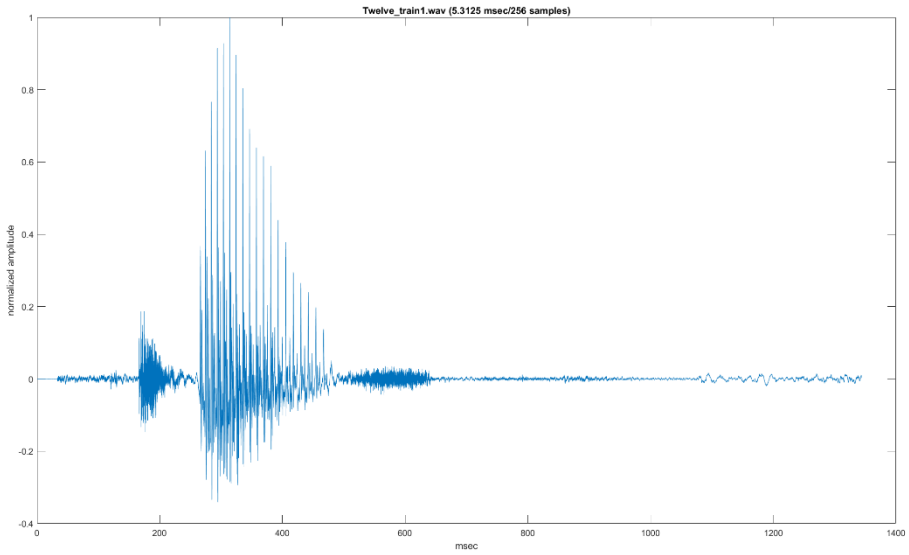
Final Project Report

Test 1

In this test, we tried to identify the speakers manually. The training data are the 11 audios in Final Project\GivenSpeech_Data\Training_Data, and the testing data are the 8 audios in Final Project\GivenSpeech_Data\Test_Data. Notice that some of the training data do not have a counterpart in the test set, for example, the three files from s9.wav to s11.wav. In order to utilize these data as a disturbance term for classification, we randomly upset the order of the data in both the training set and the test set at the same time. While manually categorizing we recorded the correspondence between the test data and the training data indices in the new order, and then mapped them back to the original indices for the precision evaluation. We performed 100% precision in manual classification.

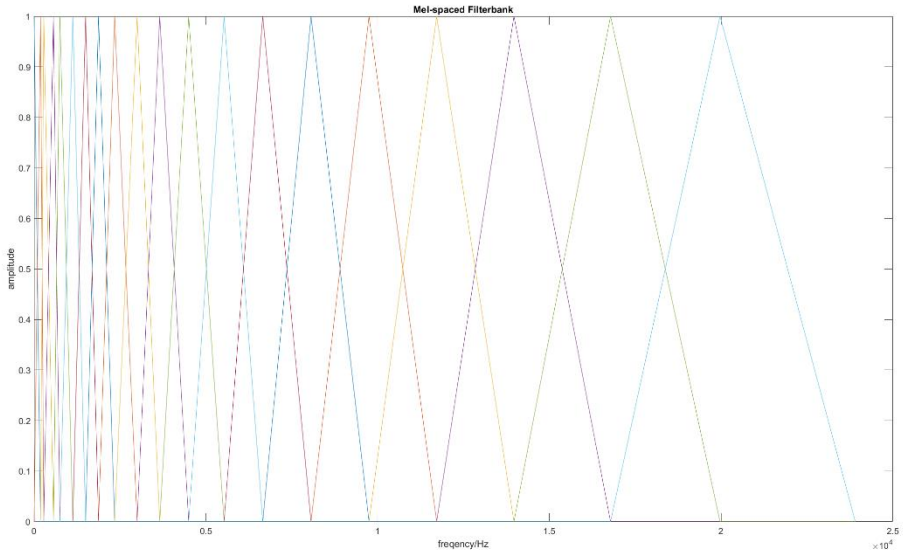
Test 2

The sampling rates of the training and testing data are 48000Hz, which means each block of 256 samples contains 5.3125 milliseconds. The signal of Twelve_train1.wav is shown below where the amplitude of the signal is normalized.

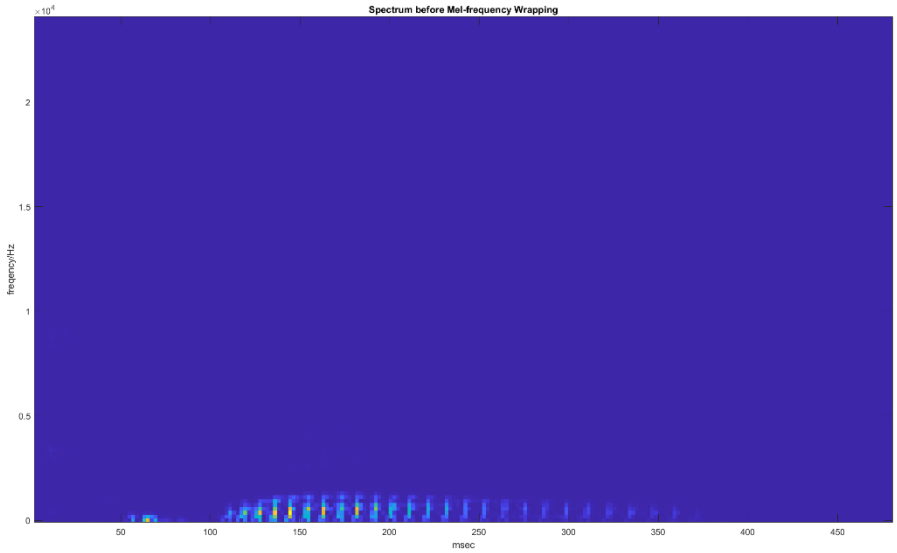


Test 3

We generated a Mel-filterbank with 20 triangular shape response filters, shown in figure. Since the sampling rate of the signal is 4.8kHz, the highest frequency our system can process is 2.4kHz, which is beyond the perception of the human ears.



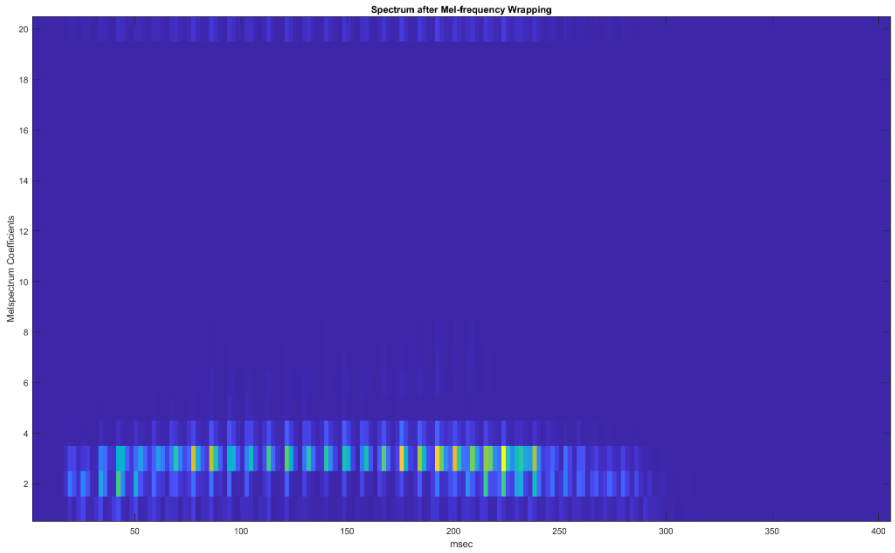
The spectrogram before Mel-filtering is shown below:



We screened the silent frames in preprocessing so the number of irrelevant frames are reduced.

Test 4

The spectrogram after Mel-filtering is shown below:



Test 5

In this test we plot the 6th and 7th components of the MFCC vectors of speaker 2 and speaker 10 and find their MFCC points are in different clusters.