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/

- Learning Music
- 2. Melody Extraction
- 3. Drum Pattern Modeling
- 4. Music Similarity



2005-10-26 р. 1/30



LabROSA Overview



Learning from Music

- A lot of music data available
 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 / classification
 - O computer (assisted) music generation
 - **O** insight into music







Ground Truth Data

- A lot of unlabeled music data available
 - manual annotation is much rarer



• Unsupervised structure discovery possible

O... but labels help to indicate what you want

- Weak annotation sources
 - o artist-level descriptions

• symbol sequences without timing (MIDI)

- errorful transcripts
- Evaluation requires ground truth

• limiting factor in Music IR evaluations?



2005-10-26 р. 4/30



Talk Roadmap





2. Melody Transcription

with Graham Poliner

COLUMBIA UNIVERSITY

Audio → Score very desirable

O for data compression, searching, learning

• Full solution is elusive

• signal separation of overlapping voices

• music constructed to frustrate!

• Simplified problem:

"Dominant Melody" at each time frame



Laboratory for the Recognition and Organization of Speech and Audio

Conventional Transcription

Pitched notes have harmonic spectra

 → transcribe by searching for harmonics

 o e.g. sinusoid modeling + grouping





• Explicit expert-derived knowledge



2005-10-26 р. 7/30



Transcription as Classification

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



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



2005-10-26 р. 8/30



Training Data

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

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



COLUMBIA UNIVERSITY

Melody Transcription Results

Trained on 17 examples
 O .. plus transpositions out to +/- 6 semitones
 O SMO SVM (Weka)

Tested on ISMIR MIREX 2005 set

O 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	Paiva 2	61.1%	1.22	58.5%	62.0%	45618
3	Poliner	61.1%	1.56	67.3%	73.4%	5471
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

O Example...



Music Info Extraction - Ellis

2005-10-26 р. 10/30



Melody Clustering

• Goal: Find 'fragments' that recur in melodies

- **O**.. across large music database
- O... trade data for model sophistication



• Data sources

o pitch tracker, or MIDI training data

Melody fragment representation
 OCT(1:20) - removes average, smoothes detail





Melody clustering results

• Clusters match underlying contour:



COLUMBIA UNIVERSITY

3. Eigenrhythms: Drum Pattern Space

with John Arroyo



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

Applications

• music categorization

- **O** ''beat box'' synthesis
- **O** insight



Music Info Extraction - Ellis

2005-10-26 р. 13/30



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



2005-10-26 р. 16/30



Eigenrhythms for Classification

• Projections in Eigenspace / LDA space



Eigenrhythm BeatBox

• Resynthesize rhythms from eigen-space





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?

O .. based on the audio waveforms?

• many aspects to subjective similarity

Applications

o query-by-example
o automatic playlist generation
o discovering new music

Problems

• the right representation• modeling individual similarity



2005-10-26 р. 19/30



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
 O .. but from a signal, not a perceptual, perspective
 O .. and not the differences between songs
- Use genre classifiers to define new space
 o prototype genres are "anchors"



Anchor Space

• Frame-by-frame high-level categorizations



Laboratory for the Recognition and Organization of Speech and Audio

COLUMBIA UNIVERSITY

'Playola' Similarity Browser

Get	Playo	la Selections: 2	o songs 📮	you rec	entry n	eard Ţ	GOI	Browse: <u>Artis</u>	ts <u>Album</u>	n <u>s Playlists</u> Rai	nge: 0-C	•
Artis	t: Th	e Woodbury Muffi	in Outbreak []	band web	page]	[Play!] Pla	ylist: -New Playlist-		÷ [Add to] [V	iew]
		Song Title The Ballad of	Artis The Woodbur	rt v Muffin	Time	Rating	Musio	-Space Browser			[What's	s This
		Tabitha	Outbreak	,	4:00			Feature		Less		Мо
	1	Monkey Dreams	The Woodbur Outbreak	y Muffin	2:57			A	tNGrunge legeRock			
	2	A Cold Dark Night	The Woodbur	y Muffin	3:13			0	Country			
		(Live)	Outbreak The Woodbur	v Muffin				D	anceRock			
	/	Leo, The Ballad of	Outbreak	y Humm	1:48			E	ectronica			
	/	Baby I Forgot To Tell You	The Woodbur Outbreak	y Muffin	4:04			1-10	NewWave			
									Rap			
								SingerS	RnBSoul			
								Singers	SoftRock			
									TradRock			
									Female			
				Simil	ar Songe: [Play this	list]		[What'	s Thi			
								Song Title		Artist	Distance	Goo
							- /	Baby I Forgot To Te You	I <u>The V</u> Outbr	Voodbury Muffin reak	0.00	-
							- /	Number five	Bizi C	hyld	0.07	
							- /	Waiting for Your Lov	e <u>Toto</u>		0.08	-
							- /	Excerpt from 'CD'	Weird	lomusic	0.08	-
b												-

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Ground-truth data

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

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

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



2005-10-26 р. 24/30



Evaluation

- Compare Classifier measures against Musicseer subjective results
 - 'triplet'' agreement percentage
 Top-N ranking agreement score:

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

• 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
 - 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'
 O then SVM finds best support for each artist





2005-10-26 р. 27/30



Artist ID Results

- ISMIR/MIREX 2005 also evaluated Artist ID
- I 48 artists, I 800 files (split train/test) from 'uspop2002'
- Song-level SVM clearly dominates
 O 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	



2005-10-26 р. 28/30



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
 o active-SVM

automatic playlist generation

	Automatic Playlist Generator	ont Timor tout/p
Rate good on	finish	00:00:30
seed pl	lay pause repeat good	bad
gabriel_peter gabriel_peter gabriel_peter springsteen_b paige_jennifer blondie / Para	/ Secret_World_Live_Disk_1_ / Blood_Of_Ed / Secret_World_Live_Disk_1_ / Red_Rain / Secret_World_Live_Disk_1_ / Steam ruce / Live_1975-1985_disc_3 / Born_To_I / Jennifer_Paige / Busted Illel Lines / Picture This	en t Run 1
john_elton / H	ere_ThereDisc_II_There_Live_at_Madisor	n_Square
led_zeppelin / depeche_mod counting_crow matthews_day coldplay / Par	/ Led_Zeppelin_I / Babe_I_m_Gonna_Leave_ e / People_Are_People / Work_Hard s / Across_A_WireVH1_Storytellers / Ang e_band / Live_at_Red_Rocks_8_15_95_Dise achutes / Shiver	You gels_of_th c_1_ / Tv
blessid union	of souls / The Singles / Let Me Re The Or	ulelaE ▼





2005-10-26 р. 29/30



Conclusions



Lab

Laboratory for the Recognition and Organization of Speech and Audio

