# “Music Information Retrieval” Community

**What:** Developing systems that retrieve **music**

**When:** Late 1990’s to Present

**Where:** ISMIR - conference started in 2000

**Why:** lots of “digital” music, lots of music lovers, lots of powerful computers

# How can we find find music?

- **Query-by-Metadata - artist, song, album, year**
  - We must know what we want
- **Query-by-(Humming, Tapping, Beatboxing)**
  - Requires talent
- **Query-by-Song-Similarity**
  - We must possess ‘acoustically’ similar songs
- **Query-by-Semantic-Description**
  - **Google** seems to work pretty well for **text**
  - **Semantic Image Labeling** is a hot topic in **Computer Vision**
  - Can it work for **music?**

# Semantic Music Annotation and Retrieval

Our goal is build a system that can

1. **Annotate** a song with meaningful words
2. **Retrieve** songs given a text-based description

**Plan:** Learn a probabilistic model that captures a relationship between the **audio content** of a song and **words** that describe the song.

# Collecting Semantic Music Data

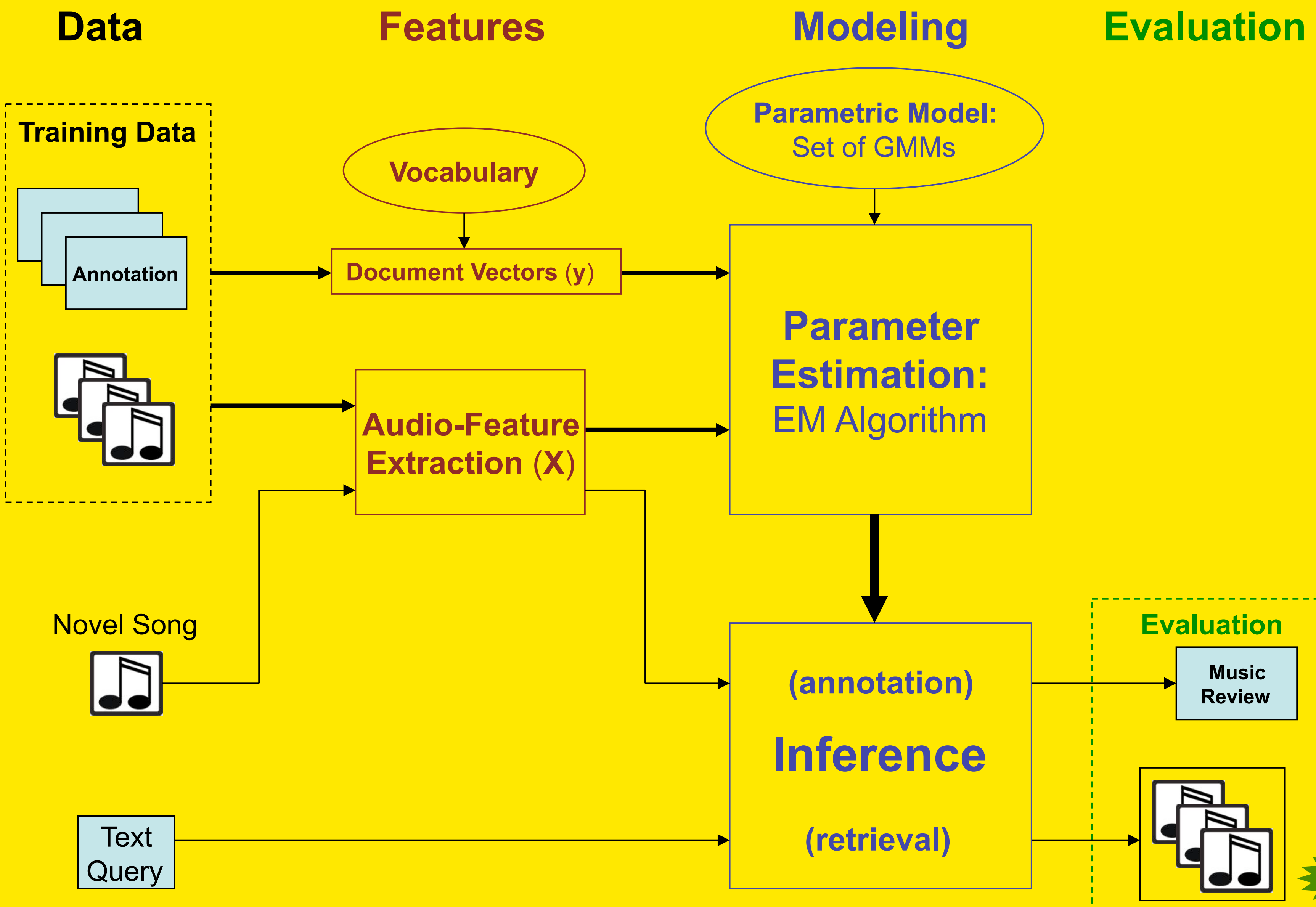
**CAL500 Data Set**: We have collected 1700 annotations for 500 'western popular' songs by having 55 individuals listen to and evaluate music.

- Each song is annotated by at least 3 individuals.

An **annotation** reflects the 'strength of association' between a song and 173 words.

- Words relate to Instrumentation, Genre, Emotion, Vocals, Usages, Quality, Tempo, …
- Collected using a standard survey

# System Overview



# Our Model

- Each song is represented by a time series **X= {$x_1$,…,$x_t$}**
  - dynamic Mel-Frequency Cepstral Coefficients [McKinney03]
- For each word w, we learn a 'word-model' **p(x|w)** using songs that are associated with the word.
  - p(x|w) is modeled using a Gaussian Mixture Model (GMM)
- **Annotation:** Given a novel song, we pick words by comparing the likelihood of the audio features under each word-model.
- **Retrieval:** Given a text query, we pick songs that are likely under the word-models associated with the words in the query.

# Annotation: Automatic Music Reviews

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

This is **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** that you might like listen to while **at a party**.

## 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** that you might like listen to while **hanging with friends**.

# Retrieval: Query-by-Semantic-Description

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

# Quantifying Annotation

Our system annotates the Cal-500 songs with 10 words from our vocabulary of 173 words.

- 'Population Annotation' Ground Truth

**Metric**: 'Word' Precision & Recall

Consider word *w*,

$$\text{Precision} = \frac{\text{\# songs correctly annotated with } w}{\text{\# songs annotated with } w}$$

$$\text{Recall} = \frac{\text{\# songs correctly annotated with } w}{\text{\# songs that should have been annotated } w}$$

**Mean** Word Recall and Word Precision are the averages over all words in our vocabulary.

# Quantifying Annotation

Our system annotates the Cal-500 songs with 10 words from our vocabulary of 173 words.

| Model | Precision | Recall |
|---|---|---|
| Random | 0.17 | 0.05 |
| ~~Upper Bound~~ | ~~1.00~~ | ~~0.30~~ |
| Our System | **0.31** | **0.14** |
| Human | 0.34 | 0.11 |

**Key point:** By pooling human annotations, our model can produce annotations that are as consistent as annotations produced by individuals, when compared against a population average.

# What's next…

## Going Bigger

- Larger Vocabulary
- More Annotation
- Novel Applications

## Modeling dependencies

- Words Correlations (e.g., 'Classic Rock' & 'Electric Guitar')
- Audio Features have temporal dependencies
- Modeling individuals rather than populations

## Comparing alternative models

- Modeling heterogeneous data is hot topic in Machine Learning
- Many models have been proposed.
- Applications: Image Labeling, Text-Document Classification

## Exploring Novel Query Paradigms

- Query-by-semantic-example
- Heterogeneous queries

# To learn more…

The mathematics (parameter estimation, inference), a description of the Cal-500 data set, evaluation of annotation and retrieval performance, and much more is available at:

**cosmal.ucsd.edu/cal**

**“Talking about music is like dancing about architecture”**

- origins unknown

# A biased view of Music Classification

2000-03: **Music classification** (by genre, emotion, instrumentation) becomes a popular MIR task

- Undergrad Thesis on Genre Classification with G. Tzanetakis

2003-04: MIR community starts to **criticize** music classification problems

- ill-posed problem due to subjectivity
- not an end in itself
- performance 'glass ceiling'

2004-06: Focus turns to **Music Similarity** research

- Recommendation
- Playlist generation

2006-07: We view **Music Annotation** as a supervised multi-class labeling problem

- Like classification but with large, less-restrictive vocabulary