CALIFORNIA STATE SCIENCE FAIR
2016 PROJECT SUMMARY

Name(s)

Ilene J. Hoffman

Project Number

J1410

Project Title

Genre Differentiation Using Timbral Analysis

Abstract

Objectives/Goals

Music has always fascinated me - especially genres. What classifies a particular song into a certain genre or subgenre? Do these subgenres actually differ? I decided to try to determine whether there was any mathematical difference between often confused genres using MATLAB. I hypothesized that when subgenres were analyzed, there would be no clustering by genre, revealing that the classifications are subjective.

Methods/Materials

I analyzed over 200 songs from 10 different genres, divided into two groups. Group 1 (learning) was made up of rock, modern pop, 80's pop, EDM, blues, and classical music. Group 2 (testing) was made up of indie rock, alternative rock, emo, and pop-punk. I extracted the MFCCs of 250ms windows throughout entire songs, recording the Multivariate Likelihood Estimate mean vector and covariance matrix of these MFCC matrices into separate files. I then read these files into another script, which compiled all of these into struct arrays by genre. I then plotted the clustered struct arrays, using the first parameter of the .m vectors for the x coordinate and the second parameter for the y coordinate. This way, I was able to plot the clustering.

Results

My hypothesis that subgenres were classified subjectively rather than mathematically was supported by the resulting plots. All four subgenres in Group 2 were overlapping and indistinct since they were clustered very closely together. However, I was surprised to find the true roots of the umbrella genres in Group 1, finding rock and 80's pop completely intermingled, while classical music was completely separate, with blues somewhere in the middle.

Conclusions/Discussion

For further comparative analysis, my recommendations would be to implement timbral vector voting and convert my algorithms into a compiler language so that they will run faster. These algorithms could be put to use in softwares such as Pandora or iTunes.

Summary Statement

I used MATLAB to analyze & plot 200 songs from various genres using the Mel frequency cepstrum to try to find the mathematical differences between musical subgenres.

Help Received

I wrote the algorithm myself with minimal help from Mrs. Gontar, a senior designer at Via Telecom. I received a lecture on k-means clustering from Professor Gontar of Ben-Gurion University.