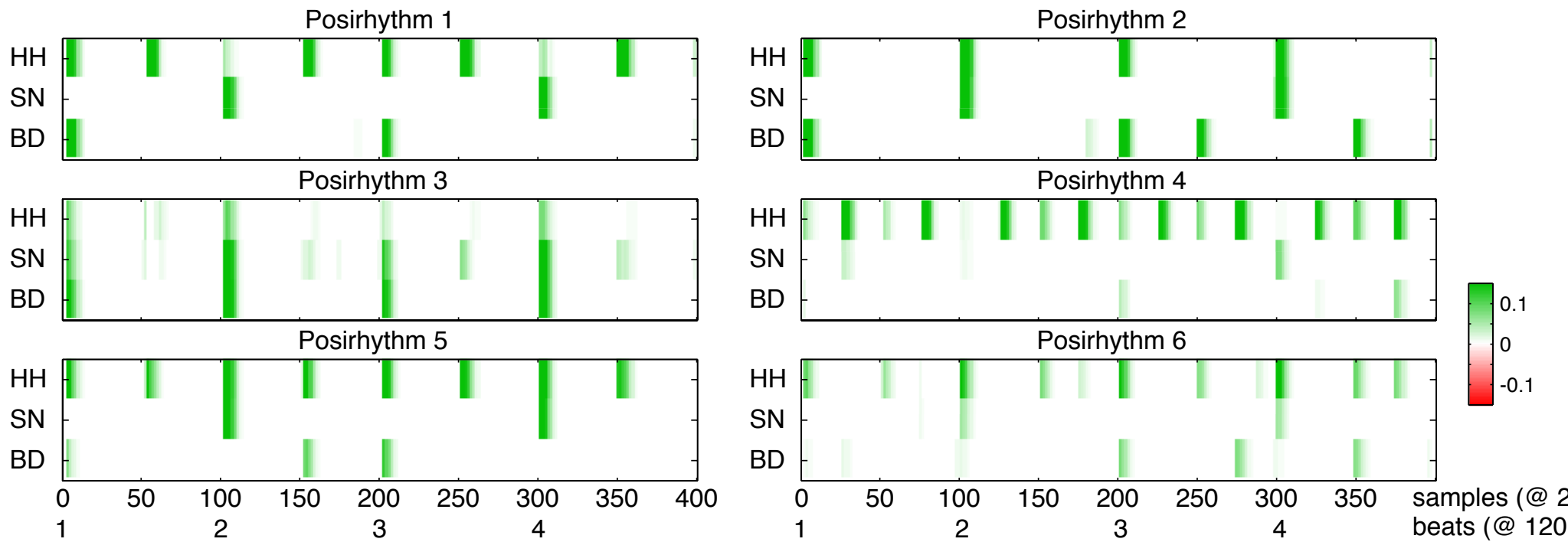


# Drum Pattern Data



- Tempo normalization + downbeat **alignment**

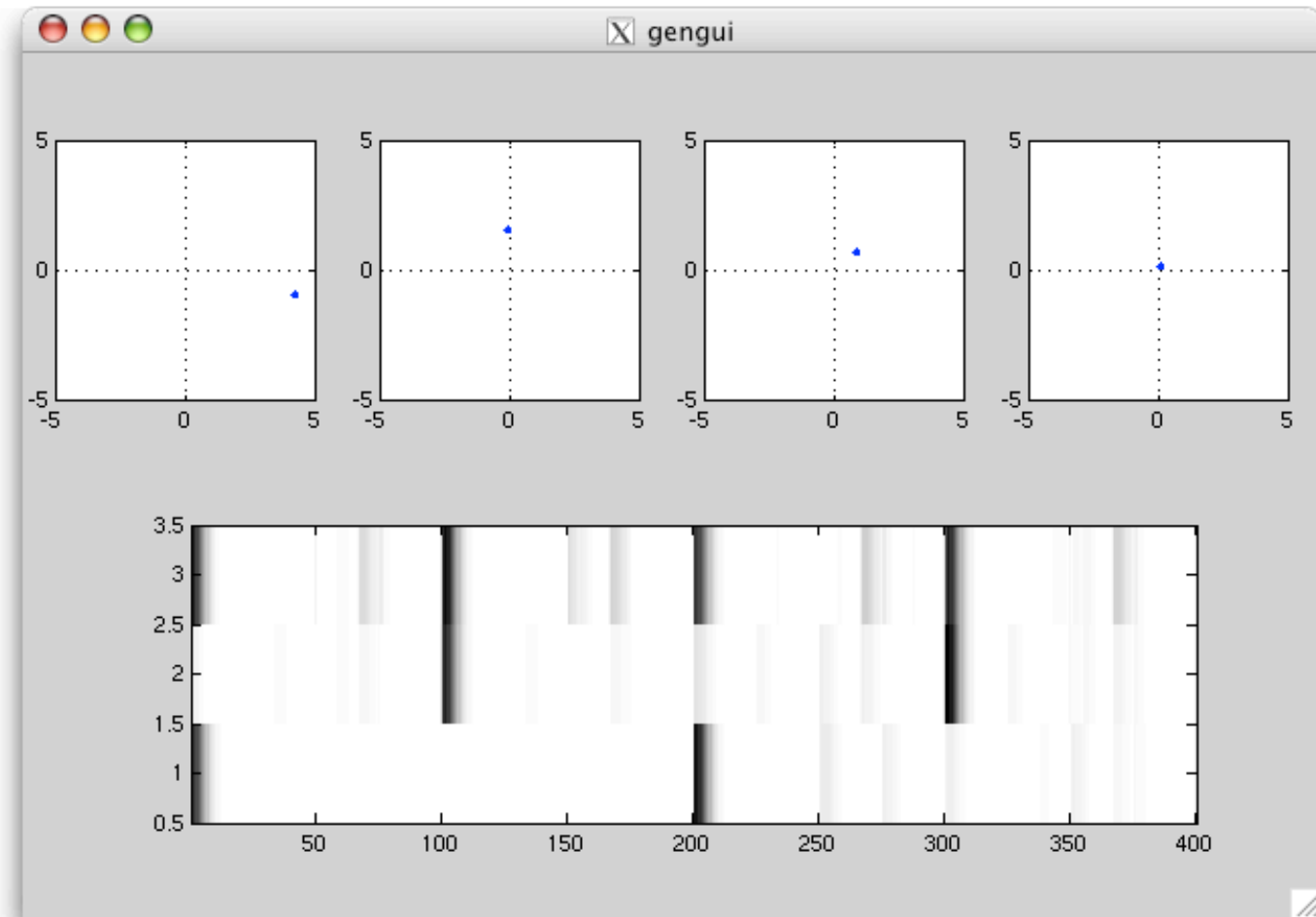
# NMF Eigenrhythms



- Nonnegative: only **add** beat-weight

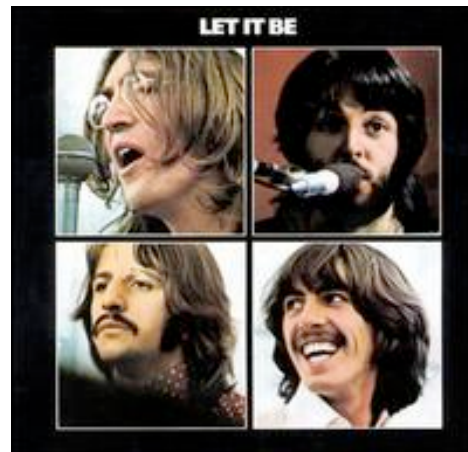
# Eigenrhythm BeatBox

- Resynthesize rhythms from eigen-space

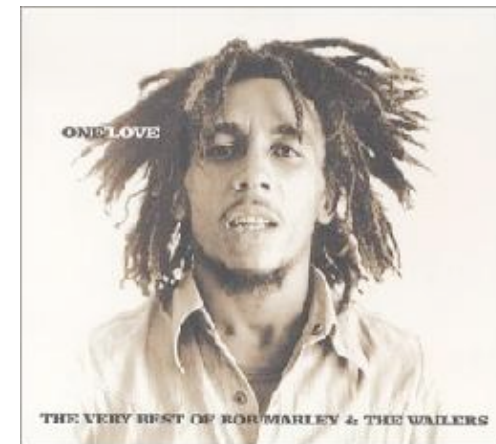


# 3. Melodic-Harmonic Fragments

- How **similar** are two pieces?



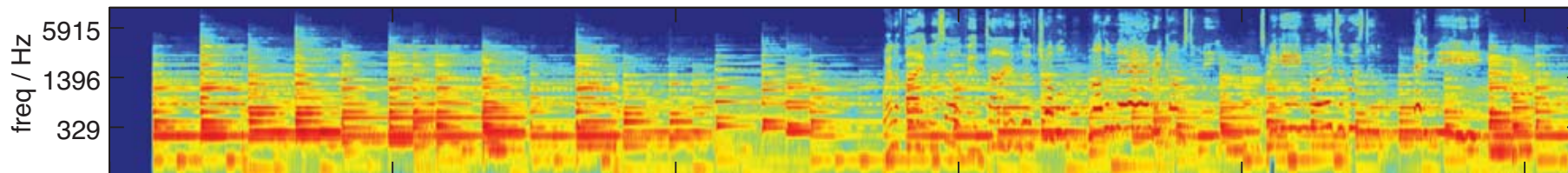
- Can we find all the pop-music **clichés**?



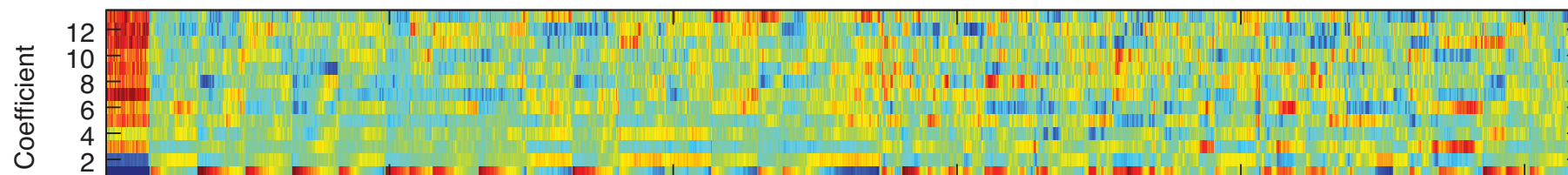
# MFCC Features

- Used in **speech recognition**

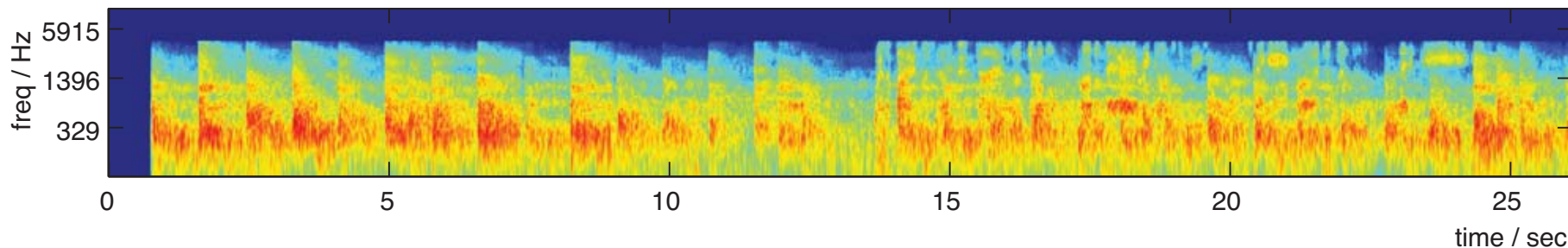
Let It Be (LIB-1) - log-freq specgram



MFCCs



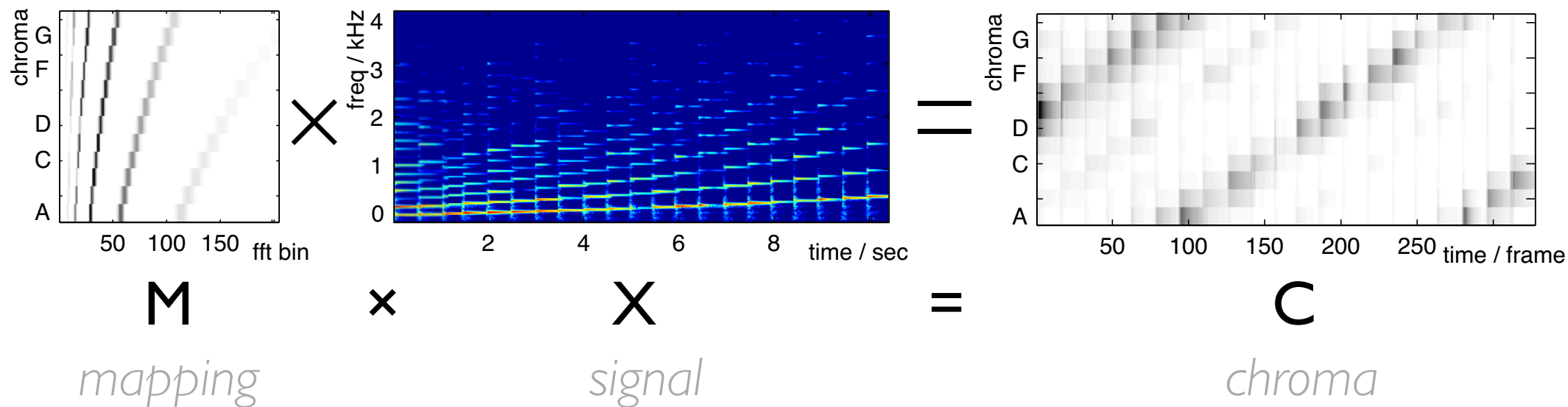
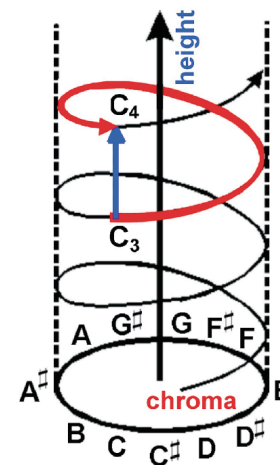
Noise excited MFCC resynthesis (LIB-2)



# Chroma Features

Fujishima 1999

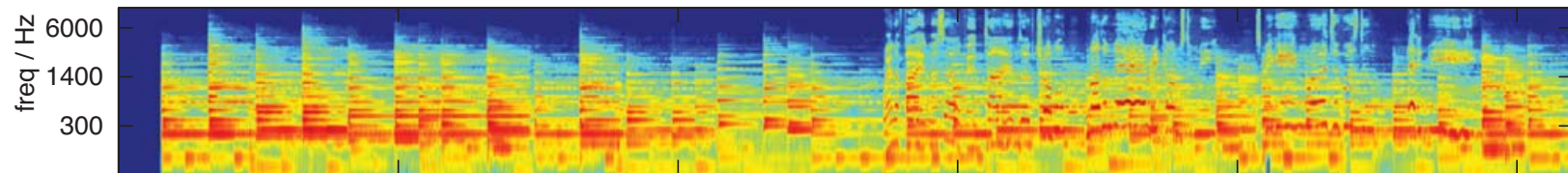
- Idea:  
Project onto **12 semitones**  
regardless of **octave**



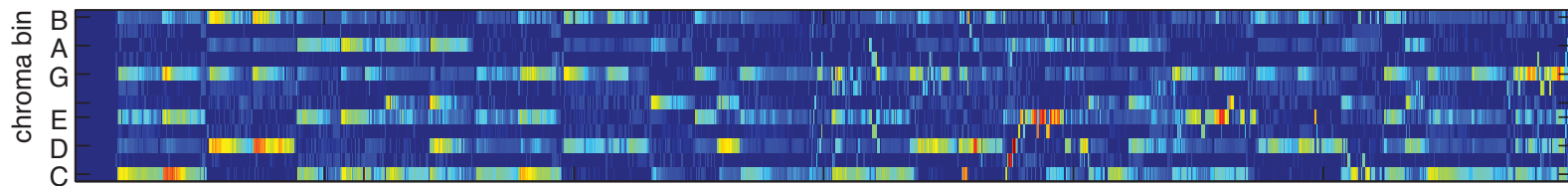
# Chroma Features

- To capture “musical” content

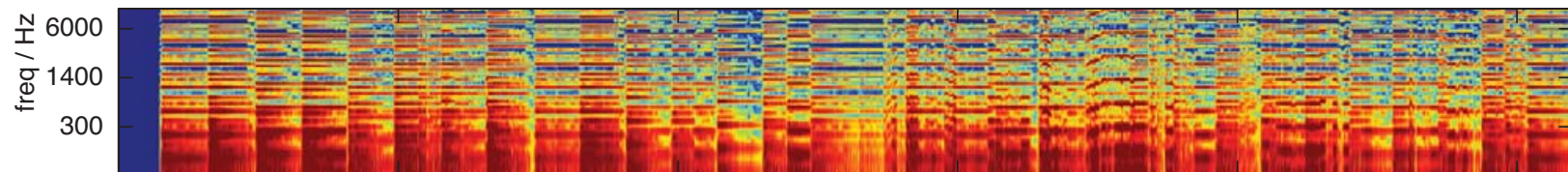
Let It Be - log-freq specgram (LIB-1)



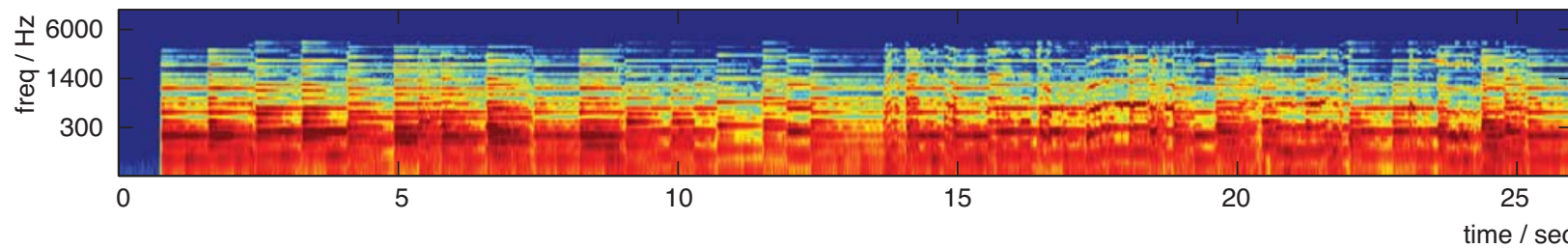
Chroma features



Shepard tone resynthesis of chroma (LIB-3)



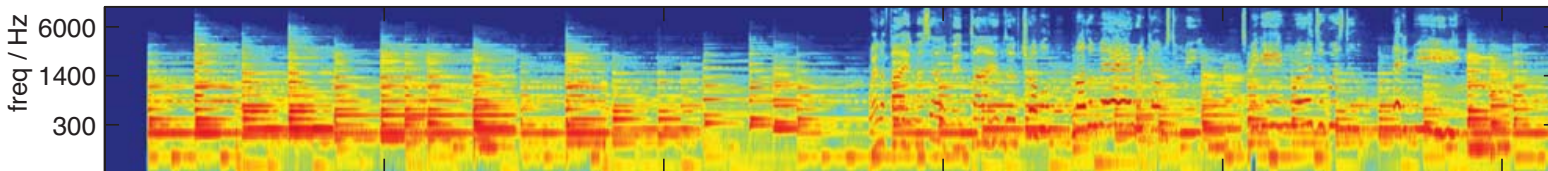
MFCC-filtered shepard tones (LIB-4)



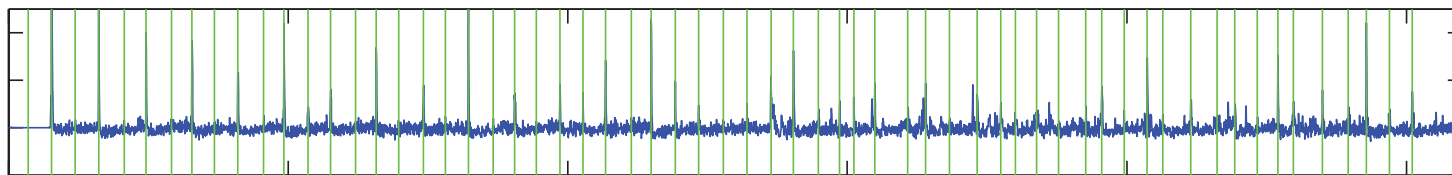
# Beat-Synchronous Chroma

- Compact representation of harmonies

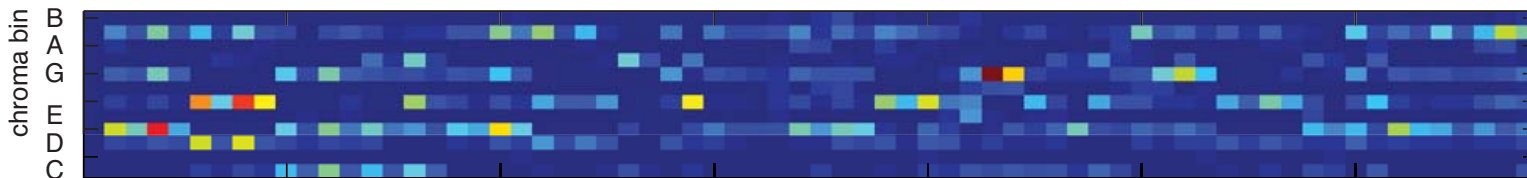
Let It Be - log-freq specgram (LIB-1)



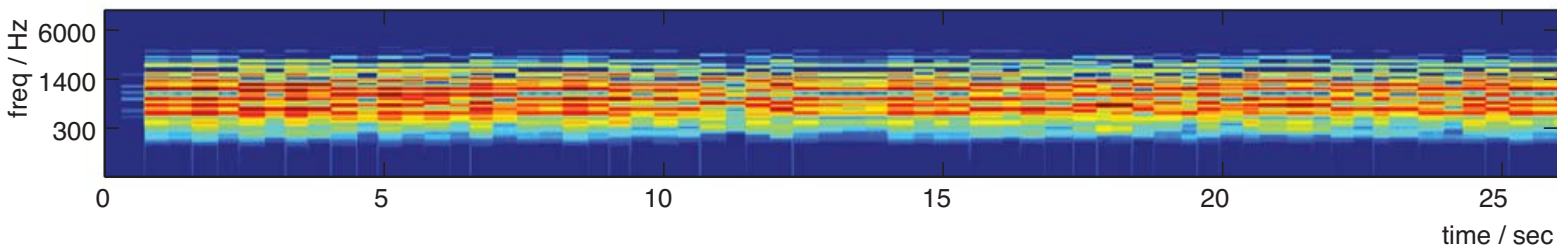
Onset envelope + beat times



Beat-synchronous chroma



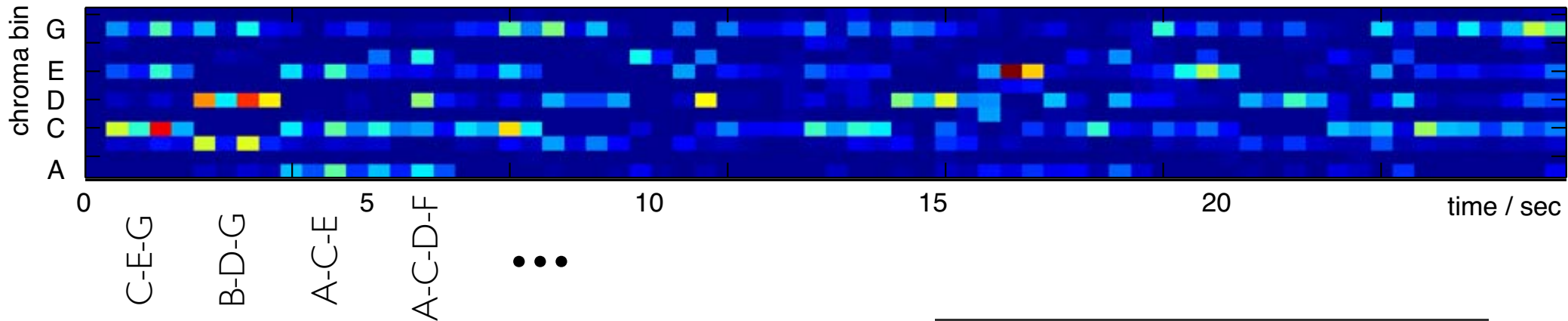
Beat-synchronous chroma + Shepard resynthesis (LIB-6)





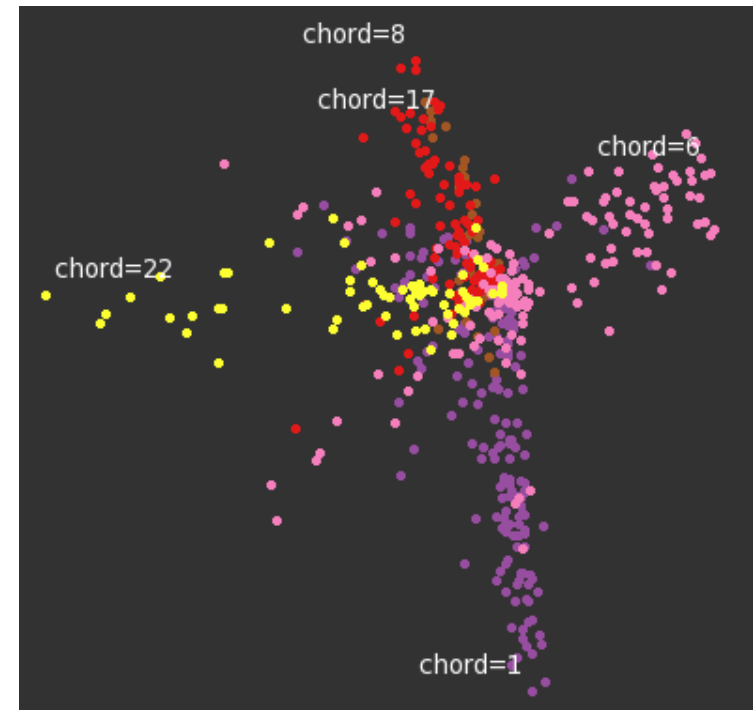
# Chord Recognition

- Beat synchronous chroma look like **chords**



- can we transcribe them?

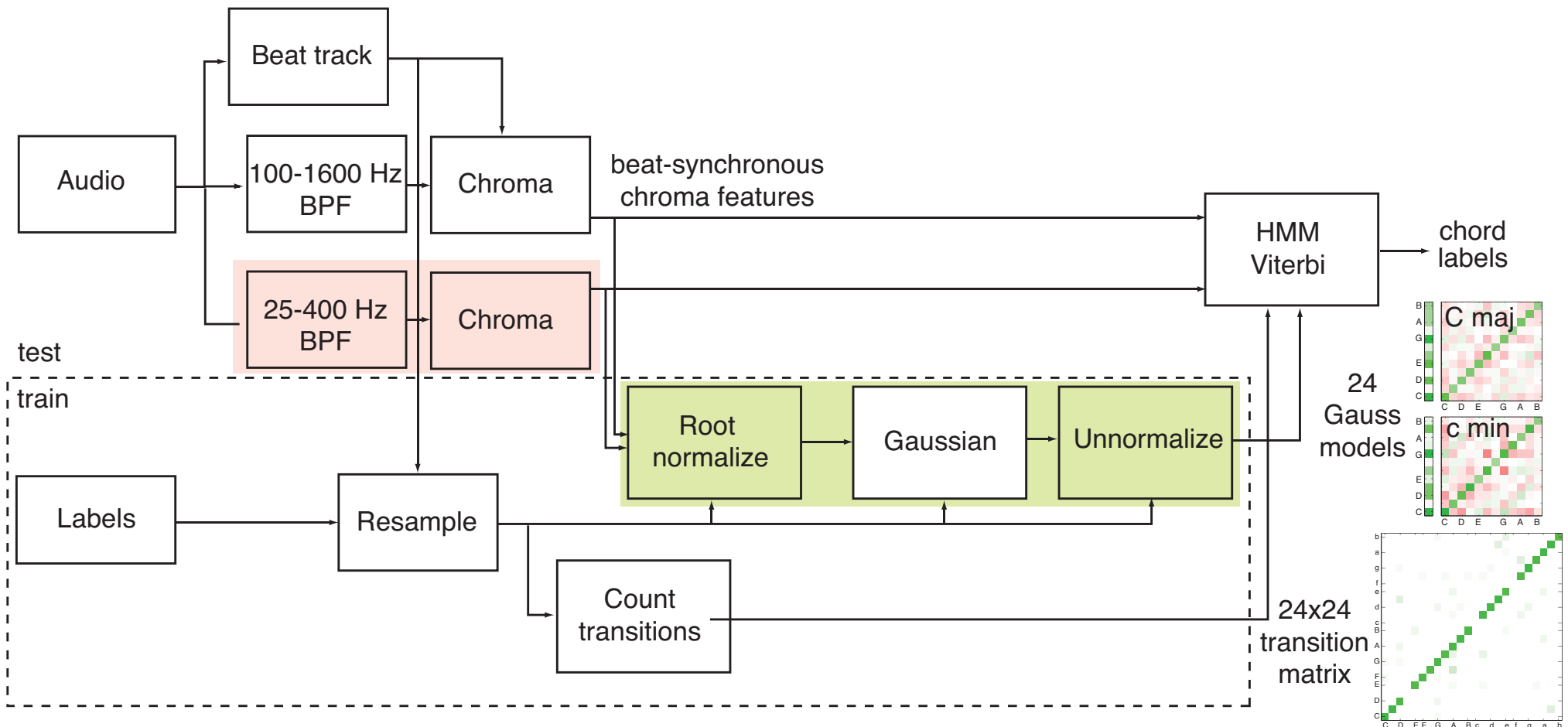
- **Two approaches**
  - **manual templates**  
(prior knowledge)
  - **learned models**  
(from training data)



# Chord Recognition System

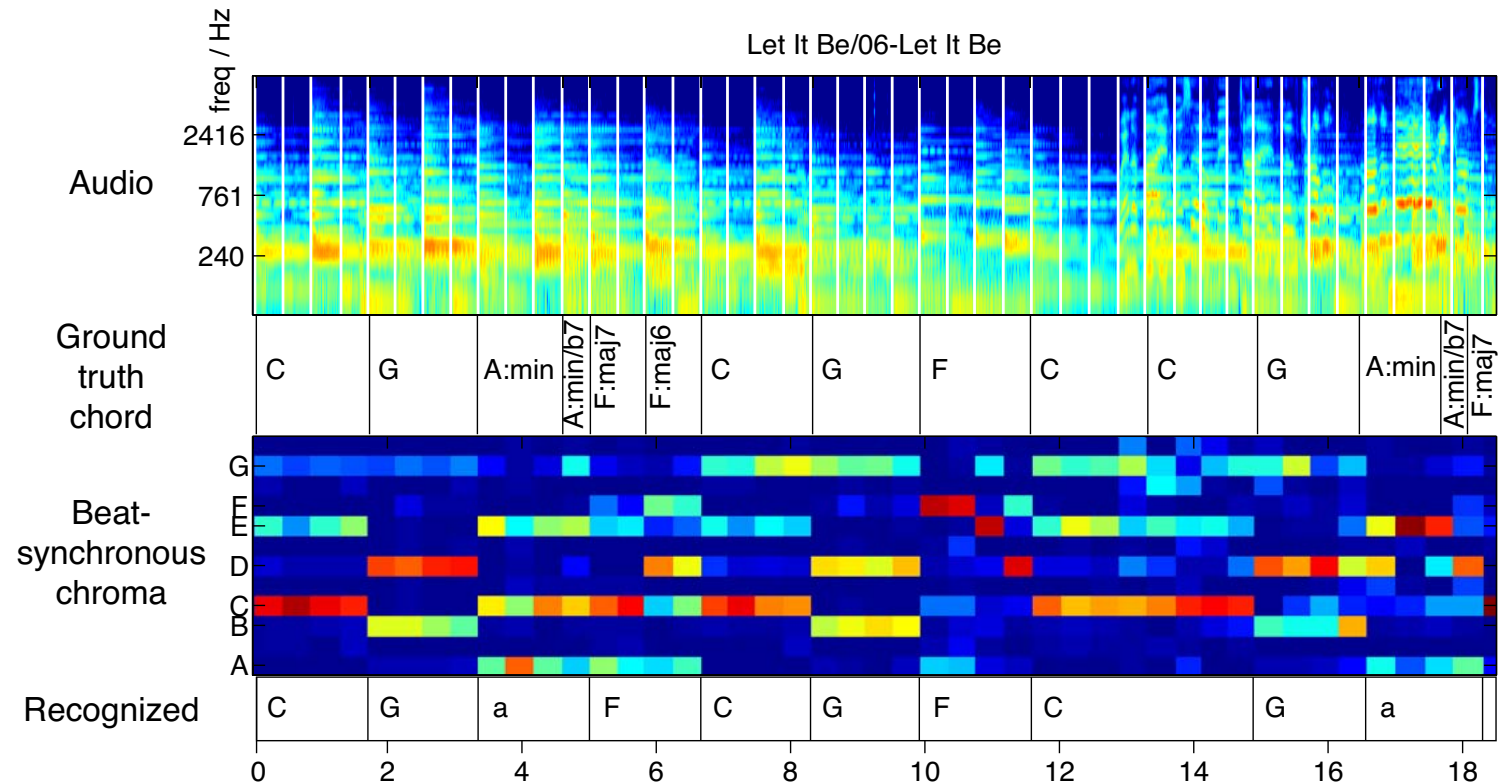
Sheh & Ellis 2003  
Ellis & Weller 2010

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



# Chord Recognition

- Often works:

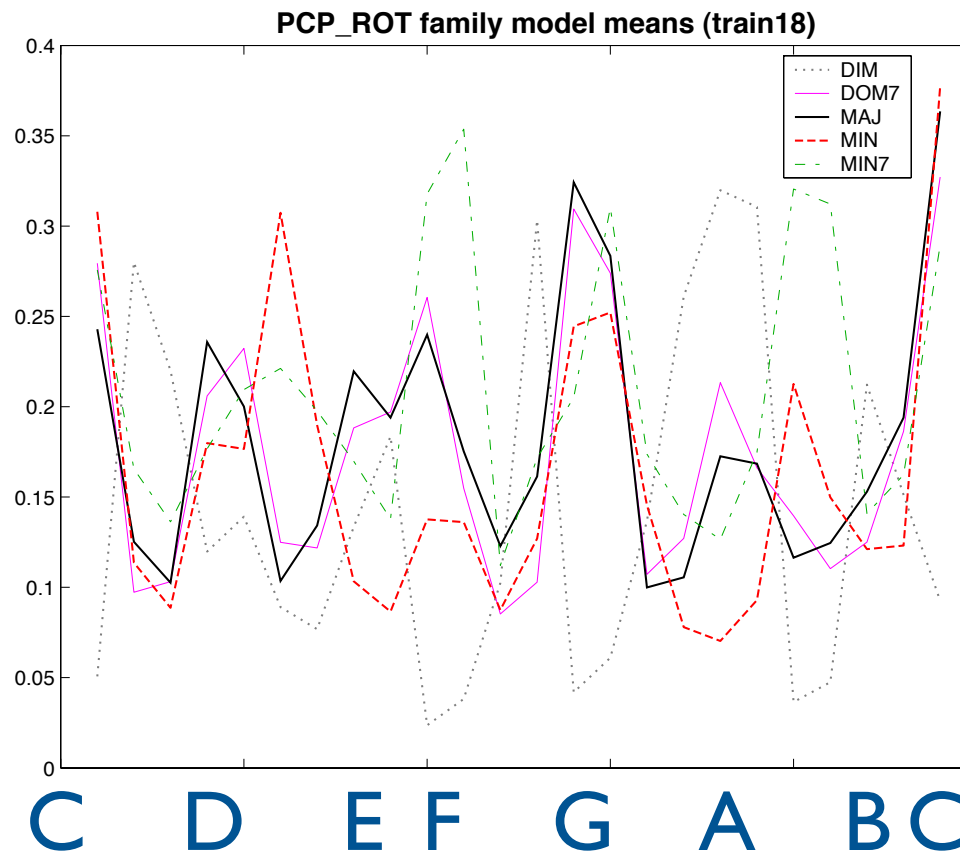


- But only 60-80% of the time

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

# What did the models learn?

- Chord model centers (**means**) indicate chord **'templates'**:



(for C-root chords)

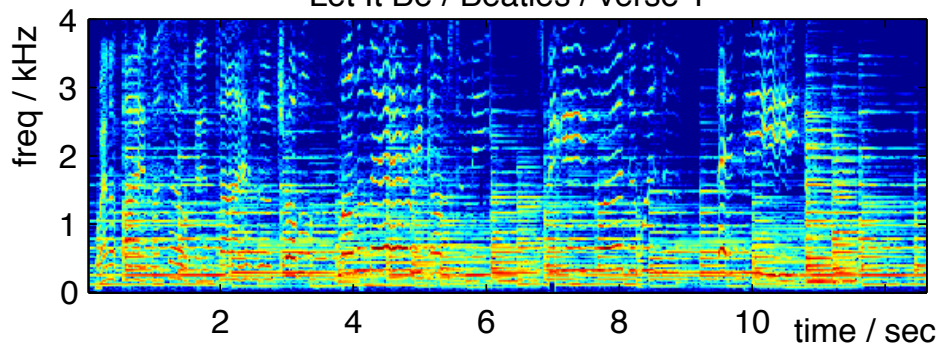
# Finding 'Cover Songs'

Ellis & Poliner '07  
Ravuri & Ellis '10

- Little similarity in surface audio...

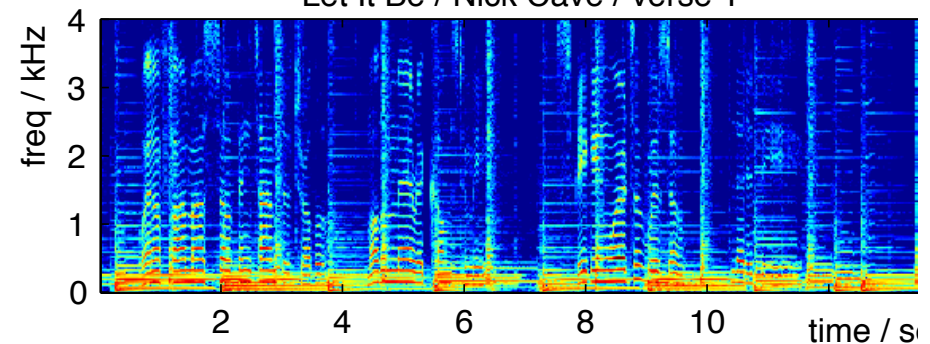
*Let It Be - The Beatles*

Let It Be / Beatles / verse 1



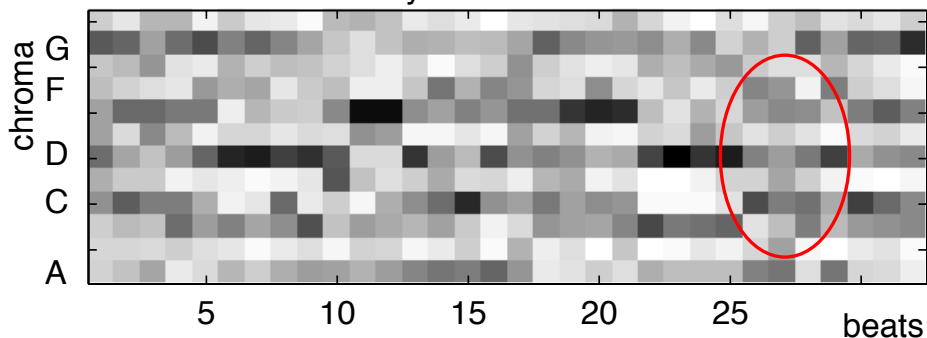
*Let It Be - Nick Cave*

Let It Be / Nick Cave / verse 1

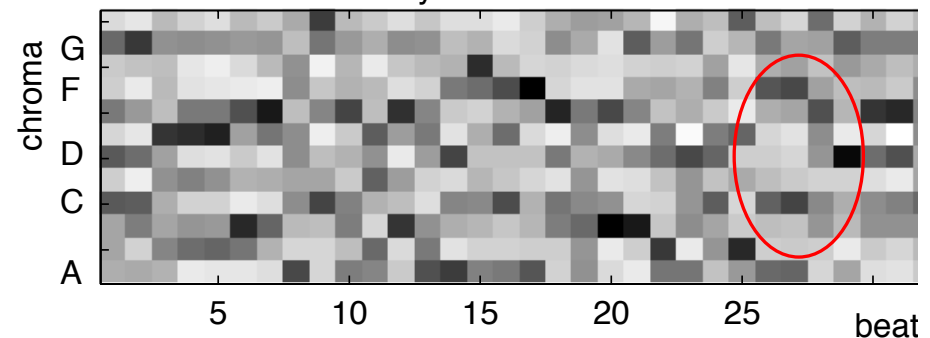


- .. but appears in beat-chroma

Beat-sync chroma features



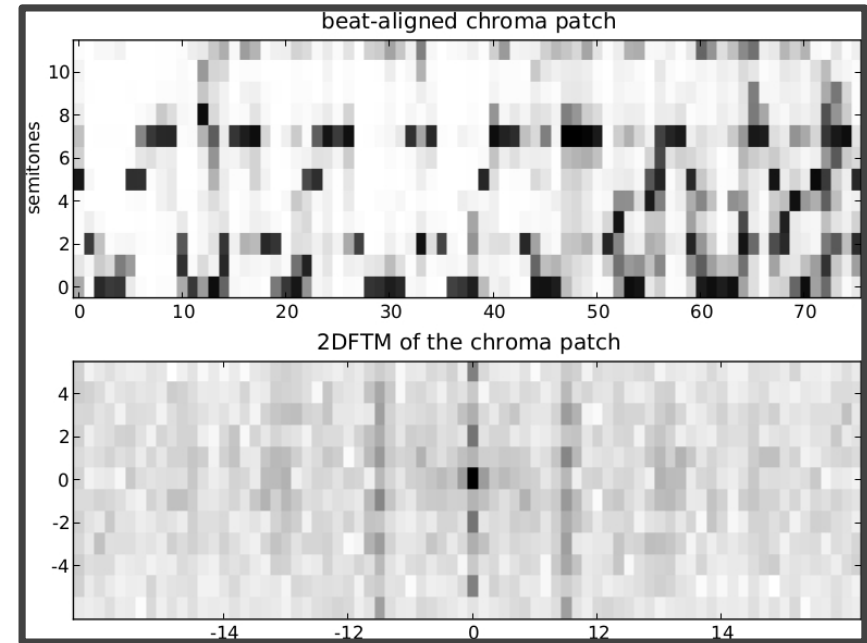
Beat-sync chroma features



# Large-Scale Cover Recognition

*Bertin-Mahieux & Ellis '12*

- 2D Fourier Transform Magnitude (2DFTM)
  - fixed-size feature to capture “essence” of chromagram:

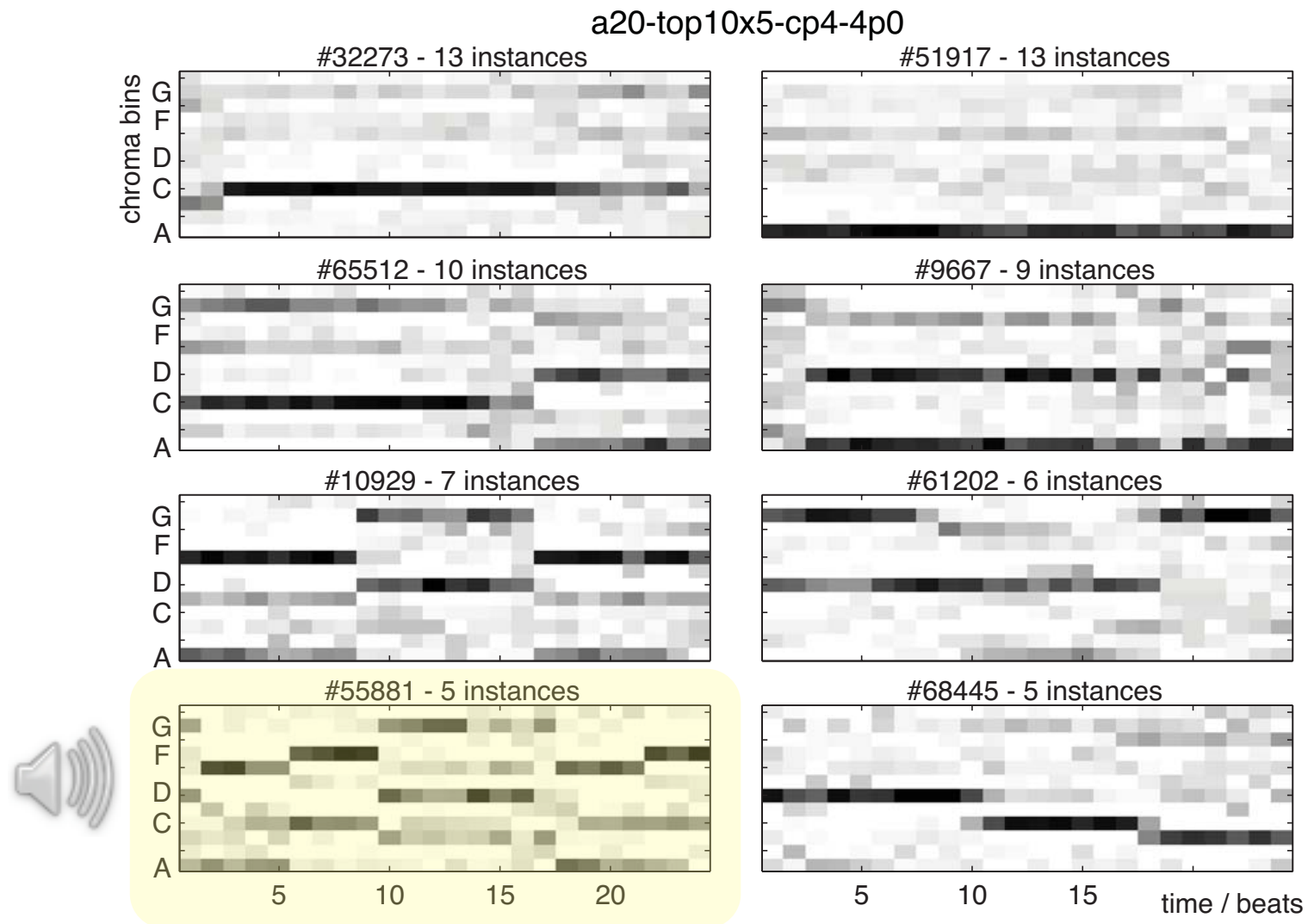


- First results on finding covers in IM songs

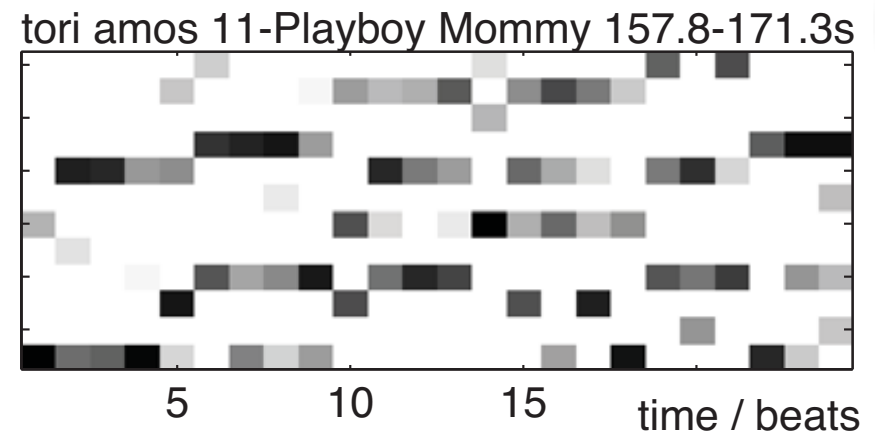
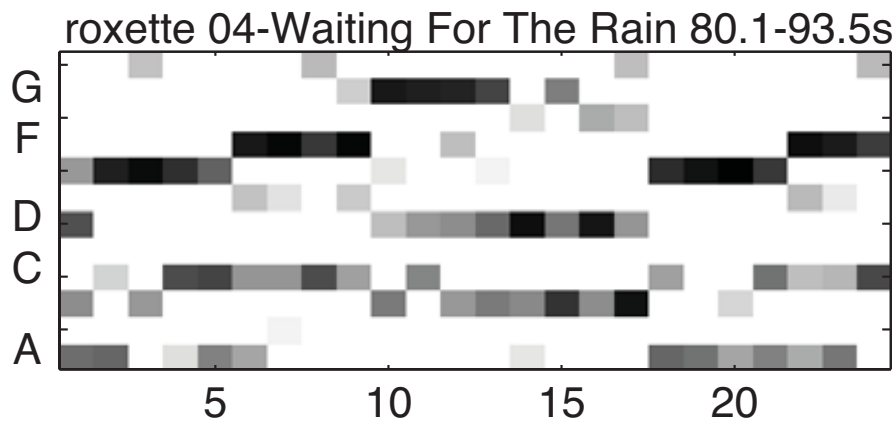
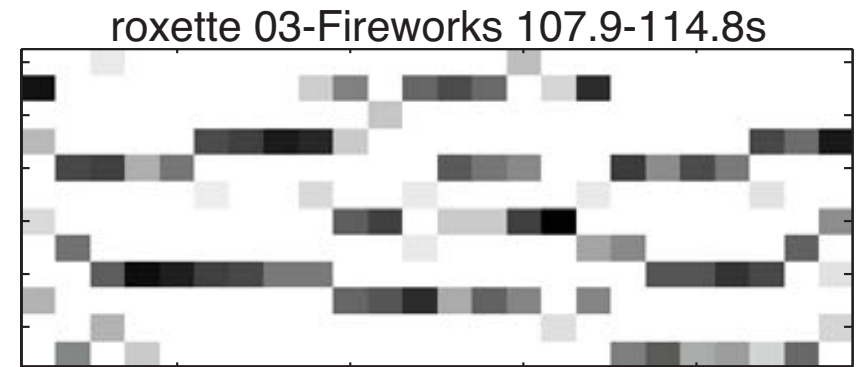
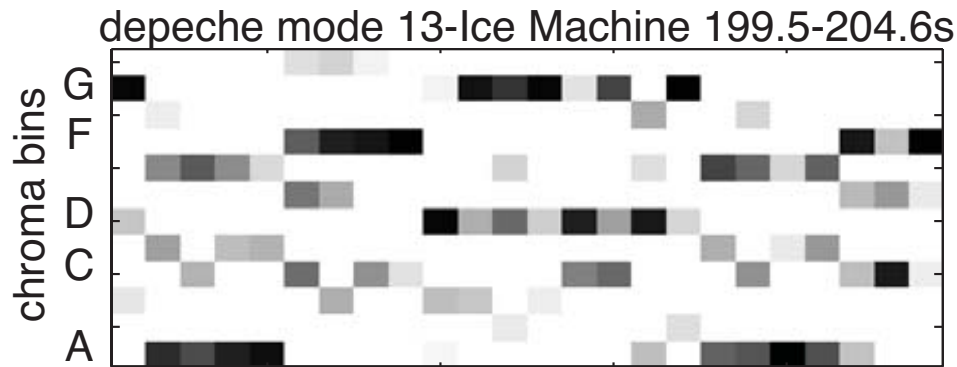
	Average rank	meanAP
random	500,000	0.000
jumpcodes 2	308,369	0.002
<b>2DFTM (50 PC)</b>	<b>137,117</b>	<b>0.020</b>

# Finding Common Fragments

- **Cluster** beat-synchronous chroma patches



# Clustered Fragments



- ... for a **dictionary** of common themes?



# 4. Example Applications: Music Discovery

Berenzweig & Ellis '03

- Connecting listeners to musicians

**Playola** Search:  Artist   
[\[About\]](#) [\[Help\]](#) [\[Turn Samples Off\]](#) [\[Turn Debug On\]](#) [\[Turn Popups Off\]](#) [\[Logout dpwe\]](#)  
 Get Playola Selections: 20 songs  you recently heard   Browse: [Artists](#) [Albums](#) [Playlists](#) Range: 0-C   
 Artist: **The Woodbury Muffin Outbreak** [\[band web page\]](#)  Playlist: -New Playlist-  [\[Add to\]](#) [\[View\]](#)

	Song Title	Artist	Time	Rating
<input type="checkbox"/>	The Ballad of Tabitha	<a href="#">The Woodbury Muffin Outbreak</a>	4:00	<input type="checkbox"/>
<input type="checkbox"/>	Monkey Dreams	<a href="#">The Woodbury Muffin Outbreak</a>	2:57	<input type="checkbox"/>
<input type="checkbox"/>	A Cold Dark Night (Live)	<a href="#">The Woodbury Muffin Outbreak</a>	3:13	<input type="checkbox"/>
<input type="checkbox"/>	Leo, The Ballad of	<a href="#">The Woodbury Muffin Outbreak</a>	1:48	<input type="checkbox"/>
<input type="checkbox"/>	Baby I Forgot To Tell You	<a href="#">The Woodbury Muffin Outbreak</a>	4:04	<input type="checkbox"/>

**Music-Space Browser** [\[What's This?\]](#)

Feature	Less	More
AltNGrunge	<input type="checkbox"/>	<input type="checkbox"/>
CollegeRock	<input type="checkbox"/>	<input type="checkbox"/>
Country	<input type="checkbox"/>	<input type="checkbox"/>
DanceRock	<input type="checkbox"/>	<input type="checkbox"/>
Electronica	<input type="checkbox"/>	<input type="checkbox"/>
MetalNPunk	<input type="checkbox"/>	<input type="checkbox"/>
NewWave	<input type="checkbox"/>	<input type="checkbox"/>
Rap	<input type="checkbox"/>	<input type="checkbox"/>
RnBSoul	<input type="checkbox"/>	<input type="checkbox"/>
SingerSongwriter	<input type="checkbox"/>	<input type="checkbox"/>
SoftRock	<input type="checkbox"/>	<input type="checkbox"/>
TradRock	<input type="checkbox"/>	<input type="checkbox"/>
Female	<input type="checkbox"/>	<input type="checkbox"/>
HiFi	<input type="checkbox"/>	<input type="checkbox"/>

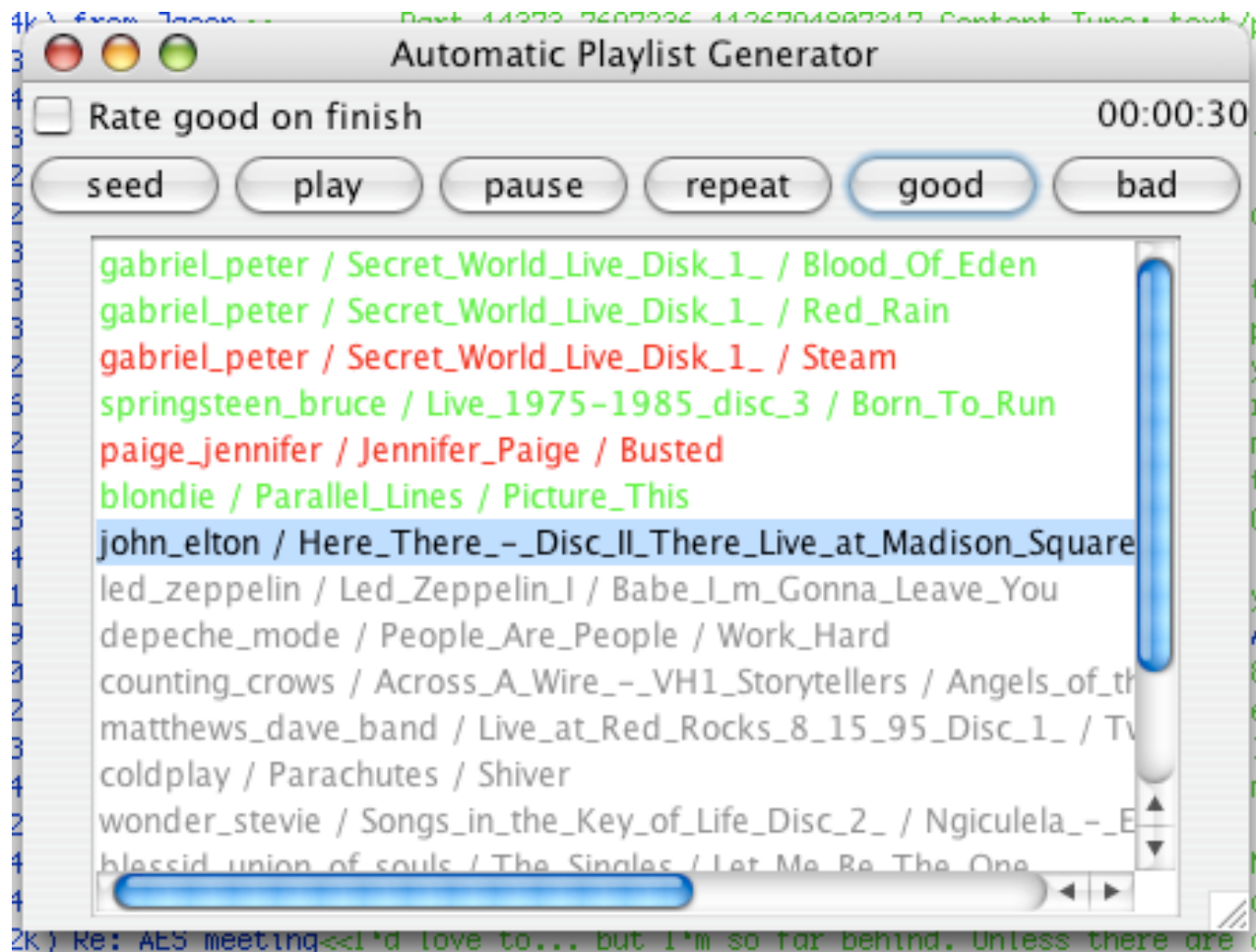
**Similar Songs:** [\[Play this list\]](#) [\[What's This?\]](#)

	Song Title	Artist	Distance	Good Match?
<input type="checkbox"/>	Baby I Forgot To Tell You	<a href="#">The Woodbury Muffin Outbreak</a>	0.00	<input type="checkbox"/>
<input type="checkbox"/>	Number five	<a href="#">Bizi Chyld</a>	0.07	<input type="checkbox"/>
<input type="checkbox"/>	Waiting for Your Love	<a href="#">Toto</a>	0.08	<input type="checkbox"/>
<input type="checkbox"/>	Excerpt from 'CD'	<a href="#">Weirdomusic</a>	0.08	<input type="checkbox"/>

# Playlist Generation

Mandel, Poliner, Ellis '06

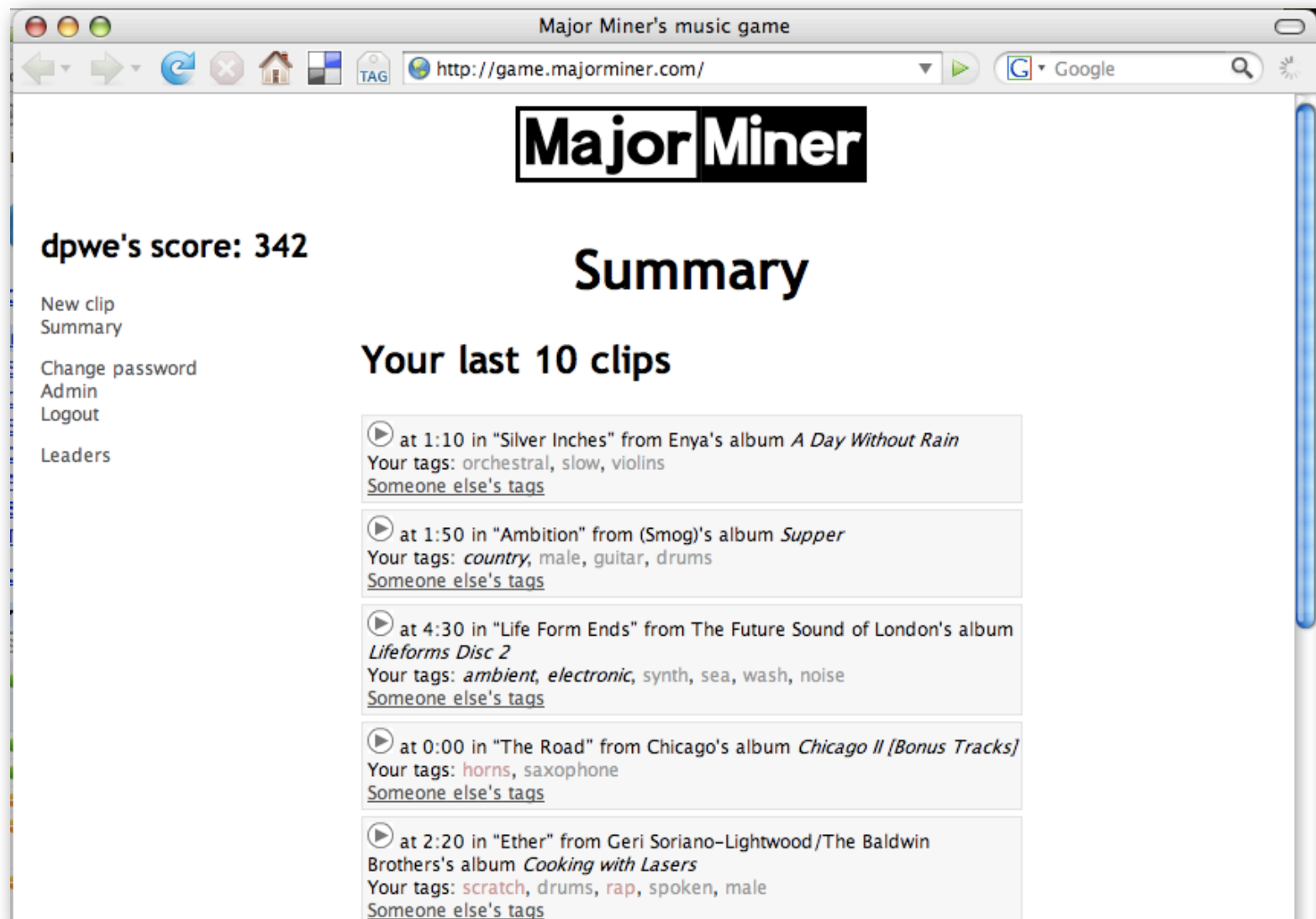
- **Incremental** learning of listeners' **preferences**



# MajorMiner: Music Tagging

Mandel & Ellis '07,'08

- Describe music using **words**



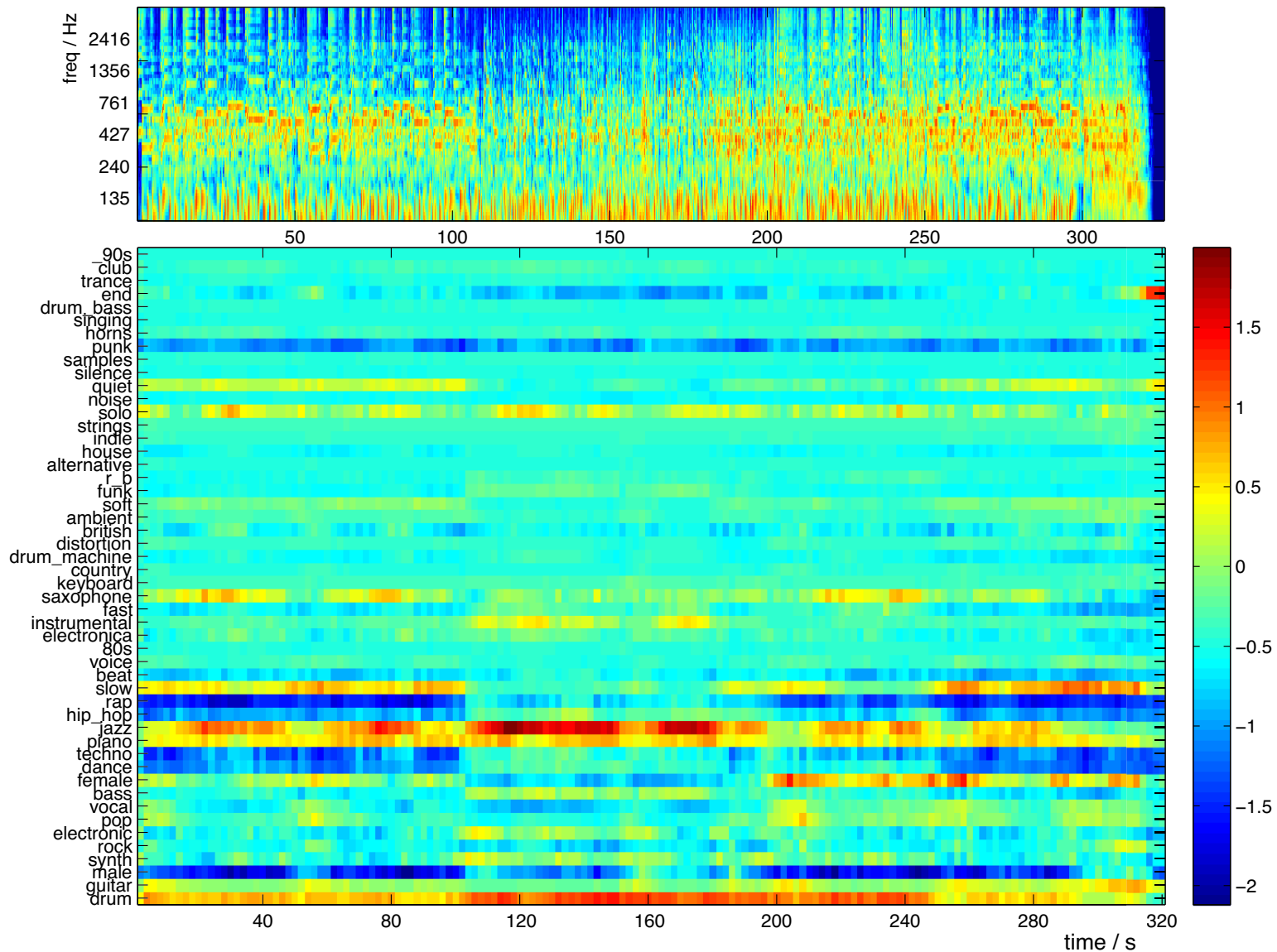
The screenshot shows a web browser window titled "Major Miner's music game" with the URL "http://game.majorminer.com/". The page features the "Major Miner" logo at the top. On the left, a sidebar lists user options: "New clip", "Summary", "Change password", "Admin", "Logout", and "Leaders". The main content area displays "dpwe's score: 342" and a "Summary" section titled "Your last 10 clips". This section lists five clips with their respective tags:

- at 1:10 in "Silver Inches" from Enya's album *A Day Without Rain*  
Your tags: *orchestral, slow, violins*  
[Someone else's tags](#)
- at 1:50 in "Ambition" from (Smog)'s album *Supper*  
Your tags: *country, male, guitar, drums*  
[Someone else's tags](#)
- at 4:30 in "Life Form Ends" from The Future Sound of London's album *Lifeforms Disc 2*  
Your tags: *ambient, electronic, synth, sea, wash, noise*  
[Someone else's tags](#)
- at 0:00 in "The Road" from Chicago's album *Chicago II [Bonus Tracks]*  
Your tags: *horns, saxophone*  
[Someone else's tags](#)
- at 2:20 in "Ether" from Geri Soriano-Lightwood/The Baldwin Brothers's album *Cooking with Lasers*  
Your tags: *scratch, drums, rap, spoken, male*  
[Someone else's tags](#)

# Classification Results

- Classifiers trained from top 50 tags

01 Soul Eyes



# Music Transcription

Poliner & Ellis  
'05,'06,'07

## Training data and features:

- MIDI, multi-track recordings, playback piano, & resampled audio (less than 28 mins of train audio).
- Normalized magnitude STFT.



## Classification:

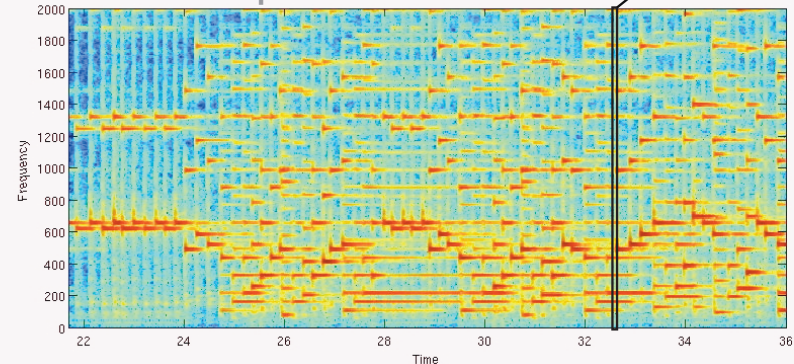
- N-binary SVMs (one for ea. note).
- Independent frame-level classification on 10 ms grid.
- Dist. to class bndy as posterior.



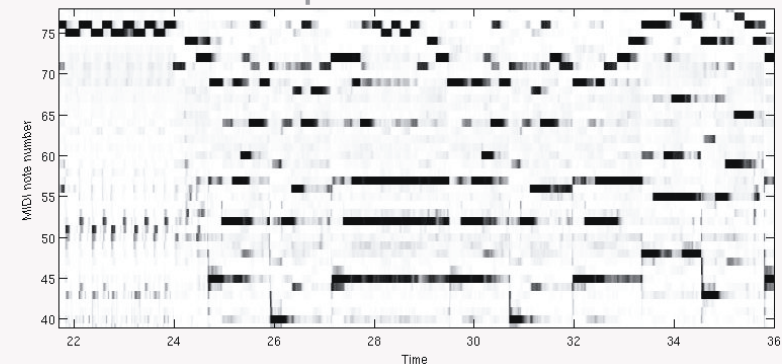
## Temporal Smoothing:

- Two state (on/off) independent HMM for ea. note. Parameters learned from training data.
- Find Viterbi sequence for ea. note.

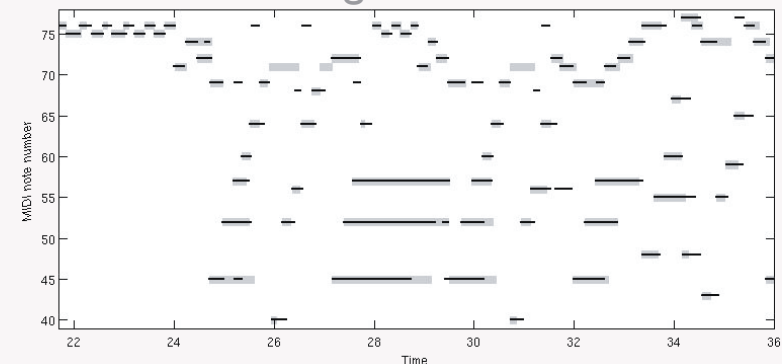
feature representation



classification posteriors



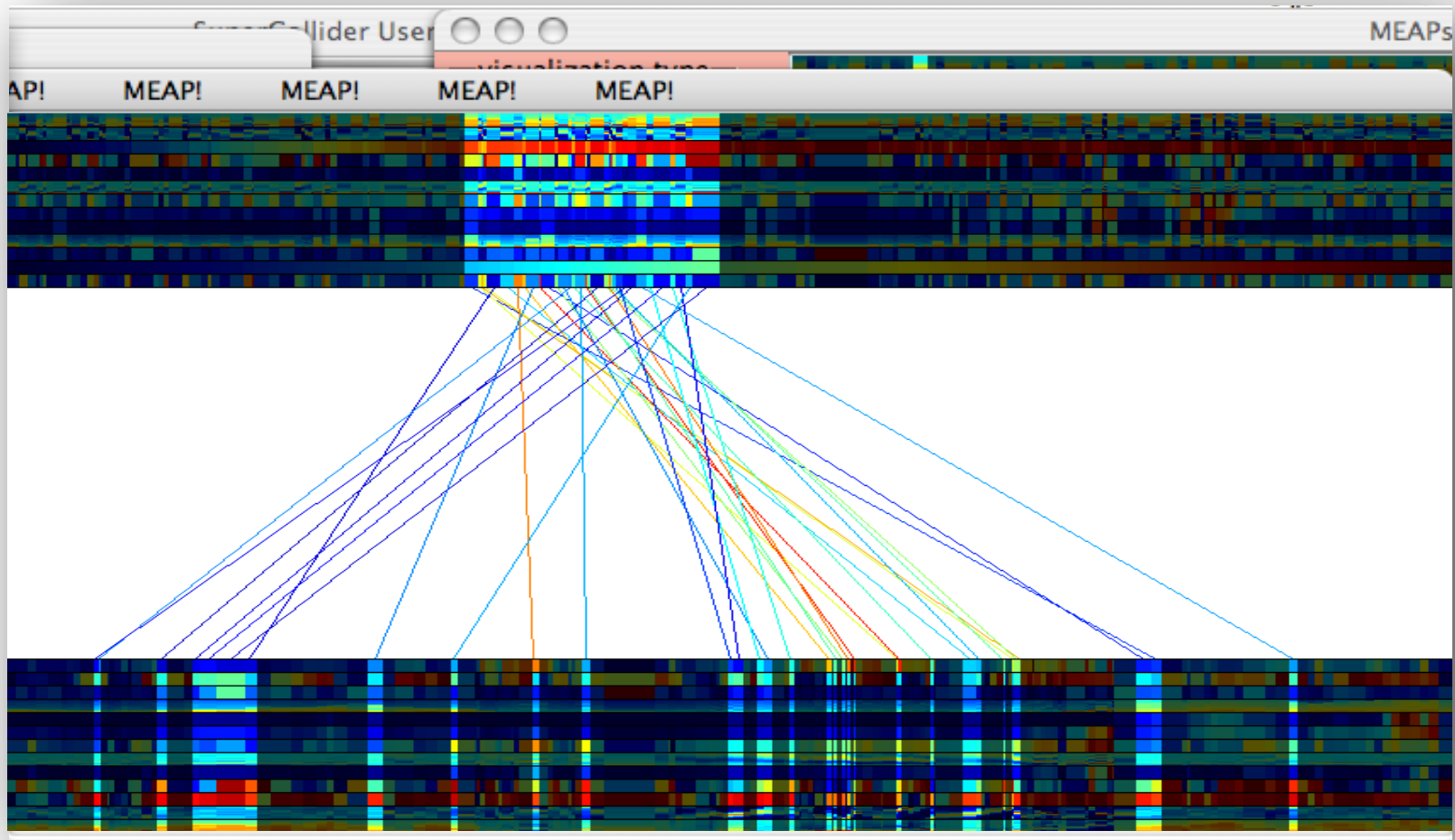
hmm smoothing



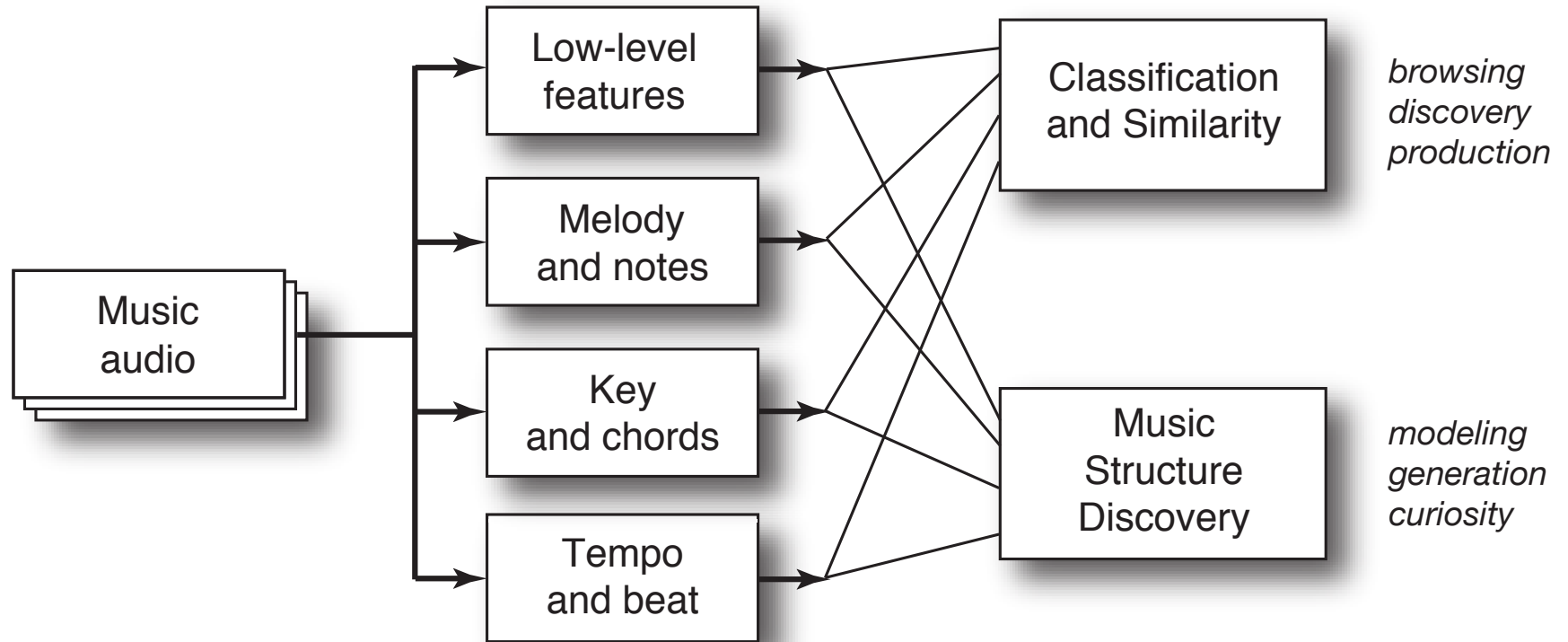
# MEAPsoft

- **M**usic **E**ngineering **A**rt **P**rojects
  - collaboration between EE and Computer Music Center

*with Douglas Repetto,  
Ron Weiss, and the rest  
of the MEAP team*



# Conclusions



- Lots of **data**  
+ noisy **transcription**  
+ weak **clustering**  
⇒ musical **insights?**