Speaker Recognition using MFCC, FFT and Eigendecomposition

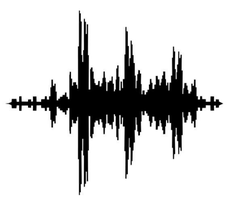
Ian Walsh & Sparsh Gupta

Olin College of Engineering

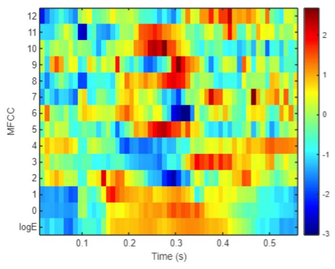
Fall 2022

Quantitative Engineering Analysis 1

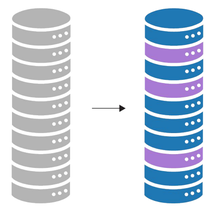
Process



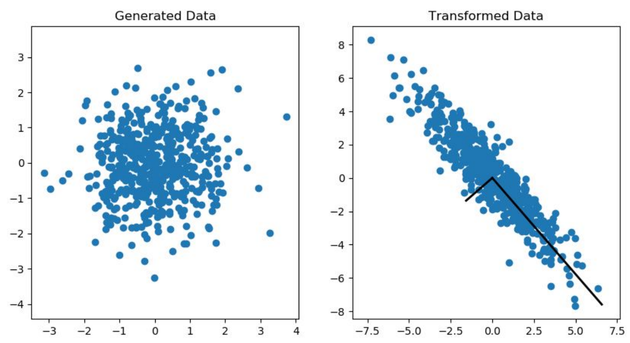
Audio Dataset
200 voice samples from 4 different subjects



MFCC
Passing data samples through MFCC to extract coefficients

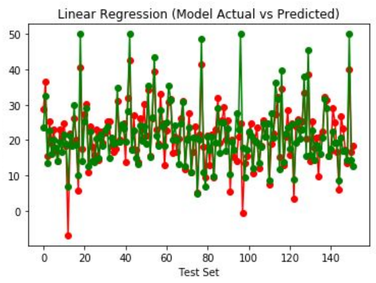


Data Split
Splitting the dataset into equal sized train/test datasets

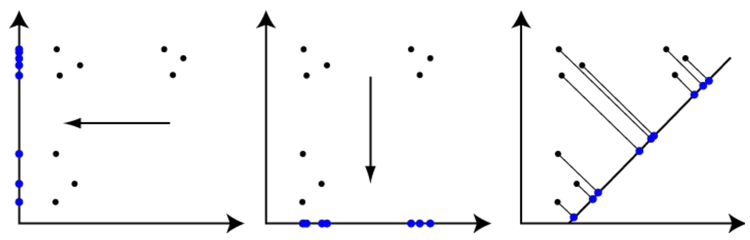


Eigendecomposition
1. Obtaining a covariance matrix (people-people)
2. Finding eigenvectors for the covariance matrix
3. Training dataset multiplied by eigenvectors to obtain important features (frequencies)

GitHub:



Classification & Accuracy
1. Finding the euclidean distance between the "train data weights" and "test data weights"
2. Classify data points into respective test data labels (4 individuals) according to a threshold value
3. Compare predicted test labels to actual test labels to compute accuracy



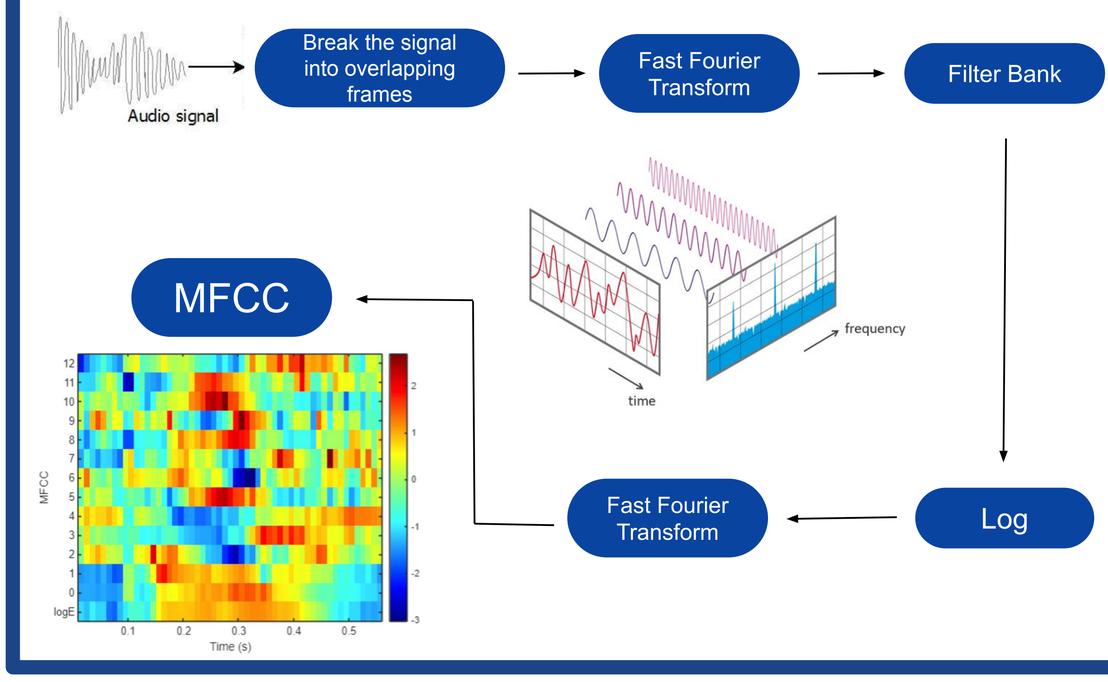
Projections & Weights
1. Projecting the training data onto the important training features (frequencies) and obtain "train data weights"
2. Projecting the test data onto the important training features (frequencies) and obtain "test data weights"

Our Data

We analyzed four different people speaking the same number for 50 samples each (fictional names below)

Mel-Frequency Cepstral Coefficients (MFCC)

A compact representation of the spectrum of an audio signal



Our Accuracy: 91%

Results

An accuracy of 91% indicates that the algorithm is working effectively on the Spoken Digit Audio dataset. As we include only four individuals in this model, the classification problem seems to be relatively easy when compared to more complex classification models.

The model was implemented in MATLAB using three significant algorithms (MFCC, FFT, & Eigendecomposition) to recognize individuals based on their voice samples.

Use Cases

Sound processing of any kind is immensely useful for automation and analysis. Speaker recognition in particular can improve on the functionality of speech analysis software by first identifying a voice, then transferring it to a recognition model custom-built for that person. It can also be used to identify multiple people talking in a conversation - a huge boon for closed caption writers the world over.

References

[1] Spoken Digit Audio Dataset, retrieved from https://github.com/Jakobovski/free-spoken-digit-dataset
[2] MATLAB MFCC, retrieved from https://www.mathworks.com/help/audio/ref/mfcc.html
[3] MFCC: https://www.youtube.com/watch?v=4_SH2nfbQZ8