1 Choice of Dataset

To provide a confidence score for the gender of a speaker based on their voice, I will use the Gender Recognition by Voice dataset, uploaded to Kaggle by Kory Becker.

The dataset provides information pertaining to a number of different acoustic variables of the human voice, providing a plethora of potential features that could be used to distinguish the gender of a voice. The variables in the dataset include

`meanfreq`: mean frequency (in kHz)

`sd`: standard deviation of frequency

`median`: median frequency (in kHz)

`Q25`: first quantile (in kHz)

`Q75`: third quantile (in kHz)

`IQR`: interquantile range (in kHz)

`skew`: skewness

`kurt`: kurtosis

`sp.ent`: spectral entropy

`sfm`: spectral flatness

`mode`: mode frequency

`centroid`: frequency centroid

`peakf`: peak frequency (frequency with highest energy)

`meanfun`: average of fundamental frequency measured across acoustic signal

`minfun`: minimum fundamental frequency measured across acoustic signal

`maxfun`: maximum fundamental frequency measured across acoustic signal

`meandom`: average of dominant frequency measured across acoustic signal

`mindom`: minimum of dominant frequency measured across acoustic signal

`maxdom`: maximum of dominant frequency measured across acoustic signal

`dfrange`: range of dominant frequency measured across acoustic signal

`modindx`: modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range

as well as `label`, which identifies whether the voice belonged to a male or female speaker.

2 Methodology

a. Data preprocessing: The dataset is already prepared for gender classification, but the application of the model to new data would require audio information processing, which I plan to accomplish using a audio analysis package for Python (Magenta, Librosa, and pyAudioAnalysis) or R (seewave and tuneR).

b. Machine learning model: I'm planning on using standard supervised learning algorithms (Logistic Regression, Naive Bayes, K-Nearest Neighbors, Decision Tree, Support Vector Machines) and selecting whichever provides the strongest model. Another approach that I'm considering would be to use deep neural networks on spectrograms of audio input (inspired by recent work from Jeremy Pinto of Mila), but such an approach would require a modified dataset.

c. Evaluation metric: To evaluate the success of my model, I'm planning to use a confusion matrix and accuracy/precision-recall/logistic loss. A conservative goal for the accepted metrics would be having the gender with the higher confidence score according to the model matching the input voice's owner >75% of the time.

3 Application

I believe that the most useful application for this project would be to track the progress of transgender individuals participating in gender-affirming voice training, the practice of modifying the gendered aspects of one's voice to be more closely aligned with the gender they identify with.

With this in mind, I'm planning on integrating the model into a web application that allows users to record and upload audio samples, and then receive a confidence score regarding the 'gender' of their voice (e.g. 23% male, 77% female). Users can use this application at regular intervals during voice training to get a feel for the progress they've made. There are other features that I'm considering, but time will tell whether my schedule cam afford me the time needed to implement these additional features.