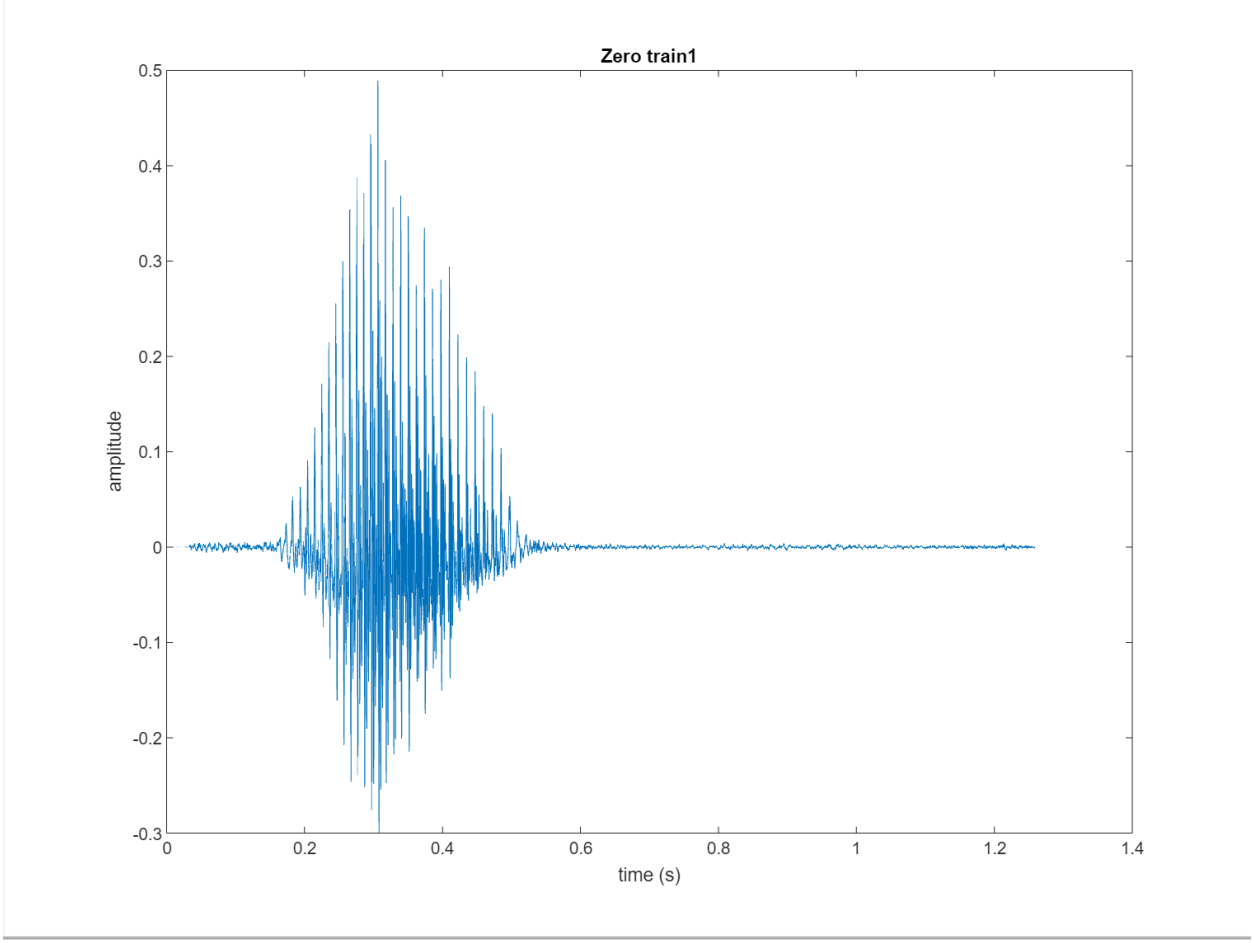
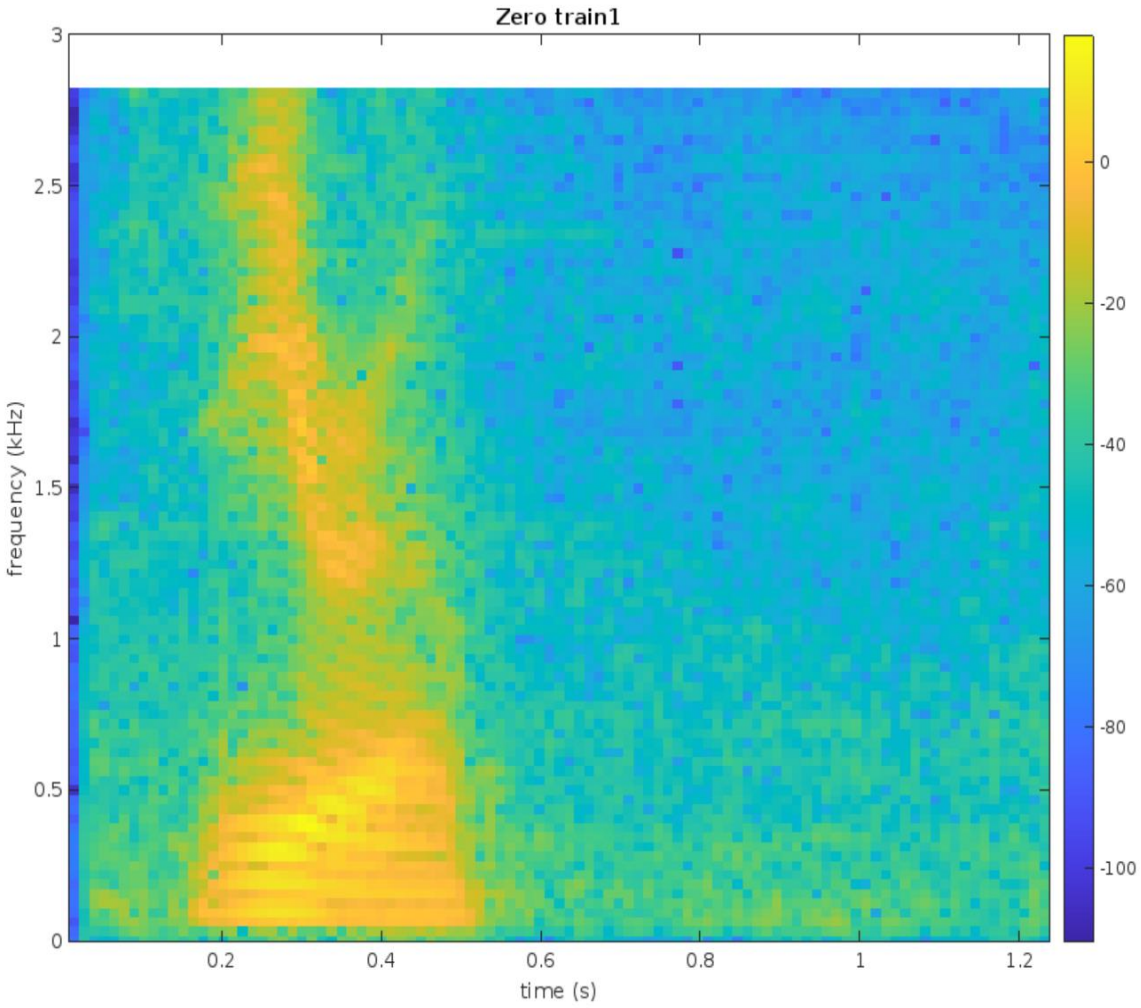
Final Project Test Results

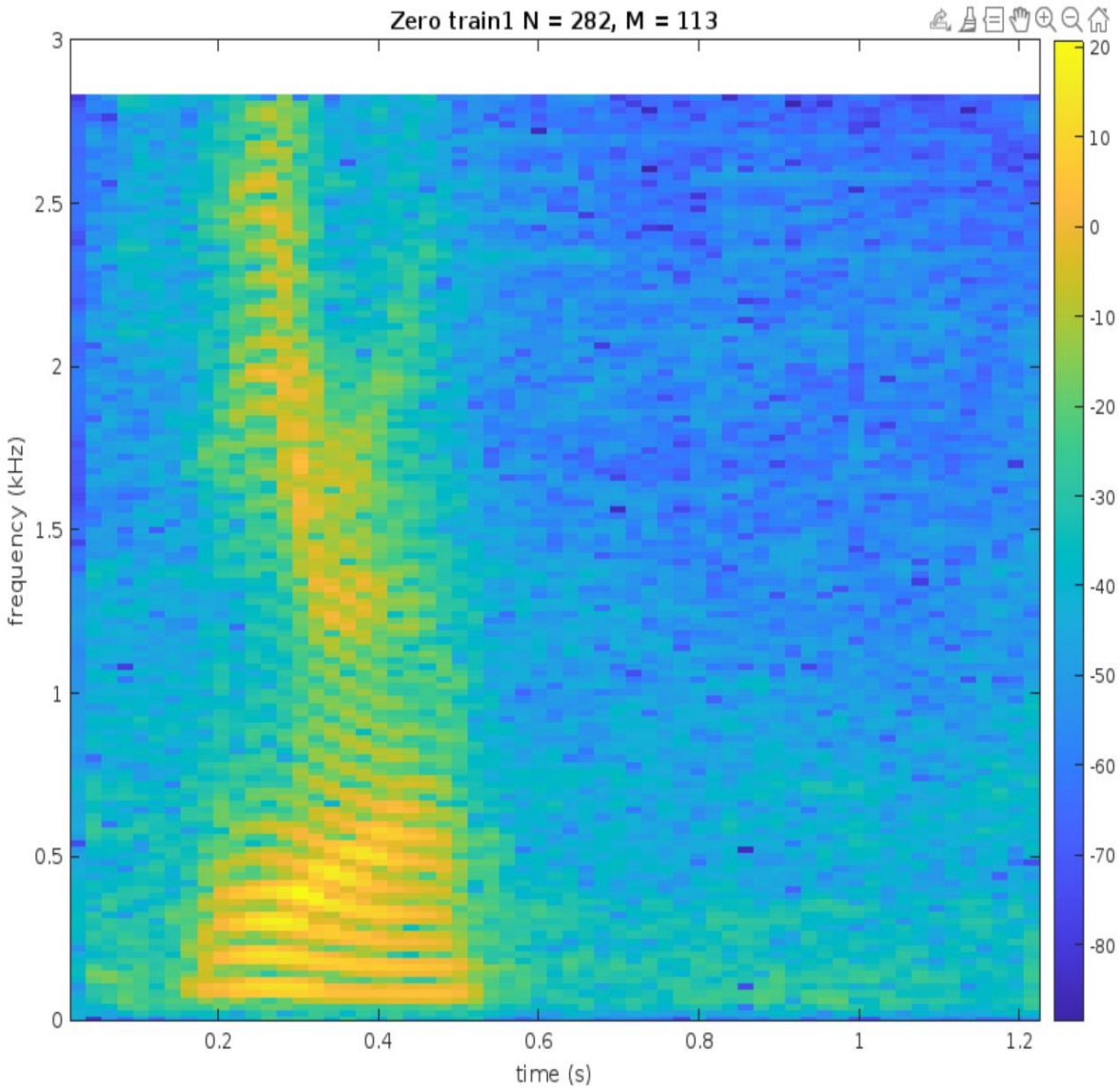
1) For Test 2 I was able to distinguish most speakers from each other, however, I had difficulty distinguishing between speakers 3, 6 and 8. When playing each file in the TEST folder, I was able to identify all speakers except speakers 3 and 8. I confused speaker 3 for speaker 6 and speaker 8 for speaker 6. So, my human performance recognition rate was 75 percent.

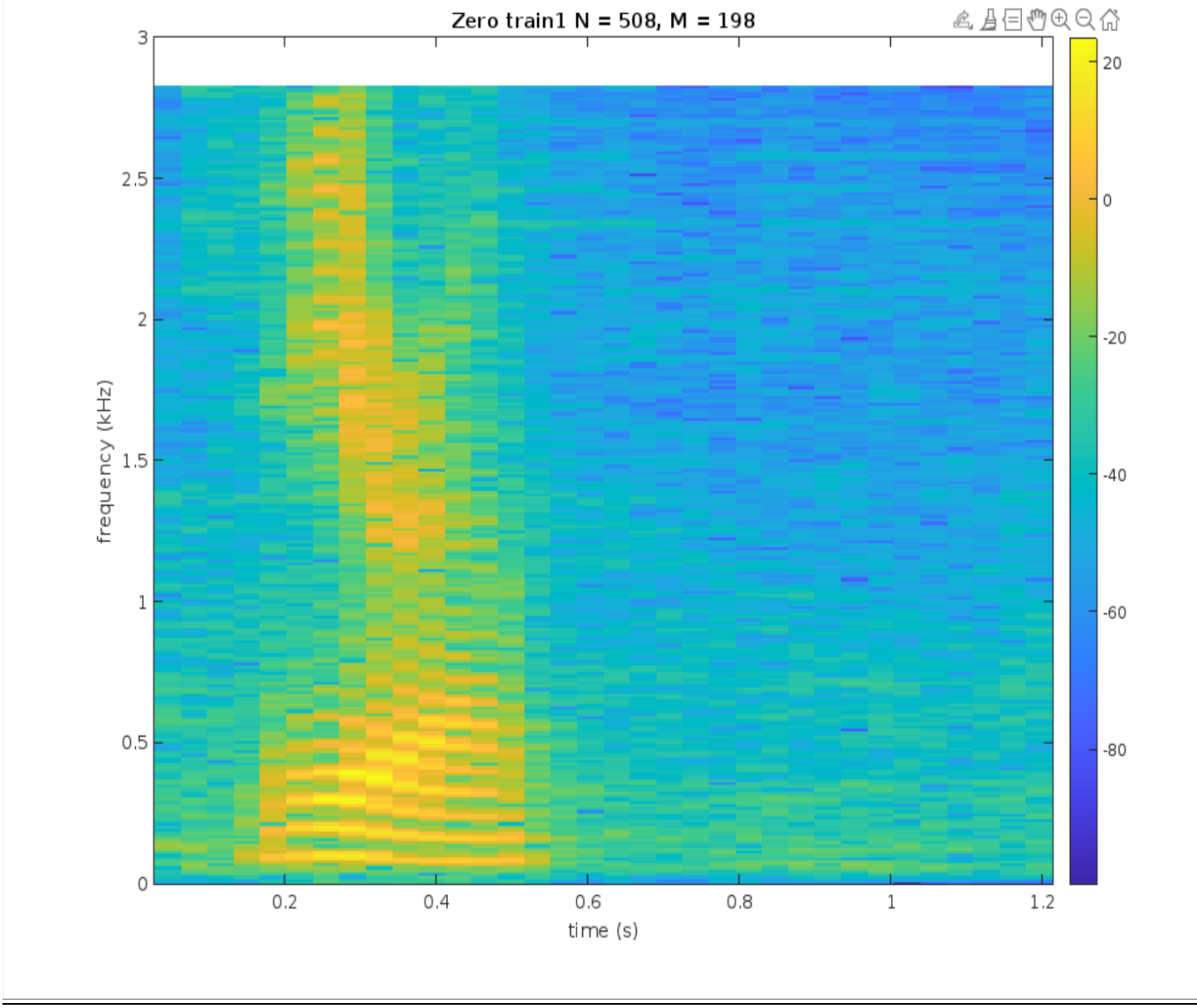
2) For test 2 we resampled the Zero_train1 wav file to get a sampling frequency of 5647 Hz. Upon doing this, a frame of 256 samples corresponds to about 45 ms.



N = 181, M = 171


Zero train1
frequency (kHz)
time (s)
0
0.5
1
1.5
2
2.5
3
0.2
0.4
0.6
0.8
1
1.2
0
-20
-40
-60
-80
-100

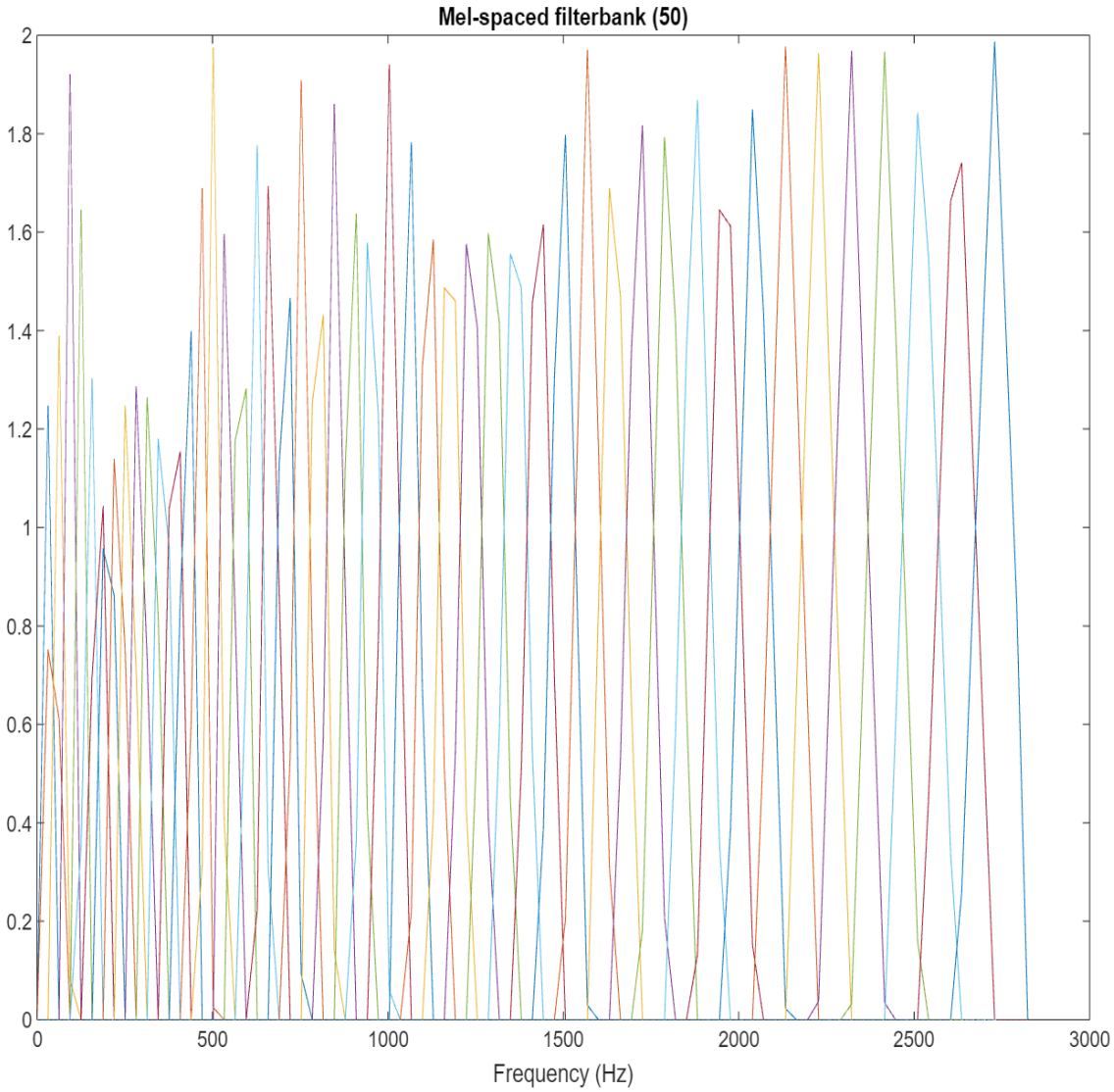

Zero train1 N = 282, M = 113
frequency (kHz)
time (s)
0
0.5
1
1.5
2
2.5
3
0.2
0.4
0.6
0.8
1
1.2
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80

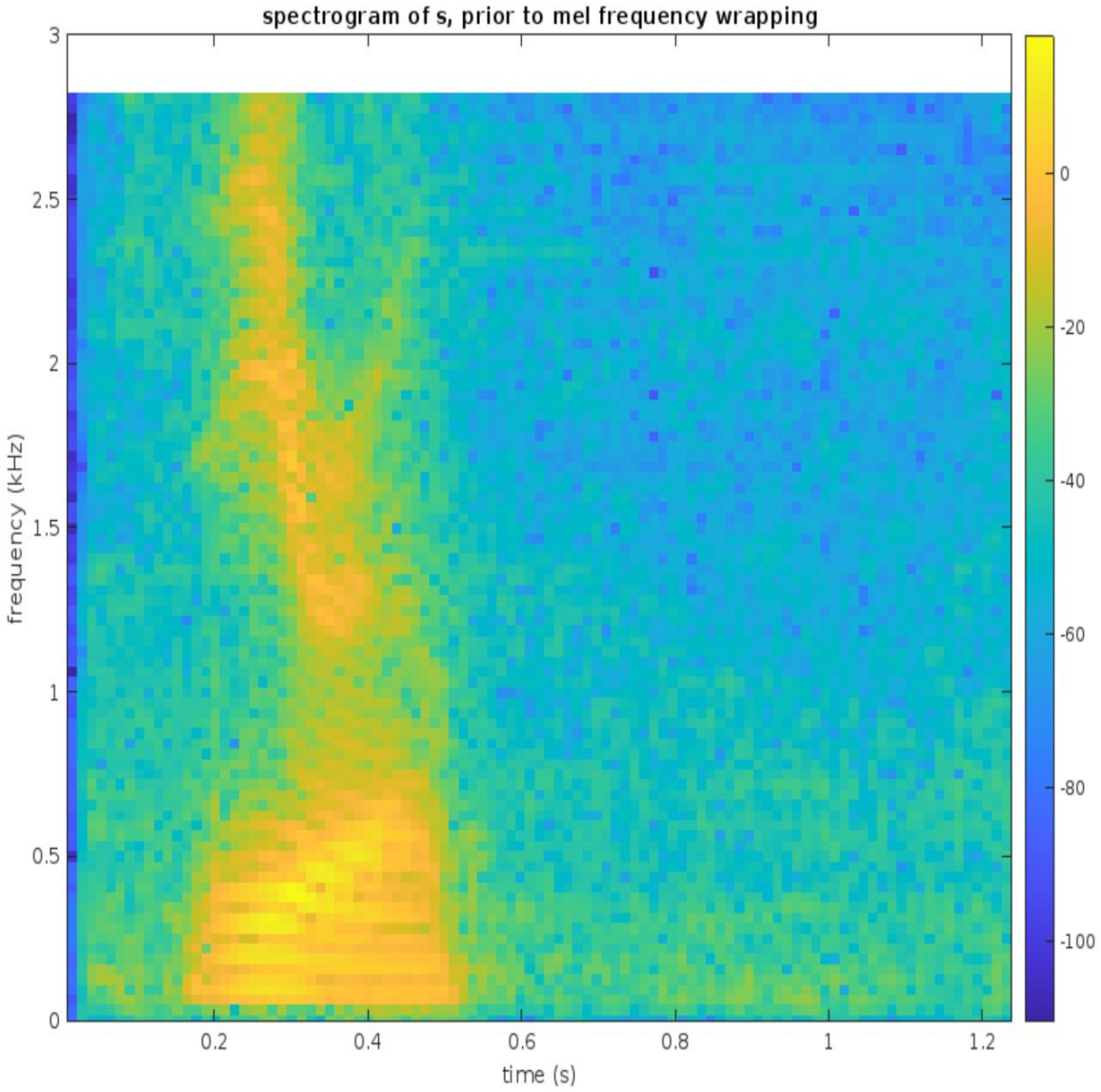


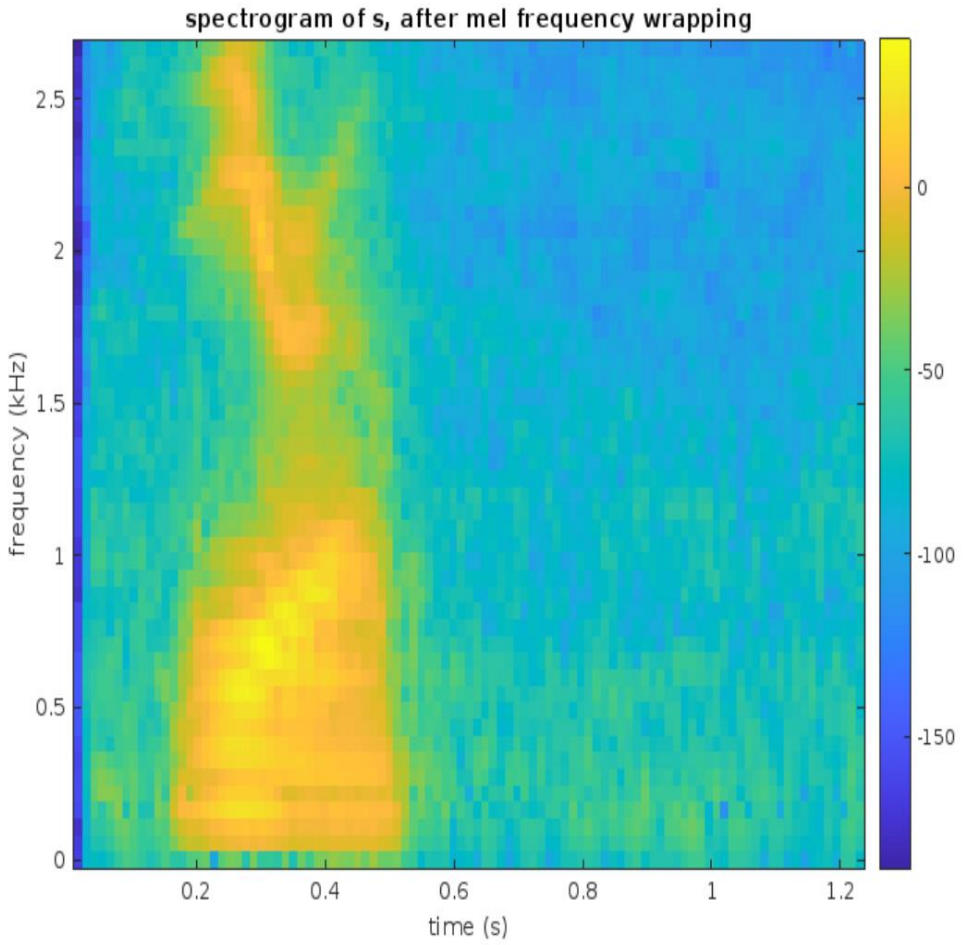
Energy Density:

In each spectrogram the greatest energy density seems to correspond to about .18 to .55 seconds, and about 0 to .8 KHz.

3)


Mel-spaced filterbank (50)
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
500
1000
1500
2000
2500
3000
Frequency (Hz)

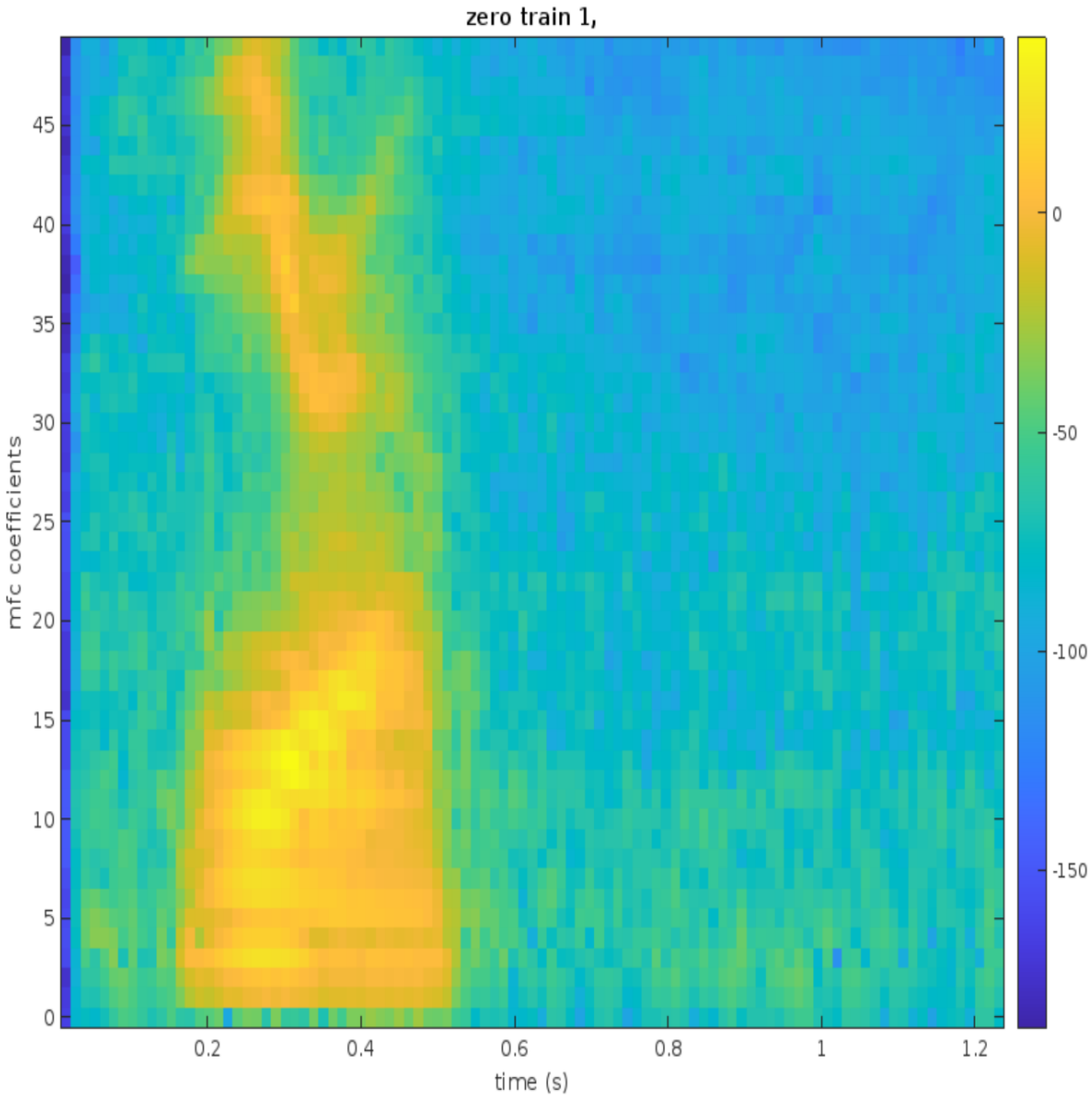

spectrogram of s, prior to mel frequency wrapping
frequency (kHz)
time (s)
0
0.5
1
1.5
2
2.5
3
0.2
0.4
0.6
0.8
1
1.2
0
-20
-40
-60
-80
-100



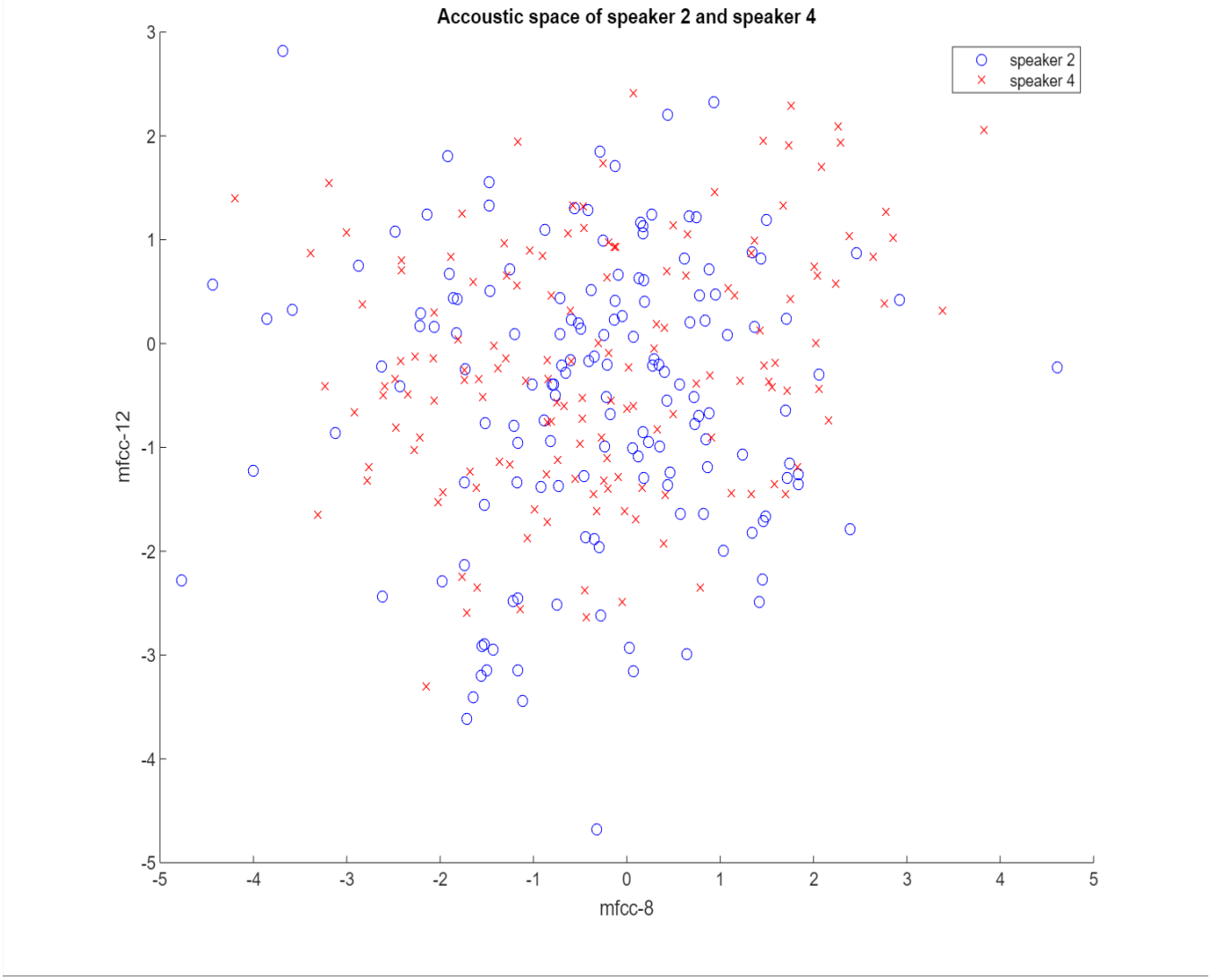
Effect:

The Mel-frequency wrapped spectrogram is smudged, giving the impression of a more even distribution of energy across frequency and time. Also worth noting, is the difference in power, observable by the color bar to the right of the plot. This change in appearance is due to the application of mel-spaced filter banks to the original audio, which is designed to scale linearly at 1000 Hz, and logarithmically above 1000 Hz; in accordance with human auditory perception of frequency.

4)

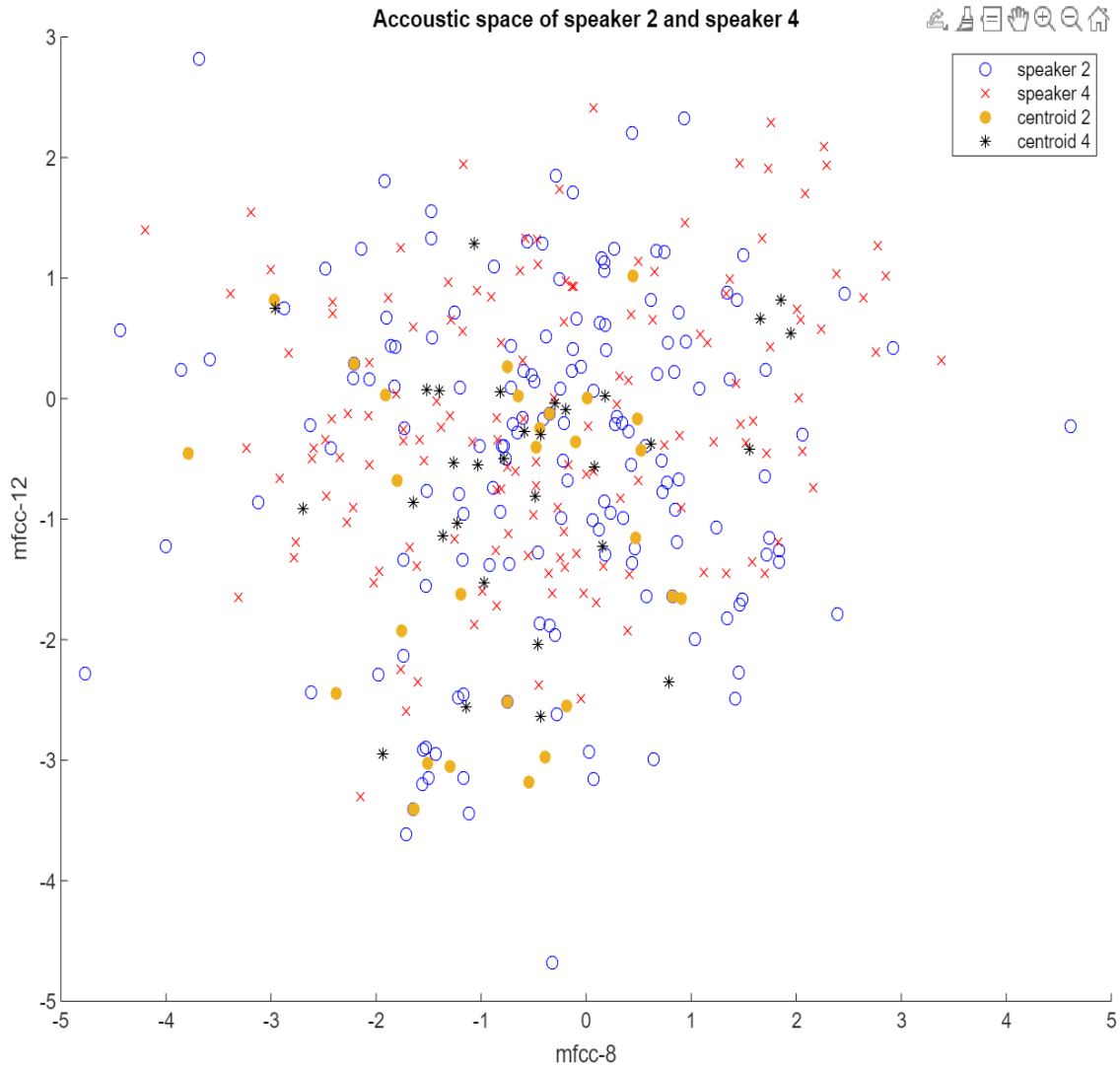

zero train 1,
mfc coefficients
time (s)
0
5
10
15
20
25
30
35
40
45
0.2
0.4
0.6
0.8
1
1.2
0
-50
-100
-150

5)



Though it might not appear to be the case at first glance, the vectors are clustered.

6)


Accoustic space of speaker 2 and speaker 4
speaker 2
speaker 4
centroid 2
centroid 4
mfcc-12
mfcc-8
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

7) The recognition rate achieved by our system for the training and testing samples provided was 100 percent. This is a whole 25 percent better than the human recognition rate.

Results of running provided testing and training samples through program

```
>> senior_design
Test 1 corresponds to train 1
Test 2 corresponds to train 2
Test 3 corresponds to train 3
Test 4 corresponds to train 4
Test 5 corresponds to train 5
Test 6 corresponds to train 6
Test 7 corresponds to train 7
Test 8 corresponds to train 8
>>
```

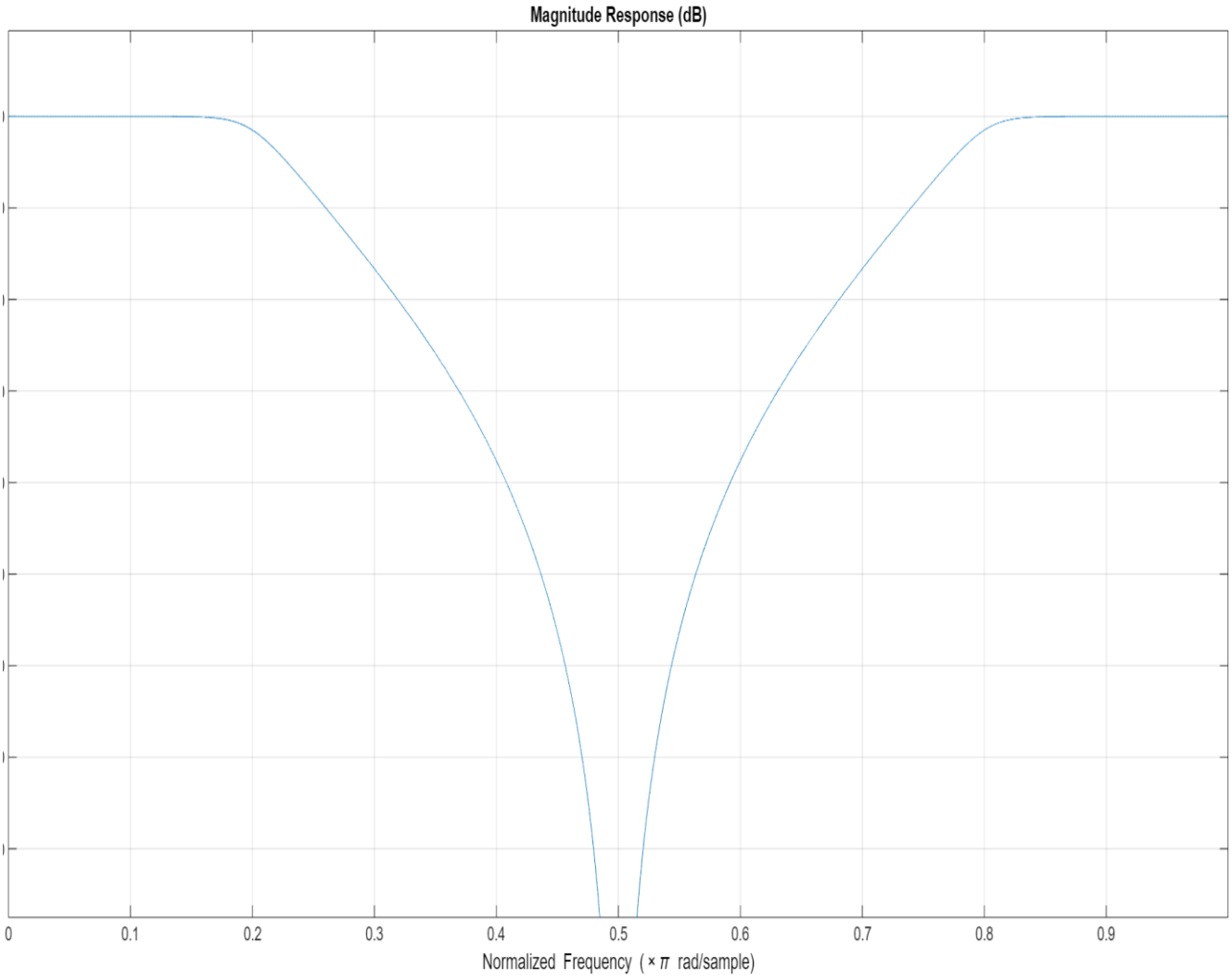
For the next part I tested the program using the audio files for the first 5 speakers for the Zero-training folder. Prior to modification, our program had a success rate of 80 percent. However, when we changed the hamming window to a hanning window, increased the centroid number to 32, and resampled the files at about 5600 Hz, we achieved a 100 percent success rate.

Results of running the first 5 speakers in the and Zero-training folder after changes

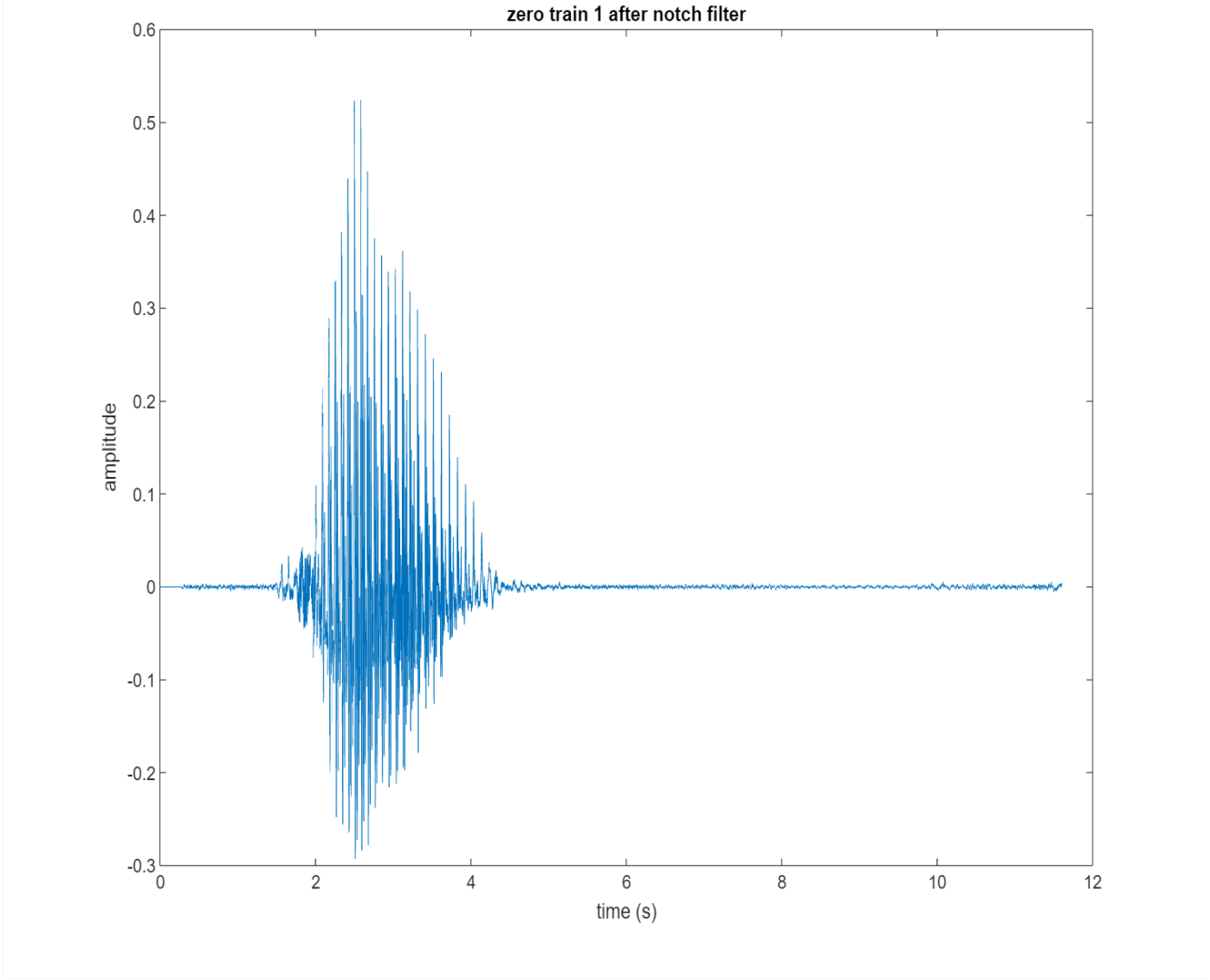
```
>> senior_design
Test 1 corresponds to train 1
Test 2 corresponds to train 2
Test 3 corresponds to train 3
Test 4 corresponds to train 4
Test 5 corresponds to train 5
>>
```

8)

Plot of notch filter used



Plot of audio file after notch filter application



Robustness:

Even after applying the notch filter, the accuracy of the system is still 100 percent when testing using the first 10 samples of the Zero-train, so it would appear that this is a very robust method of voice recognition.

```
>> senior_design
Test 1 corresponds to train 1
Test 2 corresponds to train 2
Test 3 corresponds to train 3
Test 4 corresponds to train 4
Test 5 corresponds to train 5
Test 6 corresponds to train 6
Test 7 corresponds to train 7
Test 8 corresponds to train 8
Test 9 corresponds to train 9
Test 10 corresponds to train 10
>>
```

9) The accuracy after augmenting the given training and testing zero files with an additional 10 zero training and testing files from speakers in class is still 100 percent.

Results:

```
>> senior_design
Test 1 corresponds to train 1
Test 2 corresponds to train 2
Test 3 corresponds to train 3
Test 4 corresponds to train 4
Test 5 corresponds to train 5
Test 6 corresponds to train 6
Test 7 corresponds to train 7
Test 8 corresponds to train 8
Test 9 corresponds to train 9
Test 10 corresponds to train 10
Test 11 corresponds to train 11
Test 12 corresponds to train 12
Test 13 corresponds to train 13
Test 14 corresponds to train 14
Test 15 corresponds to train 15
Test 16 corresponds to train 16
Test 17 corresponds to train 17
Test 18 corresponds to train 18
>>
```

10)

1) The accuracy of the twelve test and the accuracy of the zero test are both 100 percent.

Results for twelve and zero tests

```
>> senior_design
Test 1 corresponds to train 1
Test 2 corresponds to train 2
Test 3 corresponds to train 3
Test 4 corresponds to train 4
Test 5 corresponds to train 5
Test 6 corresponds to train 6
Test 7 corresponds to train 7
Test 8 corresponds to train 8
Test 9 corresponds to train 9
Test 10 corresponds to train 10
Test 11 corresponds to train 11
Test 12 corresponds to train 12
Test 13 corresponds to train 13
Test 14 corresponds to train 14
Test 15 corresponds to train 15
Test 16 corresponds to train 16
Test 17 corresponds to train 17
Test 18 corresponds to train 18
>>
```

2) For this part we created 1 training pool and 1 testing pool. The training pool is composed of files titled: z1 to z36. Z1 to z18 correspond to the twelve training data for speakers 1 through 18, while z19 to 36 correspond to zero training data for speakers 1-18. The testing pool is composed of files titled: tw1 to tw36. Tw1 to tw18 correspond to the twelve testing data for speakers 1 through 18, while tw19 to tw36 corresponds to zero testing data for speakers 1 through 18.

a) The system correctly identified the speaker 35/36 times

b) The system correctly identified the word 34/36 times

Results

```
>> senior_design
Test 1 corresponds to train 1
Test 2 corresponds to train 2
Test 3 corresponds to train 3
Test 4 corresponds to train 4
Test 5 corresponds to train 5
Test 6 corresponds to train 6
Test 7 corresponds to train 7
Test 8 corresponds to train 8
Test 9 corresponds to train 9
Test 10 corresponds to train 10
Test 11 corresponds to train 11
Test 12 corresponds to train 12
Test 13 corresponds to train 13
Test 14 corresponds to train 14
Test 15 corresponds to train 15
Test 16 corresponds to train 16
Test 17 corresponds to train 17
Test 18 corresponds to train 18
Test 19 corresponds to train 19
Test 20 corresponds to train 20
Test 21 corresponds to train 21
Test 22 corresponds to train 22
Test 23 corresponds to train 23
Test 24 corresponds to train 2
Test 25 corresponds to train 25
Test 26 corresponds to train 26
Test 27 corresponds to train 27
Test 28 corresponds to train 28
Test 29 corresponds to train 29
Test 30 corresponds to train 30
Test 31 corresponds to train 31
Test 32 corresponds to train 32
Test 33 corresponds to train 15
Test 34 corresponds to train 34
Test 35 corresponds to train 35
Test 36 corresponds to train 36
>>
```

Brief elaboration:

Using this test, so long as the test number and train number match, both the speaker and word are correctly identified. If you observe test 33, you can see that it corresponds to train 15. This means that a zero was confused for a twelve. However, the system correctly identified the speaker. The same cannot be said for test 24 though, in which a zero was confused for a twelve, and additionally, the speakers were confused as well.