

<http://www.xkcd.com/655/>

Audio Retrieval

David Kauchak

cs458

Fall 2012

Thanks to Doug Turnbull for some of the slides

Administrative

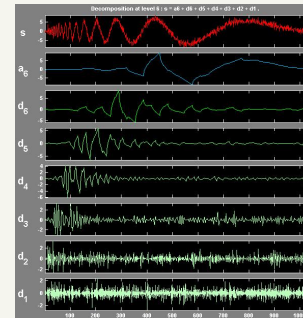
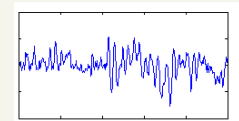
- Assignment 4 part 1 due Tuesday
- Final project instructions out soon...
 - I'll send an e-mail
- Homework 4
- No class next Tuesday!

Audio Index construction

Audio files to be indexed



audio preprocessing



slow,
jazzy,
punk

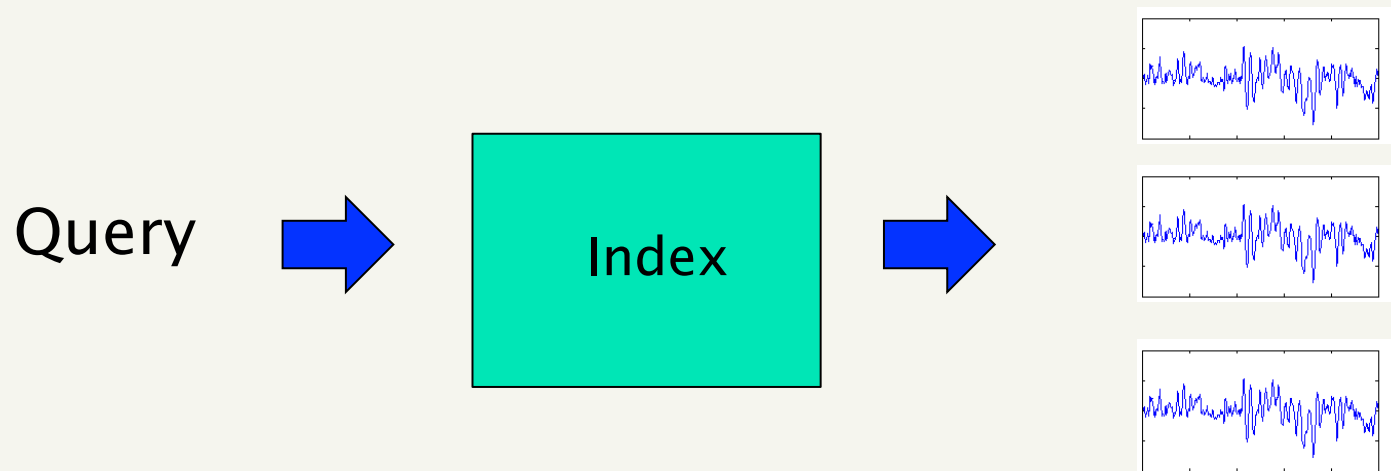
indexer

Index

may be keyed off of text

may be keyed off of audio features

Audio retrieval



Systems differ by what the query input is and how they figure out the result

Song identification

Given an audio signal, tell me what the “song” is



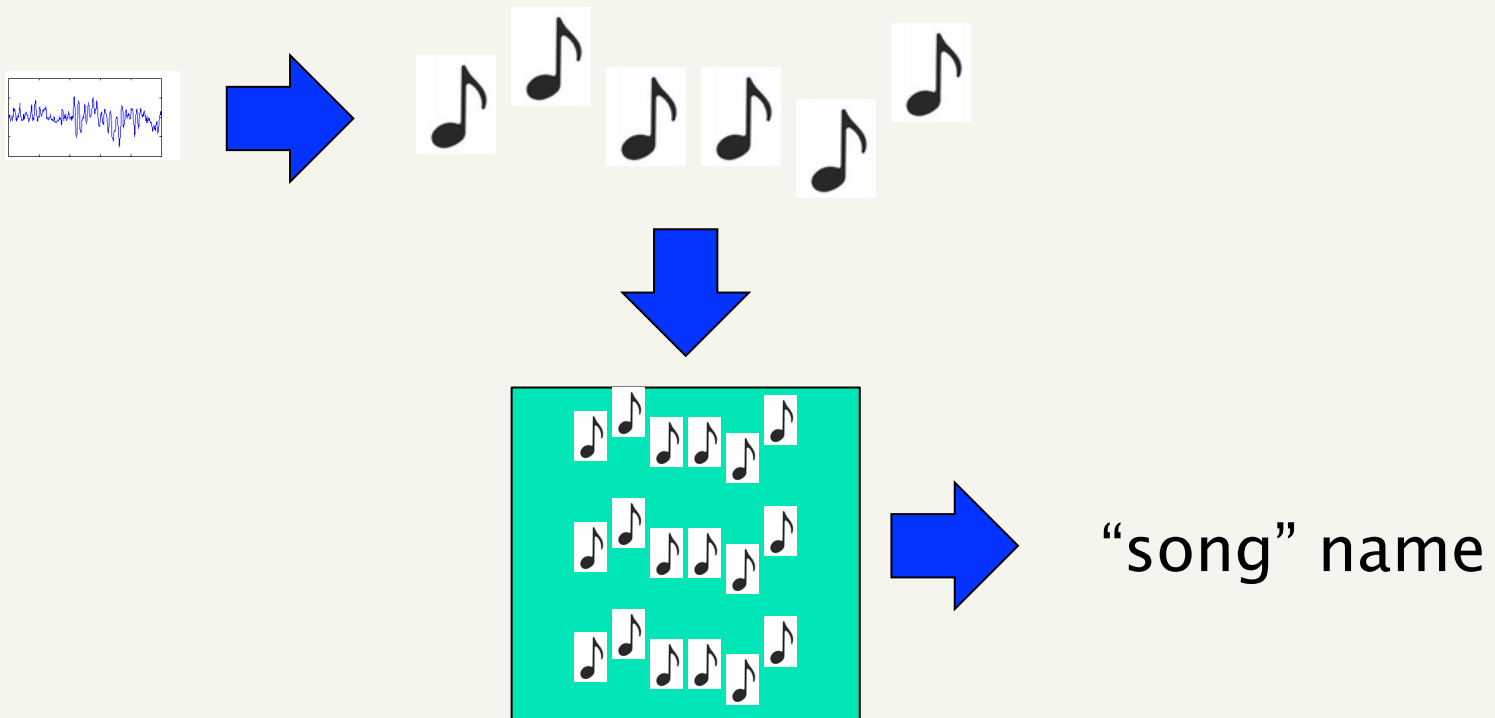
Examples:

- Query by Humming
- 1-866-411-SONG
- Shazam
- Bird song identification
- ...

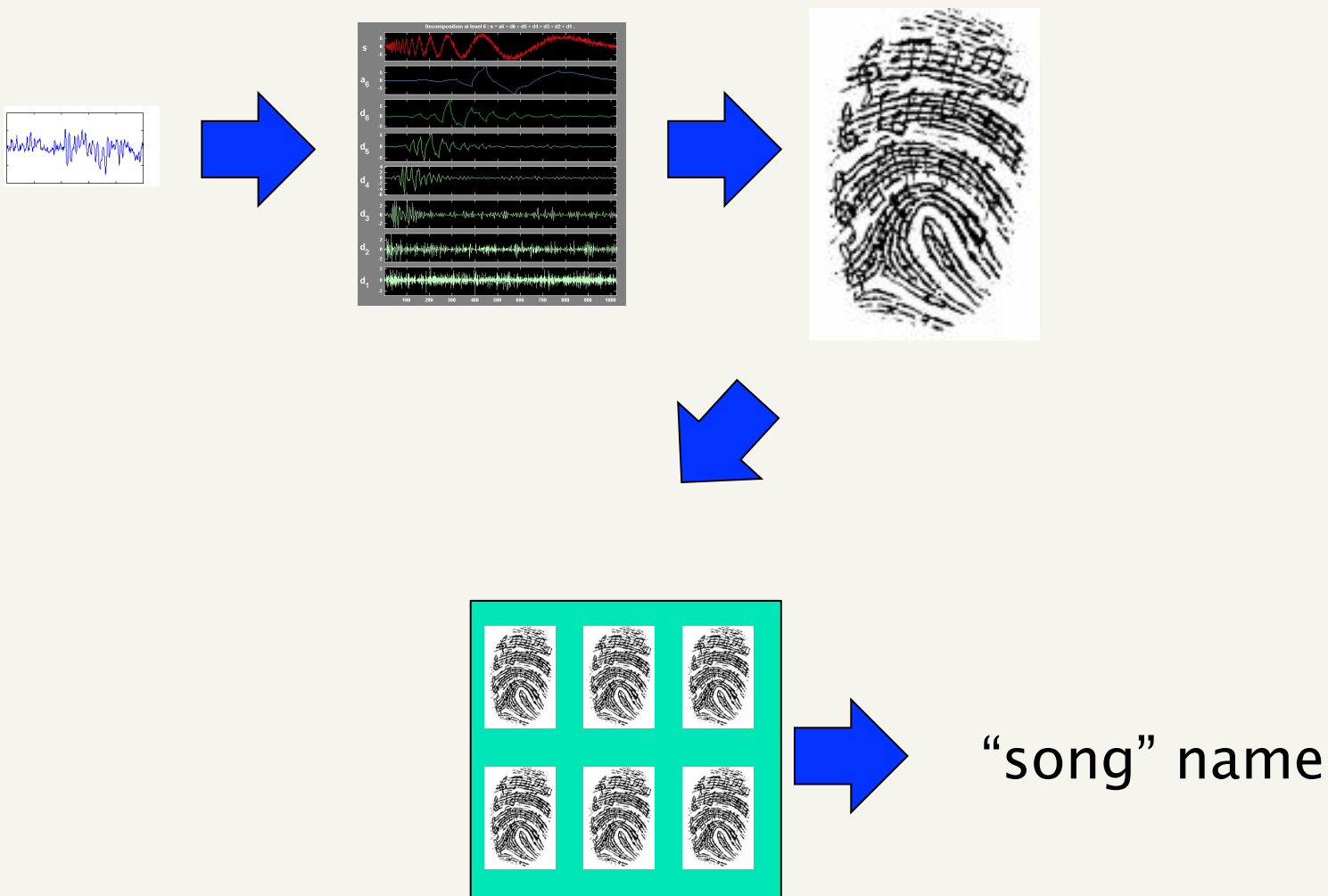
Song identification

How might you do this?

Query by humming

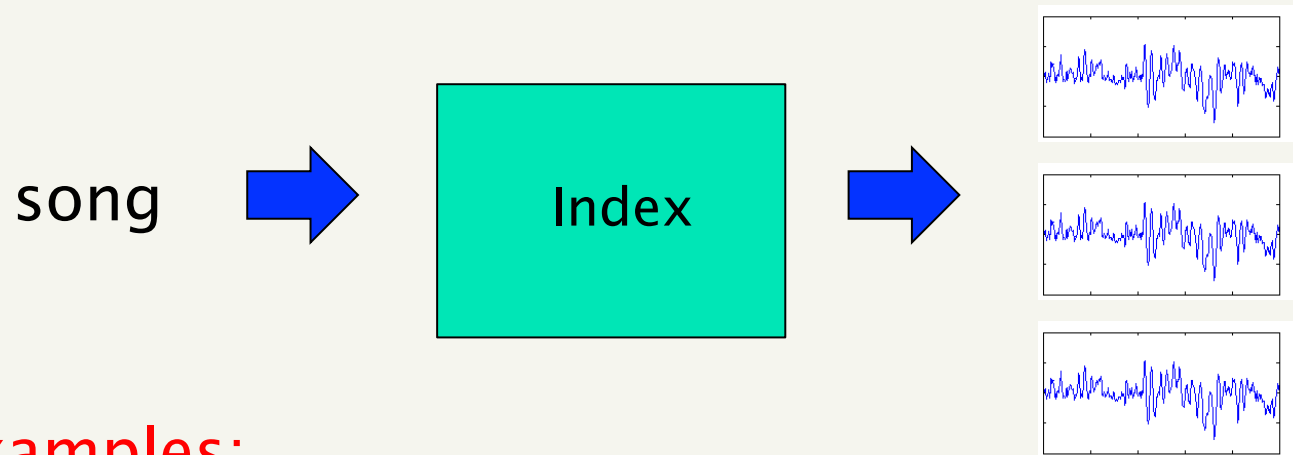


Song identification



Song similarity

Find the songs that are most similar to the input song



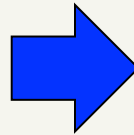
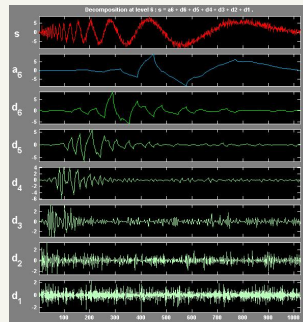
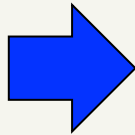
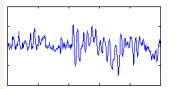
Examples:

- Genius
- Pandora
- Last.fm

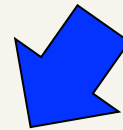
Song similarity

How might you do this?

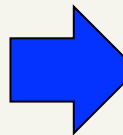
IR approach



$f_1 f_2 f_3 \dots f_n$



$f_1 f_2 f_3 \dots f_n$
 $f_1 f_2 f_3 \dots f_n$
 $f_1 f_2 f_3 \dots f_n$










$f_1 f_2 f_3 \dots f_n$

$f_1 f_2 f_3 \dots f_n$

rank by cosine sim

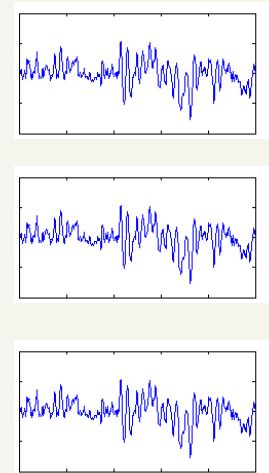
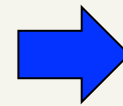
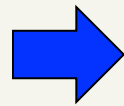
Song similarity: collaborative filtering

	<i>user₁</i>	<i>user₂</i>	<i>user₃</i>	<i>user₄</i>	<i>user_n</i>
song ₁					
song ₂					
song ₃					
⋮					
song _m					

What might you conclude from this information?

Songs using descriptive text search

jazzy,
smooth,
easy listening



Examples:

- Very few commercial systems like this ...



Meerkat- Control your radio



Meerkat

Type a tag to begin

Add Tag

ex. pop, cheerful, urban



Meerkat- Control your radio



Meerkat

Type a tag to begin

ex. pop, cheerful, urban

Add Tag





Meerkat



Spooky Lane by Sam Bush

Uncle Pen by Bill Monroe & His Bluegrass Boys

I'll Never Shed Another Tear by The Osborne Brothers

'Spooky Lane' Tags:

country influences

theme time radio hour

mandolin

seen live

bluegrass

grateful dread

Station Tags:

Add Tag

bluegrass





Meerkat



Spooky Lane by Sam Bush

Walk On Boy by Doc Watson

Mossy Cow by Yonder Mountain String Band

'Spooky Lane' Tags:

country influences

theme time radio hour

mandolin

seen live

bluegrass

grateful dread

Station Tags:

Add Tag

bluegrass ✕

mandolin ✕



Meerkat



Spooky Lane by Sam Bush

Walk On Boy by Doc Watson

Mossy Cow by Yonder Mountain String Band

Down Where The River Bends by The Osborne Brothers

'Walk On Boy' Tags:

raw music

an upbeat two-step
feel

mandolin

minor key tonality

folk

singer-songwriter

Station Tags:

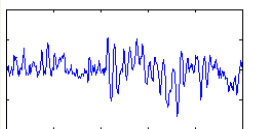
Add Tag

bluegrass ✕

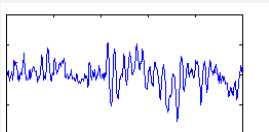
mandolin ✕

Music annotation

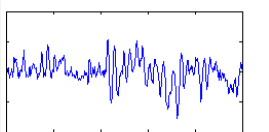
The key behind keyword based systems is annotating the music with tags



dance, instrumental, rock



blues, saxaphone, cool vibe



pop, ray charles, deep

Ideas?

Annotating music

The human approach

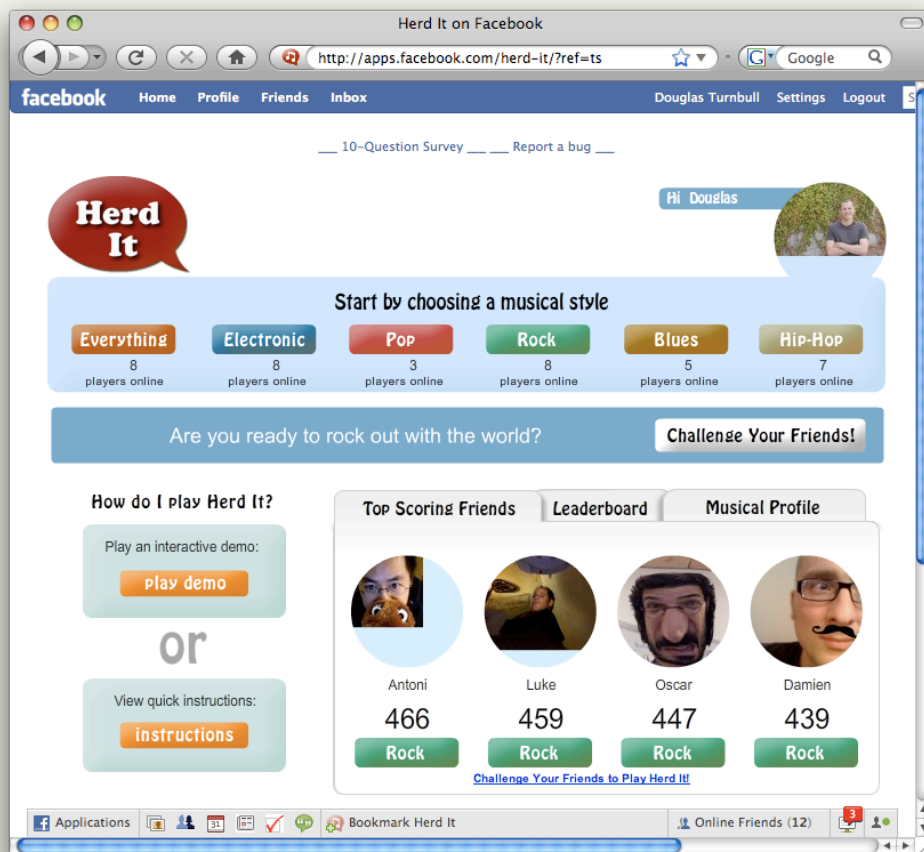


“expert musicologists”
from Pandora

Pros/Cons?

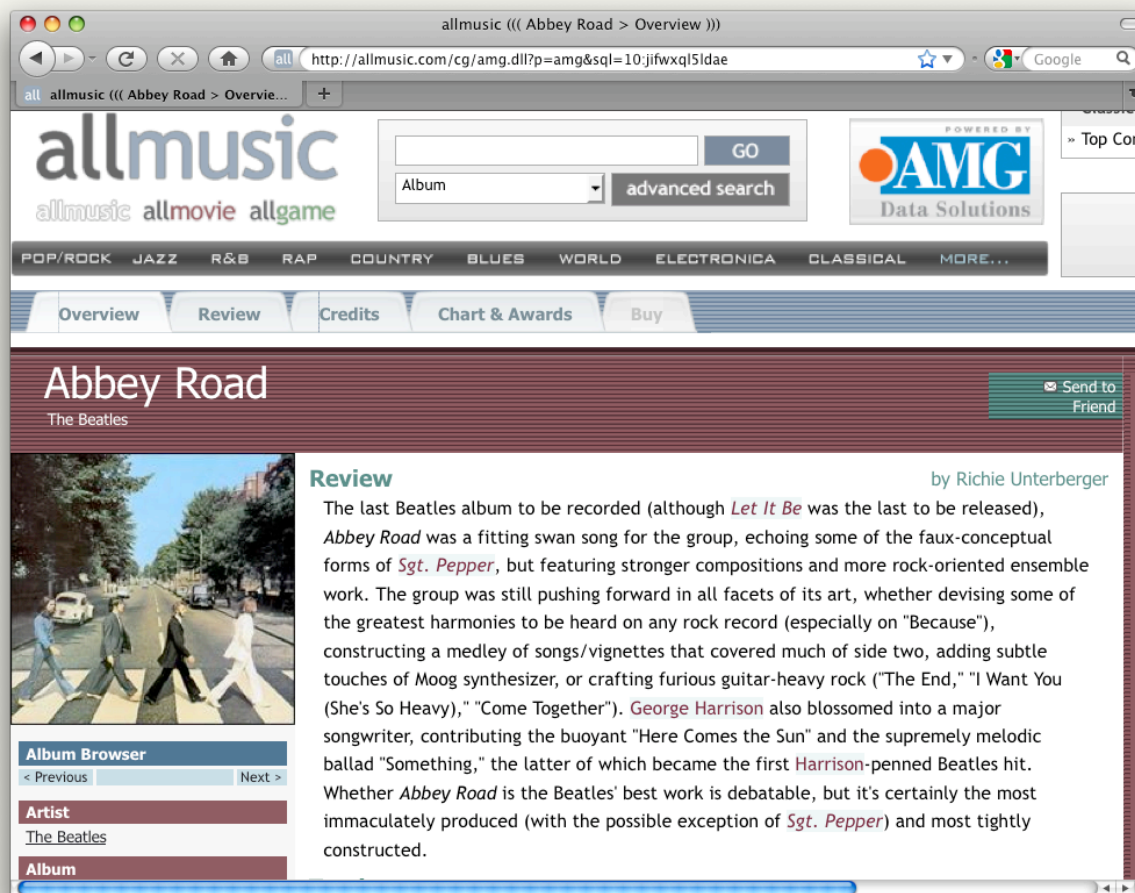
Annotating music

Another human approach: games



Annotating music

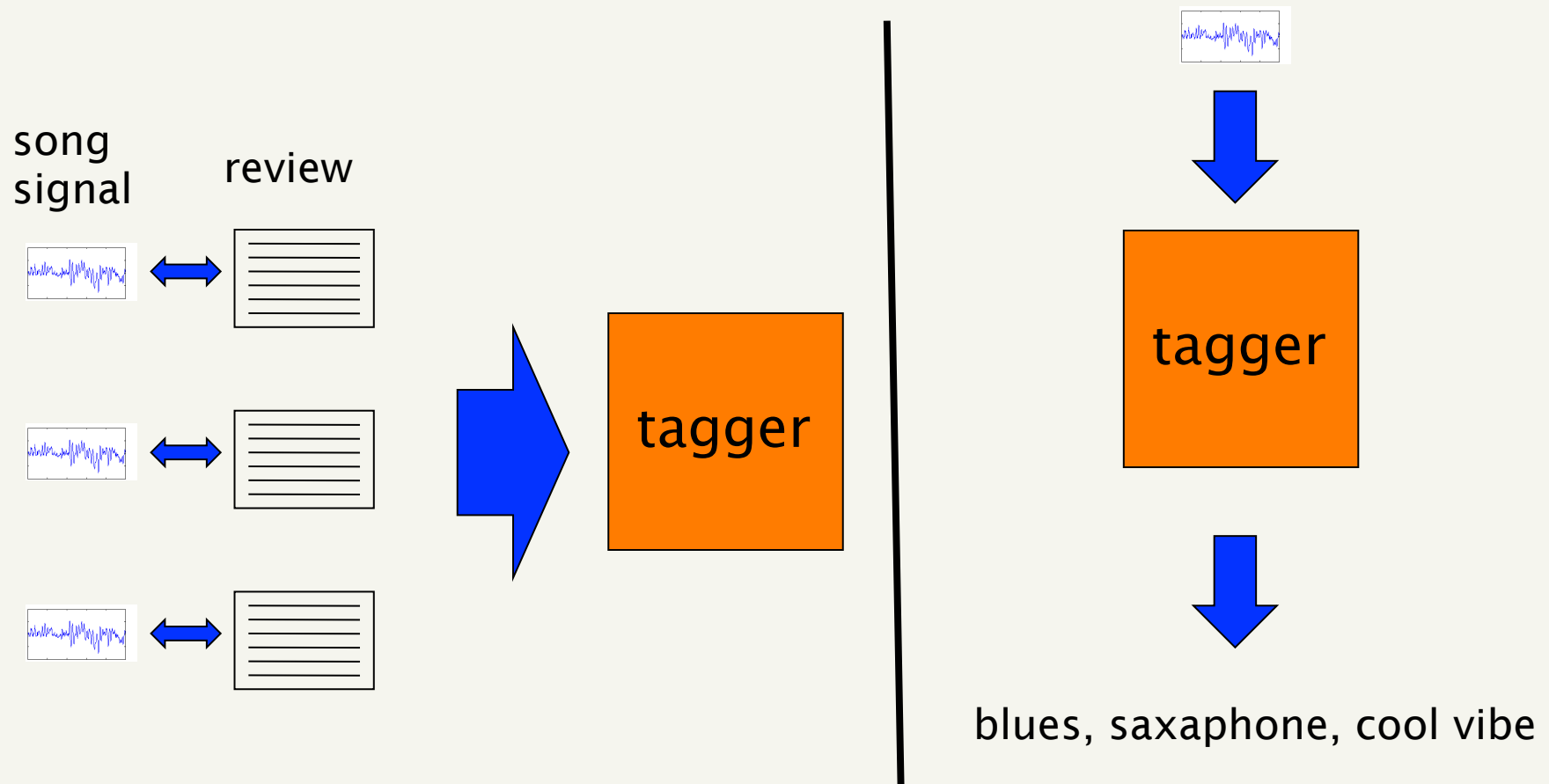
the web: music reviews



challenge?

Automatically annotating music

Learning a music tagger

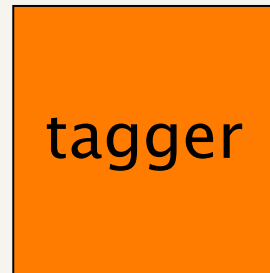
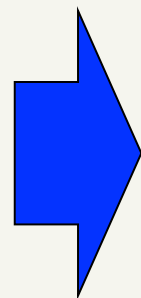
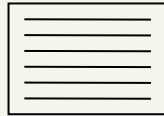
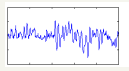
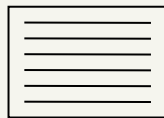
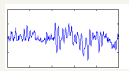
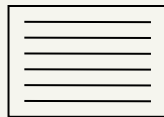
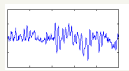


Automatically annotating music

Learning a music tagger

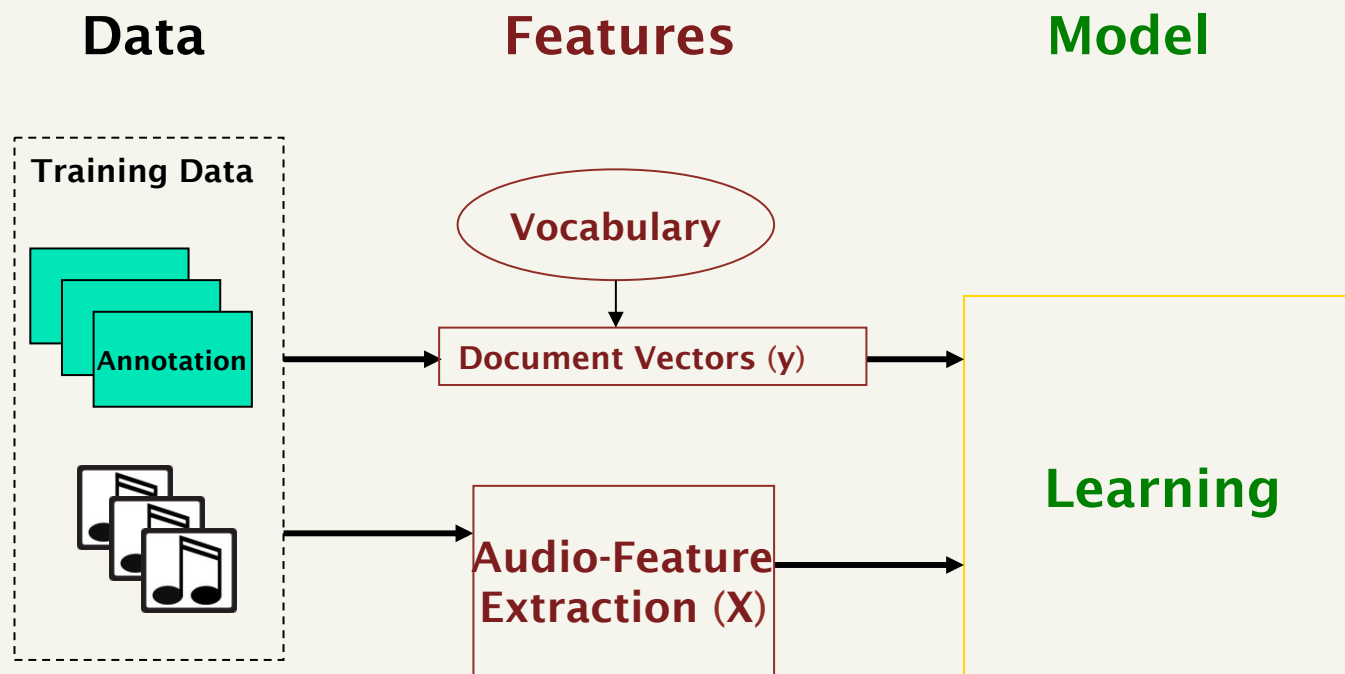
song
signal

review



What are the tasks we
need to accomplish?

System Overview

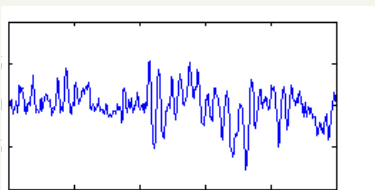


Automatically annotating music

First step, extract “tags” from the reviews

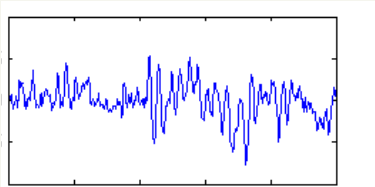
Frank Sinatra - Fly me to the moon

This is a jazzy, singer / songwriter song that is calming and sad. It features acoustic guitar, piano, saxophone, a nice male vocal solo, and emotional, high-pitched vocals. It is a song with a light beat and a slow tempo.



Dr. Dre (feat. Snoop Dogg) - Nuthin' but a 'G' thang

This is a dance poppy, hip-hop song that is arousing and exciting. It features drum machine, backing vocals, male vocal, a nice acoustic guitar solo, and rapping, strong vocals. It is a song that is very danceable and with a heavy beat.

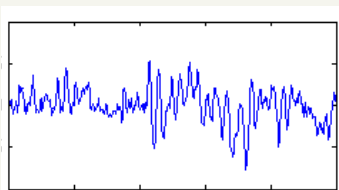


Automatically annotating music

First step, extract “tags” from the reviews

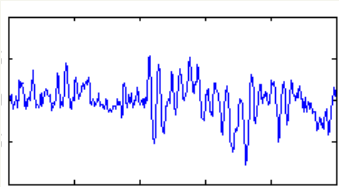
Frank Sinatra - Fly me to the moon

This is a **jazzy**, **singer / songwriter** song that is **calming** and **sad**. It features **acoustic guitar**, **piano**, **saxophone**, a nice **male vocal solo**, and **emotional**, **high-pitched** vocals. It is a song with a **light beat** and a **slow tempo**.



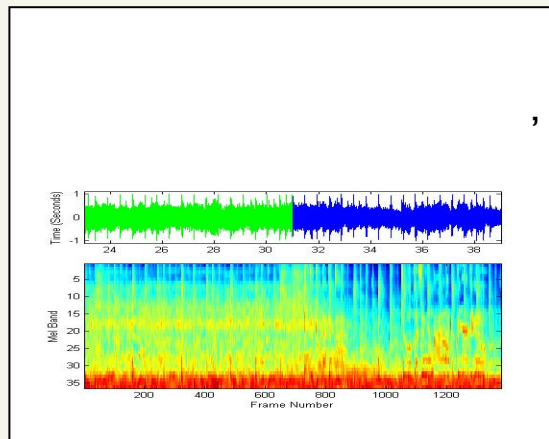
Dr. Dre (feat. Snoop Dogg) - Nuthin' but a 'G' thang

This is a **dance poppy**, **hip-hop** song that is **arousing** and **exciting**. It features **drum machine**, **backing vocals**, **male vocal**, a nice **acoustic guitar solo**, and **rapping**, **strong vocals**. It is a song that is very **danceable** and with a **heavy beat**.



Content-Based Autotagging

Learn a probabilistic model that captures a relationship between **audio content** and **tags**.

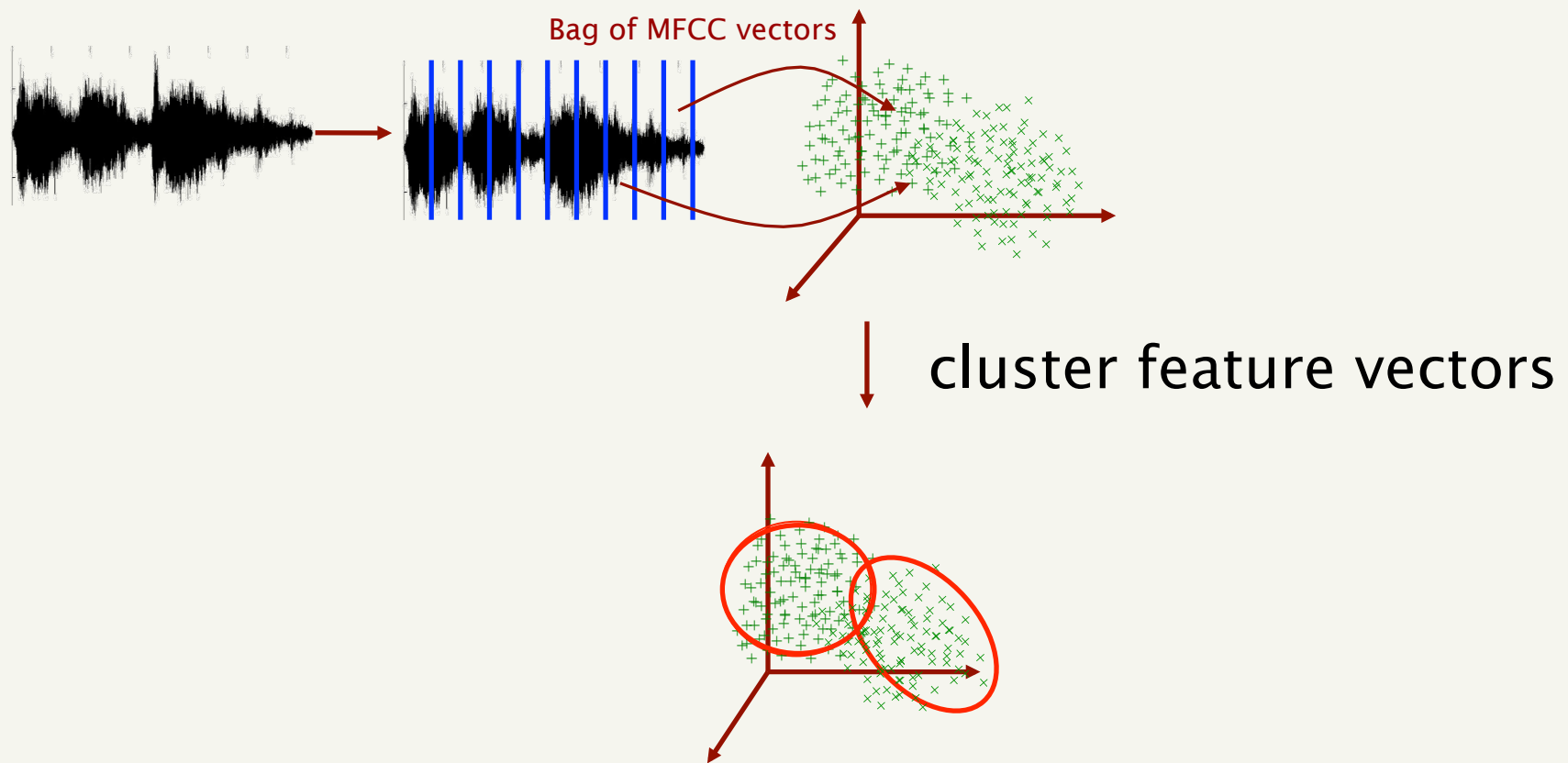


Autotagging
→

‘Jazz’
‘Male Vocals’
‘Sad’
‘Slow Tempo’

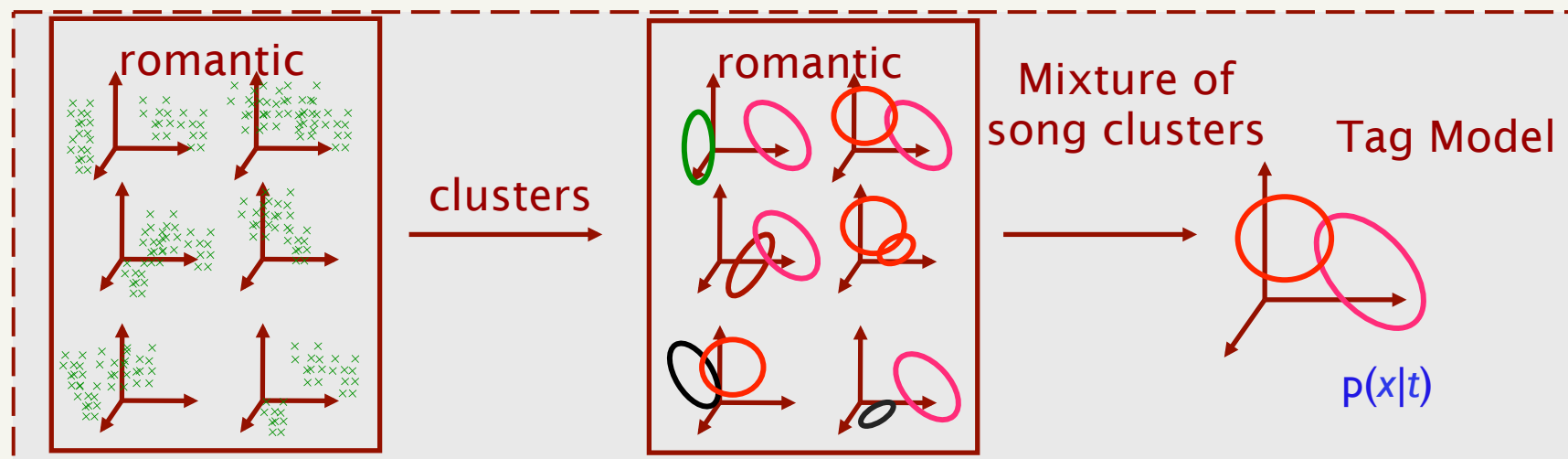
$p(\text{tag} \mid \text{song})$

Modeling a Song



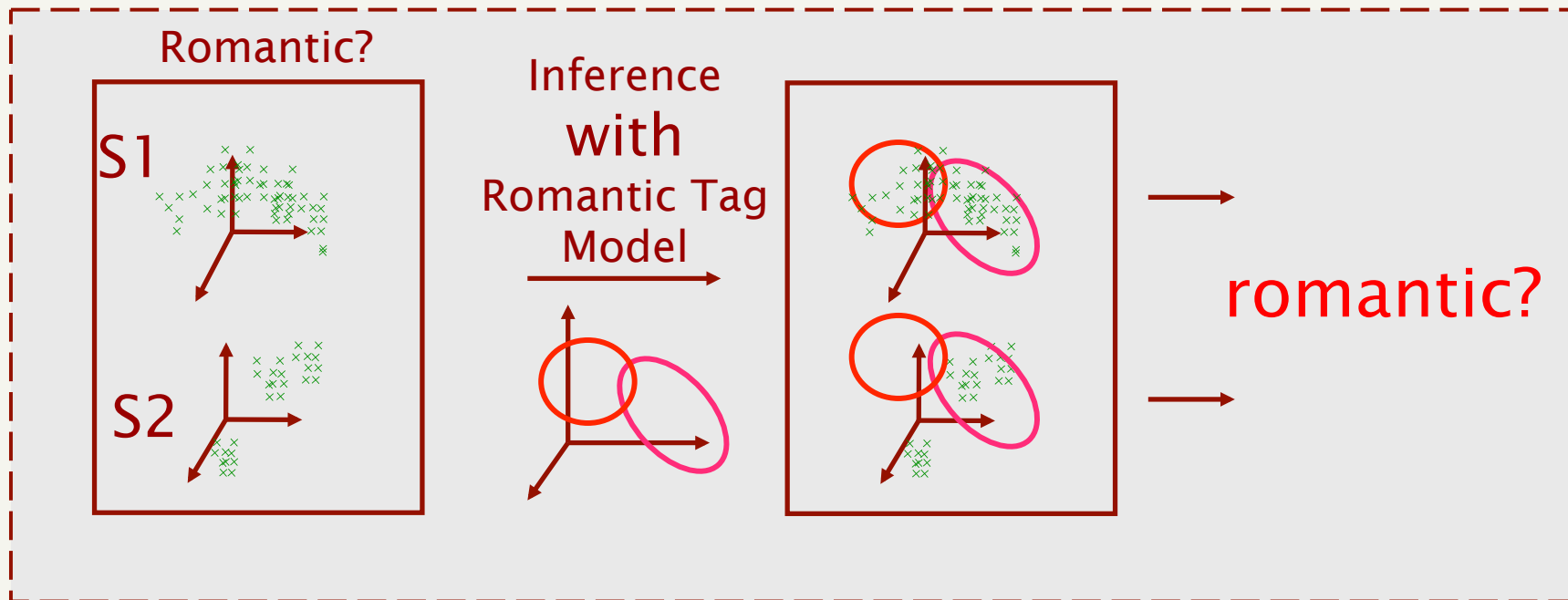
Modeling a Tag

1. Take all songs associated with tag t
2. Estimate '*features clusters*' for each song
3. Combine these clusters into a single representative model for that tag



Determining **Tags**

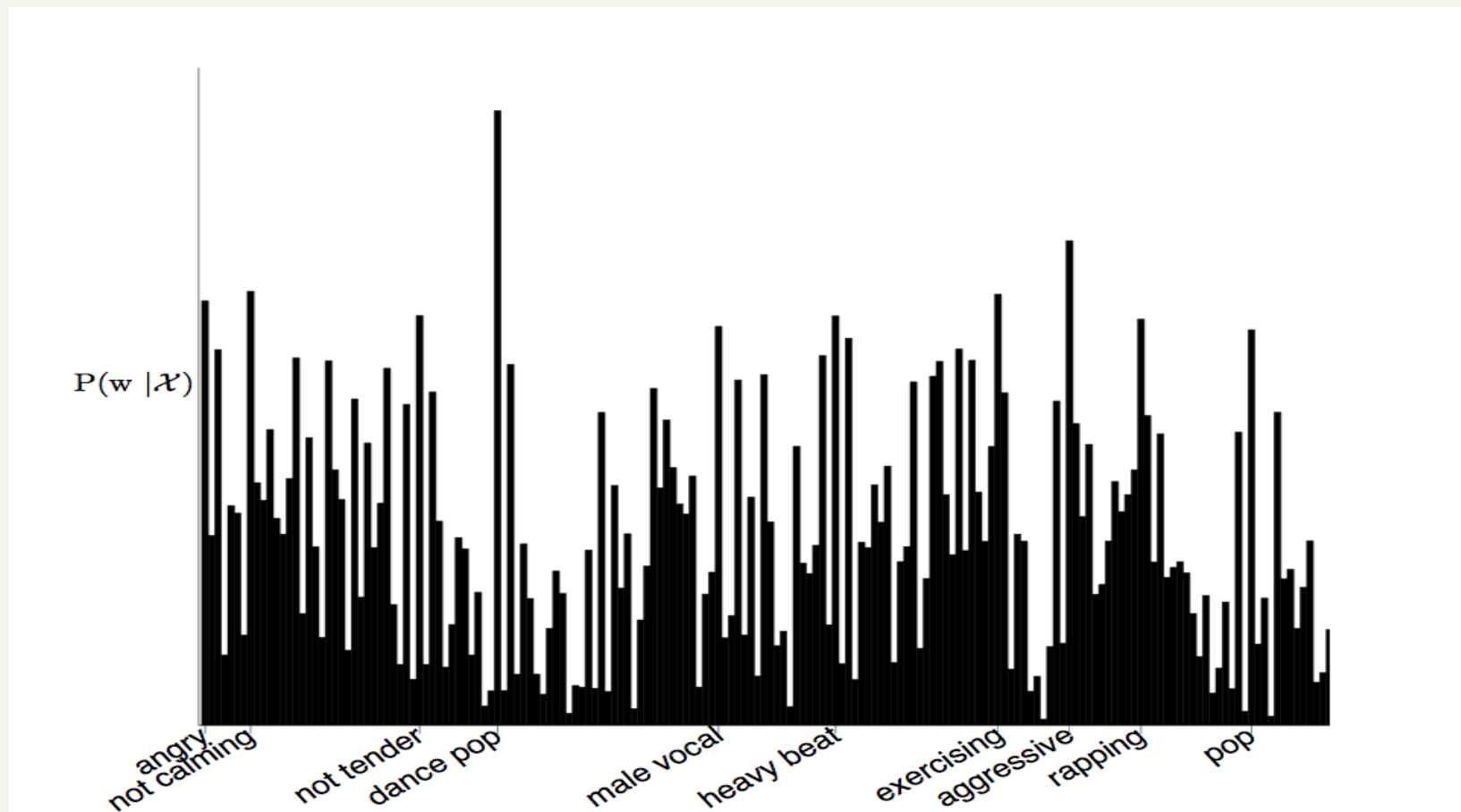
1. Calculate the likelihood of the features for a tag model





Annotation

Semantic Multinomial for “Give it Away” by the Red Hot Chili Peppers
http://www.youtube.com/watch?v=Mr_uHJPUIO8



The CAL500 data set

The [Computer Audition Lab 500-song](#) (CAL500) data set

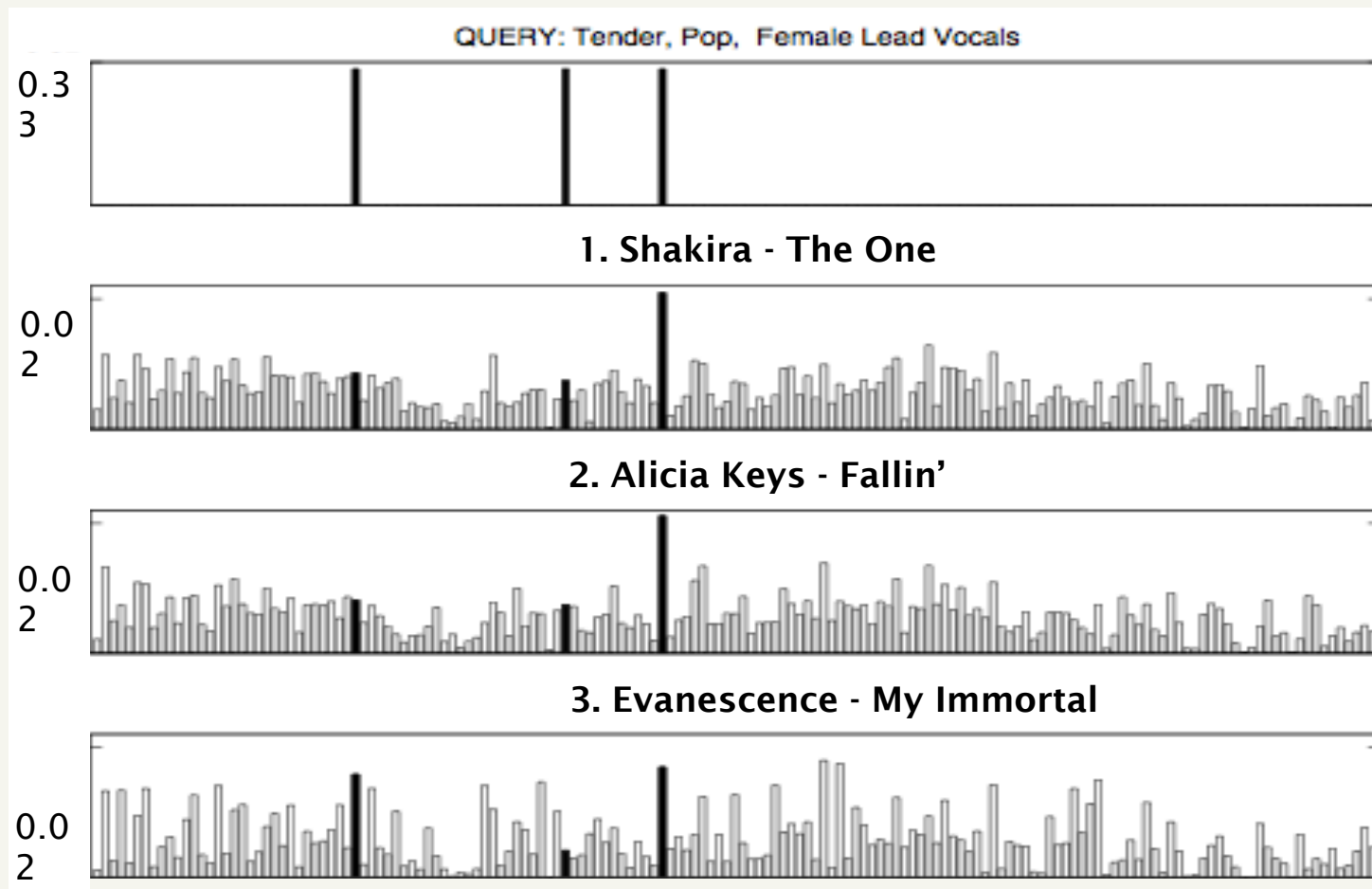
- 500 'Western Popular' songs
- 174-word vocabulary
 - genre, emotion, usage, instrumentation, rhythm, pitch, vocal characteristics
- 3 or more annotations per song
- 55 paid undergrads annotated music for 120 hours

Other Techniques

1. Text-mining of web documents
2. 'Human Computation' Games

Retrieval

The top 3 results for - “pop, female vocals, tender”



Retrieval

Query	Retrieved Songs
'Tender'	Crosby, Stills and Nash - Guinnevere Jewel - Enter from the East Art Tatum - Willow Weep for Me John Lennon - Imagine Tom Waits - Time
'Female Vocals'	Alicia Keys - Fallin' Shakira - The One Christina Aguilera - Genie in a Bottle Junior Murvin - Police and Thieves Britney Spears - I'm a Slave 4 U
'Tender' AND 'Female Vocals'	Jewel - Enter from the East Evanescence - My Immortal Cowboy Junkies - Postcard Blues Everly Brothers - Take a Message to Mary Sheryl Crow - I Shall Believe

Annotation results

Annotation of the CAL500 songs with 10 words from a vocabulary of 174 words.

Model	Precision	Recall
Random	0.14	0.06
Our System	0.27	0.16
Human	0.30	0.15

Retrieval results

Model	AROC
Random	0.50
Our System - 1 Word	0.71
Our System - 2 Words	0.72
Our System - 3 Words	0.73

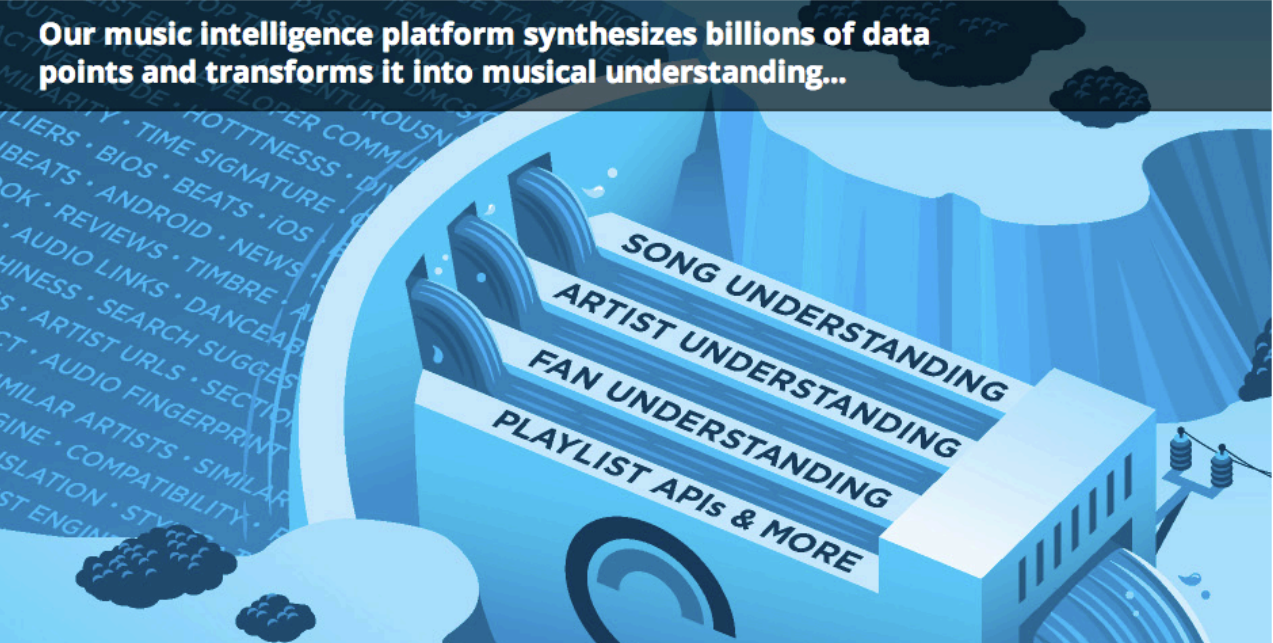
Echo nest

We Know Music...

Sign up for our newsletter for updates on new feature releases, developer meet-ups and new apps powered by our platform.

[Subscribe](#)

Our music intelligence platform synthesizes billions of data points and transforms it into musical understanding...



1,126,118,372,527	34,403,057	2,268,157	351
<small>Data points about...</small>	<small>Known Songs</small>	<small>Known Artists</small>	<small>Music Applications On Our Platform</small>

<http://the.echonest.com/>