Capstone 2 Project Proposal

This project is going to explore speech emotion recognition. The hope of this project is to develop a model that identifies the categorical emotional state of the voice data to provide better service for the speaker's needs.

Some potential uses could be:

* Call centers
* Smart devices
* Speech recognition programs
* AI powered voice apps
* Voice to text

In this particular application our goal will be to train a model to better recognize a customer’s emotional state when using a voice interacting smart device.

We will be using various python libraries, functions and methods to complete this project, some of which are: librosa, soundfile, numpy, sklearn, and pyaudio, to name a few. This should allow us to build a model using a Multi-layer Perceptron Classifier. The data set we will be using is from the Ryerson Audio-Visual Database of Emotional Speech and Song dataset. The URL is as follows: <https://drive.google.com/file/d/1wWsrN2Ep7x6lWqOXfr4rpKGYrJhWc8z7/view> This data will need to be broken down in to 3 features, the Mel Frequency Cepstral Coefficient, which represents the short-term power spectrum of a sound, the chroma, which pertains to the 12 different pitch classes, the mel, which is the Mel Spectrogram Frequency.

Once the data is successfully extracted we will categorize these into the following categories:

01':'neutral', '02':'calm', '03':'happy', '04':'sad', '05':'angry', '06':'fearful', '07':'disgust', '08':'surprised'. Once processed these categories will give us the ability to create a distribution map of the actor’s emotional states for each sample. We can then map and quantify the results with training data, followed by testing data. This will help us test the accuracy of the model.

This project could be tough since emotions are subjective and results could depend on the quality of the voice data. If successful, this model could have far reaching applications for any technology using or receiving voice data.