Formal Proposal

Adam Page (V00910612), Nathaniel Coster (V00862066), Vedant Manawat (V00904582)
Shawn Nettleton (V00906570), Jesse Lacroix (V00890406), Gregory Huber (V00862879)

Problem:

The problem we are tackling is translating musical audio files (mp3s) and identifying the notes played in order to construct sheet music for a specific instrument. The issues we will need to address when creating a solution for this problem will be:

- Identifying what instrument is being played and tune the parameters of our classifier accordingly. This will be required as although a guitar and a piano could be playing the same note, they will sound different as they are different instruments.
- Identifying which classification solution is the best for our use case. Running a comparison of what we think are the best methods will be a good way to discover our best option.
- Classifying a sequence of musical data as a specific musical note.

Initially we have chosen a guitar and piano as our starting point with the hopes to use more complex instruments such as a trumpet, French horn or other string instruments later. We have singled out these two options as they are the most popular instrument choices.

This problem is important as sheet music is a very important document for any musician. If a musician is recording their performance it could be translated into a physical copy to be easily worked on and remembered. Having this functionality could remove the need to spend time writing down every exact note so they can focus more on the music itself. This would also be useful for those who want the sheet music for themselves for a musical piece that may not have existing sheet music to go with it. Using a solution like this one could provide an individual with sheet music for a song they like so they can recreate it themselves on their instrument of choice.

Existing solutions, such as AnthemScore[4], solve the problems we have identified and show a real world demand for audio to sheet music translators. The problem of digital audio to music note classification is a complex problem, shown by the implementation of tools like AnthemScore who use neural networks trained by millions of data points. It will be interesting for us to tackle this problem and see if we are able to recreate similar results with no capital and limited computation power. This will reveal some telling characteristics of the usage of data mining techniques under our given conditions and whether or not it is possible to leverage them without high costs.

The datasets we are going to initially use are based on guitar and piano. The guitar dataset [1] has two parts, solo and comps, the solo is a simple single note lead and the comp is an actual mixture of cords and notes. Each track within the guitar set has information regarding the timing, note frequency and cords being played. The other dataset [3] we have found contains specific notes from a multitude of instruments including the piano. Using this will allow us to identify key notes and frequencies for these instruments.

The risks regarding our datasets are minimal because of their abundance and open source nature. This removes the worry of not having enough variety of data points to train our model on to hopefully avoid overfitting. There are numerous online sources of musical instrumental audio files so having backup datasets will not be an issue.

Goals:

The accuracy of the model can be tested by comparing the note sheet predicted by the model with the actual note sheet of the audio file. In general, this will be tested by looking at the Octave, Tonal and Harmonic accuracies of each note in the note sheet (these are described in depth below). Additionally, when the model has been trained, it can be subjected to evaluation based on how long it takes to parse the input audio file to make a prediction on the notes present in the notes sheet produced.

Interesting music is fundamentally complex, there can be multiple instruments playing over multiple registers and inconsistent time signatures. To match with this complexity a single measurement of success will likely not be sufficient. Possible measures for the notes themselves could be a measure of octal, tonal, or harmonic accuracy. For the pacing of the piece we could measure the accuracy of the time signature and tempo or more simply we could measure the relative pacing and length of the notes.

Octal accuracy would be a measure of how close to the correct octave the notes are, a note could be scored on [0,1] where the score is $\frac{1}{1+|x|}$ such that x is the number of octaves between the true note and the measured note.

Tonal accuracy would be a measure of how close to the correct note the measured note is, in a similar scale it could be scored on [0,1] where the score is again $\frac{1}{1+|x|}$ such that x is the number of half tones between the true and measured note. This could be measured in 2 ways, octave agnostic where every c tone is 1 step away from every c# tone, or with octave where this distance is just the pure halftone distance.

Harmonic accuracy would be a more subjective measure of how harmonically close to the correct note the measured note is. In this we would define the scores of the different intervals and use those to define the score. Something like a minor second would be scored lower than a major third as it is less “pleasant”. This would be an octave agnostic measurement, a major third would be the same whether it’s in the same octave or moving an octave. As a note all intervals do have a place for them to be “pleasant” in actual music, however when a note is a certain interval out of place is what this would be measuring.

Time signature would be more a measurement of the overall piece, measuring how close to the correct signature the model predicted. This could again be measured with the same $\frac{1}{1+|x|}$ where x is again the distance between the predicted and correct signatures where the signatures are interpreted as a reduced fraction (i.e 6/8 is equivalent to 3/4). The same sort of measurement would be used for tempo though the distance would just be euclidean over the integers.

Finally for scoring individual notes it would be some sort of measurement of relative length. That may be a measure to the nearest note, it may be a measurement to the overall time signature. For a naive approach we could just decide the detected note against the actual note, for example a half note compared to a dotted half note would be a difference of a beat. However the exact note would be dependent on the time signature so this may not work.

Project Plan:

Our initial approach to this project will be to prioritize a system that can identify single notes from the same instrument, using data from the GuitarSet [1]. This contains audio files of single notes all played on the same instrument, making it ideal for training and testing. This will then be output to sheet music. From there, different instruments will be experimented with as input. At this time, the experiment will focus on traditional solo instruments such as guitar, piano, and horn instruments such as trumpet, saxophone, etc. Upon success of this stage, multiple note input will be implemented and the experiment will be repeated.

While approaching this problem we will look at some of the research and methods developed for ISMIR [1, 2], as well as other applications that have successfully implemented similar ideas like AnthemScore [4]. Additionally, we have already consulted and hope to continue to discuss advice and approaches for this project with Dr. George Tzanetakis from the University of Victoria.

Once our initial approach is deemed successful, we will look to test the limits of our project and attempt to innovate on its shortcomings. The next step will be to explore its effectiveness with multiple notes played at once. This will again be initially performed using the GuitarSet as it also contains audio files using a mix of chords and single notes. The chords from the GuitarSet may prove more challenging to identify as unlike with a piano, the same chord can be played in multiple locations on the guitar neck, giving a slight change to the tone.

Experiments:

To evaluate the effectiveness of our tool, the following experiments will be conducted:

1. Single instrument, monophonic experiment:

   In this experiment, music tracks consisting of one instrument playing one note at a time (monophonic) will be input into the tool. The resulting sheet music for each instrument will be compared to the original audio track both visually, to the original sheet music, and audibly, using a sheet music player [5]. The comparison will focus on the follow factors:

   - Note pitch correctness - whether each note is the correct pitch (E.g. if a not that should be a C4 is actually a C4)
   - Note duration correctness - whether each note's duration is correct (E.g if a note that should be a quarter note is actually a quarter note)
   - Sentiment - how similar we perceive the input and output audio when heard sequentially

2. Single instrument, polyphonic experiment

   The experiment will mirror experiment #1 but will play a single instrument playing more than one note at a time (E.g classical piano and guitar music). The same criteria will be used for evaluation.

Task Breakdown:

Feb 14	Formal Proposal - Due
Mar 14	Minimum Viable Product (single note input), Progress Report - Due
Mar 21	Experiment #1
Mar 28	Minimal Viable Product (multi-note input), Experiment #2
Apr. 4-7	Presentations

All team members will partake in the machine learning note recognition process. Further tasks will be split as follows:

- Jesse: Preprocess and format data. Neural network analysis.
- Nathaniel: Message the MAPS dataset people for dataset. Random Forest analysis.
- Greg: Tests and output verification. From the JAMS file
- Shawn: Analysis of neural networks or other possible methods. SVM analysis
- Vedant: Analysis of neural networks or other possible methods.
- Adam: Sheet music output. Progress report writing. Decision Tree analysis.

This is subject to change depending on effort required for each task. Upon digging into the work further, we will have a more concrete understanding of who needs to be working on which task.

References:

[1] - J. P. Bello, R. Bittner, J. Pauwels, Q. Xi, and X. Ye, "Guitarset: A Dataset for Guitar Transcription", in 19th International Society for Music Information Retrieval Conference, Paris, France, Sept. 2018. [Online] Available: https://guitarset.weebly.com/

[2] - D. Eck, E. Elsen, J. Engel, C. Hawthorne, S. Oore, C. Raffel, A. Roberts, I. Simon, J. Song, "Onsets And Frames: Dual-Objective Piano Transcription", in 19th International Society for Music Information Retrieval Conference, Paris, France. [Online] Available: https://arxiv.org/pdf/1710.11153.pdf

[3] - L. Fritts, "Musical Instrument Samples". [Online] Available: https://theremin.music.uiowa.edu/MISpiano.html

[4] - "AnthemScore 4", Lunaverus, 2022. [Online] Available: https://www.lunaverus.com/

[5]- D. Zemsky, "Reading sheet music for you," *Sheet Music Scanner*. [Online]. Available: https://sheetmusicscanner.com/. [Accessed: 14-Feb-2022].