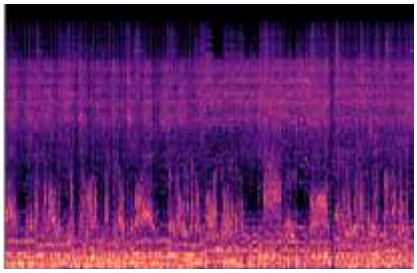
Music Genre Classification Using CNNs

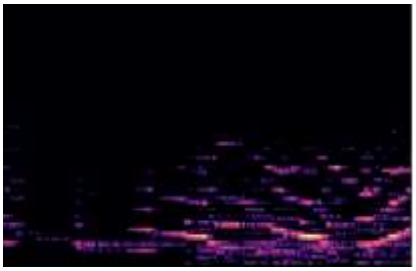
ST311 Group Project by Ben Steiner and Oliver Mapps

Dataset: GTZAN Dataset - Music Genre Classification

Our Dataset contains audio recordings of 1000 songs, each 30 seconds long and grouped evenly into 10 genres. For each song, there is associated meta-data and a mel-spectrogram. Each instance of meta-data contains auditory characteristics of the recording, such as the tempo. The mel-spectrogram is a way of visualising what frequencies are more intense on the melodic scale at a given time. Two examples are given below:



Metal Spectrogram



Classical Spectrogram

We observe that there are numerous visual differences between the spectrograms, for example, the metal spectrogram is significantly busier than the classical spectrogram. We plan to learn these differences, along with the meta-data, to classify each song into one of the ten genres.

To accomplish our goal, we plan to implement a variety of CNNs that learn the visual properties of the spectrogram, to obtain the greatest test accuracy. We will then compare these to MLP models that take the meta-data as input. Finally, we will define an ensemble model that joins the CNN and the MLP, which will maximise the predictive power we can yield from our data.

It is important to note that our dataset is relatively small, and as such we will take measures to increase the amount of data we can feed into our models:

- The dataset provides meta-data on each 3 second segment of each 30 second recording. This dramatically increases the quantity of data we can feed into our MLP.
- We can split our spectrograms into vertical slices. This preserves the frequency axis, which is more important when determining the style of music.
- There is also the possibility of adding noise to our spectrogram.