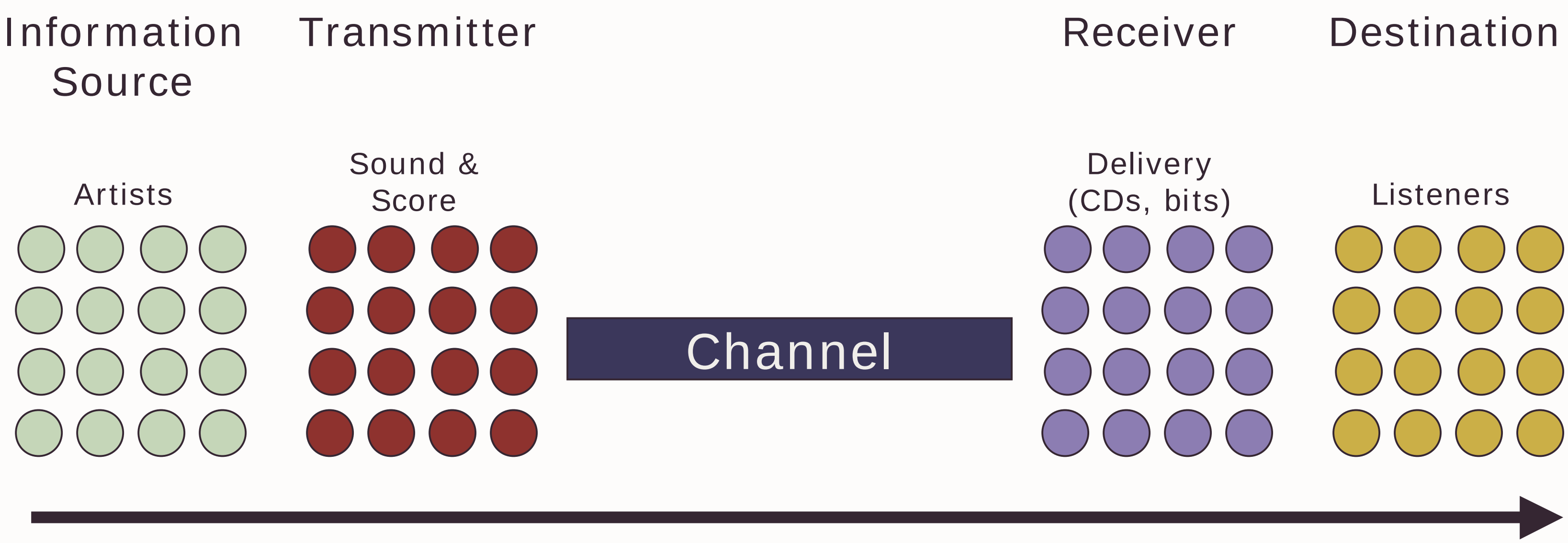
Combining Musical and Cultural Features for Intelligent Style Detection

Brian Whitman
Paris Smaragdis
MIT Media Lab
Music, Mind and Machine Group
(formerly Machine Listening)

What We're Getting At

Music Understanding

- Meyer: “Music is Information”
- We all arm a representation of music against noise

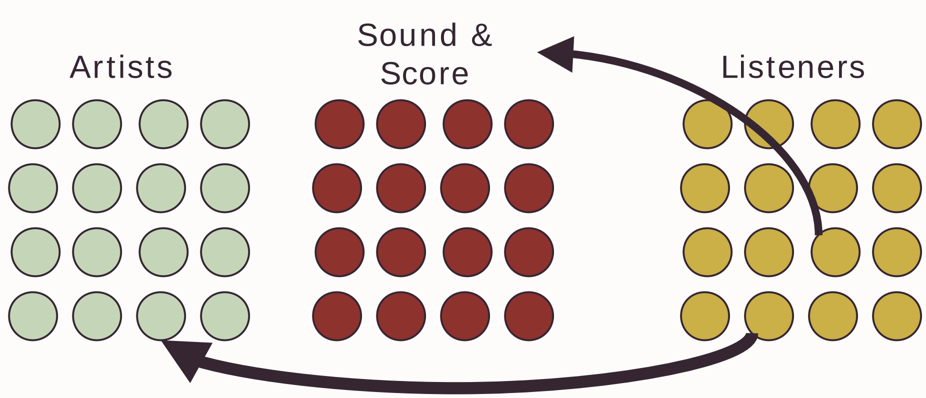


Two-Way IR

- So much going the other way!

“My favorite song”
“Timbaland produced the new Missy record”
“Uninspired electro-glitch rock”
“Reminds me of my ex-girlfriend”

P2P Collections
Online playlists
Informal reviews
Query habits



Personal vs. Community

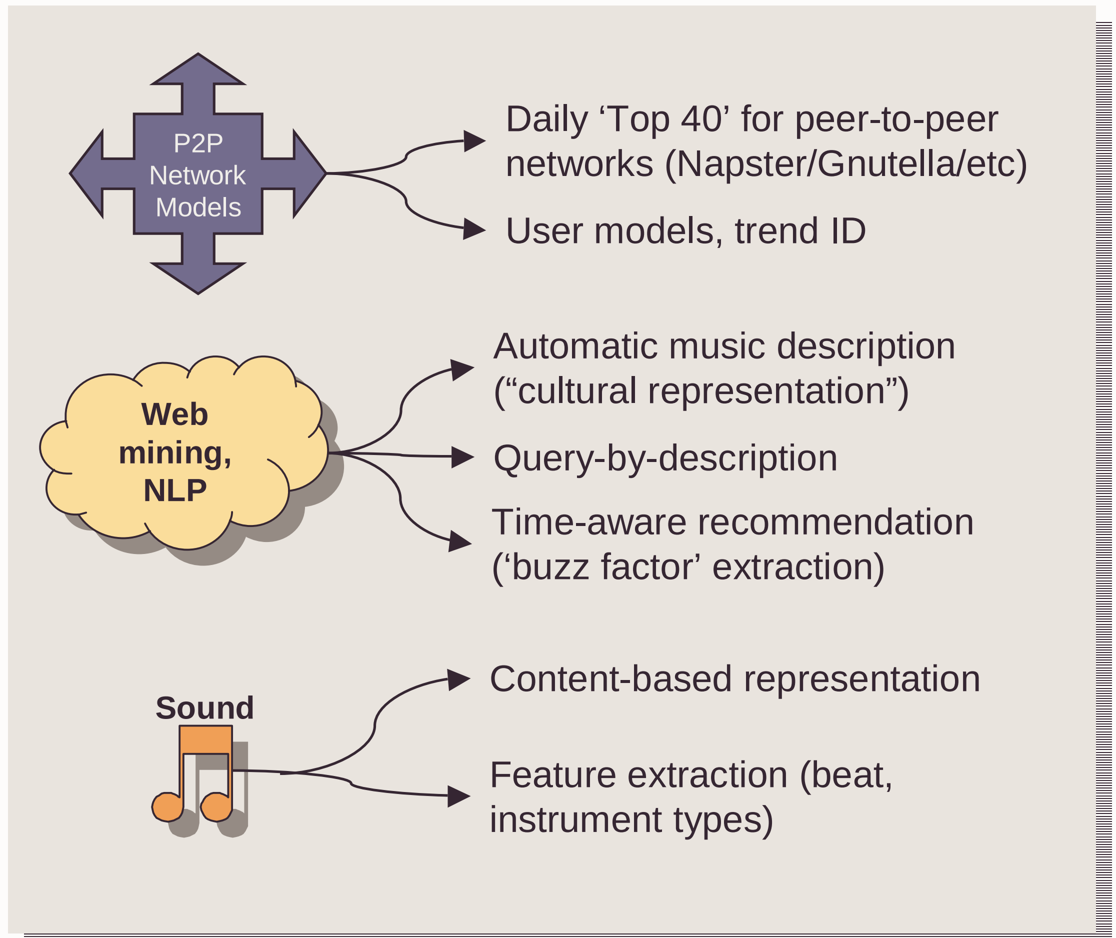
- 2 kinds of audience to artist relation
- Personal:
  - Musical memory, personal preference, local cultural noise
  - Audio sim / rec as insult!
- Community:
  - Large-scale cultural factors, “stranger recommendation” (CF)

Audio and Audience

Where does music preference come from?

Does the type of music actually matter?

Mapping personal and community musical memory



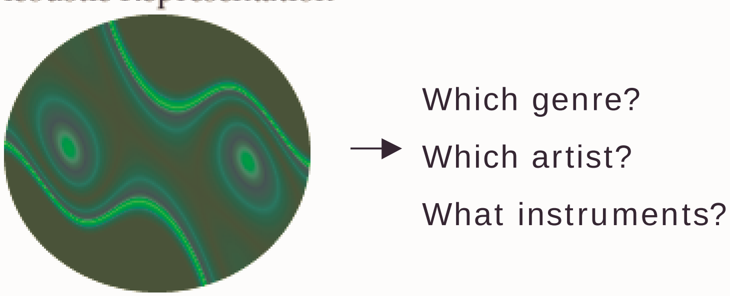
What's On Today!

- Cultural representations for music
- Bimodal acoustic/textual decision space
- Experiment: style ID task
- Cultural representations of the future

Acoustic vs. Cultural Representations

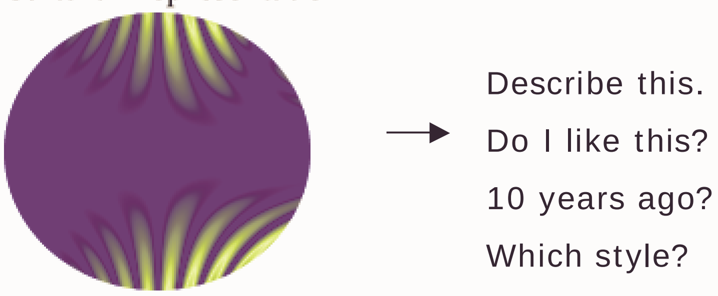
- Acoustic:
  - Instrumentation
  - Short-time (timbral)
  - Mid-time (structural)
  - Usually all we have

Acoustic Representation

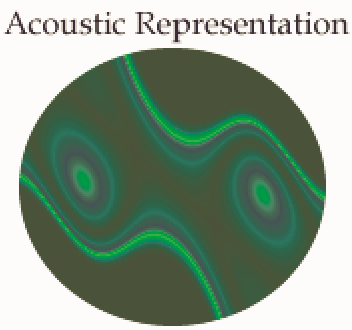


- Cultural:
  - Long-scale time
  - Inherent user model
  - Listener's perspective
  - Two-way IR

Cultural Representation



Bimodal Model



- Independent kernel hyperspaces
- Acoustic: fine-grained, frame level, short-term time-aware
- Cultural: intrinsic user model, artist level, long-term time

"Community Metadata"

- (Whitman/Lawrence ICMC2002)
- Combine all types of mined data
  - P2P, web, usenet, future?
- Long-term time aware
- One comparable representation via gaussian kernel
  - Machine learning friendly

Data Collection Overview

- Cultural Feature Extraction:
  - Web crawls for music information
  - Retrieved documents are parsed for:
    - Unigrams, bigrams and trigrams
    - Artist names
    - Noun phrases
    - Adjectives
- P2P crawl:
  - Robots watch OpenNap network for shared songs on collections.

Smoothing Function

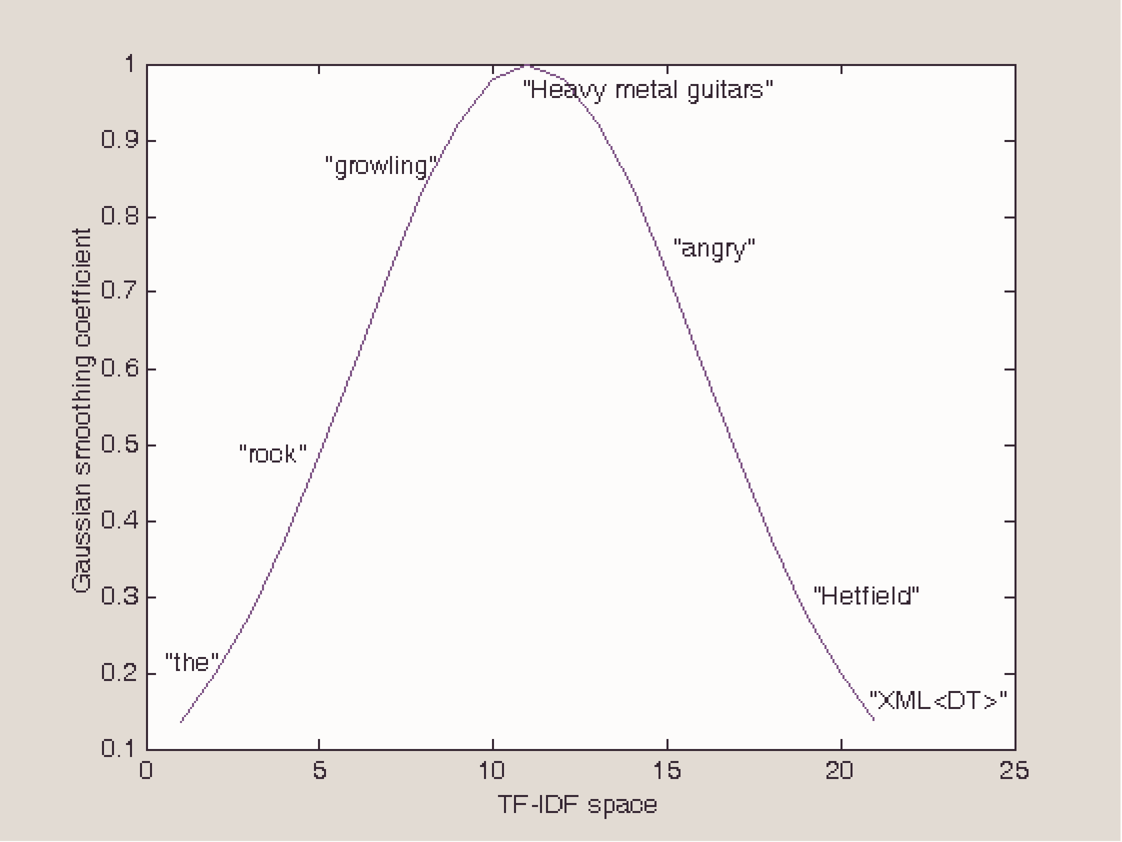
- Inputs are term and document frequency with mean and standard deviation:

$$s(f_t, f_d) = \frac{f_t e^{-(\log(f_d)-\mu)^2}}{2\sigma^2}$$

- We use mean of 6 and stdev of 0.9

Smooth the TF-IDF

- Reward ‘mid-ground’ terms



Example

- For Portishead:

n1 Term	Score
gibbons	0.0774
dummy	0.0576
displeasure	0.0498
nader	0.0490
tablets	0.0479
godrich	0.0479
irks	0.0467
corvair	0.0465
durban	0.0461
farfisa	0.0459

n2 Term	Score
beth gibbons	0.1310
sour times	0.0954
blue lines	0.0718
17 feb	0.0675
lumped into	0.0665
which come	0.0635
mellow sound	0.0573
in together	0.0519
musicians will	0.0494
enough like	0.0494

np Term	Score
beth gibbons	0.1648
trip hop	0.1581
dummy	0.1153
goosebumps	0.0756
soulful melodies	0.0608
rounder records	0.0499
dante	0.0499
may 1997	0.0499
sbk	0.0499
grace	0.0499

adj Term	Score
cynical	0.2997
produced	0.1143
smooth	0.0792
dark	0.0583
particular	0.0571
loud	0.0558
amazing	0.0457
vocal	0.0391
unique	0.0362
simple	0.0354

Style ID experiment

- AMG style prediction
  - ‘Soft’ ground truth
- Audio:
  - 10-20 songs per artist
  - Minnowmatch testbed
  - Cross album
- 25 artists, 5 styles

Cultural/Acoustic Disconnects

- Styles can be related acoustically but not culturally
  - R&B / top 40 pop (marketing)
  - Rap (substyle glut)
- Or culturally and not acoustically
  - “IDM”

What’s a Style?

- Style vs. genre
  - All styles have genres above them
  - Artists can have multiple styles
  - Albums can have styles, too
- Style as a small music cluster of cultural perception
  - = Sound + Peers + Time

Why Style?

- Recommendation within styles
  - Marketing recommendation
  - New music recommendation
  - Self-recommendation
- Creating a music hierarchy
  - Search
  - Musical synonymy / hypernymy

Artist List & Styles

Heavy Metal	Contemporary Country	Hardcore Rap	IDM	Female R&B
Guns N' Roses	Billy Ray Cyrus	DMX	Boards of Canada	Lauryn Hill
AC/DC	Alan Jackson	Ice Cube	Aphex Twin	Aaliyah
Skid Row	Tim McGraw	Wu-Tang Clan	Squarepusher	Debelah Morgan
Led Zeppelin	Garth Brooks	Mystikal	Plone	Toni Braxton
Black Sabbath	Kenny Chesney	Outkast	Mouse on Mars	Mya

Audio Representation

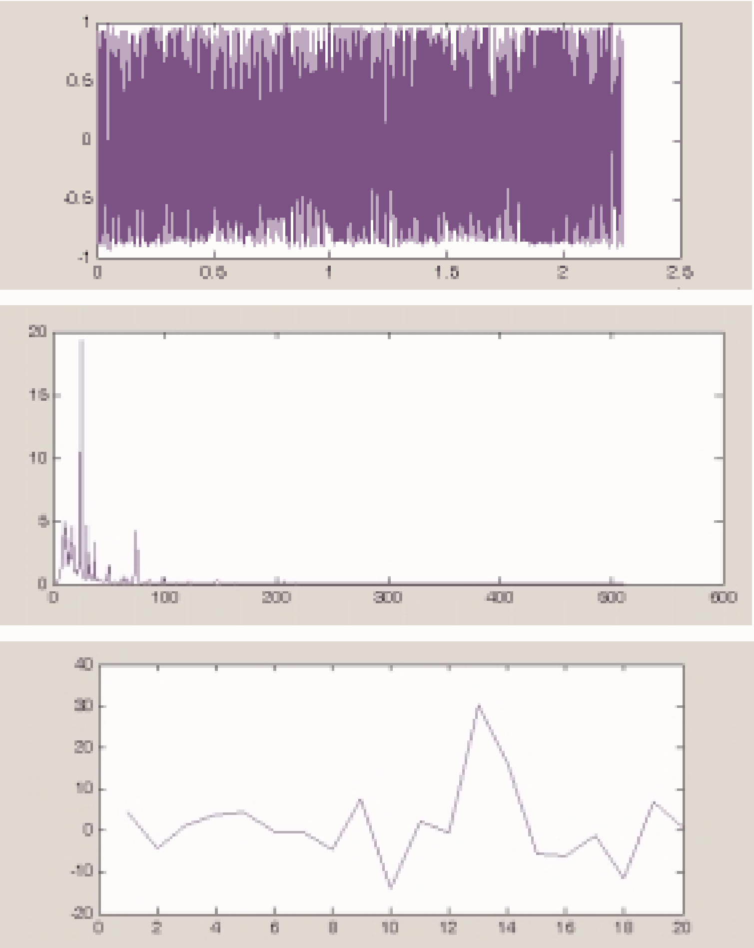
2sec audio

↓

PSD

↓

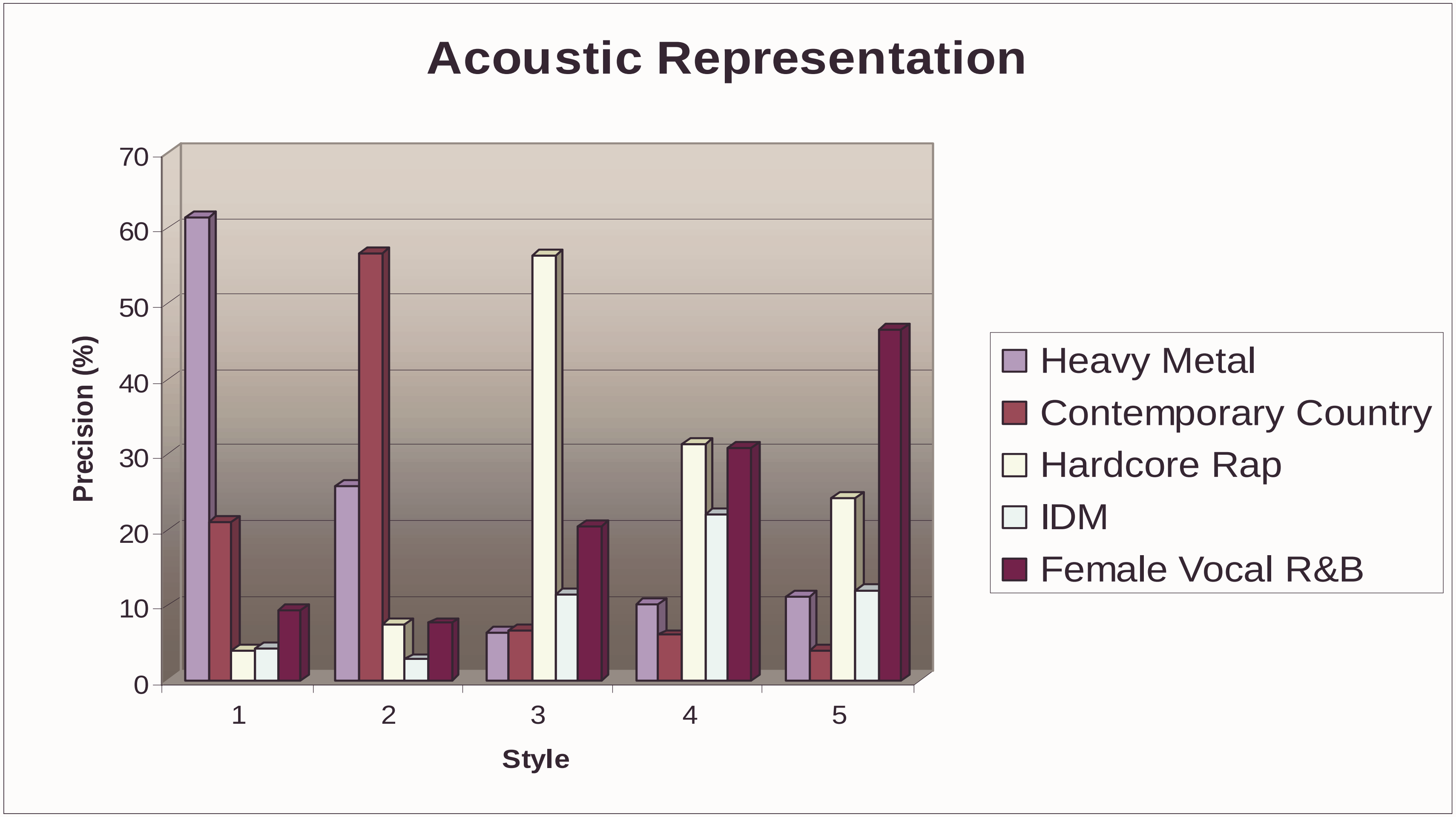
PCA weighting



Acoustic Representation Classification

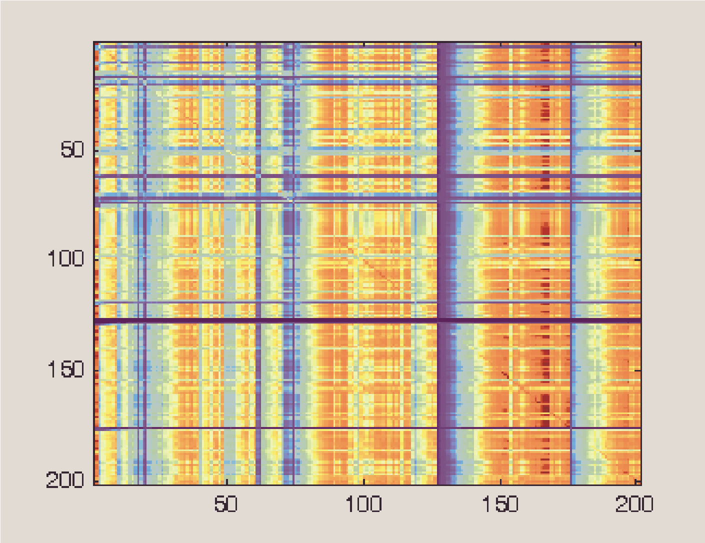
- Feedforward time-delay NN
  - 3 frame delay
- Backpropagation
- Input layer – 20 PCA coefficients
- Hidden layer of 40 nodes
- 4 train/1 test batch split

Acoustic Representation Results

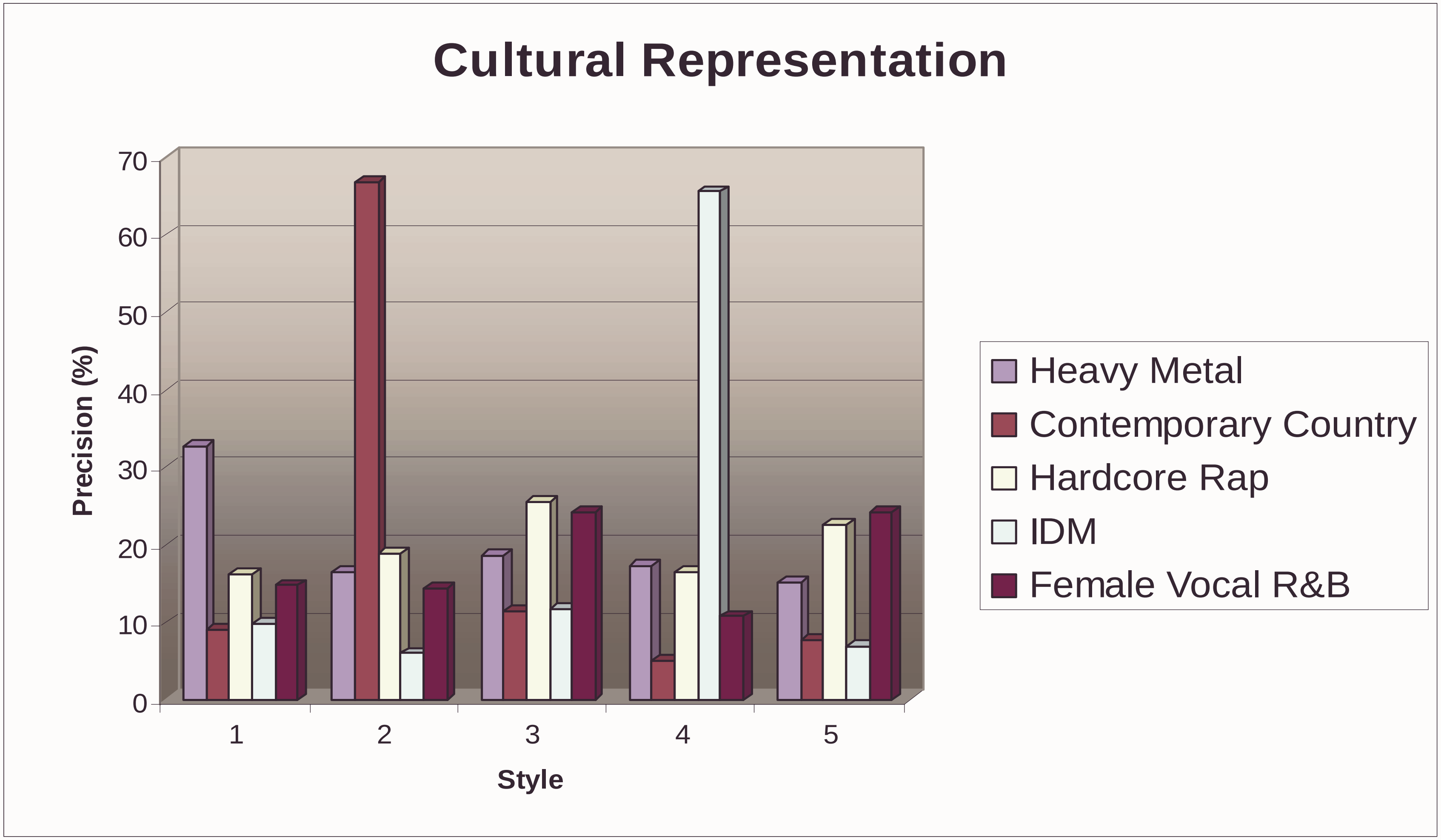


Cultural Representation Classification

- Gram matrix of CM kernel space:
  - Sum overlap of smoothing function
- K- nearest-neighbors clustering
- Given a new artist, find closest cluster in kernel space



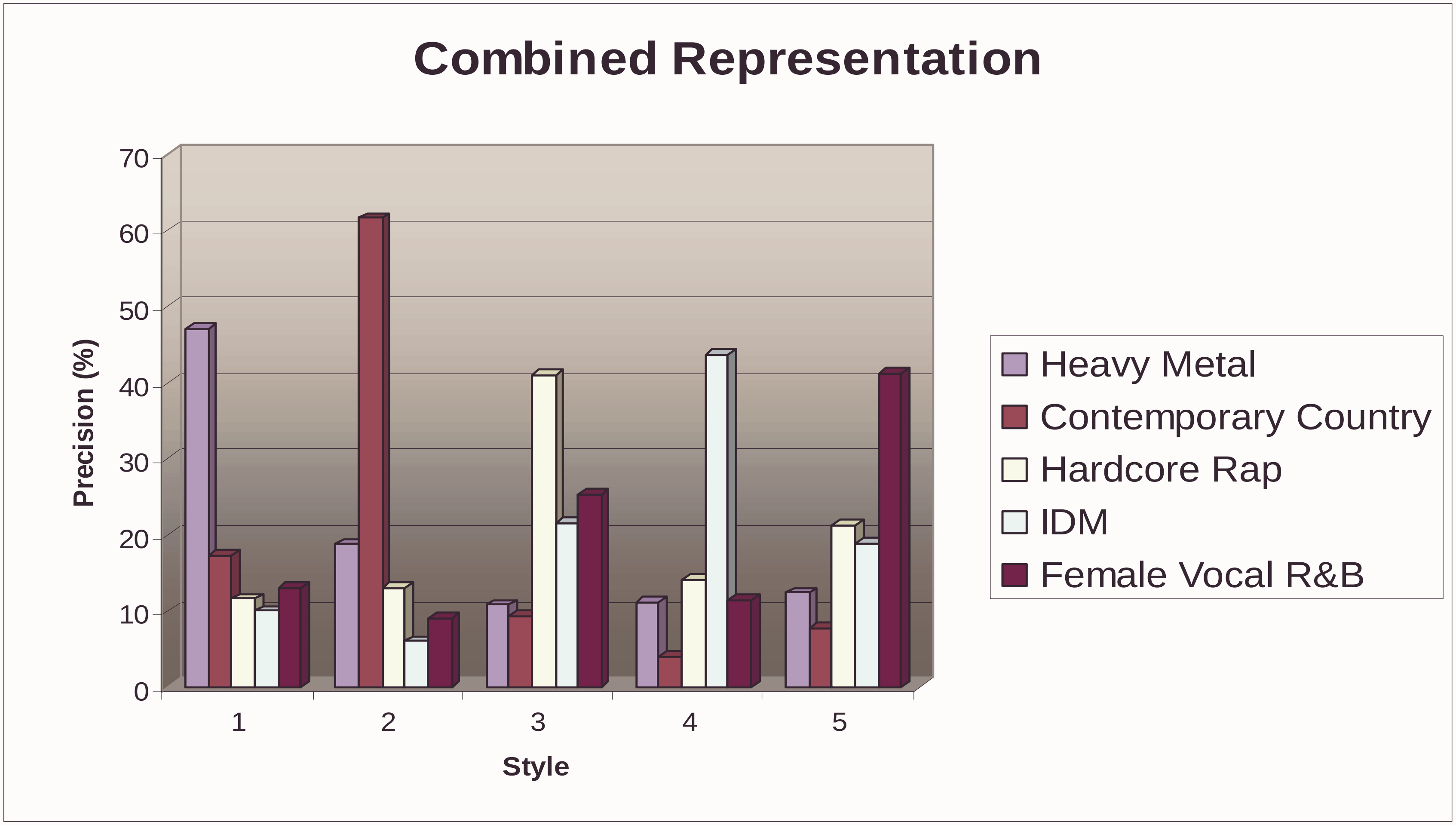
Cultural Representation Results



Combined Classification

- Can’t compare independent distance measures
- So we look at hypothesis probabilities
- Average or multiply?

Combined Classification Results



Style ID Overall

What's Next

- CM proven for artist similarity
  - Against AMG editors
    - Whitman/Lawrence (ICMC)
  - Against human evaluation
    - Ellis/Whitman/Berenzweig/Lawrence (ISMIR)
- Current IR uses of CM:
  - Recommendation / Buzz Factor Extraction
  - Query by Description
  - Grounding Sound

Time-Aware Recommendation

- CM is ‘Time-Aware:’
  - Artists change over time
  - So does audience perception
- Gauges buzz
  - Parsable content goes up during album releases, major news
- Avoids ‘stale’ recommendations
- Captures that non-audio ‘aboutness’

Query by Description

- “Play me something fast with an electronic beat!” “I’m tired tonight, let’s hear some romantic music.”
- CM vectors in time-aware QBD.
- We don’t need to label any data– the internet does that for us.

Grounding Sound

- Bimodal representation for symbol grounding of music
- Understanding sound innately

Conclusions

- Style useful and peculiar delimiter
- Test case for non-audio aboutness
- CM as cultural representation
  - Freely available
- Thanks: MMM group, Steve, Adam, Dan, Ryan Rifkin