

---

# CS 4780 Final Project Proposal: What Makes a Song "Beatlesque"?

---

## 1. Size of Team

There are four (4) members in our team.

## 2. Motivation

Music recognition is an interesting field, with well-known applications such as Shazam and Spotify. We want to apply Machine Learning techniques to song data to embark on our own introduction to the field of music recognition.

## 3. Problem Statement

Due to time constraints and to simplify our task to a binary supervised classification, we will predict if a song presented to our system was performed by the Beatles or not. Our goal is to learn the importance of various features in a songs' meta data and musical data (i.e. energy, tempo, chord progressions) that can help us classify a song as belonging to The Beatles or not.

More formally, we want to answer the following questions:

1. What combination of features of a song are unique to the subset of artists that are either the Beatles or related to the Beatles?
2. What kernel best represents the importance of each of the features determined in 1.?

## 4. General Approach

### 4.1. Background Knowledge and Data Parsing

We will begin by finding 60 artists from each of the following categories:

1. Similar to the Beatles
2. Dissimilar to the Beatles
3. Randomly chosen

Given each of those 60 artists, we will choose 4 of their songs that match their corresponding category description, for a total of 240 songs for each category to train and test on. Once we know the songs we are looking

for, we can go through our data set, the Echo Nest API and query for song meta data such as "key", "beats per measure", "danceability", "energy", etc. After doing more research on how the sound of a song is characterized by its metadata, we will narrow this down to about 50 features and begin data preparation

We also plan to use this stage to do research on kernels that could potentially help us deal with the very large number of features. We will explore kernels shown in class, and variations that are used in music classification research.

### 4.2. Training, Validation and Testing

We will break down our learning into three stages

1. Compare songs by Beatles to songs of very different sounding genres like Metal or Classical, and learn weights of features to distinguish Beatles' songs from these artists
2. Compare Beatles songs to songs of similar sounding genres like Rock or Pop, in an effort to allow for a more nuanced way of distinguishing songs that are uniquely by the Beatles
3. Fine-tune the previous 2 stages so we can ultimately distinguish Beatles songs from any random artist.

We will split our feature data so that we train and validate on 200 songs in each stage, and test on the remaining 40. The data will be converted to a format compatible with SVM light.

We will also experiment with the kernels researched during the background knowledge stage and we will pick the kernel that works best during validation.

### 4.3. Final Product and Possible Extensions

We will develop a web interface that allows users to enter a song of their choice. We will then query the Echo Nest API to obtain the song's metadata, classify whether or not it is a Beatles' song using our final weight vector, and return the probability that the song was performed by The Beatles.

## 5. Resources

We will be using the Million Song Dataset, The Echo Nest API, Python, SVMLight, and github repositories to store and host our work.

## 6. Schedule

| Date         | Milestone   |
|--------------|---|
| 11/3         | Update Proposal,                                  |
| 11/7         | Data Parsing                                      |
| 11/9         | Finish research on creating Kernels               |
| <b>11/11</b> | <b>Progress Report Due</b>                        |
| 11/15        | Testing of Various Similarity Measures (Training) |
| 11/25        | Implementation of Chosen Kernel in SVM            |
| 11/28        | Evaluation of Findings                            |
| 12/02        | Finish Final Report                               |
| 12/02        | Finish Poster                                     |
| <b>12/04</b> | <b>Poster Presentation</b>                        |
| <b>12/16</b> | <b>Final Project Report Due</b>                   |