**Introduction**

Speech recognition is widely used in current smart devices, including phones, smart speakers, as well as home controllers. Since the voice pattern of each user is different, identifying each speaker can improve the accuracy of voice recognition. Moreover, in recent Google and Siri smart speakers, smart assistance can access personal calendars, notes, contacts and reminders based on the detected speakers. Speech recognition technology can utilize the information from speaker’s voice spectrum, talking speed, punctuations or even frequently used words to make a comprehensive identification. Yet, in this project we explored the mechanism of speaker recognition with a simple 3-class classifier only based on speaker’s voice spectrum with logistic regression.

**Data Collection:**

Data Shuffling

Feature Extraction

Data Conversion

Data Acquisition

Training

Validation

Testing

Model

Evaluation

3 subjects were recruited in this test. From each subject, around 10-minute stereo voice data was recorded with numerous voice recorders with different noise backgrounds and the best recorder (R3312, Aigo) was chosen using 16kHz sampling rate and 16-bit data width in wav format which contained minimal noise levels.

**Data Analysis/Conversion and Feature Extraction:**

The .wav files couldn’t be used directly in their natural form, and these had to be converted to the frequency domain from respective time domains with the help of Fourier transforms. Post FFT, the audio samples were broken down to around 1000 data samples for each subject. The data points corresponding to the 150Hz range had the best responses with the lowest noise disturbances and as a result these samples were used for training the model.

**A screenshot of a graph

Description automatically generated**

The fundamental, first and second harmonics of the voice data from Subject 1 and 2 can be clearly observed (green arrows), while only the fundamental and first harmonic can be observed for Subject 3 (black arrows).

**Data Shuffling and Data Splitting:**

The data might be uniformly arranged and can belong to only one subject before shuffling. Due to this the model predicts all the data points to belong to a single class. To avoid this, the data points are randomly shuffled in order to make sure the model is trained well and the predictions are accurate. This means the order of the data points is randomized. The purpose is to ensure that any patterns related to the order of data do not influence the learning process, and that the model remains generalizable and doesn't learn any sequence bias.

After shuffling, 70% of the data is under the training set. The model mainly learns from this data subset, and it tries to learn patterns to get the desired results. 15% of the data is considered for validation for tuning and avoiding overfitting. The testing set also comprises of 15% for which the model is tested for new sets of data samples never encountered by the model yet.

**V0**

**V2**

**V1**

**Argmax**

**Output**

**Model:**  ​

g(w)= ₂

g(w) = ₂

This function computes the regularized cost (loss) using the cross-entropy loss function. It is used to evaluate how well the logistic regression model is performing. The regularization term, controlled by **lambda\_reg**, helps prevent overfitting by penalizing large weights.

Since logistic regression is inherently binary, this function extends it to handle multi-class problems using the one-vs-rest (OvR) approach. It trains a separate logistic regression model for each class against all the others and stores the weights for each classifier.