Final Project Tests

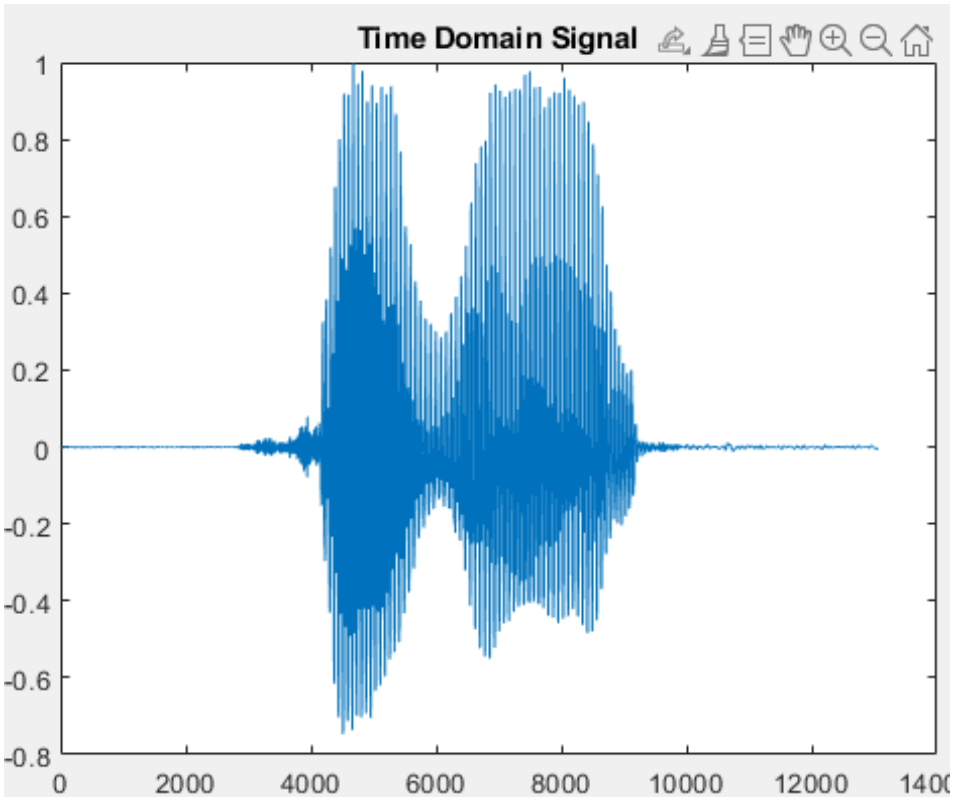
Test 1:

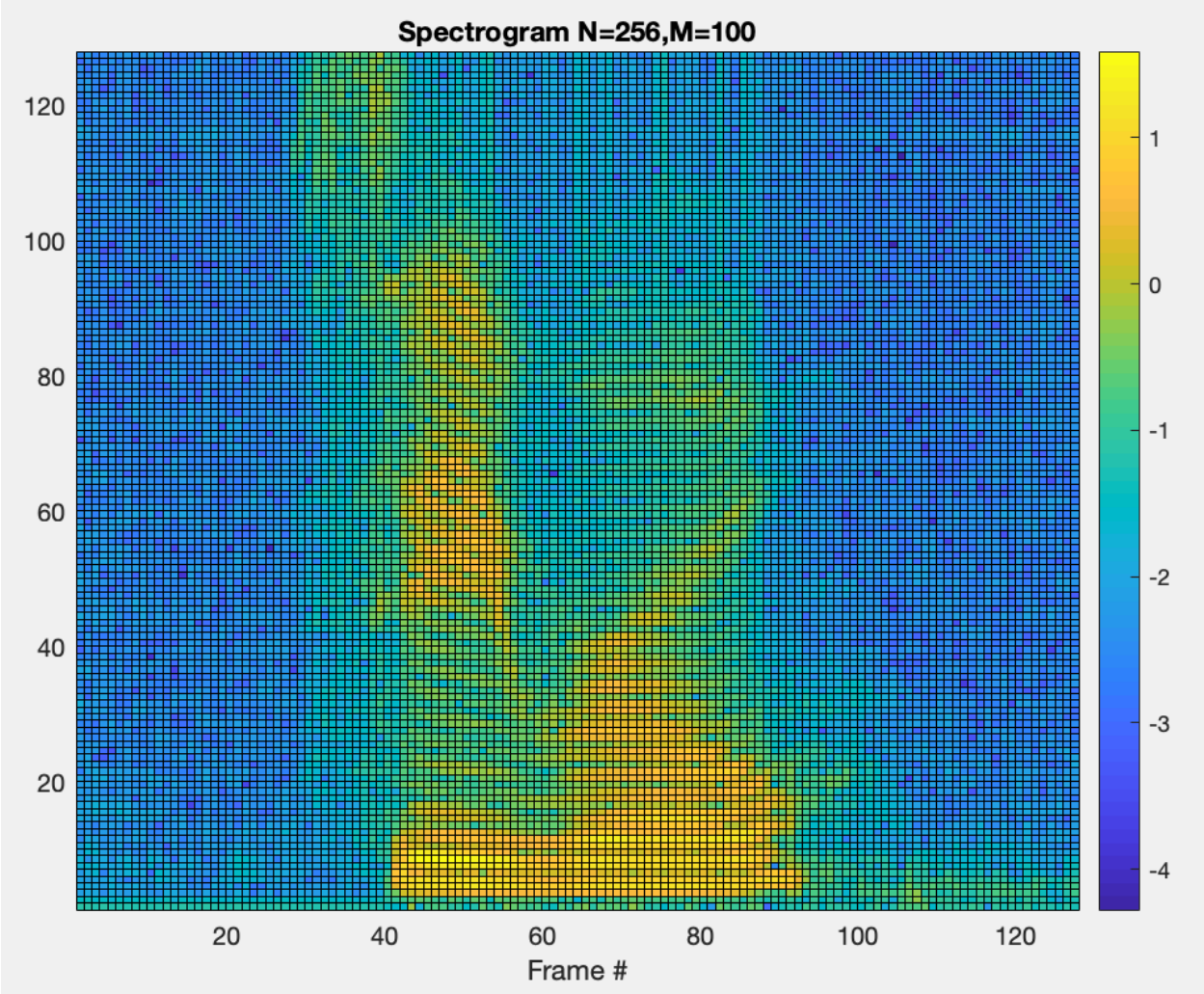
Human accuracy = 75%

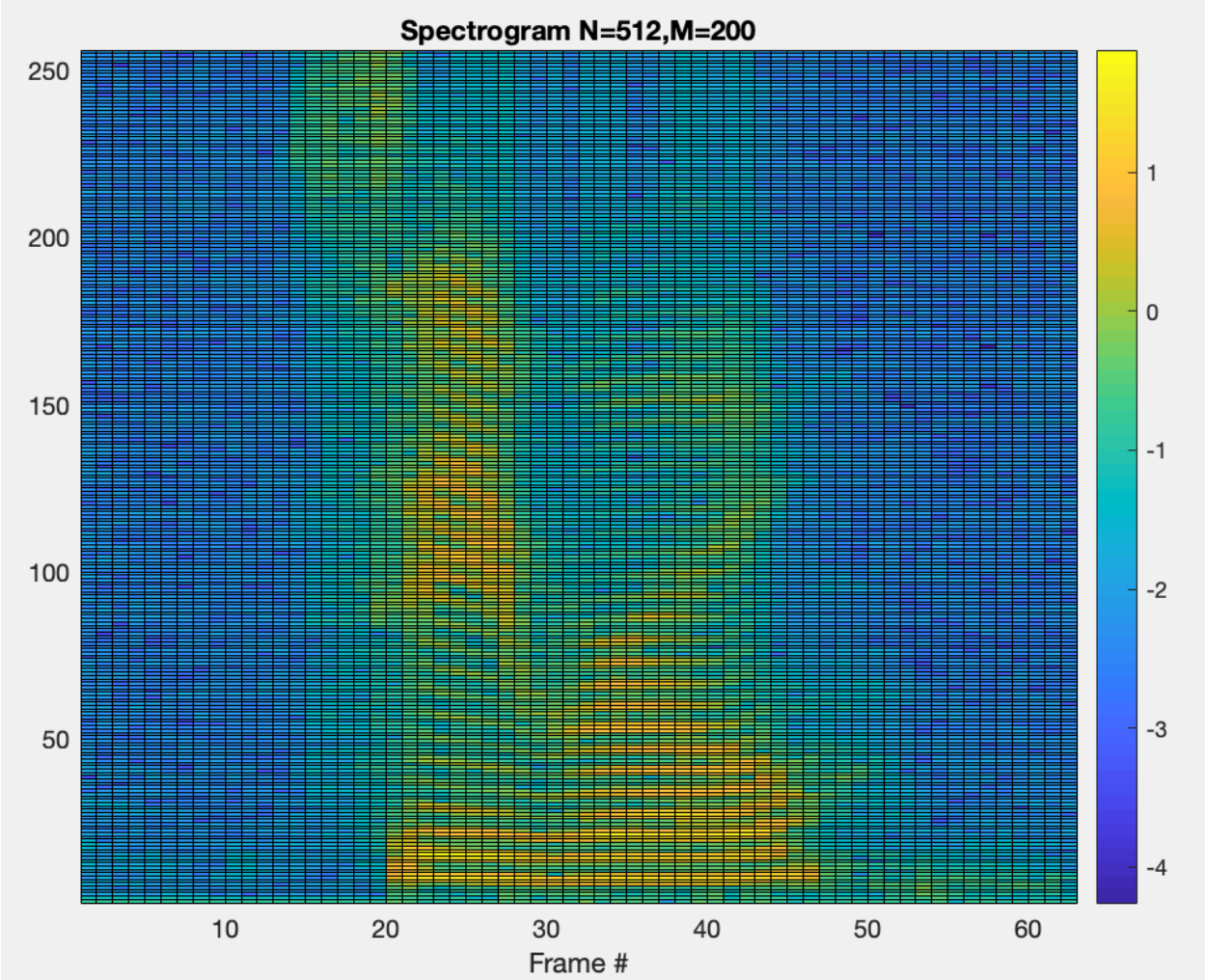
Test 2:

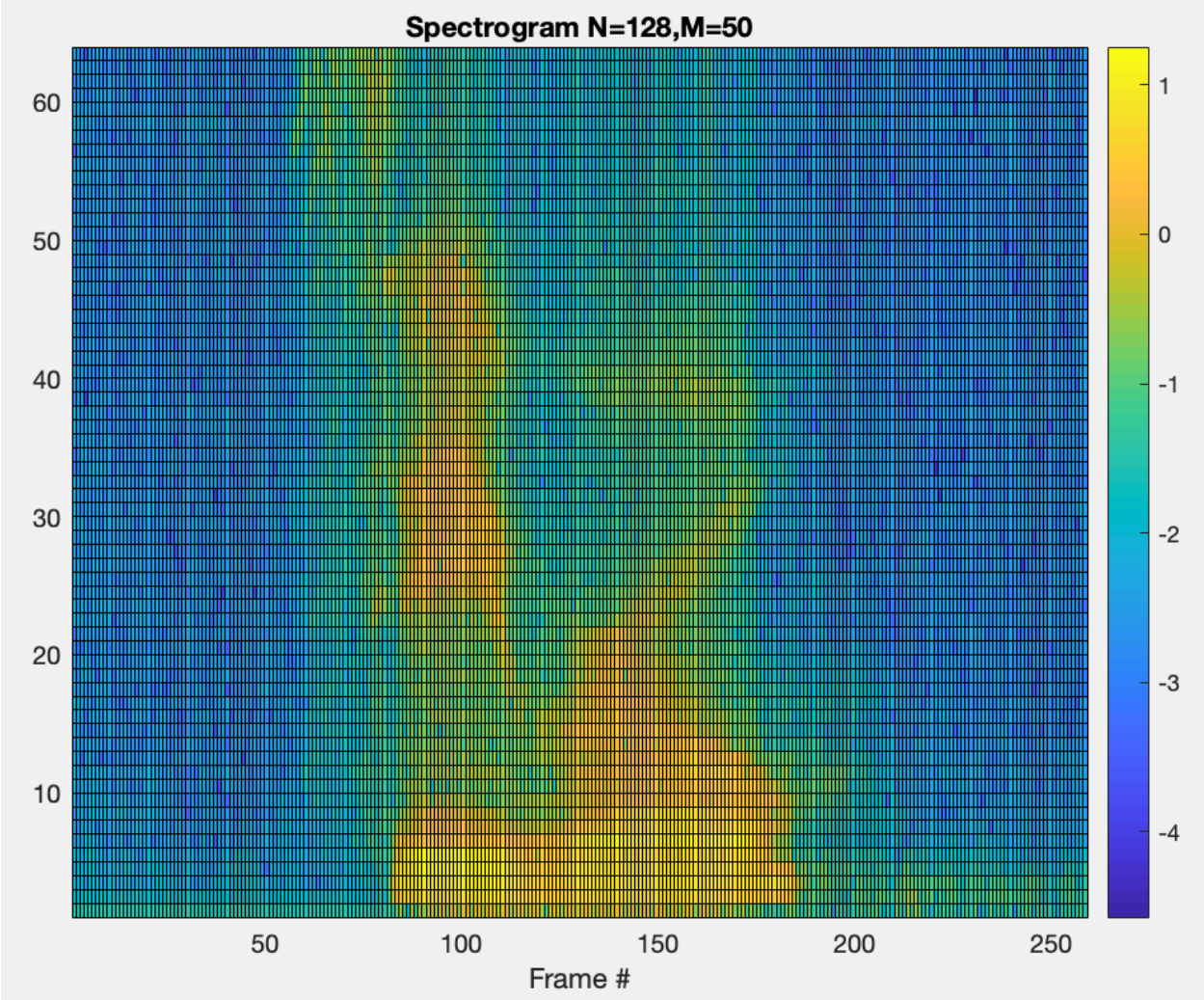
Sampling rate = 12.5 KHz

Time = 20.48 ms per frame



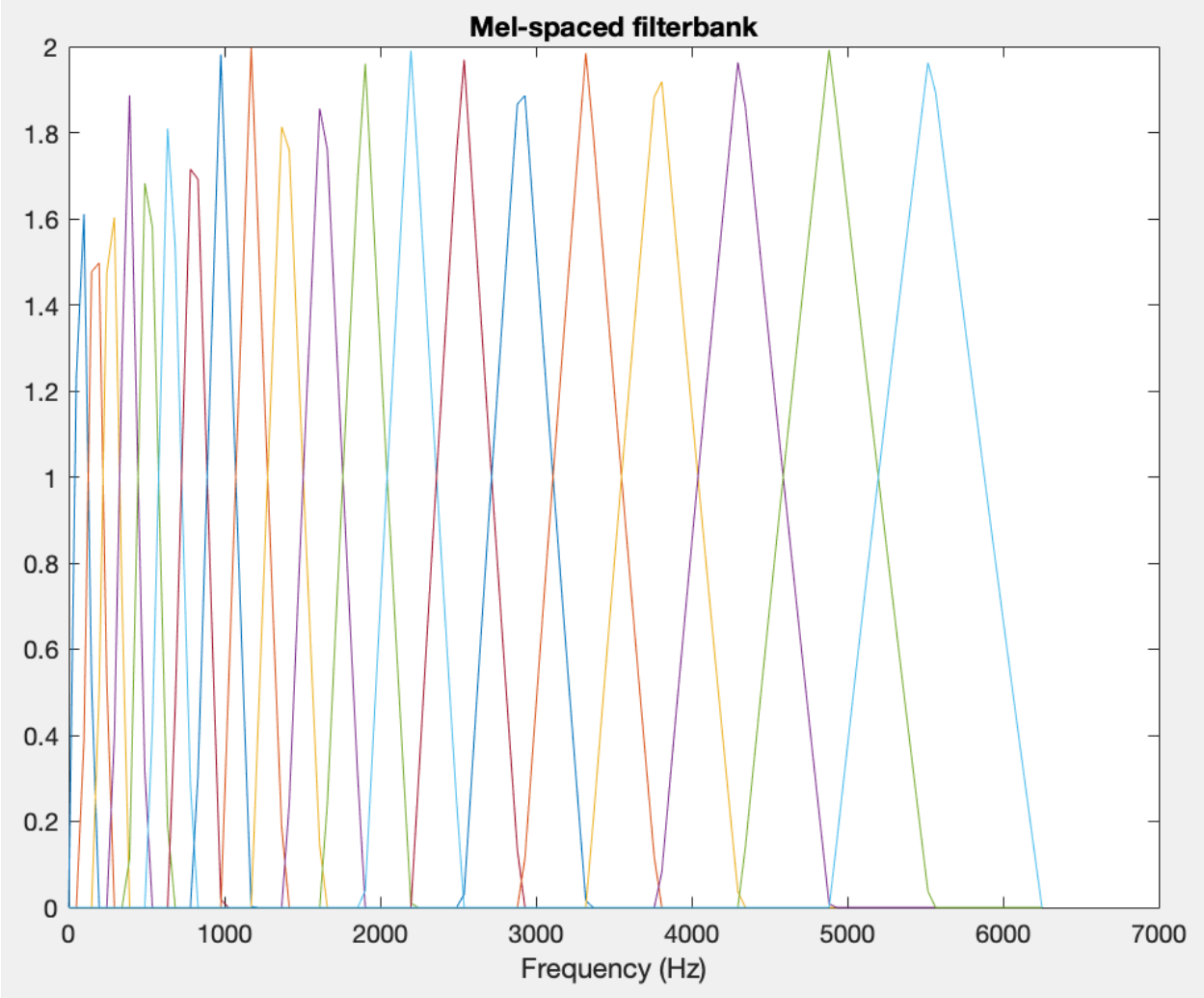

Spectrogram N=256,M=100
120
100
80
60
40
20
20
40
60
80
100
120
Frame #
1
0
-1
-2
-3
-4

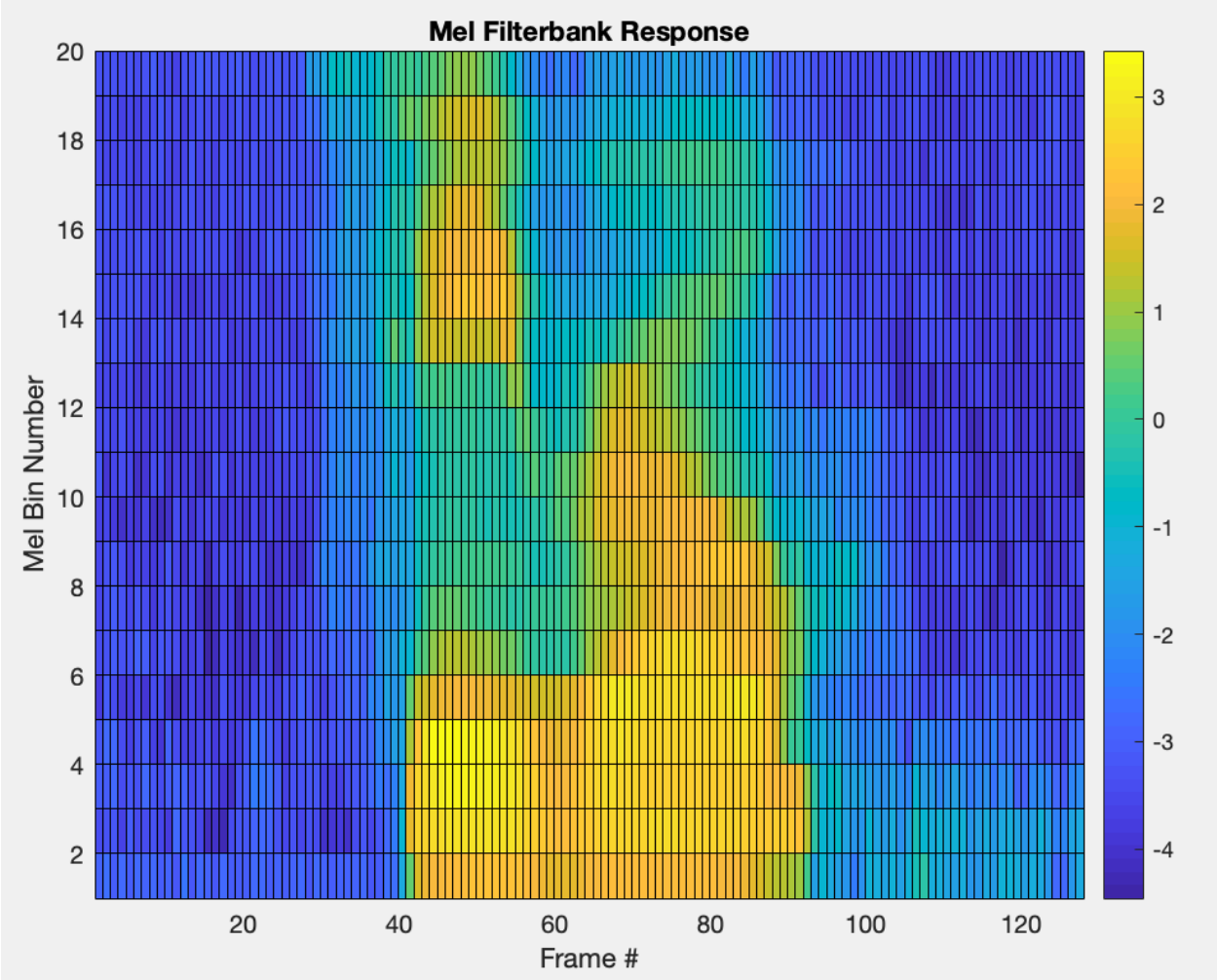

Spectrogram N=512,M=200
250
200
150
100
50
10
20
30
40
50
60
Frame #
1
0
-1
-2
-3
-4


Spectrogram N=128,M=50
60
50
40
30
20
10
50
100
150
200
250
Frame #
1
0
-1
-2
-3
-4

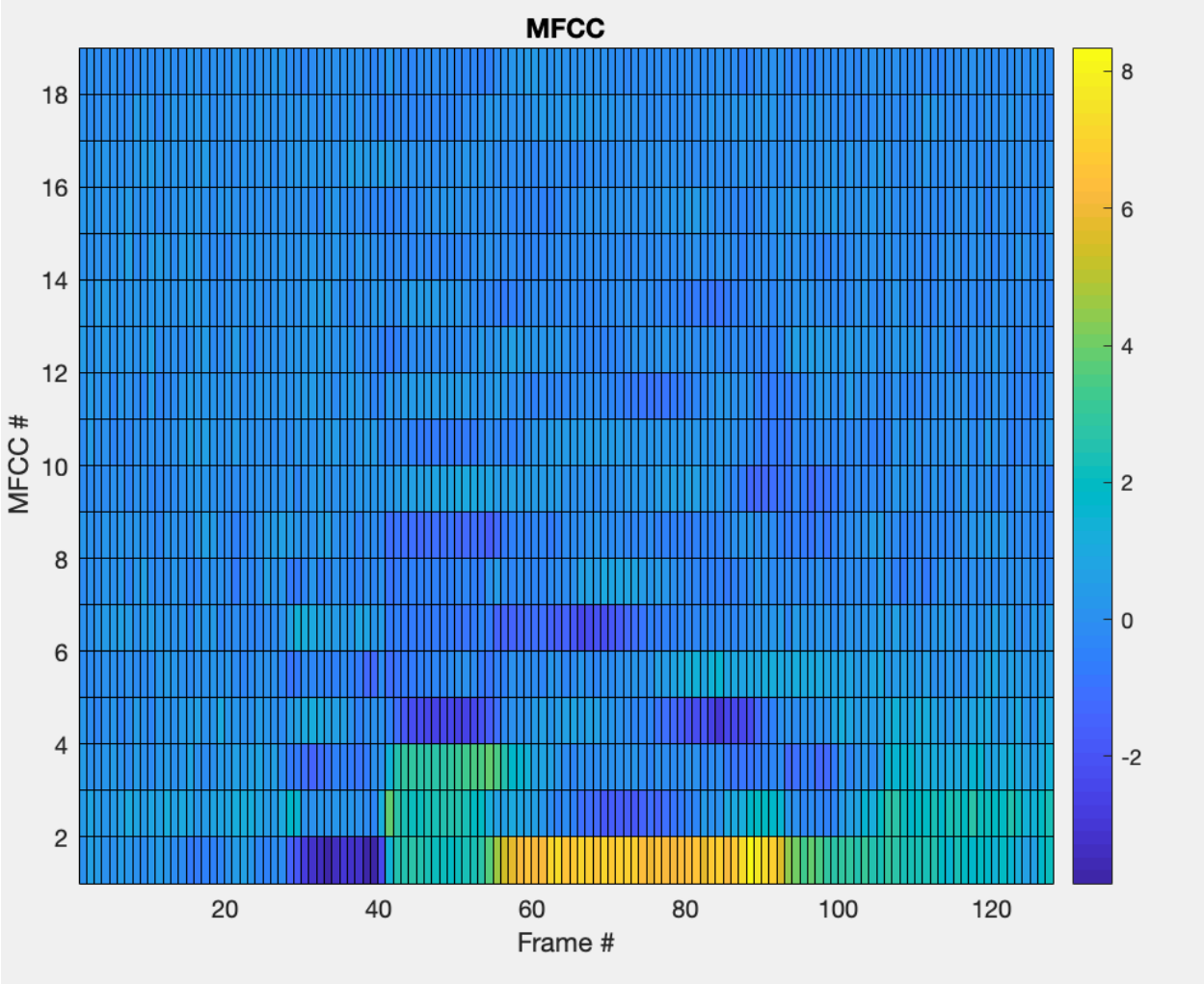
Test 3:

Mel filterbank plot looks somewhat jagged due to the discretization error of having only N = 256 points. They are not all perfectly triangular and they don't peak at the same amplitudes.

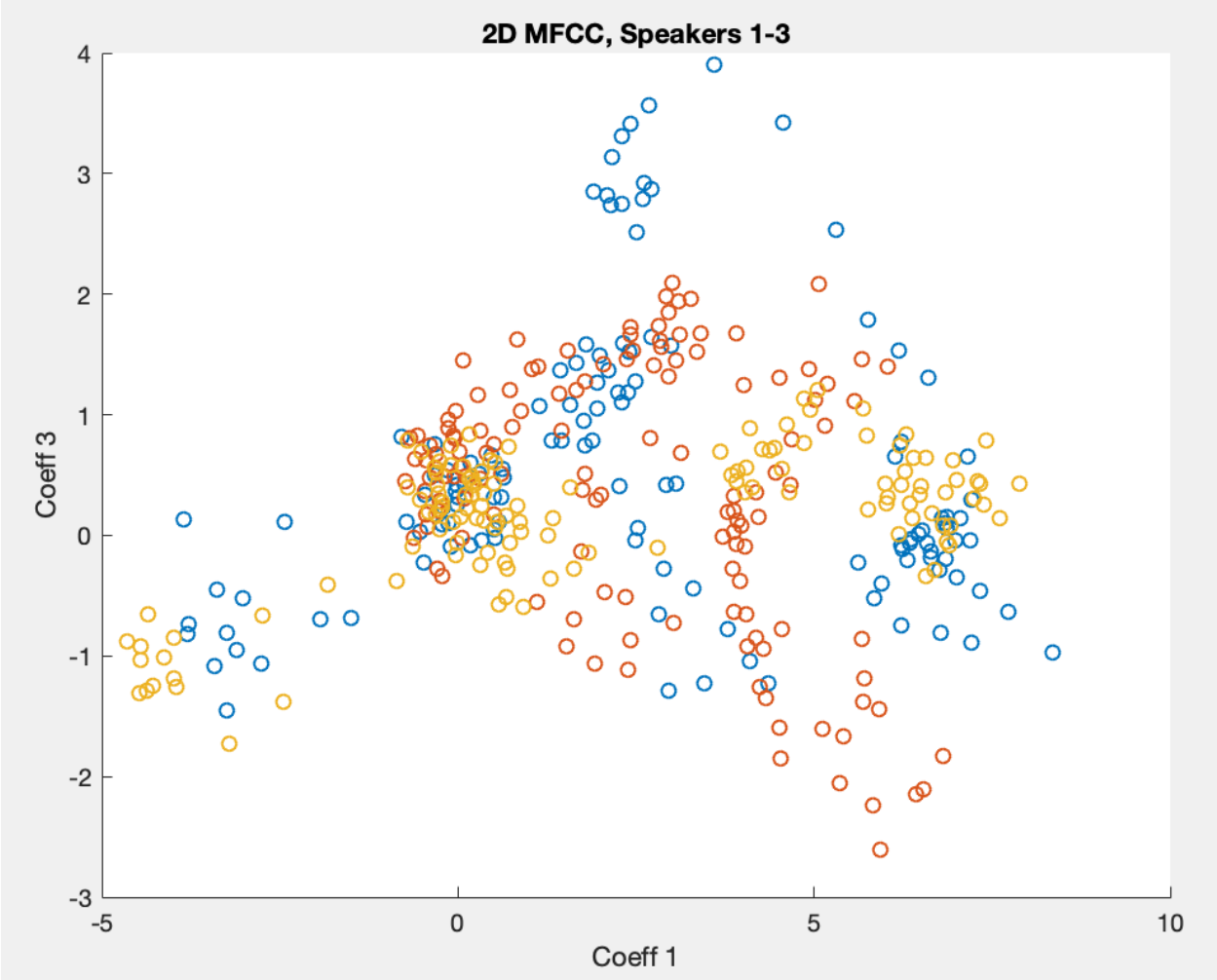



Mel Filterbank Response
Mel Bin Number
Frame #
20
18
16
14
12
10
8
6
4
2
20
40
60
80
100
120
3
2
1
0
-1
-2
-3
-4

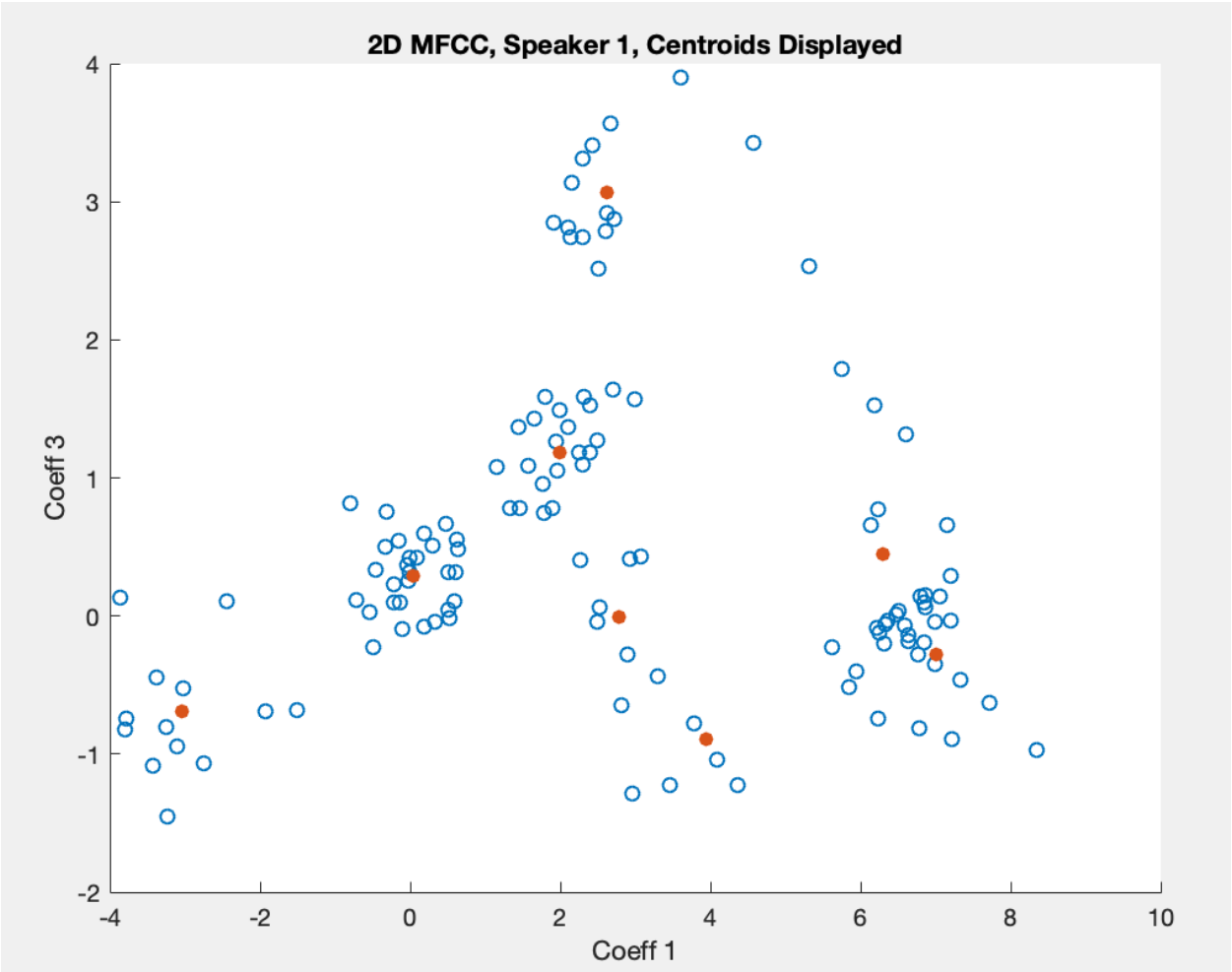
Test 4:



Test 5:



Test 6:



Test 7:

With parameters, 256,100,20,.001,.01, and 8 for N,M,P, thresh, epsilon and codeword_lim respectively, code gets 100% correct on original test audio, much better than human 75% correct

Test 8:

Reduction in accuracy due to notch filtering is highly dependent on which frequency the notch filters out, the lower frequencies where most of the speech energy is have a much larger effect. With notch at .5*pi, accuracy is reduced to 87.5% on original audio.

Test 9:

100% accuracy on group member's voices

Test 10:

82.14% overall accuracy on 28 test samples, and 16 different training voices