Earl Escobal
EEC 201 WQ 2024

Final Project Report

Approaches, Tests, and Results:

Test 1:

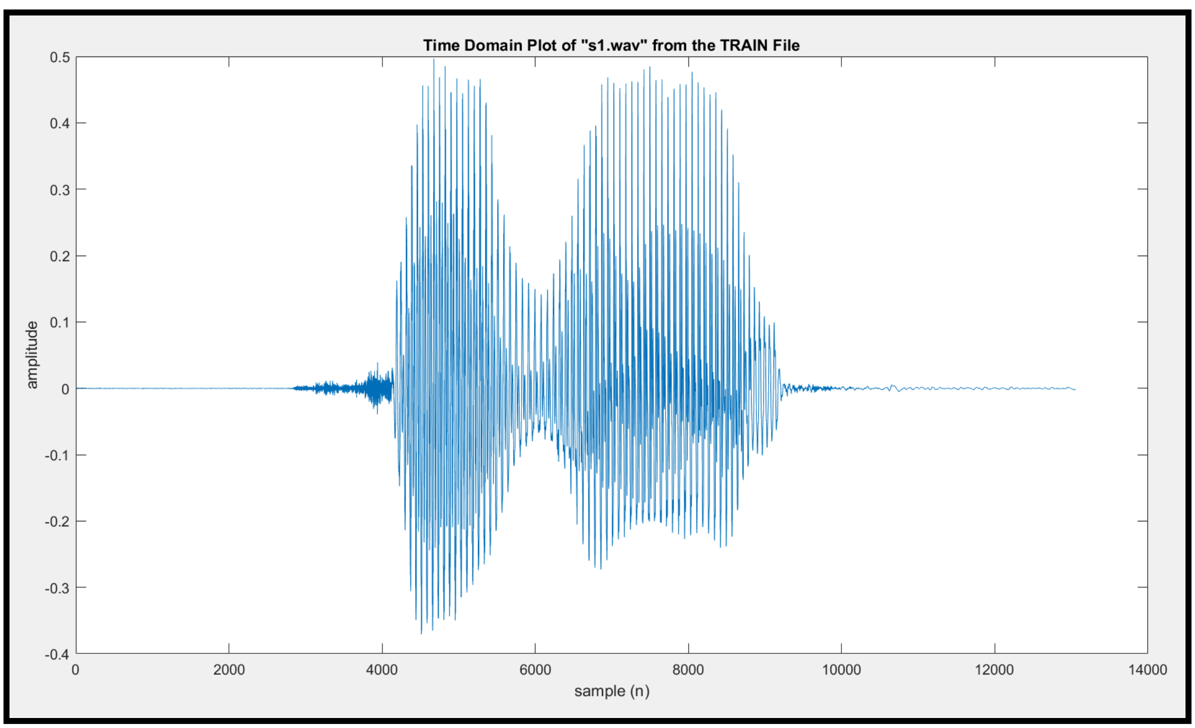
After going through each of the speakers in the TRAIN file, I would say that I can distinguish between the 11 speakers based on the differences in their voices.

After playing the speakers in the TEST file in a random order, I only correctly identified 2/8 of the speakers, meaning that my recognition rate was 25%.

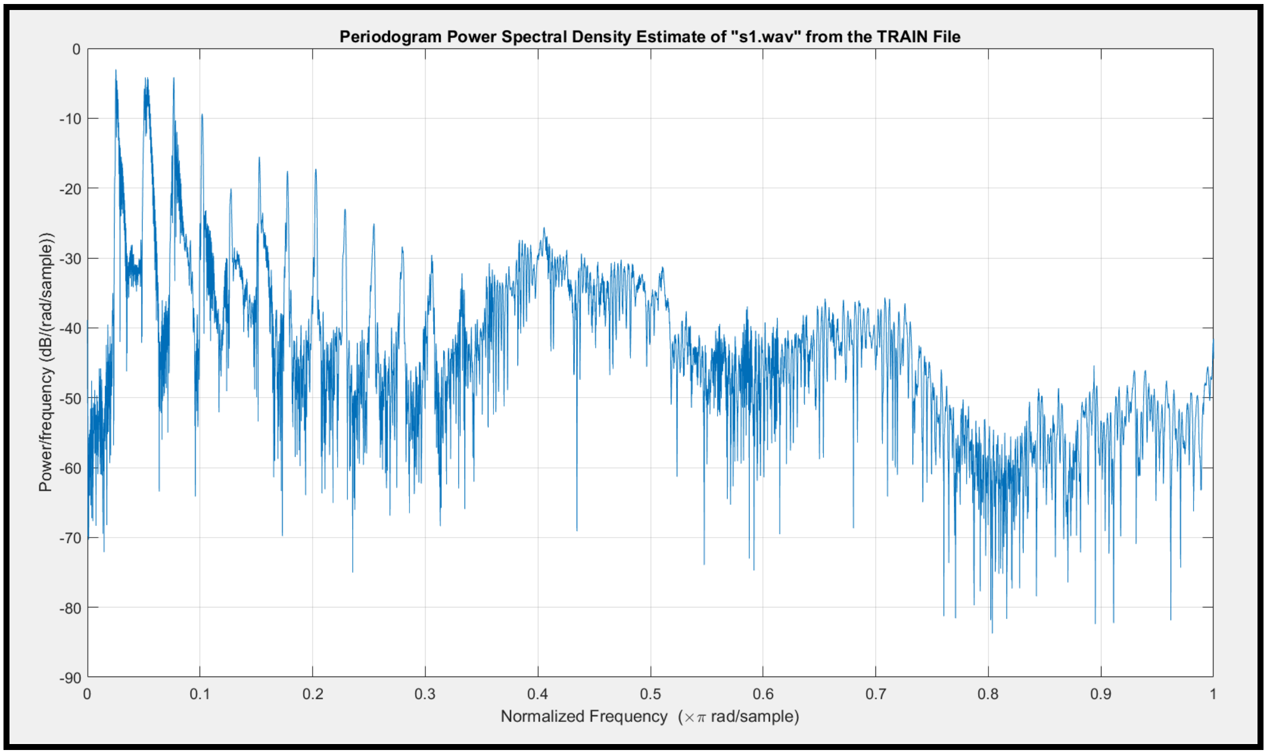
Test 2:

Using the “audioread” function in MATLAB, I determined that the sampling rate of audio files in the TRAIN and TEST folders is 12.5 kHz. This means that there are approximately 20.5 ms of speech in every 256 samples.

A time domain plot and periodogram of “s1.wav” from the TRAIN file are shown below.



Time Domain Plot of “s1.wav” from the TRAIN File



Periodogram of “s1.wav” from the TRAIN File

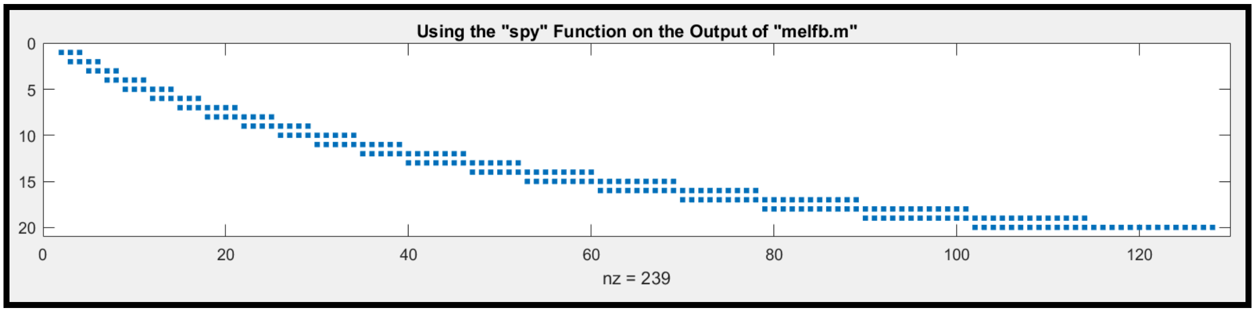
Upon inspection of the time domain plot, it can be seen that most of the signal’s energy is from n = 4,000 to n = 9,000, meaning that most of the signal’s energy is from 320 ms to 720 ms.

Likewise, upon inspection of the periodogram, it can be seen that most of the signal’s energy is at frequencies below 0.3π, meaning that most of the signal’s energy is at frequency components less than 1.875 kHz, which makes sense since speech primarily consists of lower frequencies.

Test 3:

I was not able to plot the mel-spaced filterbank responses because I did not know how to extract them from the matrix output of “melfb.m”, but I did use the “spy” function to generate some graphical representation of the mel-spaced filterbank responses, and the result of that is shown below.

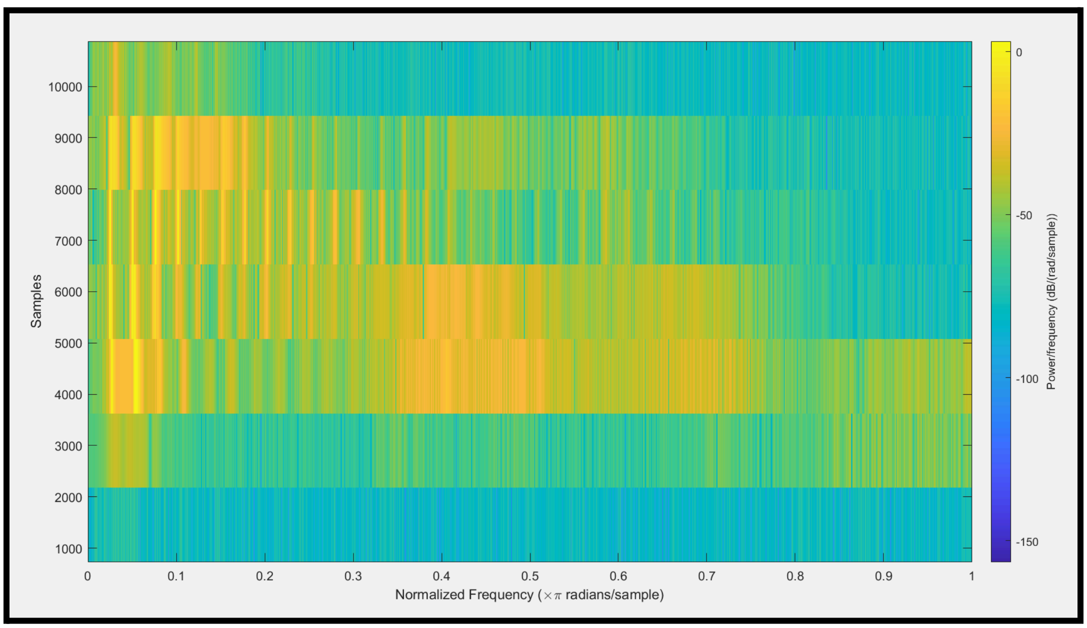
For higher values of k, the amount of samples appears to increase, which makes sense since the “bins” that are centered at higher frequencies should have larger widths, meaning this agrees with the theoretical response.



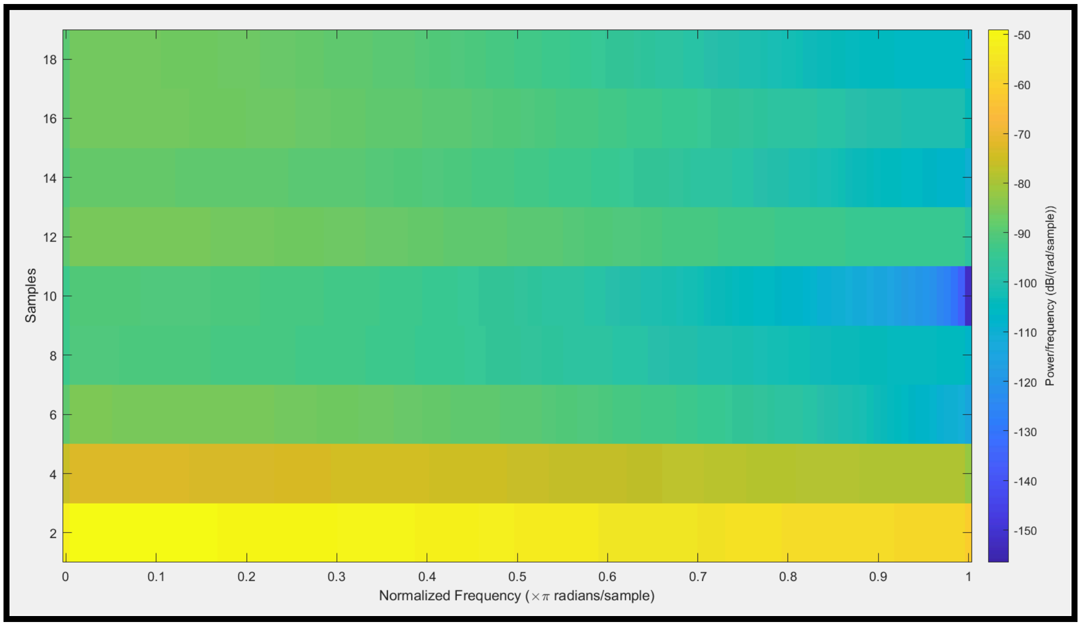
Visual Result of Using the “spy” Function on the Output of “melfb.m”

Spectrograms for“s1.wav” from the TRAIN file before and after mel-frequency wrapping are shown below.

Upon visual inspection of these two spectrograms, it appears that the mel-frequency wrapping (the use of “melfb.m”) causes lower frequencies to become the dominant frequencies of the signal. This makes sense because speech primarily consists of lower frequencies, so if one wants to extract the features of speech, these lower frequencies should be emphasized.



Spectrogram of “s1.wav” from the TRAIN File Before Mel-frequency Wrapping



Spectrogram of “s1.wav” from the TRAIN File After Mel-frequency Wrapping

Test 4:

Using a DCT, I completed the “cepstrum” step. The 20 mel cepstrum coefficients for “s1.wav” from the TRAIN file are shown below.

```
>> c_train1

c_train1 =

    0.0059
    0.0079
    0.0076
    0.0071
    0.0065
    0.0058
    0.0051
    0.0044
    0.0037
    0.0031
    0.0026
    0.0021
    0.0017
    0.0013
    0.0010
    0.0007
    0.0005
    0.0003
    0.0002
    0.0001
```

Mel Cepstrum Coefficients for “s1.wav” from the TRAIN File

Test 5:
(Incomplete)

Test 6:
(Incomplete)

Test 7:
(Incomplete)

Test 8:
(Incomplete)

Test 9:
(Incomplete)

Unique Efforts:

Since I did this project individually, all efforts in this project were my own.