Extracting Information from 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/

- I. Motivation: Learning Music
- 2. Notes Extraction
- 3. Drum Pattern Modeling
- 4. Music Similarity



2006-02-13 р. 1/32



LabROSA Overview





Laboratory for the Recognition and Organization of Speech and Audio 2006-02-13 р. 2/32



Learning from Music

- A lot of music data available
 o e.g. 60G of MP3
 ≈ 1000 hr of audio, 15k tracks
- What can we do with it?
 o implicit definition of 'music'
- Quality vs. quantity
 - Speech recognition lesson:

Motivating Applications

o music similarity (recommendation, playlists)

o computer (assisted) music generation

• insight into music





	<u>é</u>		Q				
4			5	rarch			
Source	Artist		Albur				
D Lbrary	All (113 Artists)	All (167	Albums)				
V 💦 Music Store	'Til Tuesday	prakir	na estacion Esp	eranz			
'w Shopping Cart	ABC	01 Every	thing Must Go				
Purchased Music	Aimee Mann	01 Little	Earthquakes				
Shared Music	Alanis Morissette	01 What	ever				
My Top Rated	Angelique Kidjo (Benin)	02 Gaue	ho				
· Faranth Plauad	Anna Domino	02 I'm With Stupid					
Top 15 Mars Novad	Aura Msimang (South Africa) 02 Under the Pink						
Top 23 Most Playeo	8-52's	a 03 Bachelor No. 2 kulolo and the Nigerian Brothers (7 + 03 Boys For Pele					
Tem s/s for steve	Baba Ken Okulolo and the Nigerian Brothers (7 +						
2 getrunes	-	e					
Music Together - Drum	Song Name	Time	Artist				
Music Together - Fiddle	Passion Play (When All The Slaves Are Fr	5:25	Joni Mitchell				
Nathan sings	Cherokee Louise	4:32	Joni Mitchell				
	😸 The Windfall	5.16	Joni Mitchell				
	Slouching Towards Bethlehem	6:55	Joni Mitchell				
	S Come In From The Cold	7:31	Joni Mitchell				
		4:54	Joni Mitchell				
	of The O		Mitchell				
	P		- 11				
				h			
	1055	- 70	O CD				
	1955 songs, 5.1 day	s, 7.9	10 GB				
+ X 0 X							

Ground Truth Data

 A lot of unlabeled music data available
 manual annotation is expensive and rare



- Unsupervised structure discovery possible
 - •.. but labels help to indicate what you want
- Weak annotation sources
 - artist-level descriptions
 - symbol sequences without timing (MIDI)
 - errorful transcripts
- Evaluation requires ground truth
 - limiting factor in Music IR evaluations?





Talk Roadmap





2. Notes Extraction

with Graham Poliner

Audio → Score very desirable

• for data compression, searching, learning

• Full solution is elusive

signal separation of overlapping voicesmusic constructed to frustrate!

Maybe simplify problem:
 "Dominant Melody" at each time frame



Music Information Extraction - Ellis

Laboratory for the Recognition and Organization of Speech and Audio 2006-02-13 р. 6/32



Conventional Transcription

Pitched notes have harmonic spectra

 → transcribe by searching for harmonics

 e.g. sinusoid modeling + grouping





• Explicit expert-derived knowledge



Music Information Extraction - Ellis

2006-02-13 р. 7/32



Transcription as Classification

- Signal models typically used for transcription
 harmonic spectrum, superposition
- But ... trade domain knowledge for data
 transcription as pure classification problem:



single N-way discrimination for "melody"
per-note classifiers for polyphonic transcription





Melody Transcription Features

 Short-time Fourier Transform Magnitude (Spectrogram)



Training Data

- Need {data, label} pairs for classifier training
- Sources:

pre-mixing multitrack recordings + hand-labeling?
 synthetic music (MIDI) + forced-alignment?



Melody Transcription Results

Trained on 17 examples

• ... plus transpositions out to +/-6 semitones

• All-pairs SVMs (Weka)

Tested on ISMIR MIREX 2005 set

• includes foreground/background detection

Rank	Participant	Overall Accuracy	Voicing d'	Raw Pitch	Raw Chroma	Runtime / s
1	Dressler	71.4%	1.85	68.1%	71.4%	32
2	Ryynänen	64.3%	1.56	68.6%	74.1%	10970
3	Poliner	61.1%	1.56	67.3%	(73.4%)	5471
3	Paiva 2	61.1%	1.22	58.5%	62.0%	45618
5	Marolt	59.5%	1.06	60.1%	67.1%	12461
6	Paiva 1	57.8%	0.83	62.7%	66.7%	44312
7	Goto	49.9%*	0.59*	65.8%	71.8%	211
8	Vincent 1	47.9%*	0.23*	59.8%	67.6%	?
9	Vincent 2	46.4%*	0.86*	59.6%	71.1%	251
10	Brossier	3.2%* †	0.14 * †	3.9% †	8.1% †	41

• Example...





Music Information Extraction - Ellis

2006-02-13 р. 11/32



Polyphonic Transcription

- Train SVM detectors for every piano note
 same features & classifier but different labels
 88 separate detectors, independent smoothing
- Use MIDI syntheses, player piano recordings



• about 30 min training data



2006-02-13 р. 12/32



Piano Transcription Results

Significant improvement from classifier:
 o frame-level accuracy results:

Algorithm	Errs	False Pos	False Neg	d'
SVM	43.3%	27.9%	15.4%	3.44
Klapuri&Ryynänen	66.6%	28.1%	38.5%	2.71
Marolt	84.6%	36.5%	48.1%	2.35





3. Eigenrhythms: Drum Pattern Space

with John Arroyo



Eigen-analysis (or ...) to capture variations?
 by analyzing lots of (MIDI) data, or from audio

Applications

- music categorization
- "beat box" synthesis
- insight

Lab ROSA Laboratory for the Recognition and Organization of Speech and Audio

2006-02-13 р. 14/32



Aligning the Data

• Need to align patterns prior to modeling...



Eigenrhythms (PCA)



- Need 20+ Eigenvectors for good coverage of 100 training patterns (1200 dims)
- Eigenrhythms both add and subtract





Posirhythms (NMF)



• Nonnegative: only adds beat-weight

• Capturing some structure



2006-02-13 р. 17/32



Eigenrhythms for Classification

• Projections in Eigenspace / LDA space



Eigenrhythm BeatBox

• Resynthesize rhythms from eigen-space



Lab

Laboratory for the Recognition and Organization of Speech and Audio



4. Music Similarity

with Mike Mandel and Adam Berenzweig

• Can we predict which songs "sound alike" to a listener?

... based on the audio waveforms?
many aspects to subjective similarity

Applications

query-by-example
automatic playlist generation
discovering new music

Problems

the right representationmodeling individual similarity



2006-02-13 р. 20/32



Music Similarity Features

Need "timbral" features: Mel-Frequency Cepstral Coeffs (MFCCs)



COLUMBIA UNIVERSITY

Lab

Timbral Music Similarity

- Measure similarity of feature distribution
 i.e. collapse across time to get density p(x_i)
 compare by e.g. KL divergence
- e.g. Artist Identification
 - learn artist model p(x_i | artist X) (e.g. as GMM)
 classify unknown song to closest model



"Anchor Space"

- Acoustic features describe each song
 - •... but from a signal, not a perceptual, perspective
 - .. and not the differences between songs
- Use genre classifiers to define new space
 prototype genres are "anchors"



Anchor Space

• Frame-by-frame high-level categorizations



COLUMBIA UNIVERSITY

Laboratory for the Recognition and Organization of Speech and Audio

Lab

'Playola' Similarity Browser

	A	😔 http://www.playola.or	g/index	.php					
Pla	ay	1013 Search:		Artis	t 💌	Search [About] [Help] [Turn Samples Off] [Loqout	<u>dpwe]</u>
Get Sel	ectio	ons: 20 songs 💌 recen	tly hea	rd 💌	G	ol Browse: <u>Artists</u> <u>All</u>	<u>bums Playlists</u> Rang	je: 0-	c 🔳
Artist:	Beat	tles [band web page]	[Play!]	Playlis	t: -N	lew Playlist-	A1 🔽	dd to]	[View]
		Song Title	Artist	Time	Music	-Space Browser			
Album:	Magi	cal Mystery Tour			Feat	ure 1	0.00		More
	1	Baby You're a Rich Man	Beatles	3:03	reat	Althicmunes			HOTE
	9	Plue Jay Way	Reptice	2.56		CollegePock			
		blue Jay way	Deatles	3:50		Country			
	1	Penny Lane	Beatles	3:03		DanceRock			
	1	Magical Mystery Tour	Beatles	2:51		Electronica			
		The Feel on the 1999	Deset las	2.00		MetalNPunk			
		The Fool on the Hill	Beatles	3:00		NewWave			
	/	I Am the Walrus	Beatles	4:37		Rap			
	1	Flying	Beatles	2:17		SingerSongwriter			
	,	Your Mother Chauld Know	Repties	2:20		SoftRock			
		Tour Mother Should Know	Deatles	2:29		TradRock			
	/	Strawberry Fields Forever	Beatles	4:10		Female			
Album:	Yello	w Submarine				HIFI			
	1	All You Need Is Love	Beatles	3:52	Simila	r Songs: [Play this list]			
	1	Yellow Submarine	Beatles	2:40		Song Title	Artist	Distar	Good Match?
	1	All Together Now	Beatles	2:10	- /	Let It Be	Beatles	0.00	P
	1	Hey Bulldog	Beatles	3:11	Þ 🦊	Double Hockey Sticks	Adam The Gimbel	0.06	7
xog ch a 🔽 🕨	/	It's All Too Much	Beatles	6:25	Þ 🦊	Light in Your Eyes	Blessid Union of Sou	0.06	7
		Only a Northern Sana	Bootles	2.24	Þ 🥒	Mori	Tranzas	0.07	🍋 🦊

Ground-truth data

- Hard to evaluate Playola's 'accuracy'
 - o user tests...
 - ground truth?

"Musicseer" online survey:
ran for 9 months in 2002
1,000 users, > 20k judgments
http://labrosa.ee.columbia.edu/ projects/musicsim/ Which artist is most similar to: Janet Jackson?

- 1. R. Kelly
- 2. <u>Paula Abdul</u>
- 3. <u>Aaliyah</u>
- 4. Milli Vanilli
- 5. En Vogue
- 6. <u>Kansas</u>
- 7. Garbage
- 8. <u>Pink</u>
- 9. Christina Aguilera



2006-02-13 р. 26/32



Evaluation

- Compare Classifier measures against Musicseer subjective results
 - "triplet" agreement percentage • Top-N ranking agreement score: $\sum_{k=1}^{N} \alpha^{k} \alpha^{k} q^{k} q^{k$

$$s_i = \sum_{r=1}^{n} \alpha_r^r \alpha_c^{k_r}$$
 $\alpha_r = \left(\frac{1}{2}\right)^3$ $\alpha_c = \alpha_r^2$

- First-place agreement percentage
 - simple significance test

_ap

Laboratory for the Recognition and Organization of Speech and Audio





Using SVMs for Artist ID

- Support Vector Machines (SVMs) find hyperplanes in a high-dimensional space
 - relies only on matrix of distances between points
 much 'smarter' than nearest-neighbor/overlap
 want diversity of reference vectors...



COLUMBIA UNIVERSITY



Song-Level SVM Artist ID

 Instead of one model per artist/genre, use every training song as an 'anchor'
 then SVM finds best support for each artist





Organization of Speech and Audio

2006-02-13 р. 29/32



Artist ID Results

- ISMIR/MIREX 2005 also evaluated Artist ID
- I48 artists, I800 files (split train/test) from 'uspop2002'
- Song-level SVM clearly dominates

 using only MFCCs!

MIREX 05 Audio Artist (USPOP2002)

Rank	Participant	Raw Accuracy	Normalized	Runtime / s
1	Mandel	68.3%	68.0%	10240
2	Bergstra	59.9%	60.9%	86400
3	Pampalk	56.2%	56.0%	4321
4	West	41.0%	41.0%	26871
5	Tzanetakis	28.6%	28.5%	2443
6	Logan	14.8%	14.8%	?
7	Lidy	Did not co	omplete	



Music Information Extraction - Ellis

2006-02-13 р. 30/32



Playlist Generation

- SVMs are well suited to "active learning"
 o solicit labels on items closest to current boundary
- Automatic player with "skip"
 Ground truth data collection
 active-SVM
 automatic playlist generation

000	Automatic Playlist Generator	[
📃 Rate good on fini	ish	00:00:30
seed play	pause repeat good	bad
gabriel_peter / Se gabriel_peter / Se gabriel_peter / Se springsteen_bruce paige_jennifer / Je blondie / Parallel_ john_elton / Here led_zeppelin / Lee depeche_mode / counting_crows / / matthews_dave_b coldplay / Parache wonder_stevie / S	cret_World_Live_Disk_1_ / Blood_Of_Eden cret_World_Live_Disk_1_ / Red_Rain cret_World_Live_Disk_1_ / Steam e / Live_1975-1985_disc_3 / Born_To_Run ennifer_Paige / Busted Lines / Picture_This ThereDisc_II_There_Live_at_Madison_Squ d_Zeppelin_I / Babe_I_m_Gonna_Leave_You People_Are_People / Work_Hard Across_A_WireVH1_Storytellers / Angels_c and / Live_at_Red_Rocks_8_15_95_Disc_1_ utes / Shiver ongs_in_the_Key_of_Life_Disc_2_ / Ngiculela_ souls / The_Singles / Let_Me_Re_The_One	



2006-02-13 р. 31/32



Conclusions



Lab

Laboratory for the Recognition and Organization of Speech and Audio

