

# Math 404 Project: Harmonization

Jessica Morrise

For my project, I want to try to use an algorithm to predict the harmonic progression for a given melody.

## The Problem

Musical harmony is mathematically based. For any melody, the harmonizing notes have to follow certain patterns (if the music is going to sound any good). These patterns are flexible, yet predictable enough that two people who know music theory will often harmonize the same melody in the nearly same way. I want to explore whether it is possible to predict these patterns using a model.

To make the problem more tractable, I will train the model on songs that are all similar in style. Folk songs and hymns, for example, tend to be fairly similar harmonically and melodically; the same is true of pop songs or of classical preludes.

## Method

I plan to train a Hidden Markov Model on known melodies and their corresponding harmonic progressions. It's natural in music to think of the harmony as an underlying state. In my HMM I will model the harmony as the hidden state, and the melody as observations which are emitted by the hidden state. This will hopefully allow me to estimate harmony given a melody (if I'm understanding HMM correctly, that is).

## Data

This is the reason why this project might not be feasible. I'm not sure where to get data. Ideally I'd like to have at least a couple hundred short pieces of music, recorded as a numerical representation of pitches and durations. I've found a large database of lead sheets (melodies and chords) in MusicXML format, and I'm determining whether these will be useful. MIDI files might be an option but are hard to parse. Creating my own data would be time intensive.

## Other Ideas

More ideas to fall back on if gathering data proves impossible:

- Airplane arrivals: Use a machine learning model to predict whether a flight will arrive late, early, or on time (lots of data available from the ASA)
- Public transportation: Analyze pickup and drop-off data from both yellow taxis and Uber in NYC to model how the two services interact and predict which service a user will choose (data for both available online)
- STEM education: Use a public dataset on universities, possibly in conjugation with census data, to predict what is effective at influencing students to go into STEM fields (data available from Kaggle and other sources)