CS472 Final Project Proposal

Alea Minar

5/8/2025

The Problem

I will use features of music (tempo, zero crossing rate, brightness, timbre, harmony) to predict the emotional qualities of songs and classify them as focus-friendly (likely to support concentration and studying) or distracting.
This topic is of particular interest to me as I am conducting a related small-scale experiment for my psychology class (PSY348 Music and the Brain), where I am exploring how music affects cognitive performance on small tasks related to numeracy.

Datasets

I will be using the **DEAM dataset** (http://cvml.unige.ch/databases/DEAM/), which is distributed under a Creative Commons license and available for academic use.
The dataset consists of audio, annotations, and features.

Audio

Consists of 1802 music clips in WAV/MP3 file format. The clips are available in both randomly sampled 45-second excerpts and full-length versions.

Annotations

Valence refers to how pleasant or unpleasant an emotion is.
Arousal refers to how activating or deactivating an emotion is.
Valence and arousal are measured from -1 to 1.

The dataset has **static** and **dynamic** annotations of valence and arousal scores.
Static annotations are averaged valence and arousal scores for each song.
Dynamic annotations are per-second valence and arousal scores.

Features

I will extract the following features using librosa:

- Tempo → speed of the music
- Zero Crossing Rate → noisiness/harshness
- Spectral Centroid → brightness/sharpness
- MFCCs → Timbre/texture and color
- Chroma → Harmony and tonality

I think these will be sufficient features to classify the valence and arousal and make a focus-friendliness prediction.
I have already downloaded the dataset.

Models

I plan to experiment with two types of models: one simple and interpretable, and one more complex and powerful.

Decision Trees

A decision tree is a natural fit for a simple model that makes feature importance easy to identify. It will also be fast to train and easy to implement using scikit-learn.

Neural Network

While a neural network will be less interpretable and require more tuning, it works well with high-dimensional data like MFCCs and chroma vectors, which may provide a more accurate prediction if the relationships are nonlinear.

Other Analysis

In addition to comparing model performance, I am interested in analyzing model behavior and feature contributions.
For the decision tree, I will examine feature importances to see which musical attributes are most predictive of focus-friendliness.
For the neural network, I will experiment with tuning hyperparameters:

- Regularization → Prevent overfitting and assess generalization
- Number of hidden layers and neurons → Explore model capacity
- Batch size → Investigate learning rate

I am hoping to gain insight into:

- The complexity of the relationship between the audio features and perceived focus-friendliness.
- The trade-off between model complexity and interpretability.
- How regularization and network architecture affect the model's ability to generalize.