Task 1: Code submitted (task1.m)

Task 2: Feature Extraction

For each type of feature extracted do the following things,
a) write an explanation on how the feature is extracted.

1. Standard deviation was calculated using MATLAB’s sd function excluding zero values
   Syntax: std(m1(m1~=0)) where m1 is the input matrix
   Formula:

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}}.$$

2. Min - The minimum of each feature was calculated using the min function of MATLAB
   Syntax: min(m1(m1~=0)) where m1 is the input matrix

3. Max - The minimum of each feature was calculated using the min function of MATLAB
   Syntax: max(m1(m1~=0)) where m1 is the input matrix

4. FFT was calculated using MATLAB’s FFT function
   Syntax: fft(action) where action is the input matrix

FFT takes each row of the IMU and EMG data matrix as input and computes FFT value. This is what we get as the output from the fft function in MATLAB.

5. DWT was calculated using MATLAB’s DWT function with input parameters : ‘db1’
   Syntax: dwt(action(i,:), 'db1') where action(i,:) is a single vector representing one row(for eg.eating action 1 acc x)

FFT takes each row of the IMU and EMG data matrix as input and computes FFT value. This is what we get as the output from the fft function in MATLAB.

b) Write an intuition on why you use such a feature

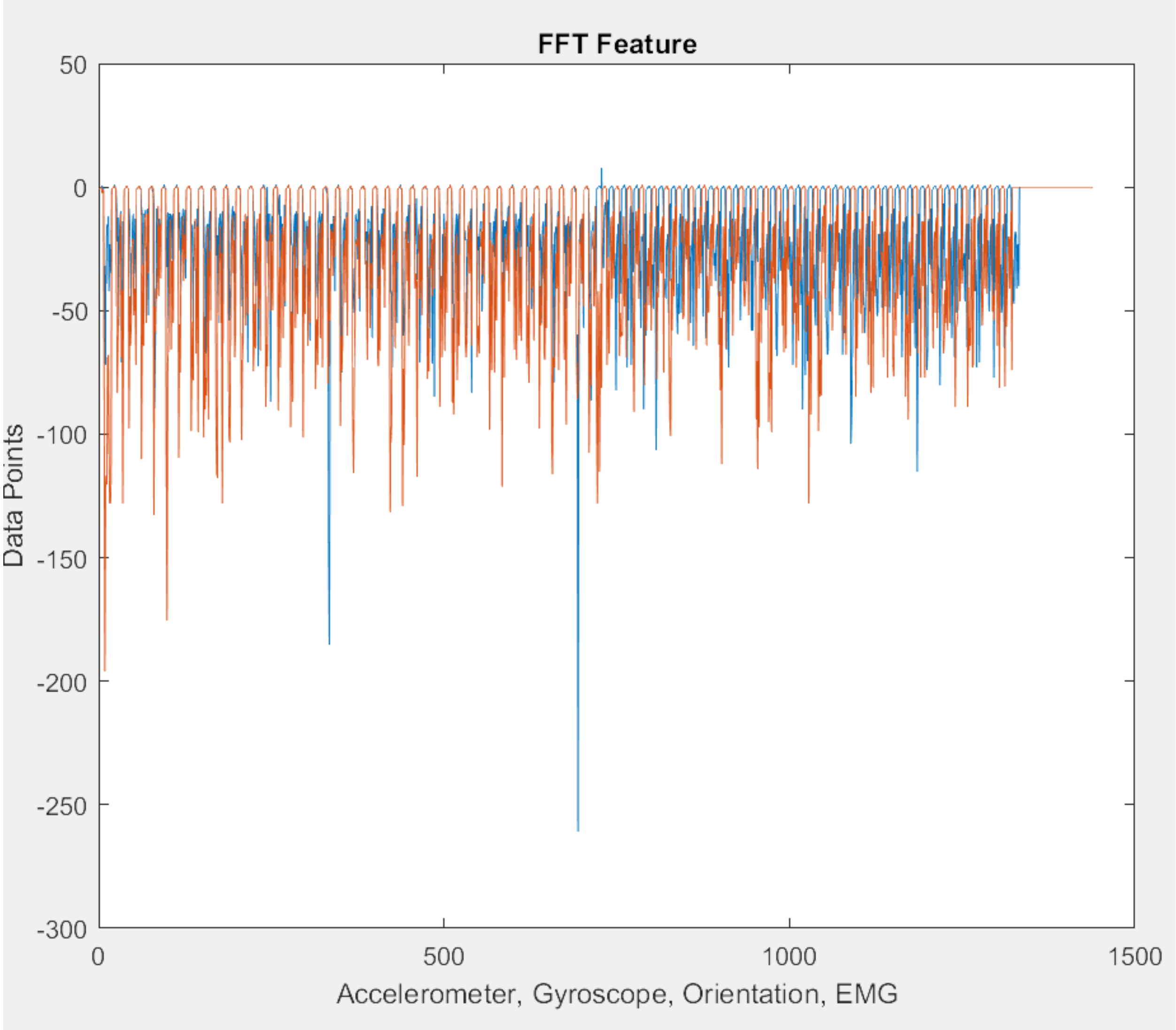
1. We chose to calculate five features from the eating and non eating csv files consisting of IMU and EMG values-Min,Max,FFT,DWT and Standard deviation.
2. Standard deviation was chosen because it is not sensitive to outliers,whereas mean is very sensitive to outliers.
3. Min and Max are important since these are required to compute the range of the values.
4. FFT was chosen because the IMU and EMG values are collected for a finite time interval and they vary with time.FFT is an apt representation for values/signal in a finite time interval and vary with time.
5. DWT is used because it is a good feature for samples that vary with location and time.

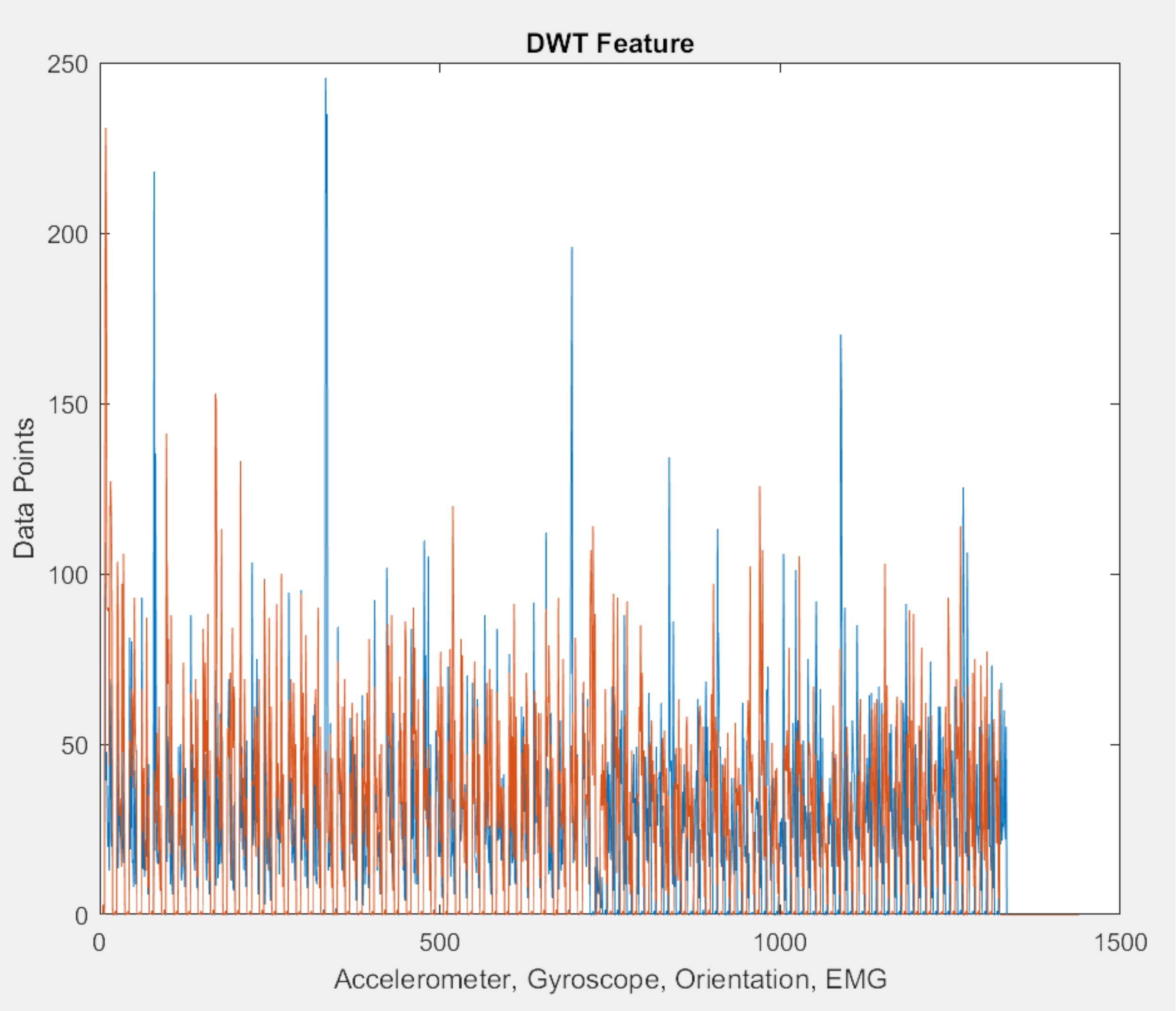
c) Write a matlab code to extract that feature from each time series stored in the csv files created in task 1.

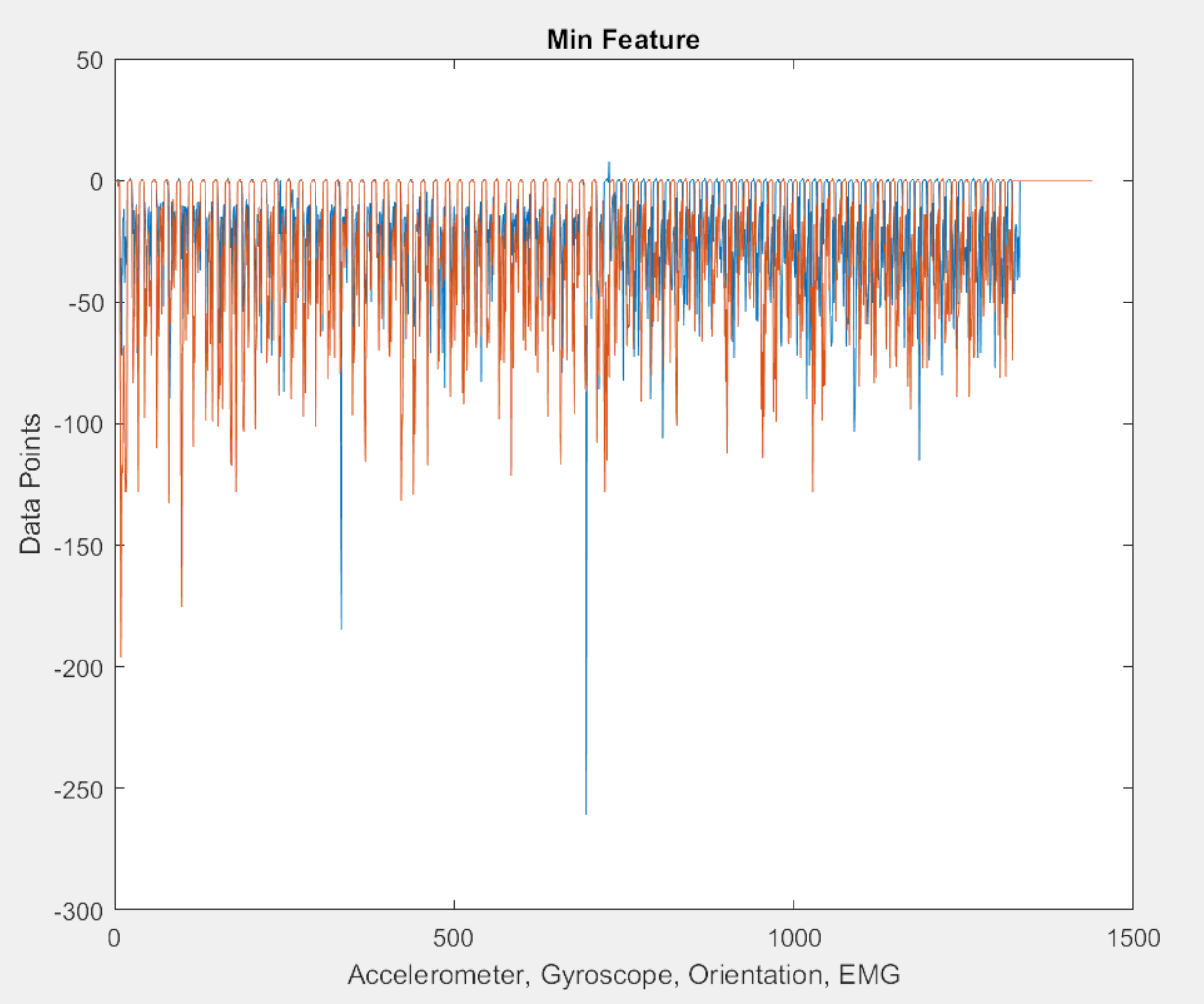
d) Generate two plots: i) features extracted from all eating actions, and ii) features extracted from all non-eating actions.

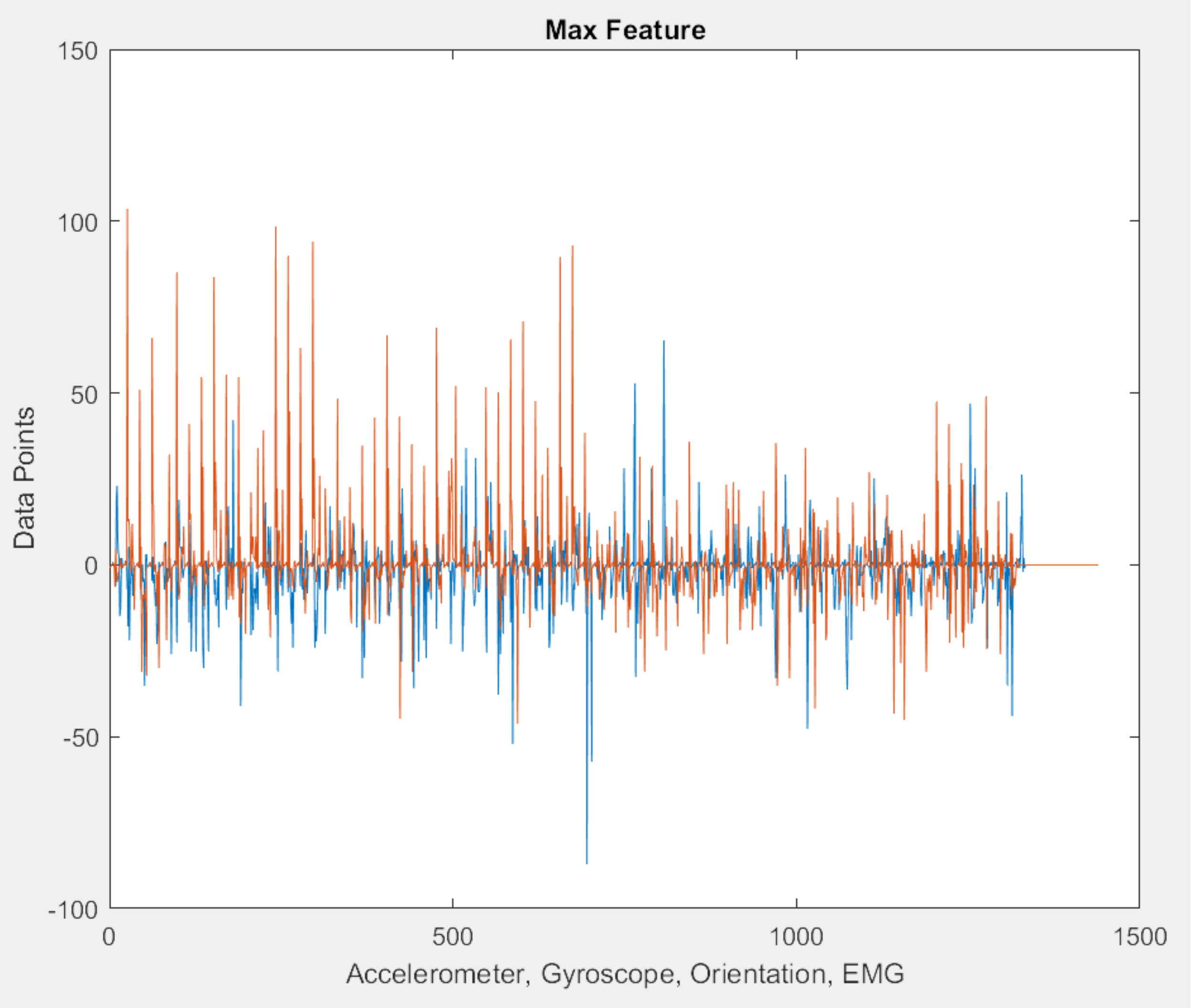
Legend:

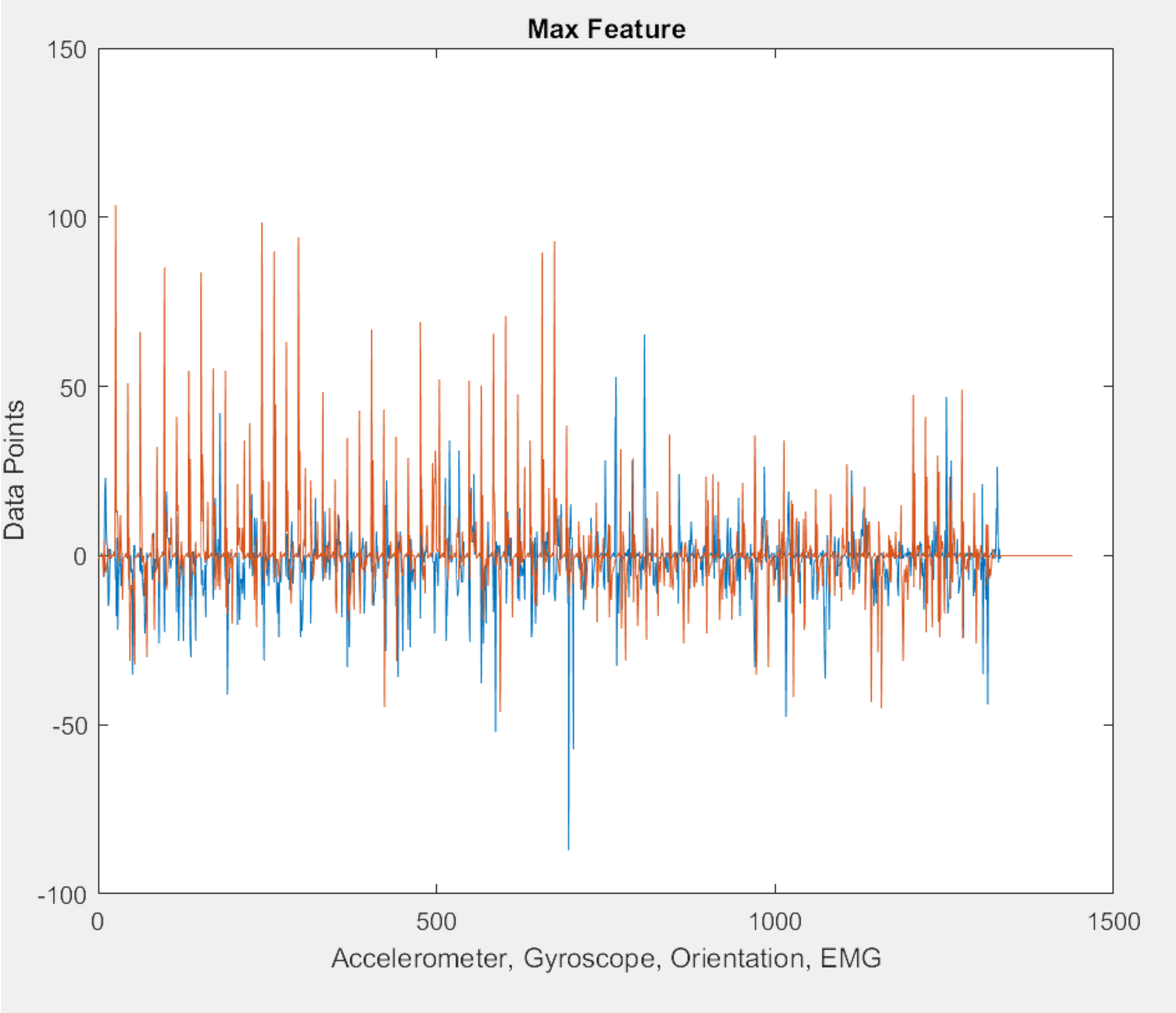
Blue-Eating
Red -Non-Eating

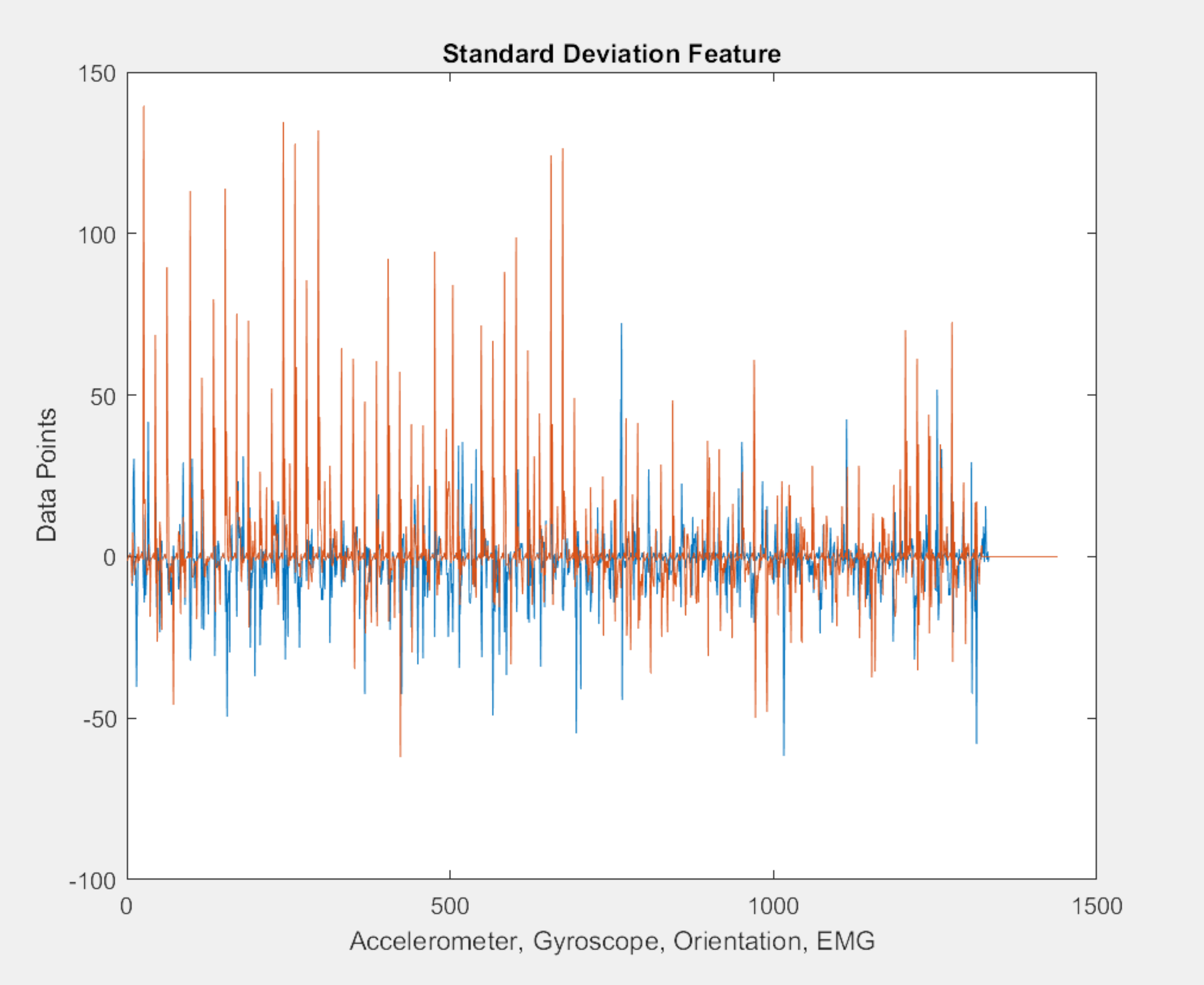



DWT Feature
Data Points
0
50
100
150
200
250
0
500
1000
1500
Accelerometer, Gyroscope, Orientation, EMG


Min Feature
Data Points
50
0
-50
-100
-150
-200
-250
-300
0
500
1000
1500
Accelerometer, Gyroscope, Orientation, EMG


Max Feature
Data Points
150
100
50
0
-50
-100
0
500
1000
1500
Accelerometer, Gyroscope, Orientation, EMG


Max Feature
150
100
50
0
-50
-100
0
500
1000
1500
Data Points
Accelerometer, Gyroscope, Orientation, EMG



e) Discuss whether your initial intuition about the features that you selected holds true or not.

From the graph we can infer that there is clear distinction between eating and non eating classes. This shows that the features that we selected are good enough to distinguish if a person is eating or not eating.

Task 3: Feature Selection

Subtask 1: Arranging the feature matrix

-You know PCA only takes one matrix. How will you arrange all sensors and their corresponding features into a single matrix such that the eigenvectors of the covariance matrix directly makes sense to your data set? This means that if the PCA results gives you a eigenvector then the new feature matrix can be obtained by simply multiplying the eigenvector with the old feature matrix. (You might need two matrices corresponding to the two classes)

-Write your logic of feature matrix arrangement.

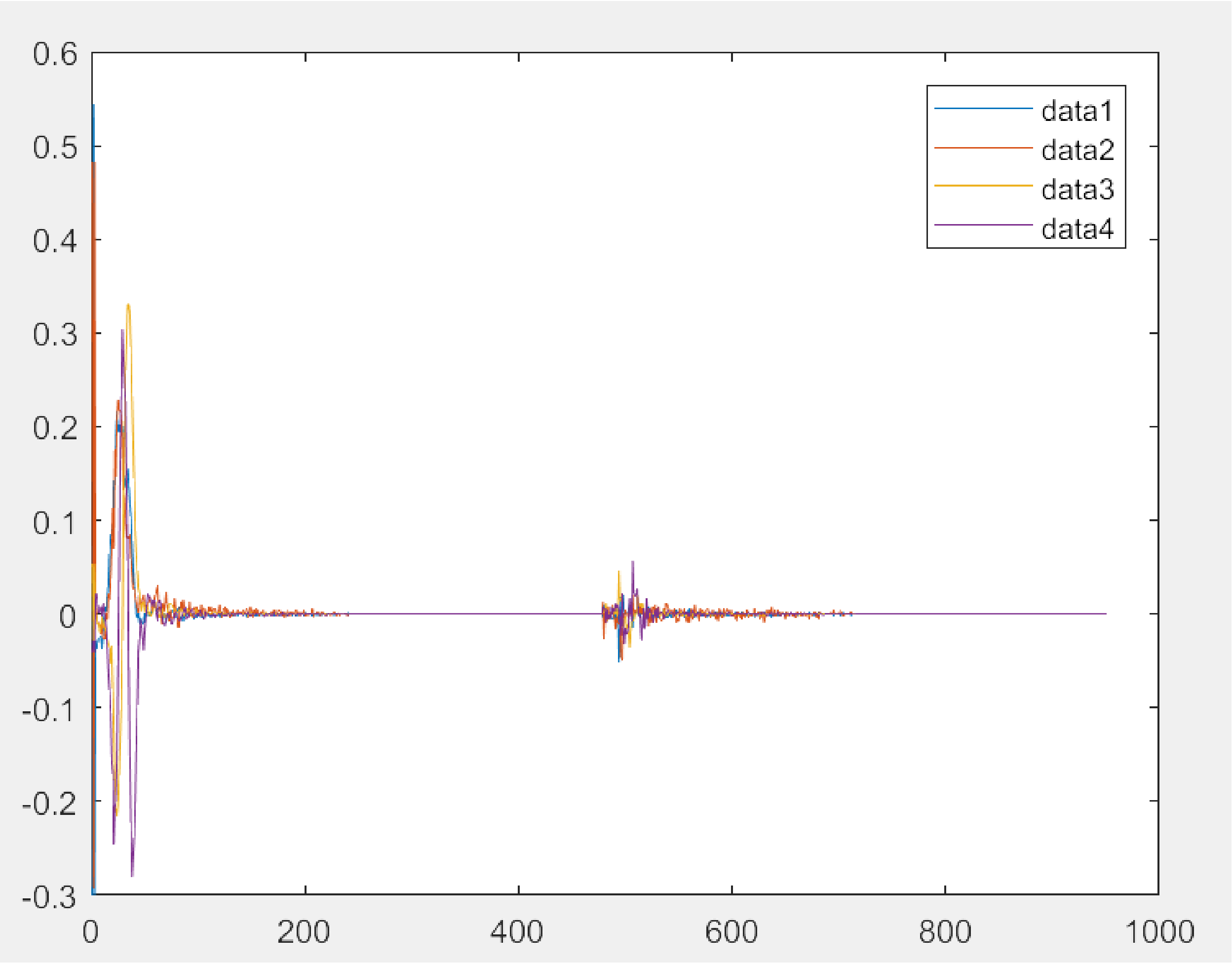
Logic of Feature Matrix Arrangement

Min,Max,Standard deviation,DWT and FFT were calculated for these values and stored in different columns in a new feature matrix.Then this matrix is given to the PCA function.

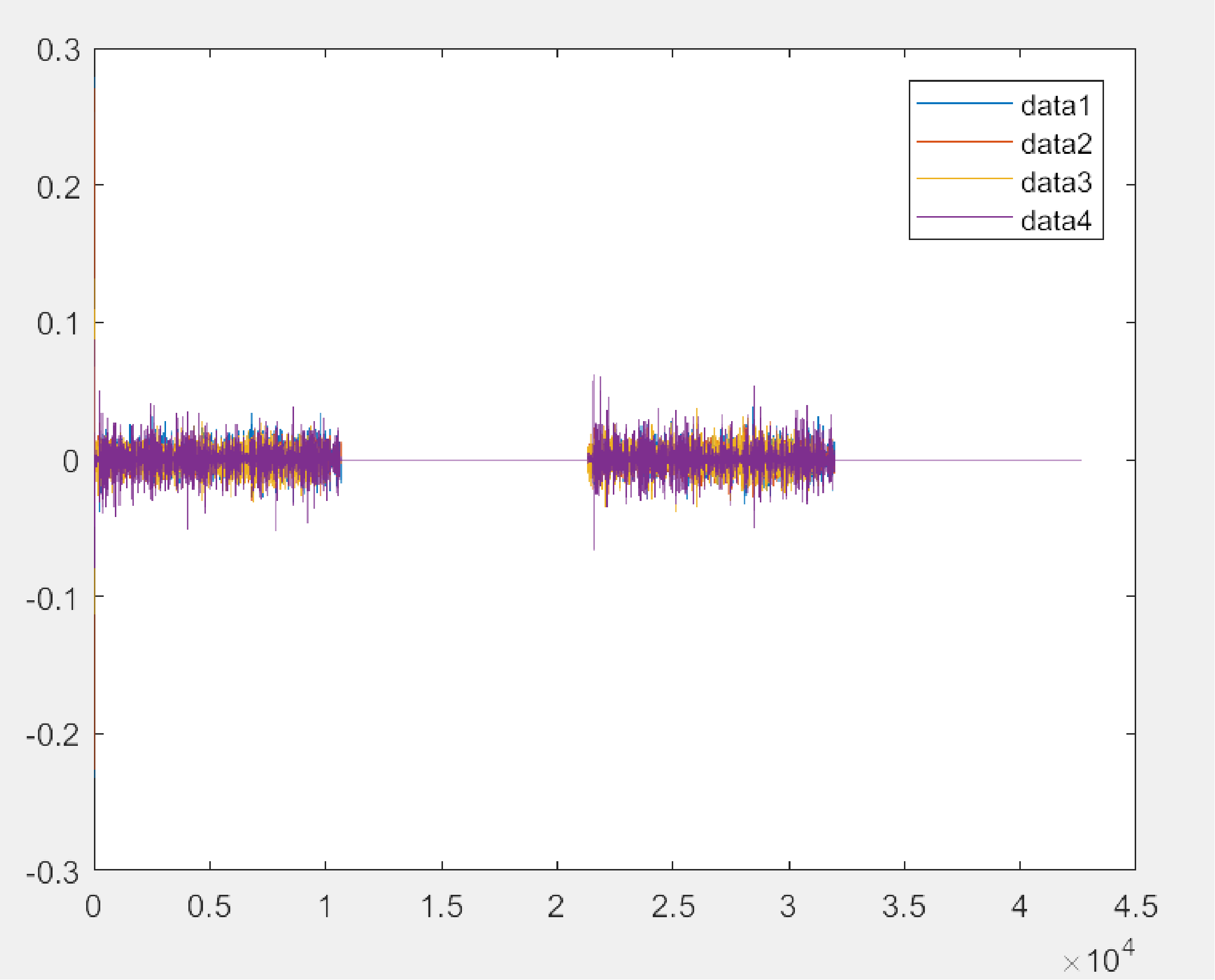
Subtask 2: Execution of PCA

Use Matlab's PCA function to run PCA on your feature matrix. Show all the eigenvectors in a plot.

EATING ACTION:



Non-eating



Subtask 3: Make sense of the PCA eigenvectors
Write an explanation on the reason why the eigenvectors turned out the way they did.

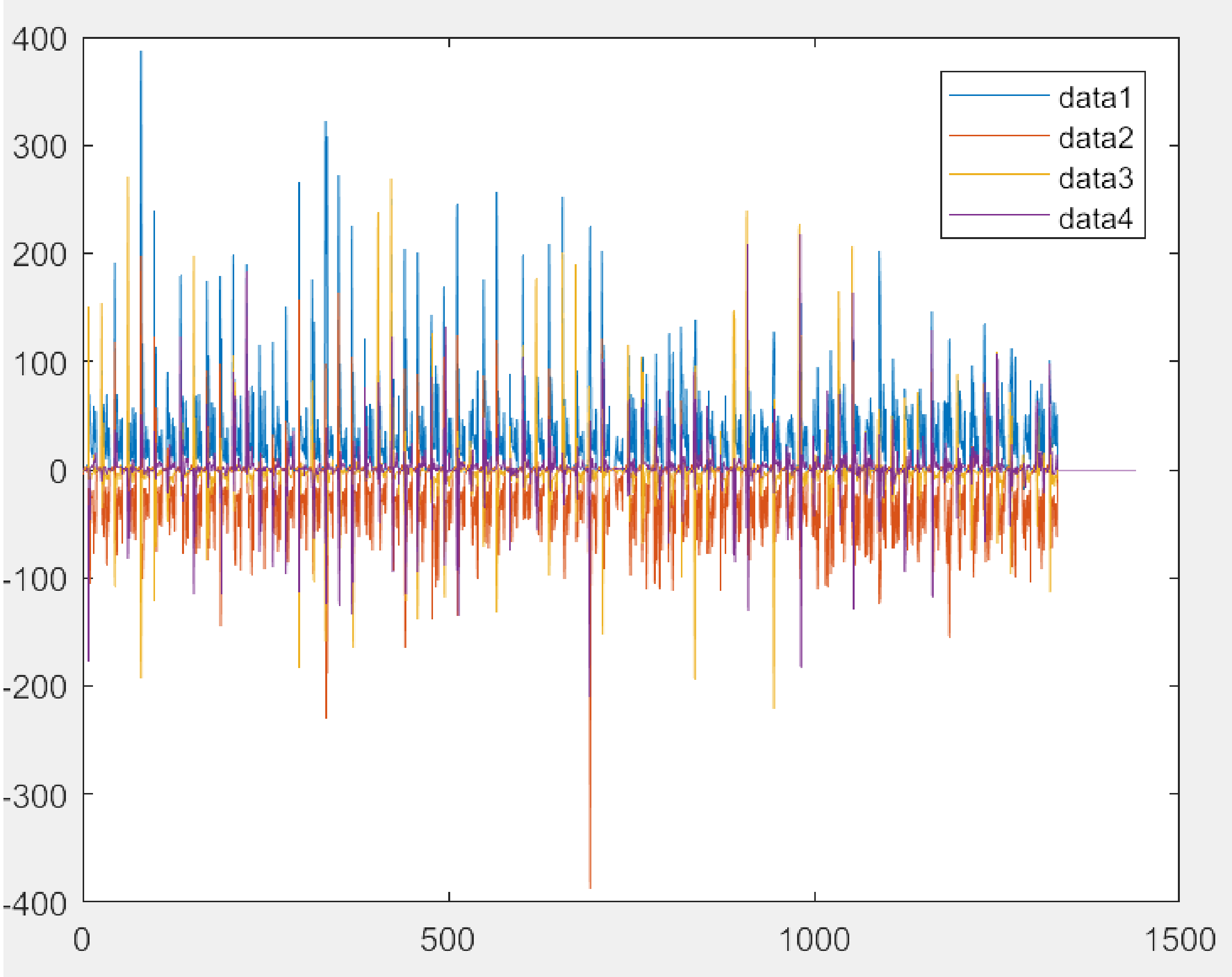
Eating:- Since the user puts more energy while eating and since we have used EMG and IMU sensor values which are measures of energy and orientation, we can infer that the user puts more energy and exerts more movement while eating. We can observe a lot of variation in the EMG values from the graph.

Non eating: When compared to the eating graph, the non eating graph has less variation in it. This illustrates that the user puts comparatively less movement and spends less energy.

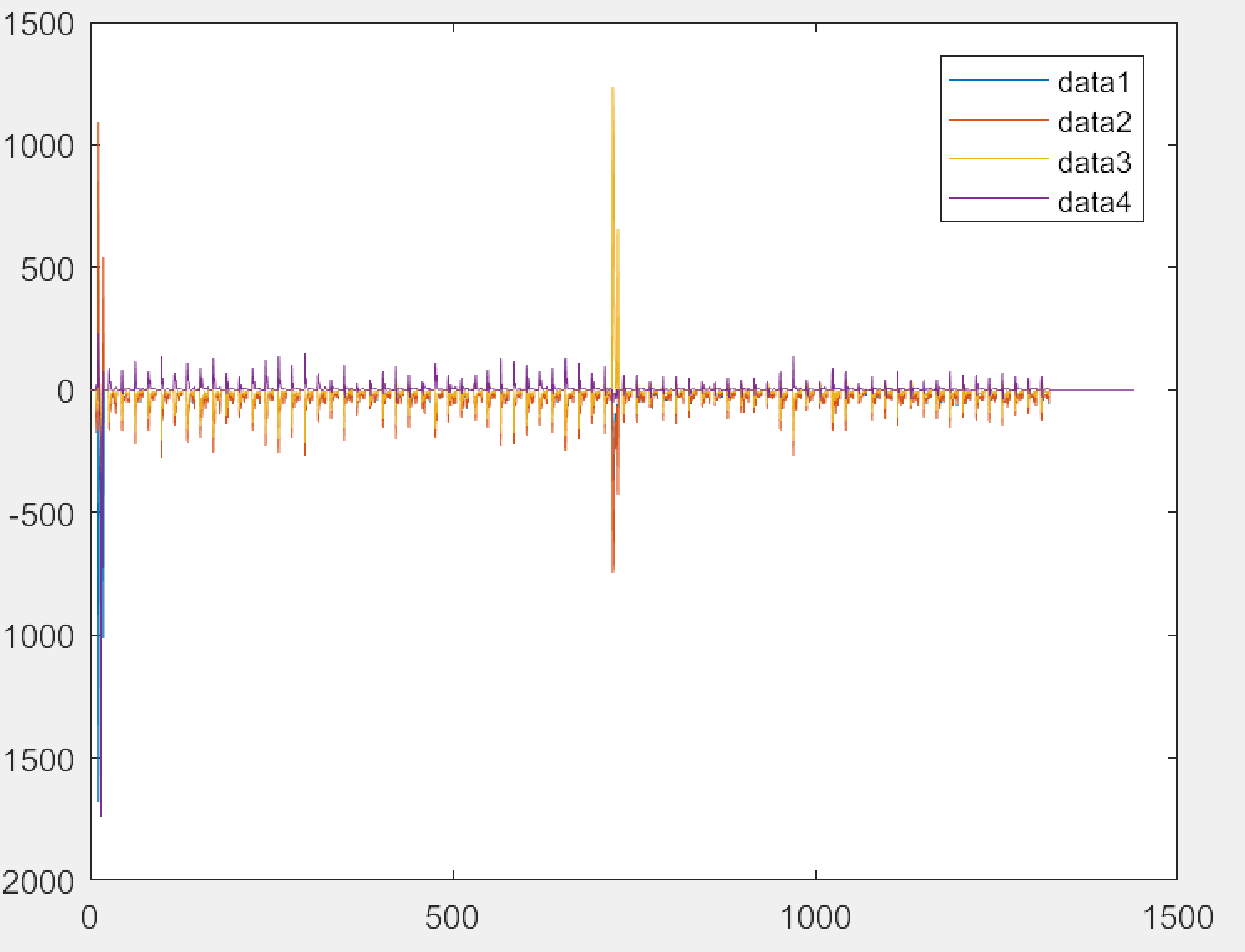
Reason for the initial jump in the non-eating action could be the “snap” action that the user makes before he/she starts eating.

Subtask 4: Results of PCA
Create the new feature matrix. Generate two plots: i) features extracted from all eating actions, and ii) features extracted from all non-eating actions. For multiple eating or non-eating actions you can choose to overlap the plots.

<u>Eating action</u>



Non-eating



Subtask 5: Argue whether doing PCA was helpful or not. May be compare the plots generated from subtask d of task 2 and subtask 4 of Task 3.

PCA for the dataset has been useful to an extent in order to reduce the dimensionality of the feature matrix. We can observe that the feature 2 (orange) & feature 1(blue) has more variance and hence, data is more inclined towards these dimensions (features).