# Assignment 2 Report

## Data Mining

## CSE 572

## Spring 2018

**Submitted to:**

**Prof. Ayan Banerjee**

**Ira A. Fulton School of Engineering**

**Arizona State University**

**Submitted by:**

**Lakshmi Sneha Kandukuri (**lkanduku@asu.edu**)**
**Manasa Pola (**mpola@asu.edu**)**
**Himaja Tirumalasetti (**htirumal@asu.edu**)**
**Hitesh Kumar Sannithi (**hsannithi@asu.edu**)**
**Kushal Reddy Papakannu (**kpapakan@asu.edu**)**

**March 23, 2018**

# Table of Contents

# 1. Introduction

The assignment is to identify the actions representing 10 gestures (about, and, can, cop, deaf, decide, father, find, go out, hearing). The data of each gesture is collected using wristband sensors. The data is also collected using Kinect. The information is pre-processed in such a way that, the useful features for each gesture is used to analyze the data.

# 2. Team Members

Following are the group member of this project

Lakshmi Sneha Kandukuri (lkanduku@asu.edu)

Manasa Pola (mpola@asu.edu)

Himaja Tirumalasetti (htirumal@asu.edu)

Hitesh Kumar Sannithi (hsannithi@asu.edu)

Kushal Reddy Papakannu (kpapakan@asu.edu)

# 3. Assignment 1

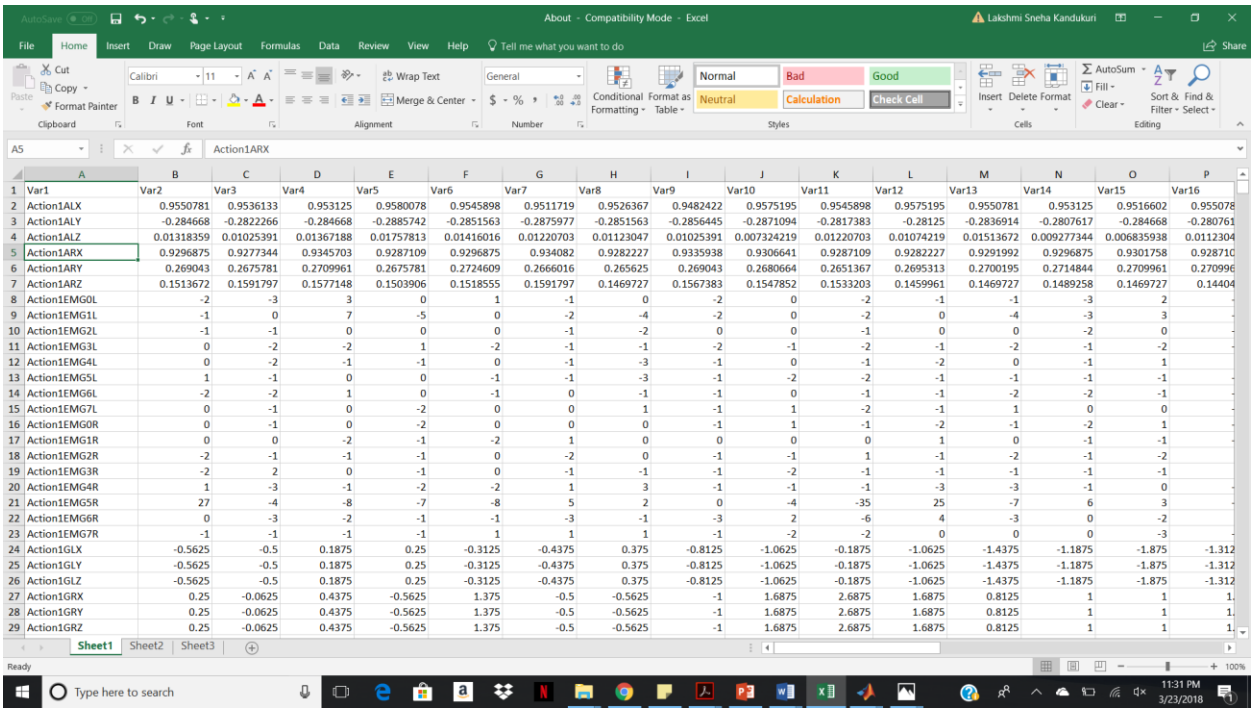
The assignment 1 is for collecting the sensor data. All the team members should go to Impact lab for collecting the data. Any one of the team member should volunteer for gesturing in front of the screen wearing 2 wristbands, one on right and one on left hand. The gestures are identified by the sensors using the hand movement of the person wearing the wristbands. The gestures are captured using 4 sensors Gyroscope, Accelerator, EMG Sensor, Orientation. According to movement of the hand, each sensor captures the data, for example, if the gesture consists of rotations, orientation sensor data is useful to identify the gesture.

# 4. Assignment 2

The Assignment 2 Consists of 3 tasks.

Task 1: To process the raw data collected by the sensors into 10 separate classes, each class corresponding to a gesture. We have taken the data sets of 6 persons for analyzing the data. Each gesture has been performed 20 times. Data of all actions corresponding to a gesture of 6 persons is formatted, combined and stored in a csv file using MATLAB code. This results in 10 csv files, one for

each gesture. The formatted csv file consists of a matrix data which has rows and columns, rows representing types of sensors and columns representing the time series when the data was collected. Below is the sample of the formatted data matrix:



| | Var1 | Var2 | Var3 | Var4 | Var5 | Var6 | Var7 | Var8 | Var9 | Var10 | Var11 | Var12 | Var13 | Var14 | Var15 | Var16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Action1ALX | 0.9550781 | 0.9536133 | 0.953125 | 0.9580078 | 0.9545898 | 0.9511719 | 0.9526367 | 0.9482422 | 0.9575195 | 0.9545898 | 0.9575195 | 0.9550781 | 0.953125 | 0.9516602 | 0.955078 |
| 3 | Action1ALY | -0.284668 | -0.2822266 | -0.284668 | -0.2885742 | -0.2851563 | -0.2875977 | -0.2851563 | -0.2856445 | -0.2871094 | -0.2817383 | -0.28125 | -0.2836914 | -0.2807617 | -0.284668 | -0.280761 |
| 4 | Action1ALZ | 0.01318359 | 0.01025391 | 0.01367188 | 0.01757813 | 0.01416016 | 0.01220703 | 0.01123047 | 0.01025391 | 0.007324219 | 0.01220703 | 0.01074219 | 0.01513672 | 0.009277344 | 0.006835938 | 0.0112304 |
| 5 | Action1ARX | 0.9296875 | 0.9277344 | 0.9345703 | 0.9287109 | 0.9296875 | 0.934082 | 0.9282227 | 0.9335938 | 0.9306641 | 0.9287109 | 0.9282227 | 0.9291992 | 0.9296875 | 0.9301758 | 0.928710 |
| 6 | Action1ARY | 0.269043 | 0.2675781 | 0.2709961 | 0.2675781 | 0.2724609 | 0.2666016 | 0.265625 | 0.269043 | 0.2680664 | 0.2651367 | 0.2695313 | 0.2700195 | 0.2714844 | 0.2709961 | 0.270996 |
| 7 | Action1ARZ | 0.1513672 | 0.1591797 | 0.1577148 | 0.1503906 | 0.1518555 | 0.1591797 | 0.1469727 | 0.1567383 | 0.1547852 | 0.1533203 | 0.1459961 | 0.1469727 | 0.1489258 | 0.1469727 | 0.14404 |
| 8 | Action1EMG0L | -2 | -3 | 3 | 0 | 1 | -1 | 0 | -2 | 0 | -2 | -1 | -1 | -3 | 2 | |
| 9 | Action1EMG1L | -1 | 0 | 7 | -5 | 0 | -2 | -4 | -2 | 0 | -2 | 0 | -4 | -3 | 3 | |
| 10 | Action1EMG2L | -1 | -1 | 0 | 0 | 0 | -1 | -2 | 0 | 0 | -1 | 0 | 0 | -2 | 0 | |
| 11 | Action1EMG3L | 0 | -2 | -2 | 1 | -2 | -1 | -1 | -2 | -1 | -2 | -1 | -2 | -1 | -2 | |
| 12 | Action1EMG4L | 0 | -2 | -1 | -1 | 0 | -1 | -3 | -1 | 0 | -1 | -2 | 0 | -1 | 1 | |
| 13 | Action1EMG5L | 1 | -1 | 0 | 0 | -1 | -1 | -3 | -1 | -2 | -2 | -1 | -1 | -1 | -1 | |
| 14 | Action1EMG6L | -2 | -2 | 1 | 0 | -1 | 0 | -1 | -1 | 0 | -1 | -1 | -2 | -2 | -1 | |
| 15 | Action1EMG7L | 0 | -1 | 0 | -2 | 0 | 0 | 1 | -1 | 1 | -2 | -1 | 1 | 0 | 0 | |
| 16 | Action1EMG0R | 0 | -1 | 0 | -2 | 0 | 0 | 0 | -1 | 1 | -1 | -2 | -1 | -2 | 1 | |
| 17 | Action1EMG1R | 0 | 0 | -2 | -1 | -2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | -1 | -1 | |
| 18 | Action1EMG2R | -2 | -1 | -1 | -1 | 0 | -2 | 0 | -1 | -1 | 1 | -1 | -2 | -1 | -2 | |
| 19 | Action1EMG3R | -2 | 2 | 0 | -1 | 0 | -1 | -1 | -1 | -2 | -1 | -1 | -1 | -1 | -1 | |
| 20 | Action1EMG4R | 1 | -3 | -1 | -2 | -2 | 1 | 3 | -1 | -1 | -1 | -3 | -3 | -1 | 0 | |
| 21 | Action1EMG5R | 27 | -4 | -8 | -7 | -8 | 5 | 2 | 0 | -4 | -35 | 25 | -7 | 6 | 3 | |
| 22 | Action1EMG6R | 0 | -3 | -2 | -1 | -1 | -3 | -1 | -3 | 2 | -6 | 4 | -3 | 0 | -2 | |
| 23 | Action1EMG7R | -1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | -2 | -2 | 0 | 0 | 0 | -3 | |
| 24 | Action1GLX | -0.5625 | -0.5 | 0.1875 | 0.25 | -0.3125 | -0.4375 | 0.375 | -0.8125 | -1.0625 | -0.1875 | -1.0625 | -1.4375 | -1.1875 | -1.875 | -1.312 |
| 25 | Action1GLY | -0.5625 | -0.5 | 0.1875 | 0.25 | -0.3125 | -0.4375 | 0.375 | -0.8125 | -1.0625 | -0.1875 | -1.0625 | -1.4375 | -1.1875 | -1.875 | -1.312 |
| 26 | Action1GLZ | -0.5625 | -0.5 | 0.1875 | 0.25 | -0.3125 | -0.4375 | 0.375 | -0.8125 | -1.0625 | -0.1875 | -1.0625 | -1.4375 | -1.1875 | -1.875 | -1.312 |
| 27 | Action1GRX | 0.25 | -0.0625 | 0.4375 | -0.5625 | 1.375 | -0.5 | -0.5625 | -1 | 1.6875 | 2.6875 | 1.6875 | 0.8125 | 1 | 1 | 1. |
| 28 | Action1GRY | 0.25 | -0.0625 | 0.4375 | -0.5625 | 1.375 | -0.5 | -0.5625 | -1 | 1.6875 | 2.6875 | 1.6875 | 0.8125 | 1 | 1 | 1. |
| 29 | Action1GRZ | 0.25 | -0.0625 | 0.4375 | -0.5625 | 1.375 | -0.5 | -0.5625 | -1 | 1.6875 | 2.6875 | 1.6875 | 0.8125 | 1 | 1 | 1. |

Task 2: Feature Extraction: In this task, we have applied 5 feature extraction techniques to find the useful features to identify the gestures in the formatted data. Those 5 feature extraction techniques used are

1. Discrete Wavelet Transform
2. Fast Fourier Transform
3. Power Spectral Density
4. Moving Mean
5. Auto Correlation

Task 3: Feature Selection: In this task, we reduce the features extracted in the previous step

## 4.1. Feature Extraction

a) Features are extracted using 5 Feature extraction techniques:

a. Discrete Wavelet Transform
b. Fast Fourier Transform
c. Power Spectral Density
d. Moving Mean
e. Auto Correlation

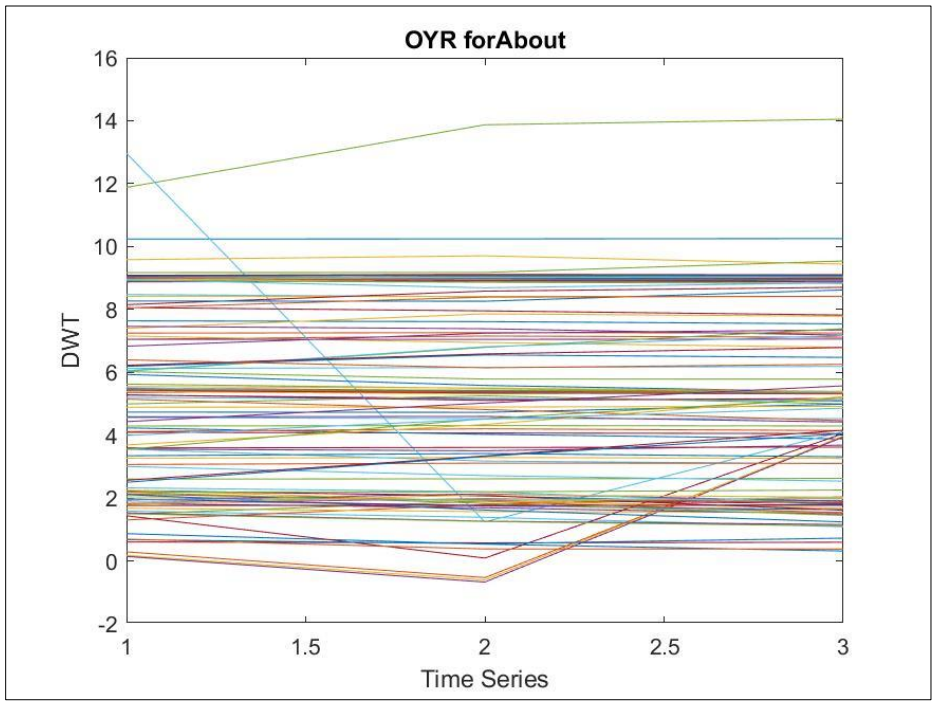
As we have considered the data of 6 persons, the number of actions per feature are 120(6 persons x 20 actions per person). There are total of 34 features. Therefore, 34 x 120 = 4080 records in the matrix. We have built a temporary matrix from the original matrix in such a way that all the records corresponding to same feature are stacked together, therefore we have 34 temporary matrices. We have applied the above-mentioned feature extraction techniques on each of the temporary matrix.

c) We have written a MATLAB code to apply the feature extraction techniques, we have used the built in functions for each techniques in the code like fft() for Fast Fourier transform, autocorr() for Auto Correlation, movmean() for moving mean with a window of 3, welch() for Power spectral density and dwt() for Discrete Wavelet Transform. The code file with filename Task2.m has been attached with the report submitted.

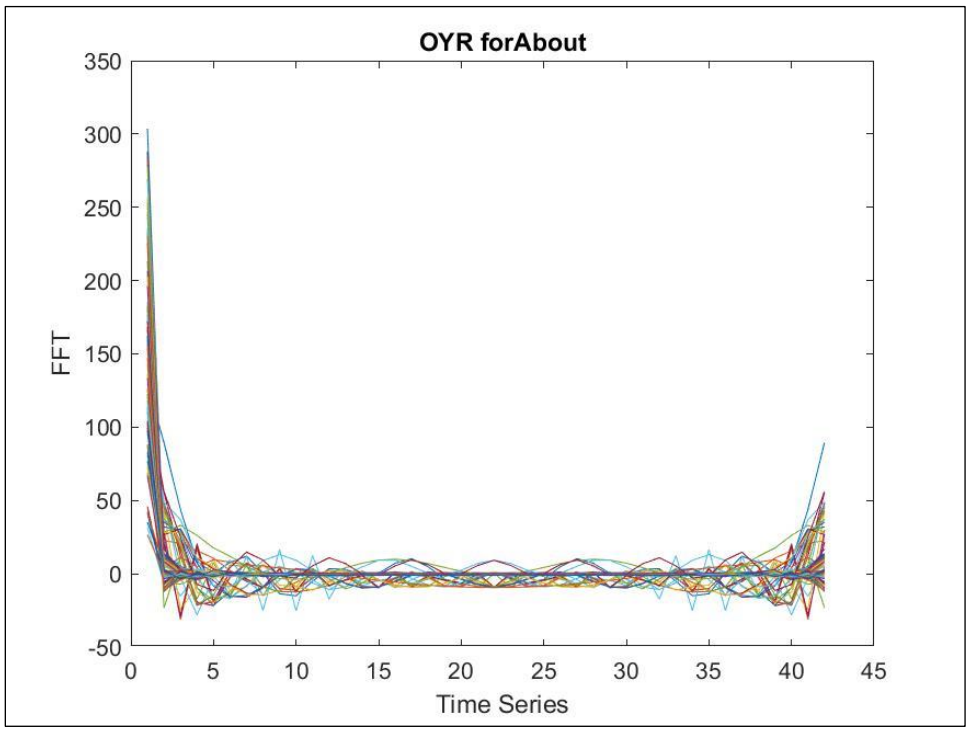
### 4.1.1. ABOUT

b) In About gesture, the left hand remains in constant position and the right hand rotates around the left hand. Hence left hand should not have any variation while right hand should produce more variations. As per our intuition Orientation around Right Hand(OYR) sensor, should produce significant difference when compared with other features.

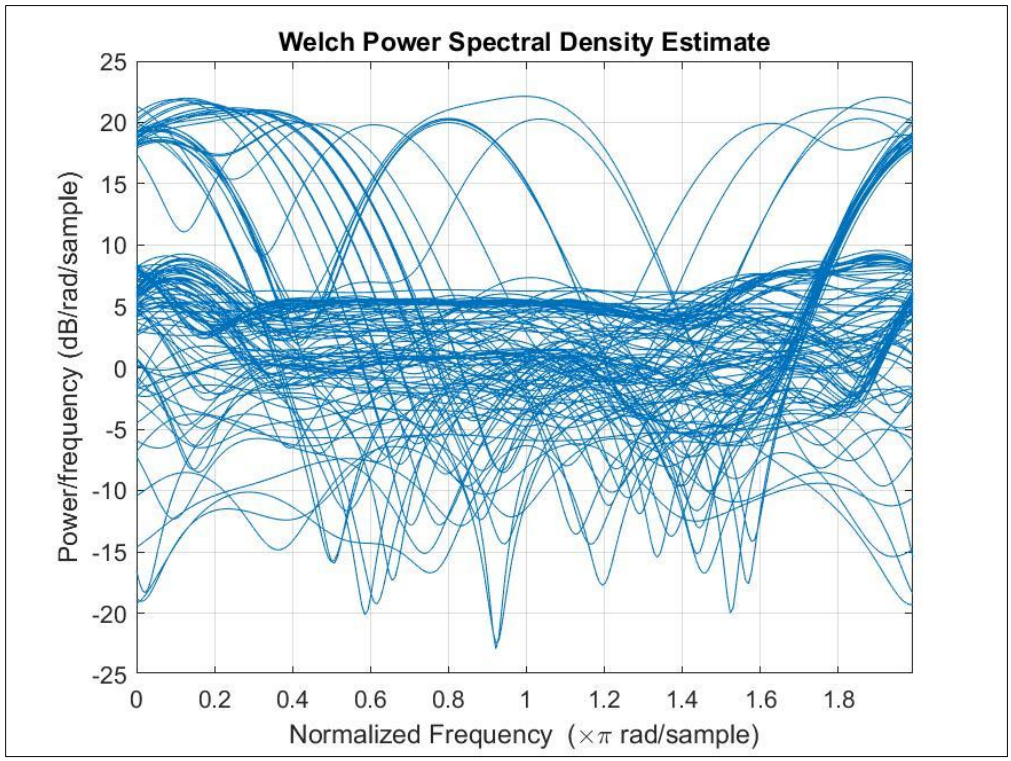
d) Below is graph plotted for Discrete Wavelet Transform against time series:



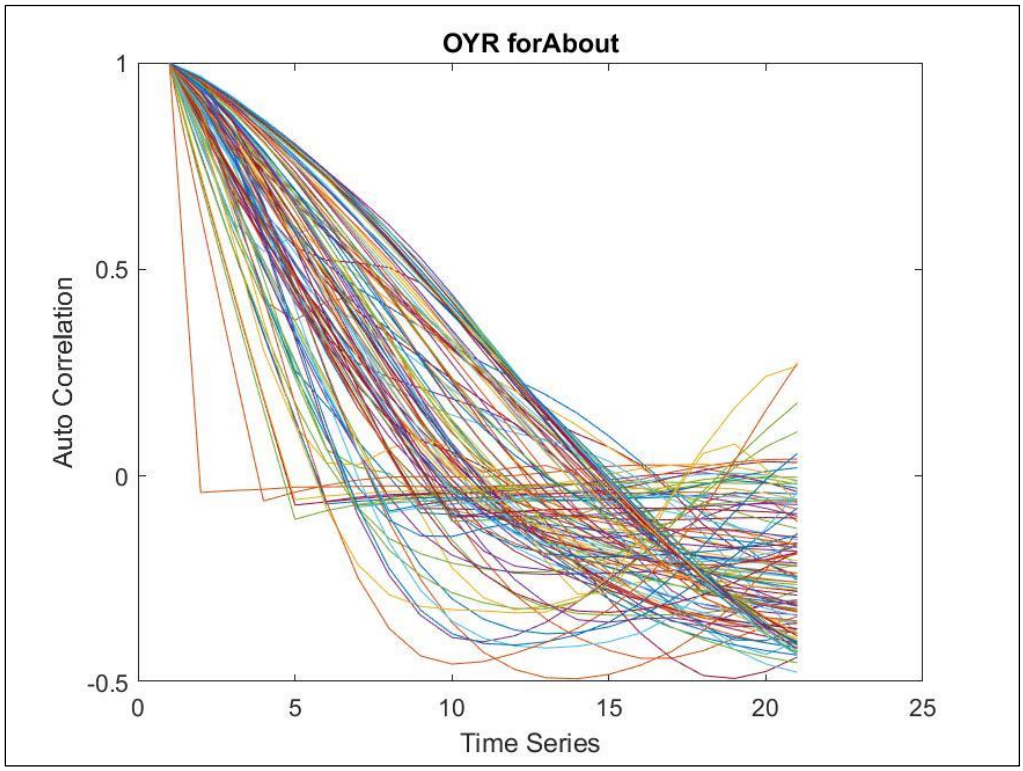

Below is graph plotted for Fast Fourier Transform against time series:



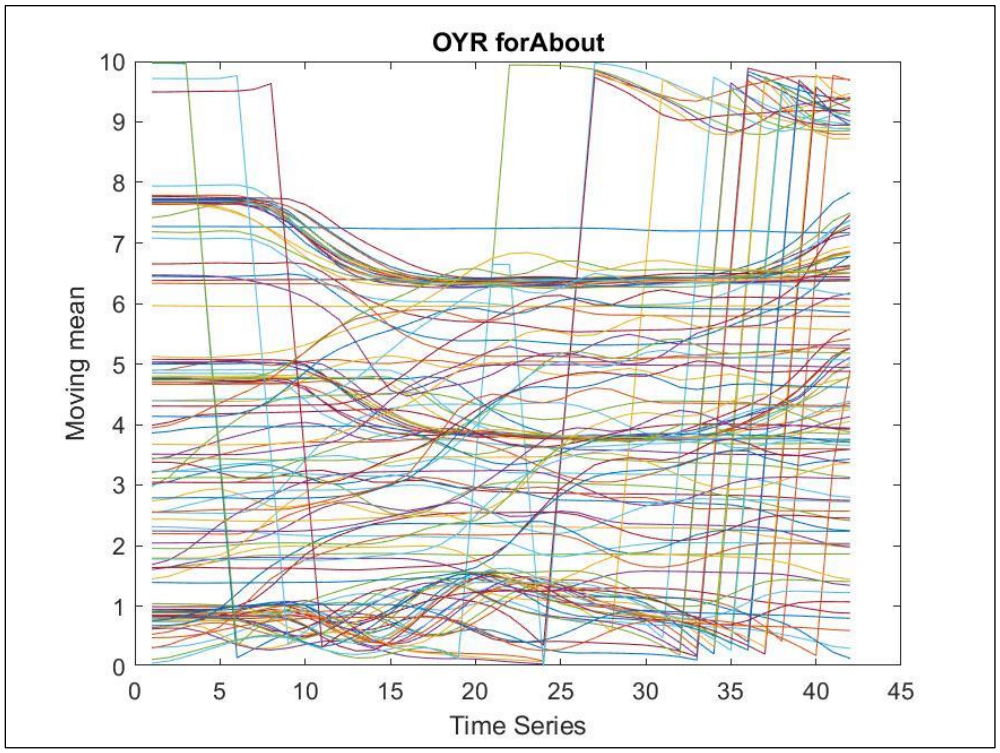
Below is graph plotted for Power Spectral Density against time series




Below is graph plotted for Power Spectral Density against time series:



Below is graph plotted for Moving Mean against time series:

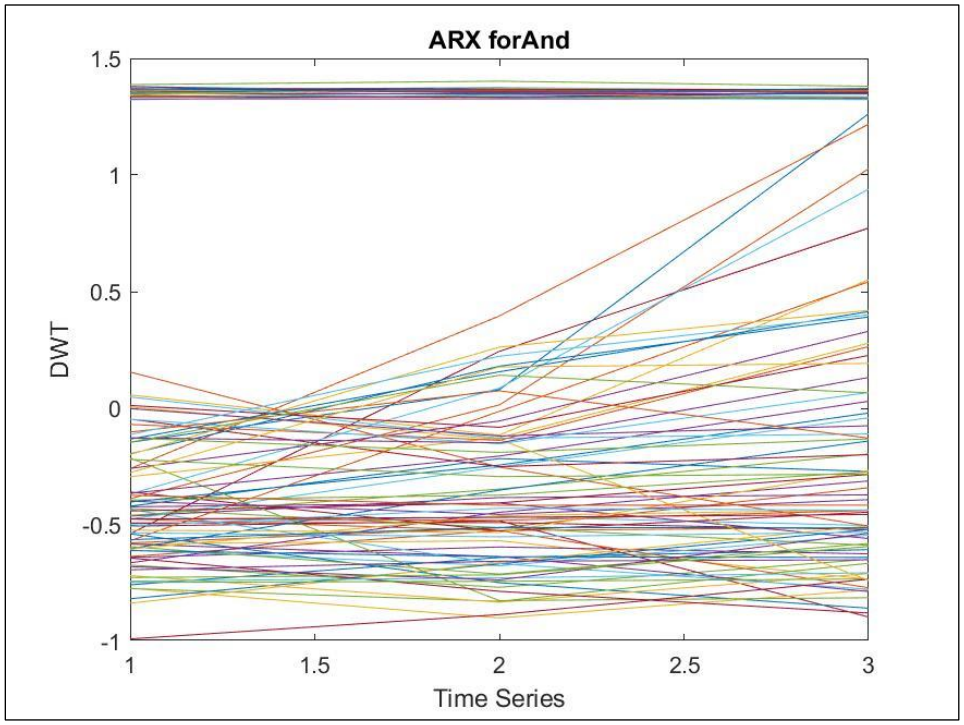



e) Now, Our Initial Intuition holds true based on the features extracted using different techniques. The waves in the graph clearly indicate a high variation across OYR sensor when each extraction method is plotted against time series.

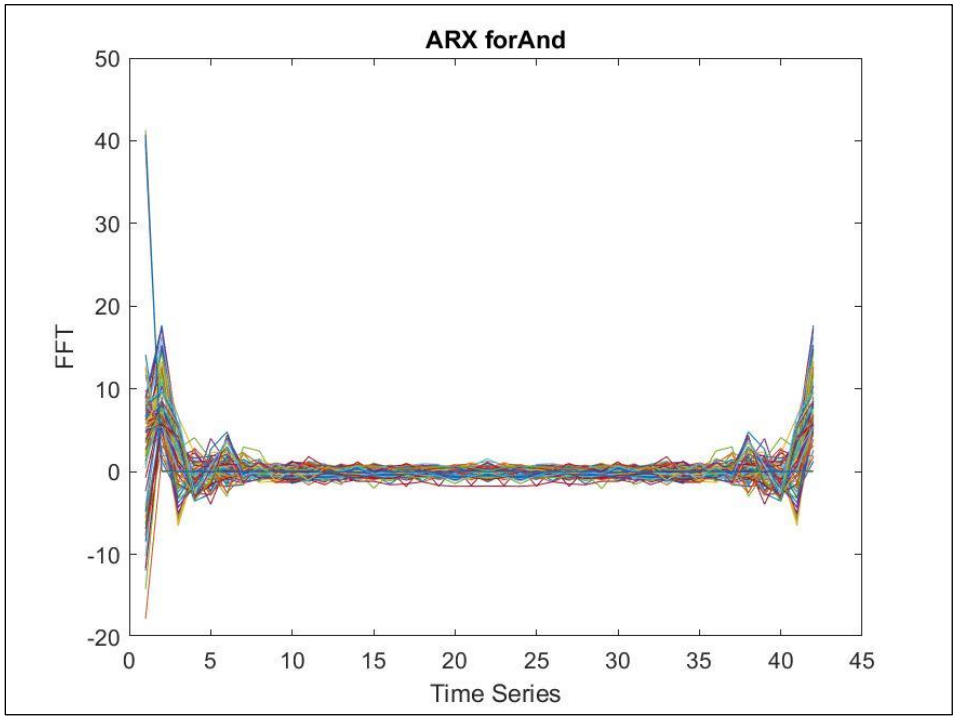
### 4.1.2. AND

b) In AND gesture, it only involves movement of right hand along X axis and left hand remains constant so there should not be in any variation. As there is change in acceleration of right hand along X axis that corresponding sensor should record the variations.

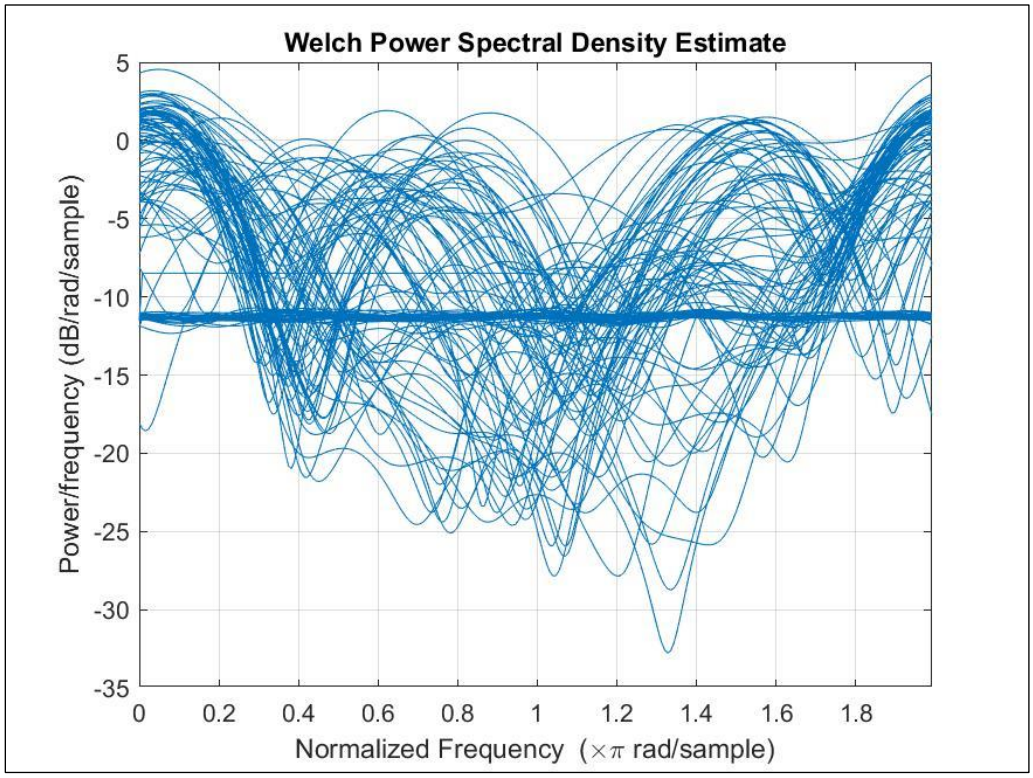
d) Below is graph plotted for Discrete Wavelet Transform against time series:



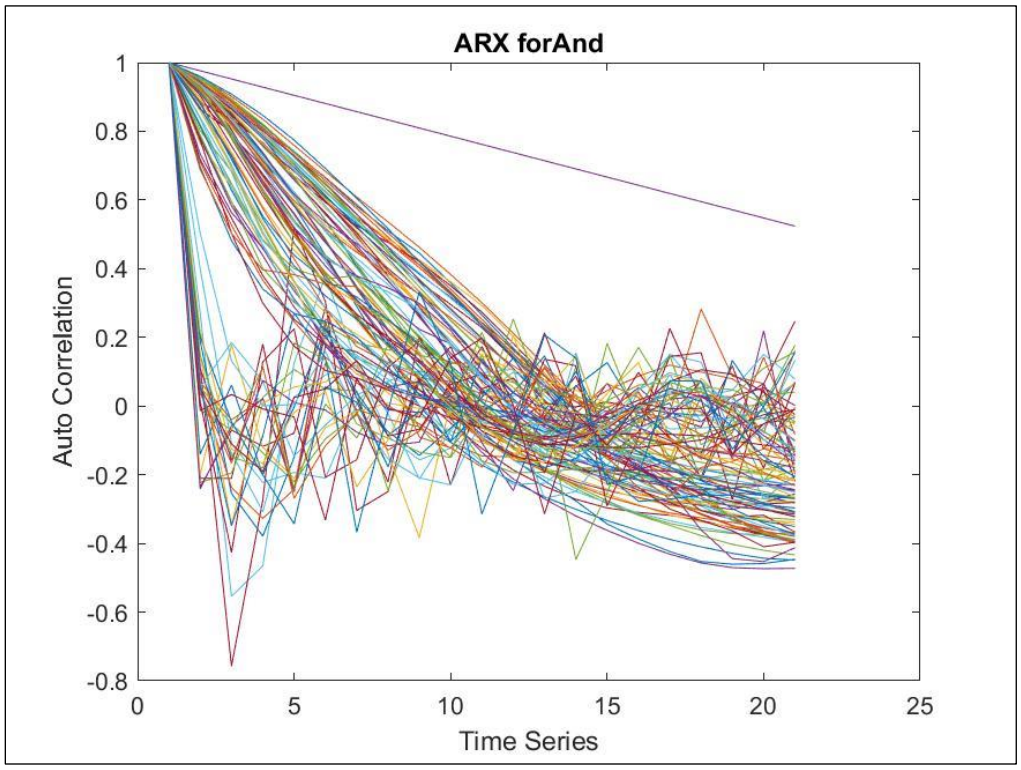

Below are graphs plotted for Fast Fourier Transform against time series:



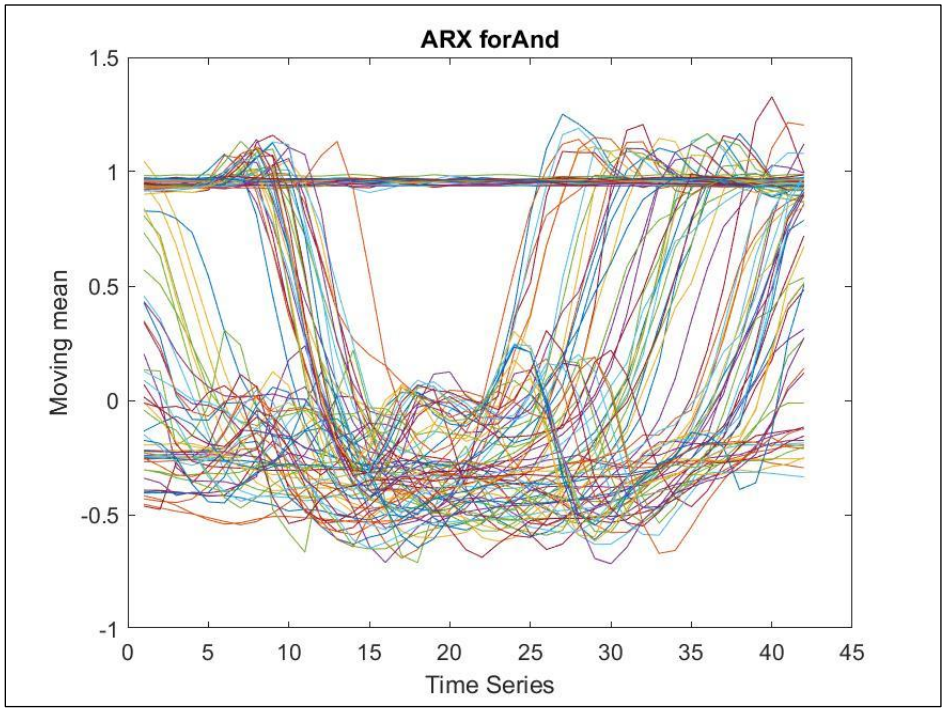
Below are graphs plotted for Power Spectral Density against time series:



Below are graphs plotted for Auto correlation against time series:




Below are graphs plotted for Moving Mean against time series:

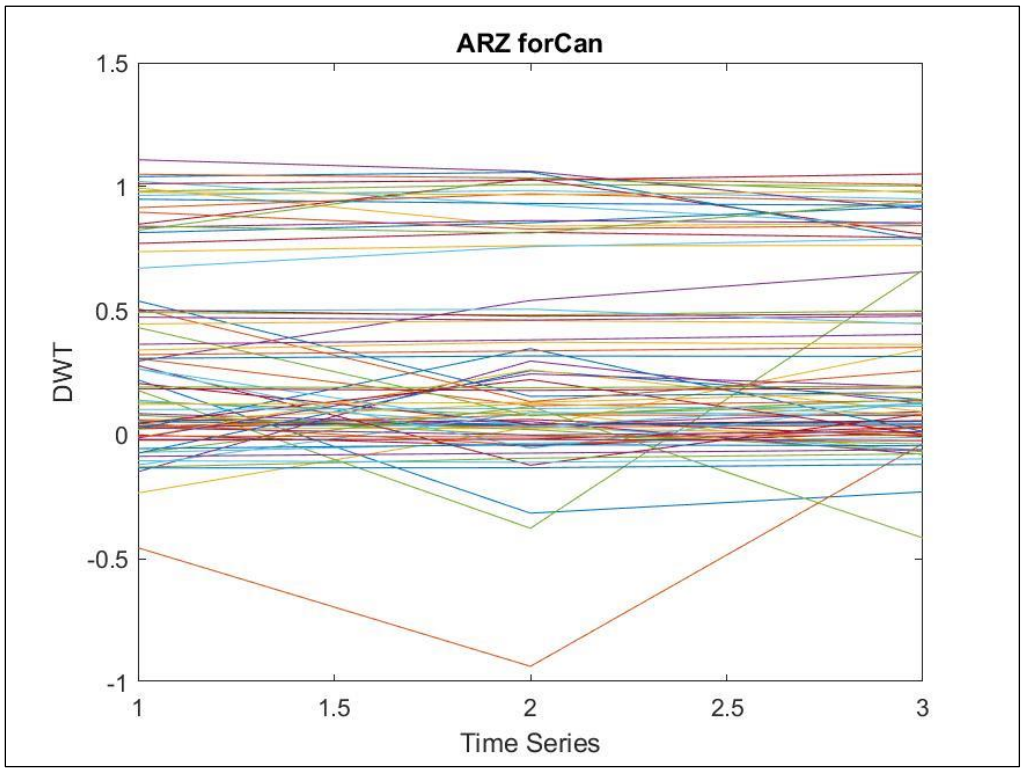


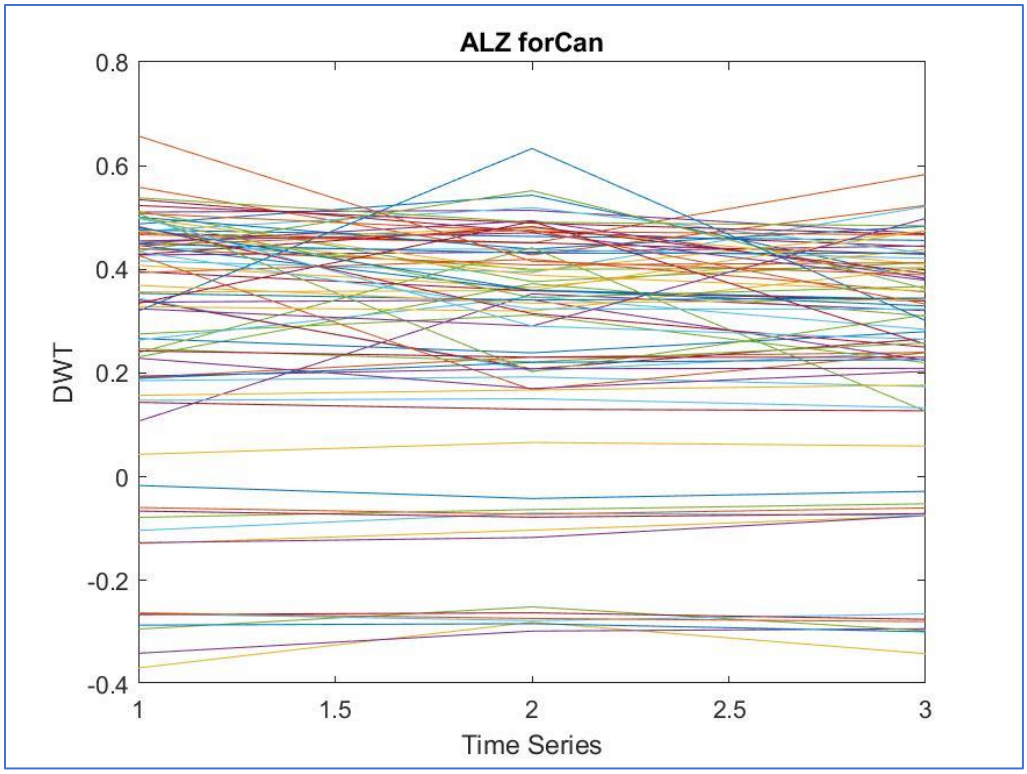
e) The plots obtained using different feature extraction methods clearly depicts a high variation across accelerometer along X axis and less difference in orientation. Now, initial intuition of variations in ARX Sensor for AND holds true.

### 4.1.3. CAN:

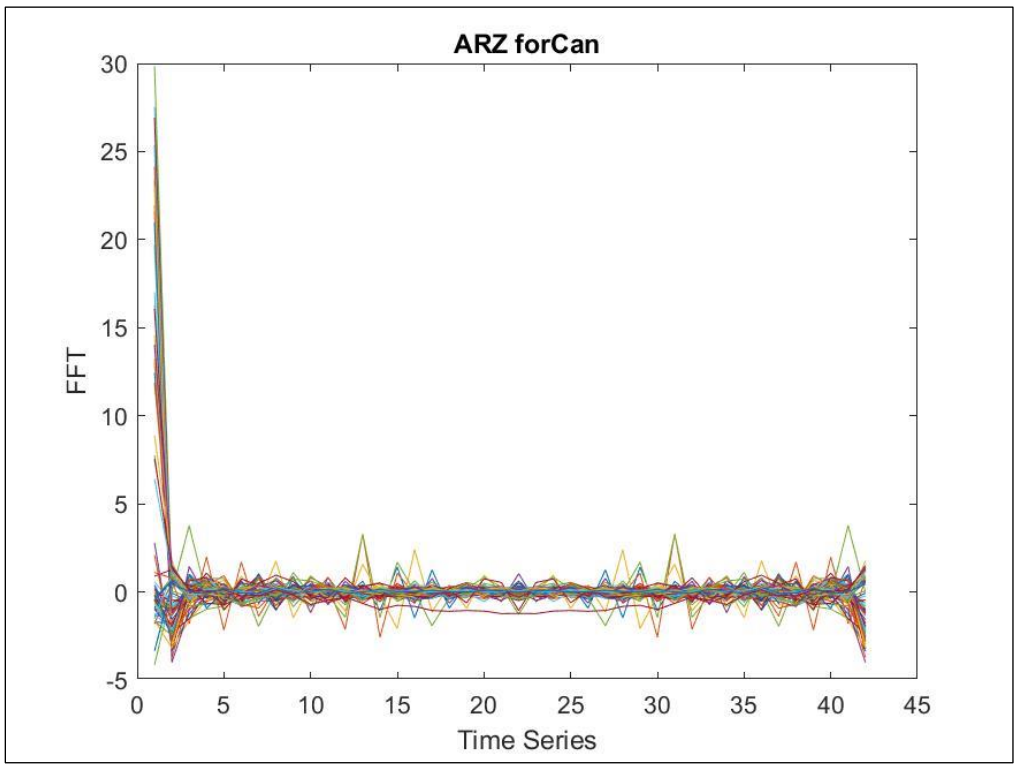
b) The Can gesture involves movement of two hands lowering together at the same time respectively with some force. These variations are produced along Z- axis. As per our intuition the Accelerometer sensor should produce significant variations along Z-axis and should be able to differentiate between various features.

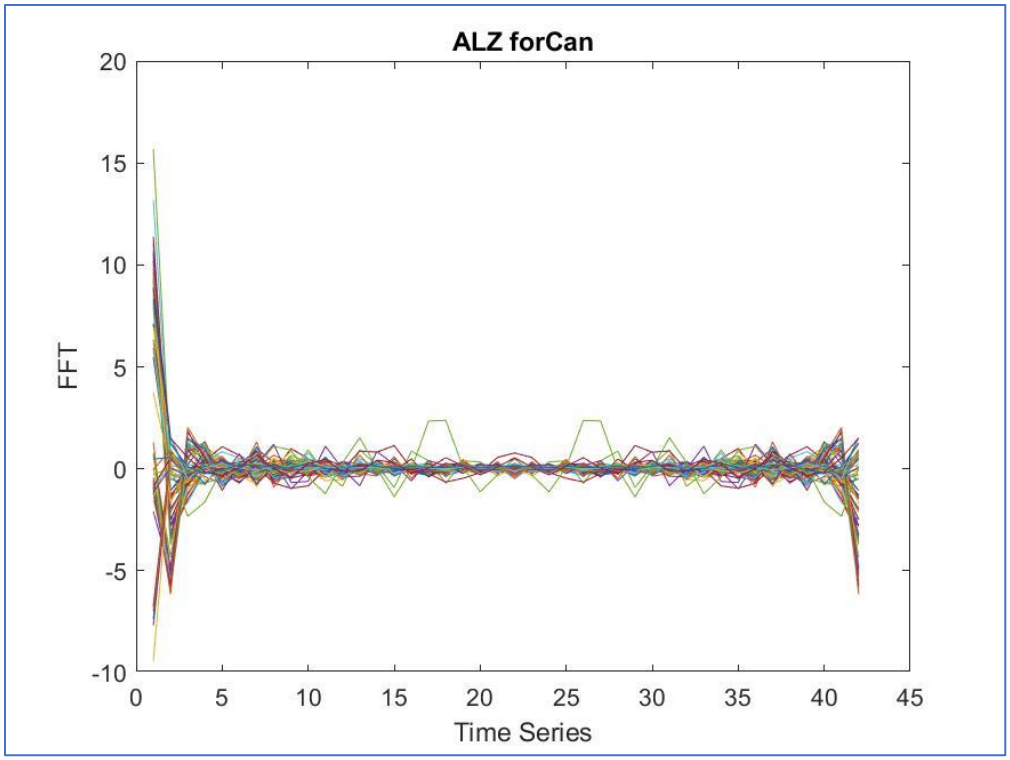
d) Below are graphs plotted for Discrete Wavelet Transform against time series:



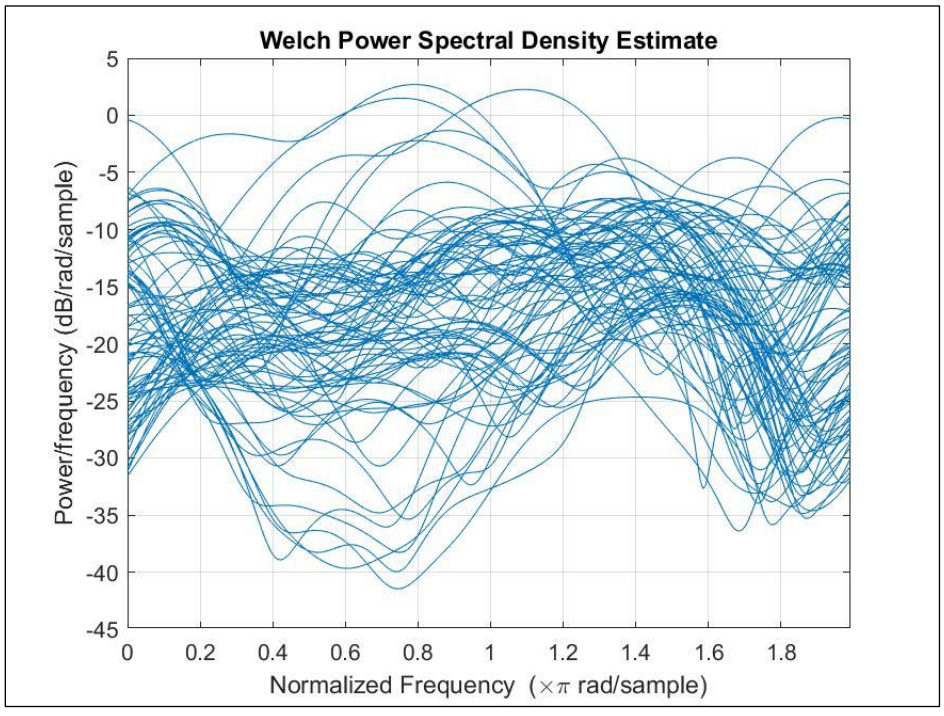


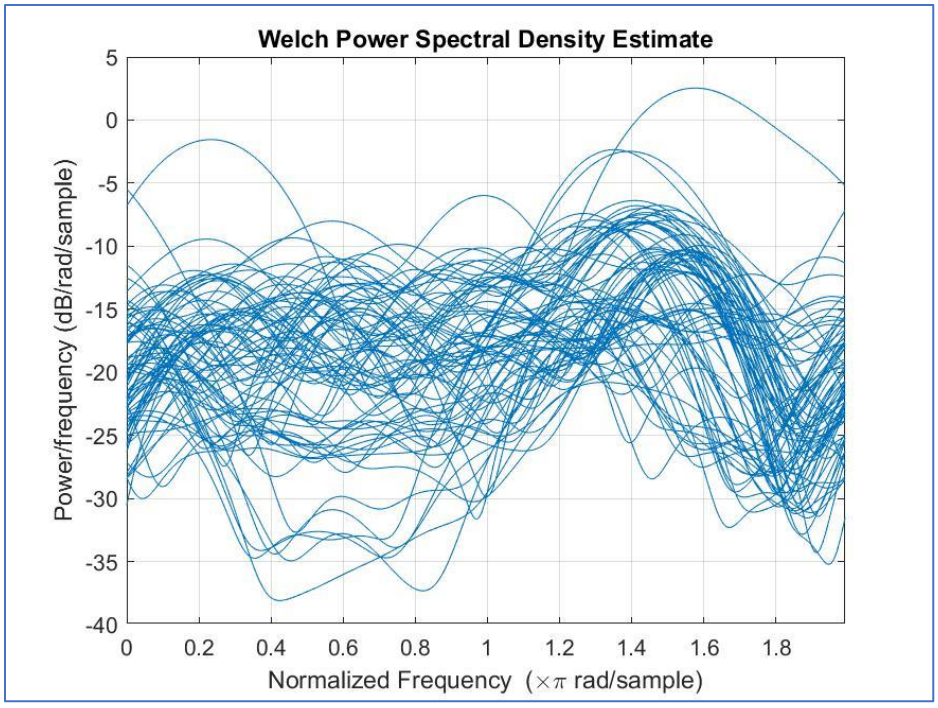

Below are graphs plotted for Fast Fourier Transform against time series:



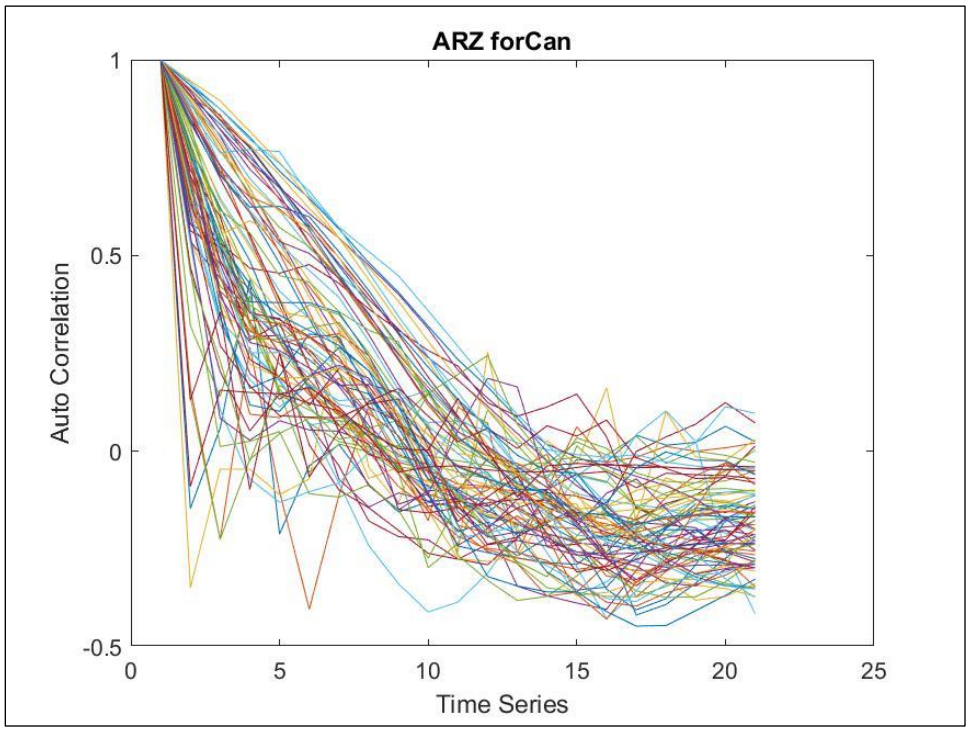


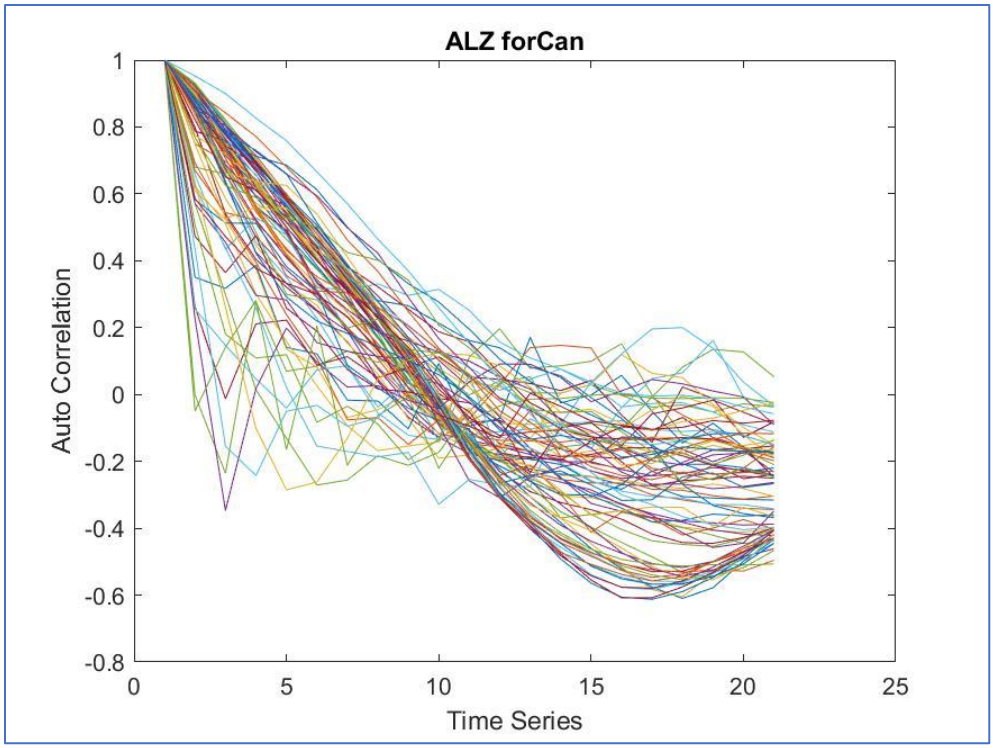

Below are graphs plotted for Power Spectral Density against time series:



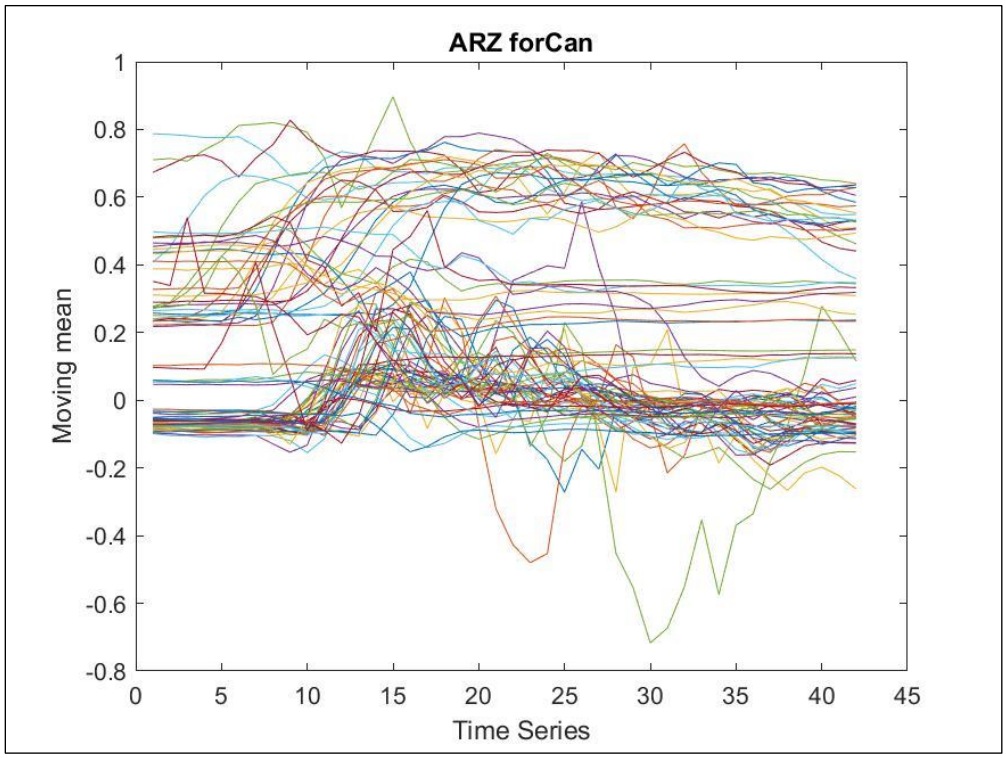


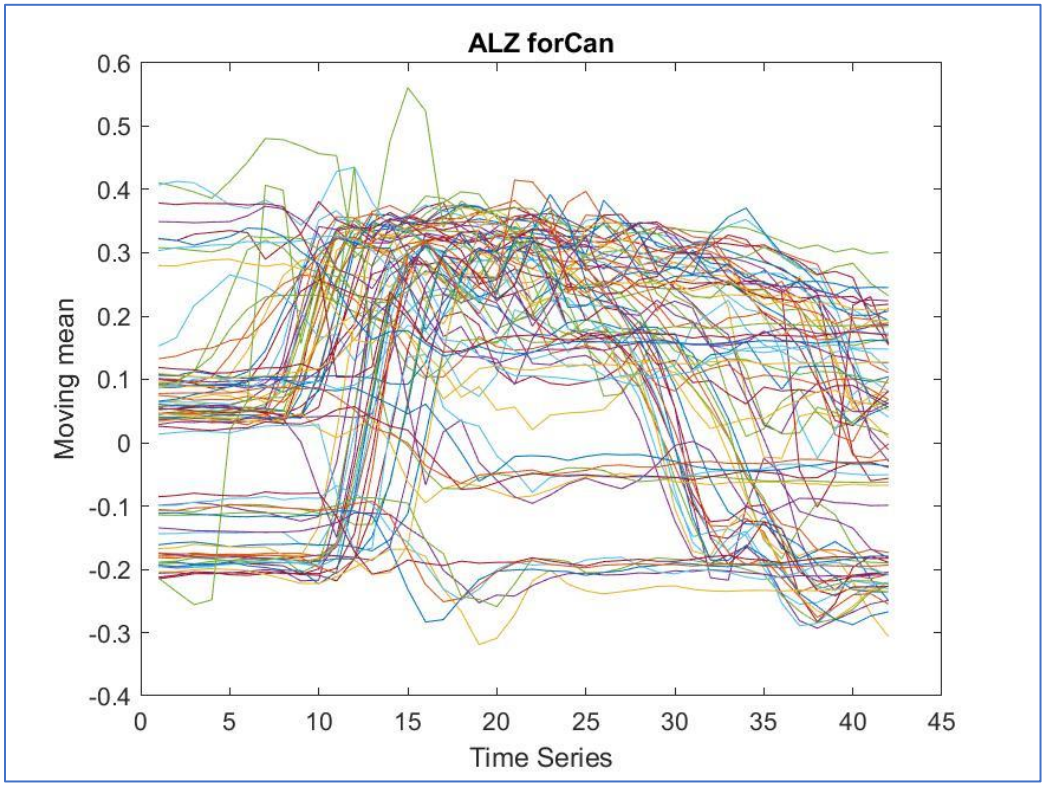

Below are graphs plotted for Auto correlation against time series:




ALZ forCan
Auto Correlation
Time Series
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
5
10
15
20
25

Below are graphs plotted for Moving Mean against time series:



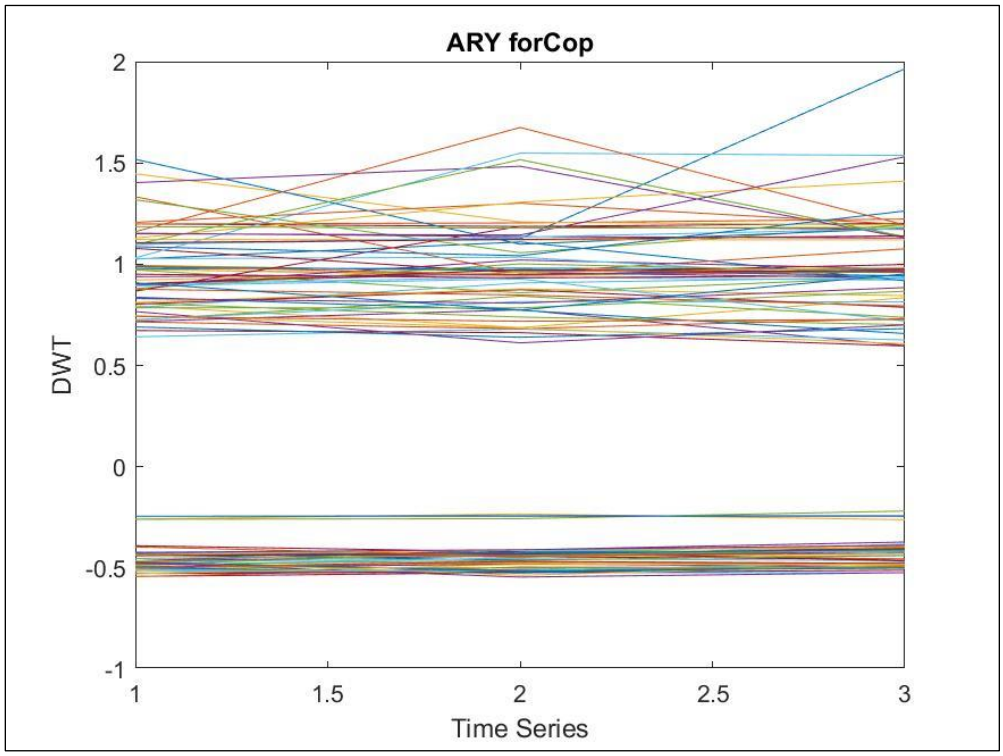



e) The graphs produced from feature extraction reveal that the gesture produces significant variation in accelerometer sensor along Z-axis. Also, the data from other sensors such as gyroscope and EMG produce values that are similar or overlapping and hence does not produce required variation. So, it is safe to assume that our intuition is true based on the graphs obtained.

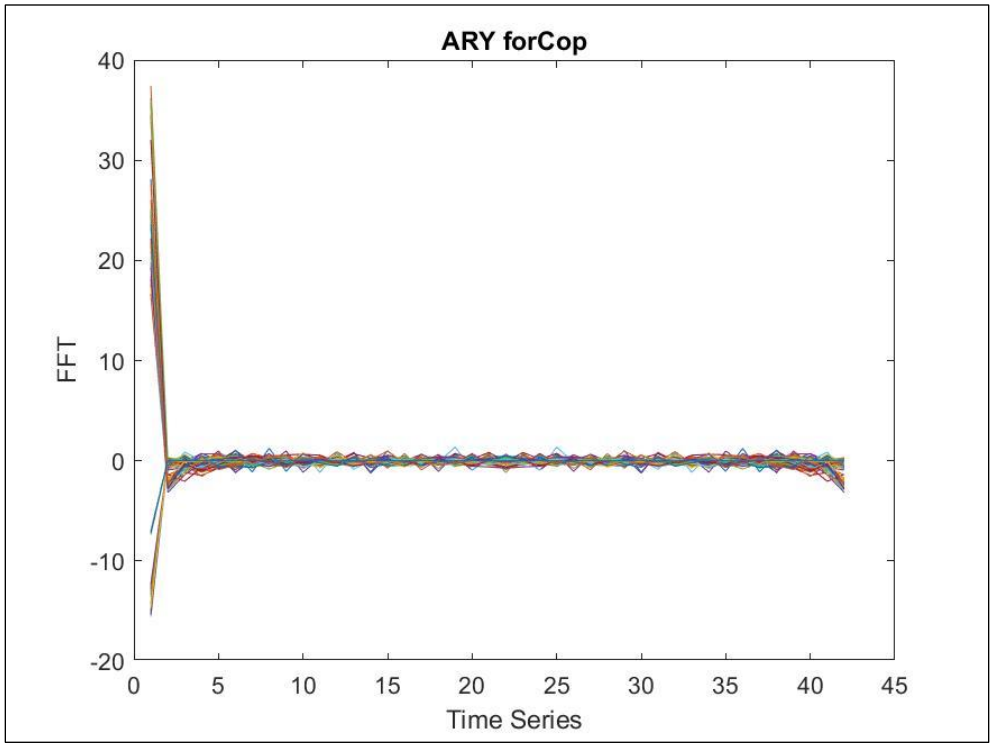
### 4.1.4. COP:

b) COP sign involves slight to and fro movement of right hand above the chest and no movement of left hand. So, variations should be observed in the right hand but not in the left hand. Our intuition is EMG because we could feel a muscular movement of right hand.

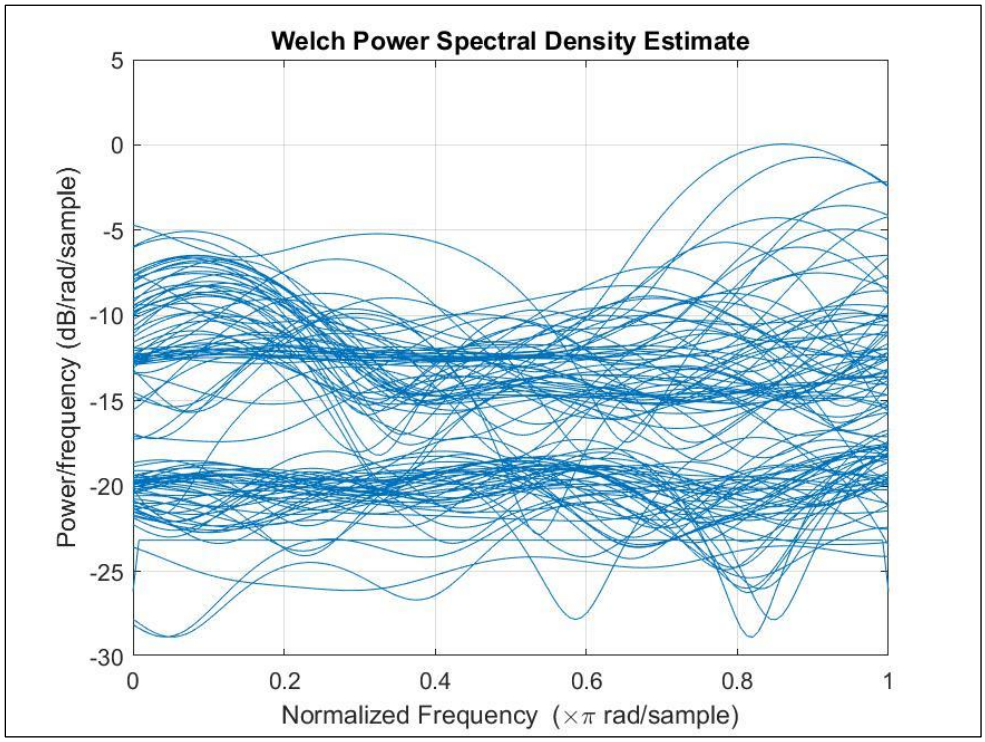
d) Below are graphs plotted for Discrete Wavelet Transform against time series:



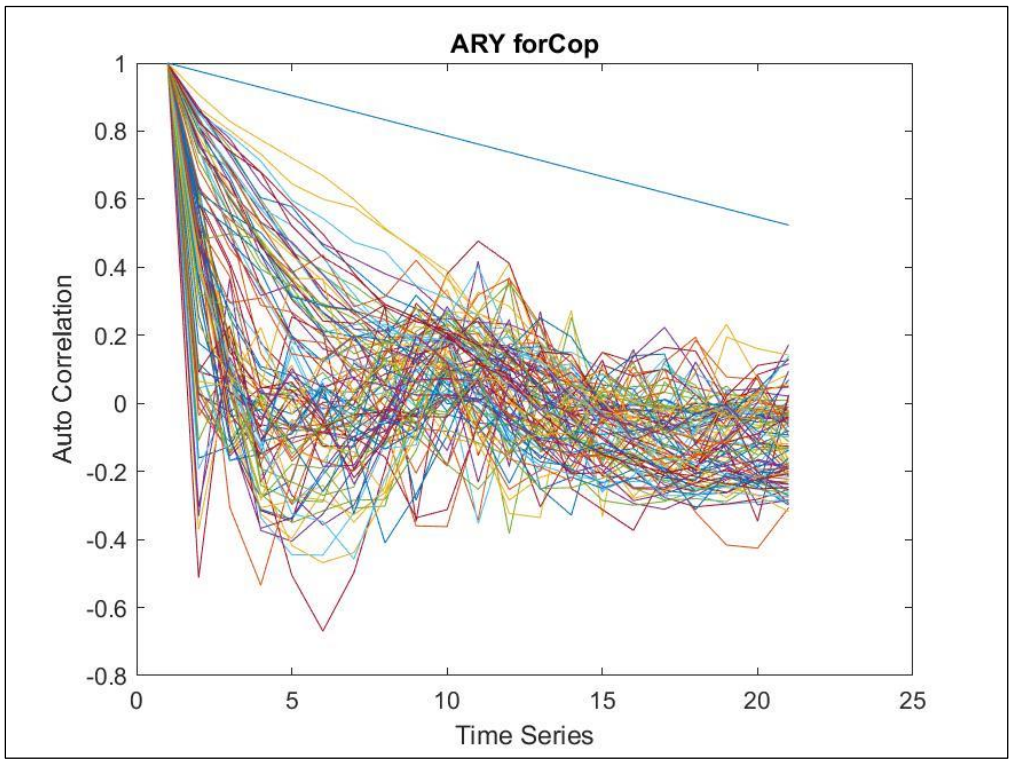
Below are graphs plotted for Fast Fourier Transform against time series:



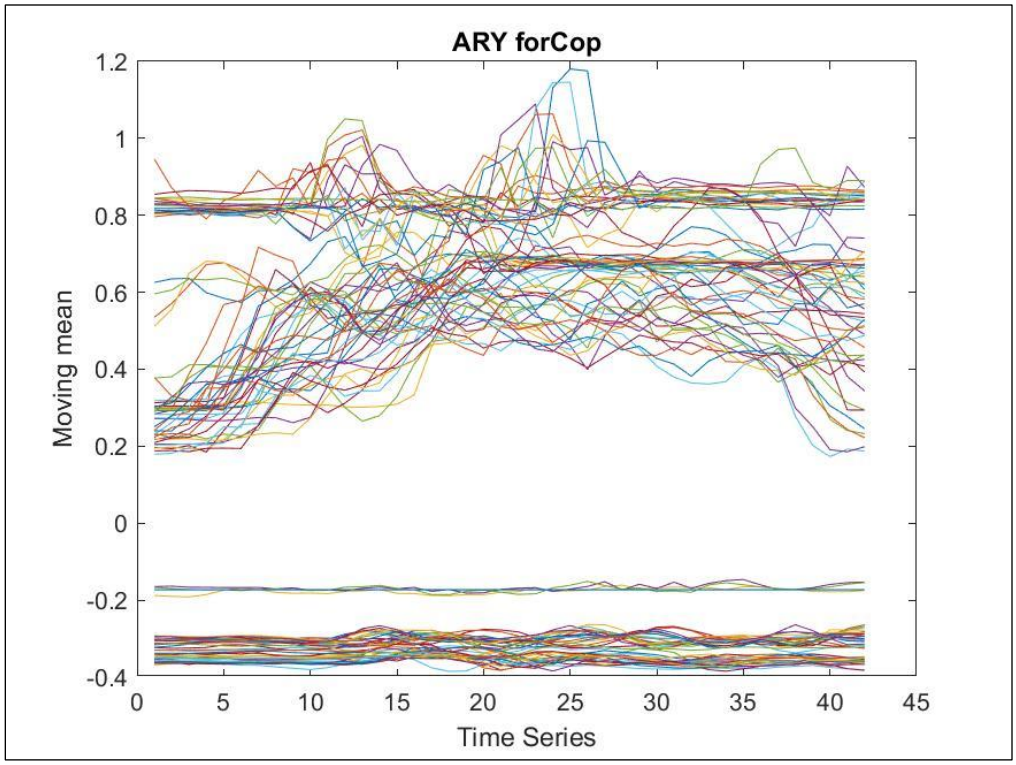
Below are graphs plotted for Power Spectral Density against time series:




Below are graphs plotted for Auto correlation against time series:



Below are graphs plotted for Moving Mean against time series:

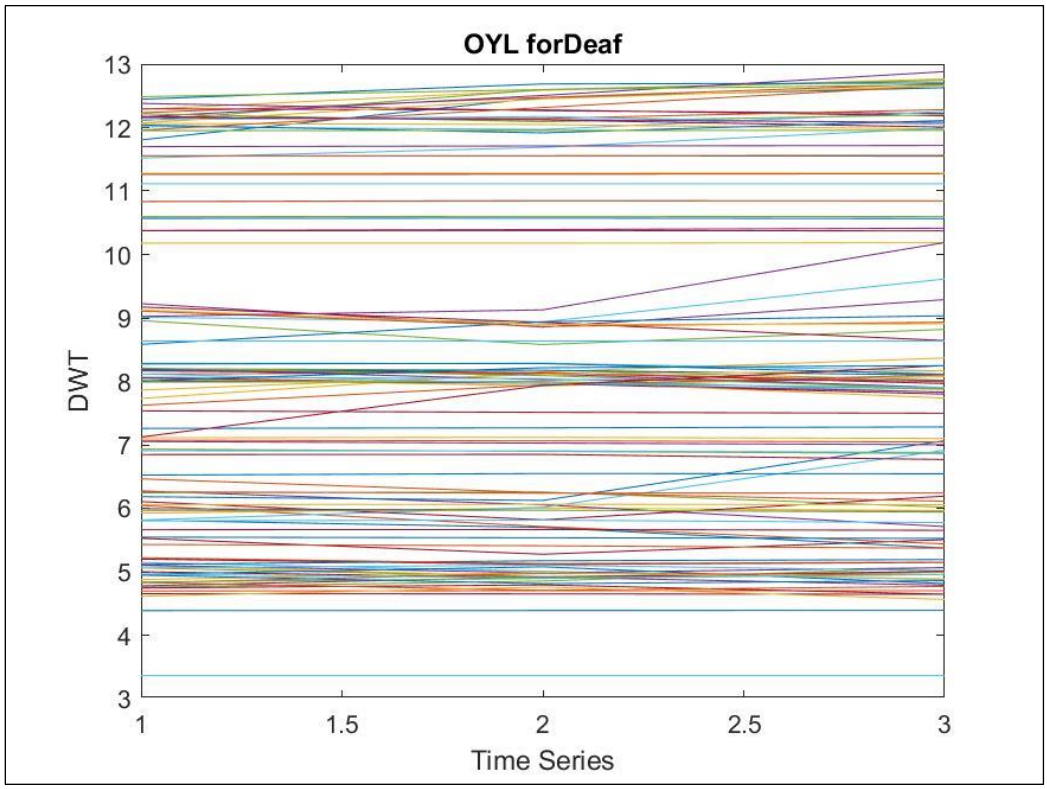



e) Though our initial intuition is EMG but there as per the feature extraction techniques variation is observed in accelerometer of Y direction.

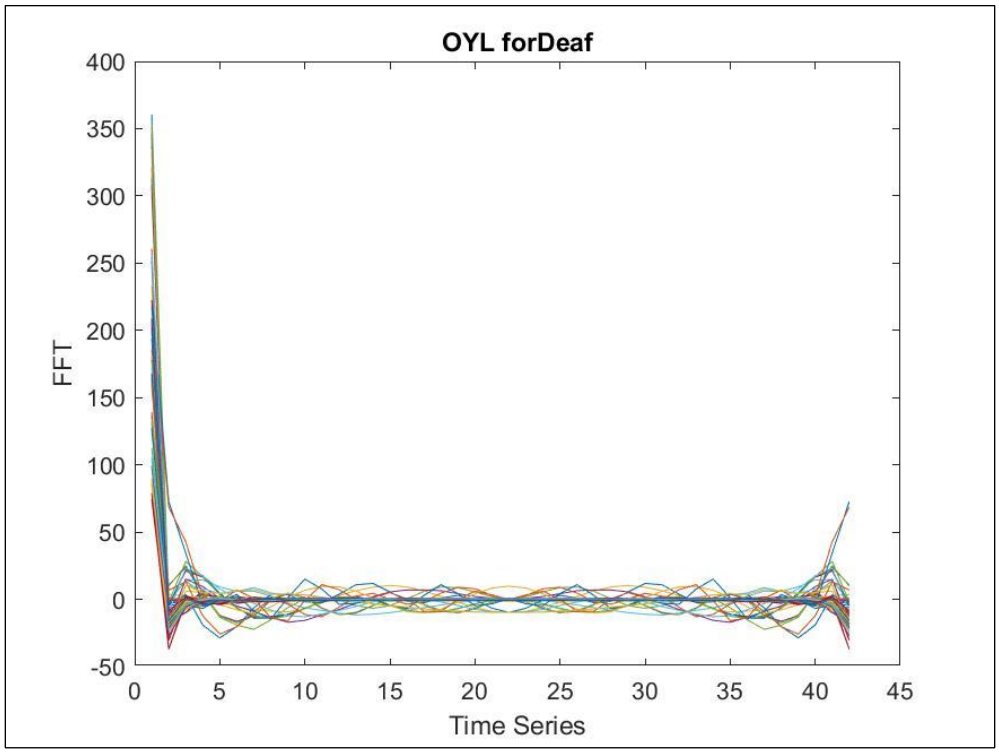
### 4.1.5. DEAF

b) The Deaf gesture involves the to and fro angular movement of right hand around the face from the lips to ears. As per our intuition gyroscope should be able to identify the variations obtained from the angular movement of the right hand. Since only the angle is going to vary in the gesture we think it is safe to assume gyroscope would pick up the difference.

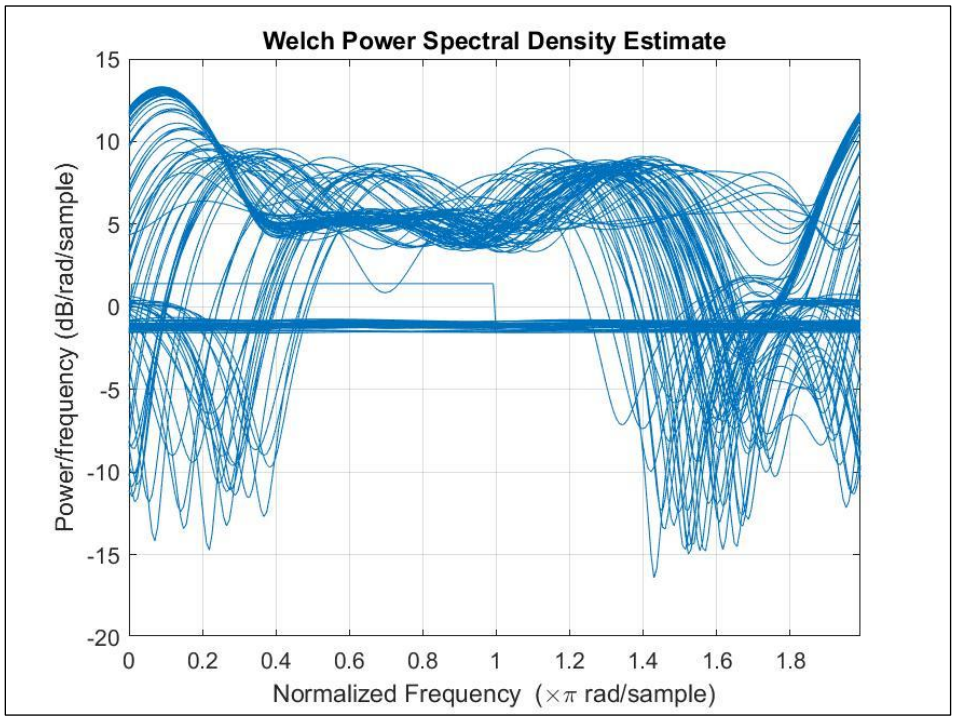
d) Below are graphs plotted for Discrete Wavelet Transform against time series:



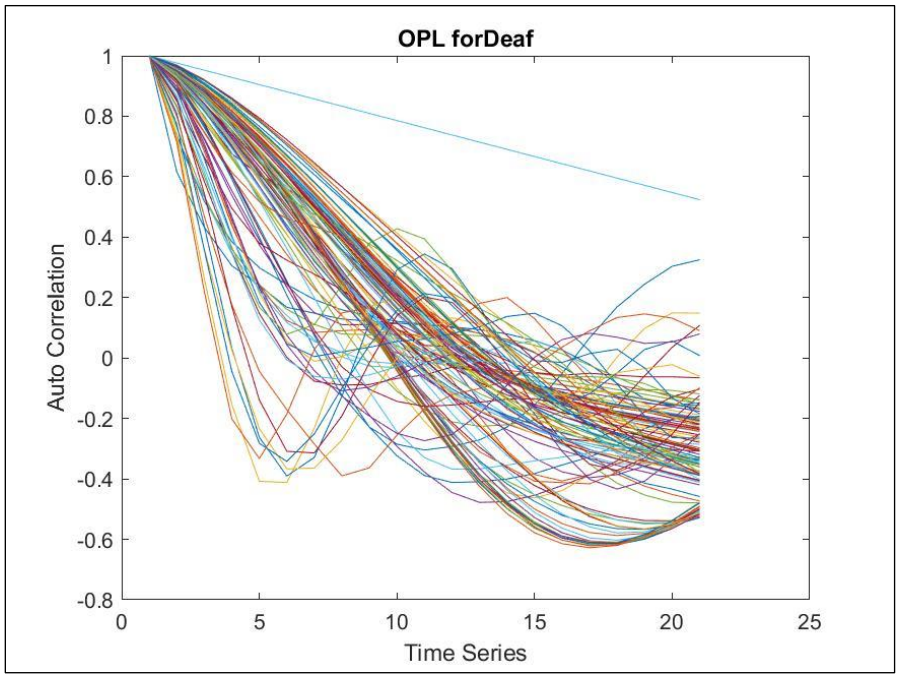
Below are graphs plotted for Fast Fourier Transform against time series:



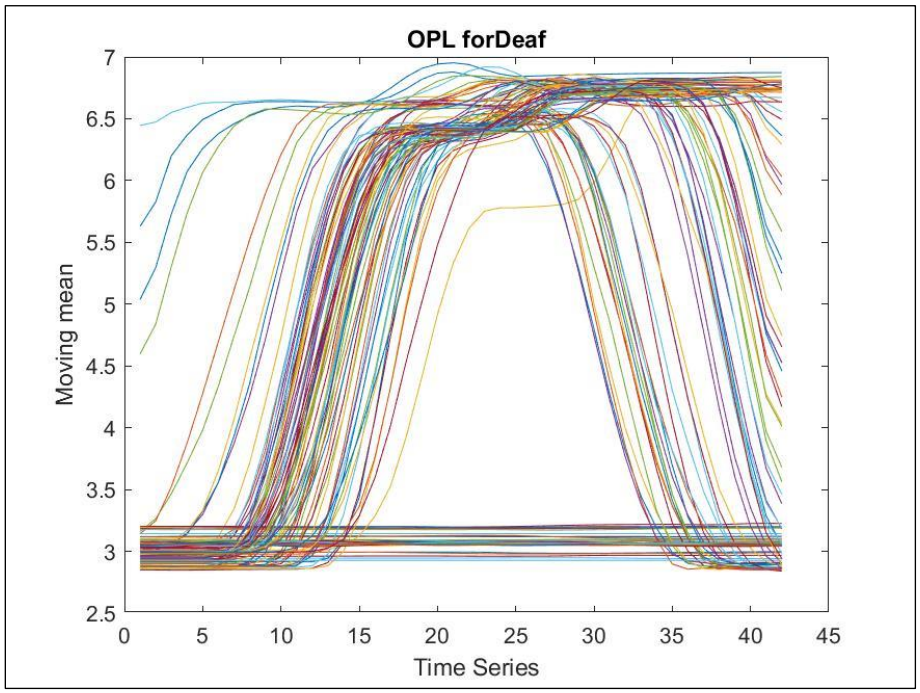

Below are graphs plotted for Power Spectral Density against time series:



Below are graphs plotted for Auto correlation against time series:




Below are graphs plotted for Moving Mean against time series:



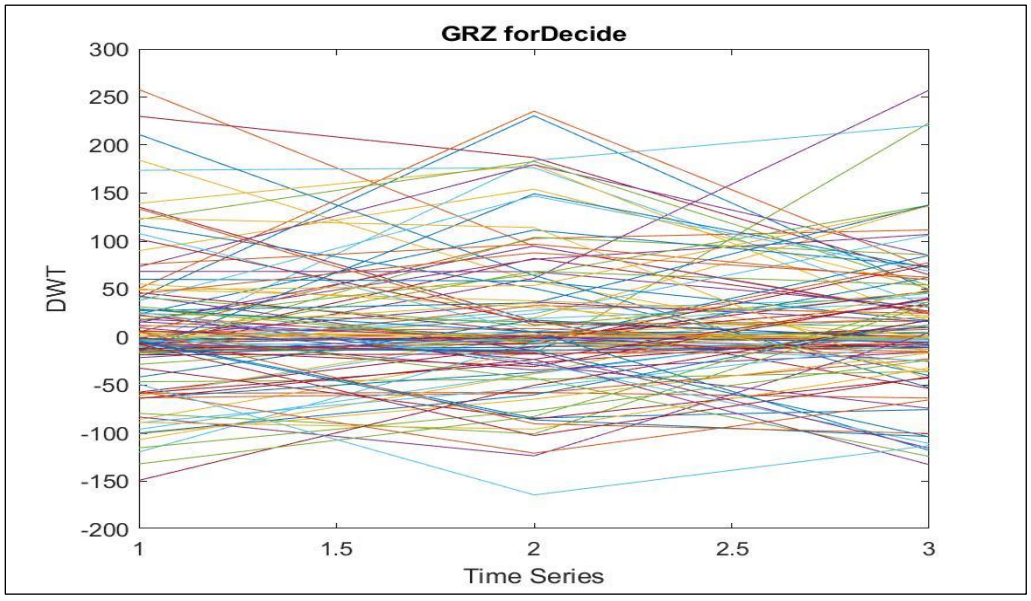
e) Based on the data from the graph our intuition turned out to be partially correct. However, OYR (Orientation Y-axis Right hand) sensor produces significant difference over other sensors like gyroscope, EMG, Accelerometer etc. OYR calculates the change in orientation for the right hand and plots the data in the graph accordingly. Gyroscope, however, measures only the angle change but not the orientation along Y- axis. EMG and Accelerometer do provide a subtle difference in the features, but it wasn't significant when compared with Orientation Sensor.

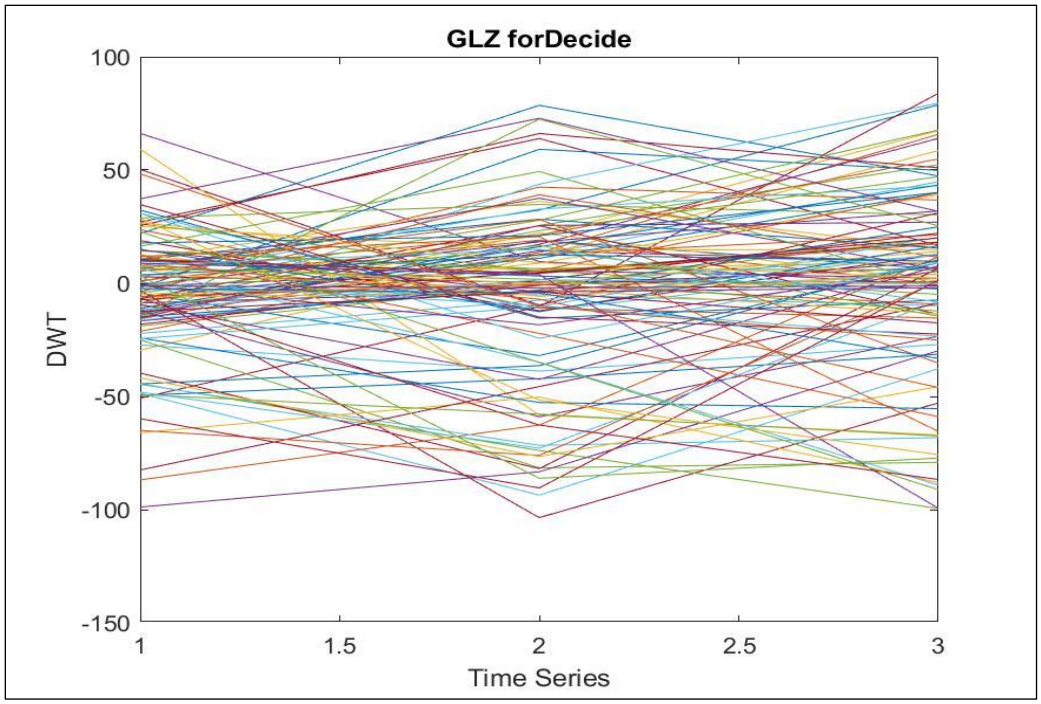
### 4.1.6. DECIDE

b) Our intuition is that, the feature which can be useful in determining the gesture would be gyroscope. As we look at the gesture, we identify that the right has an angular motion and the gesture ends with a jerk on both the hands. So, we can see the change in the angle of the hands leading to change in angular velocity of the hands.

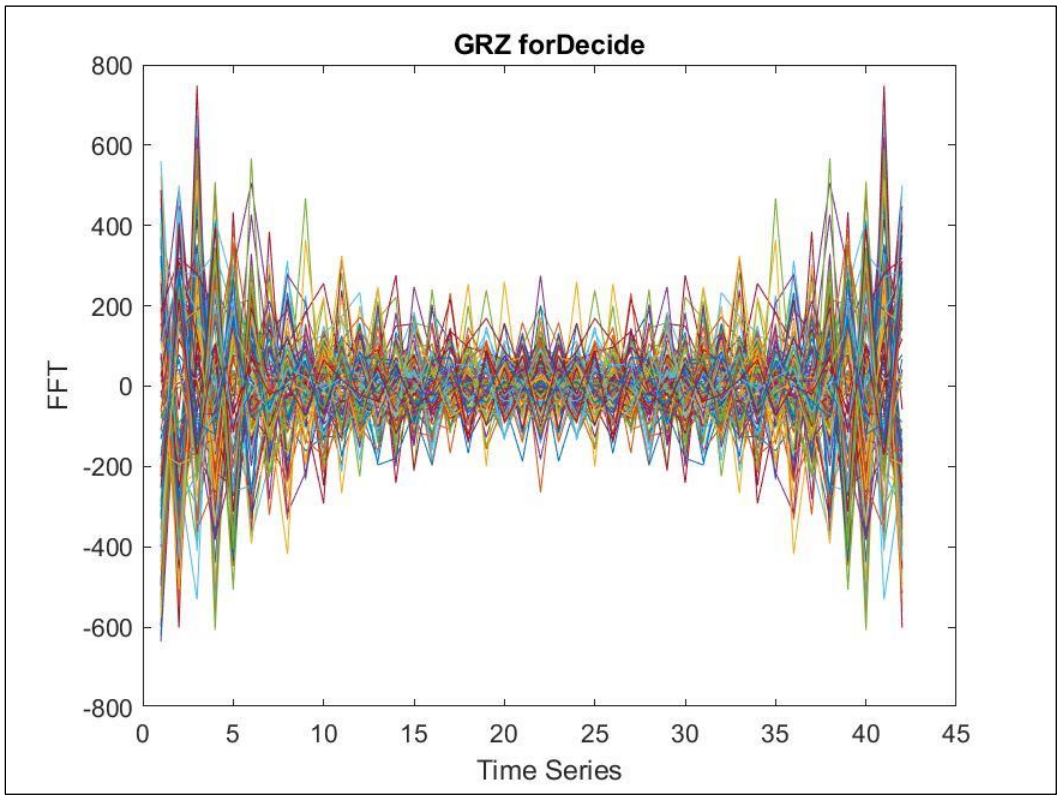
d) The feature with maximum variation for the gesture Decide is Gyroscope of right hand in the direction of Z axis and left hand in the direction of Z axis.

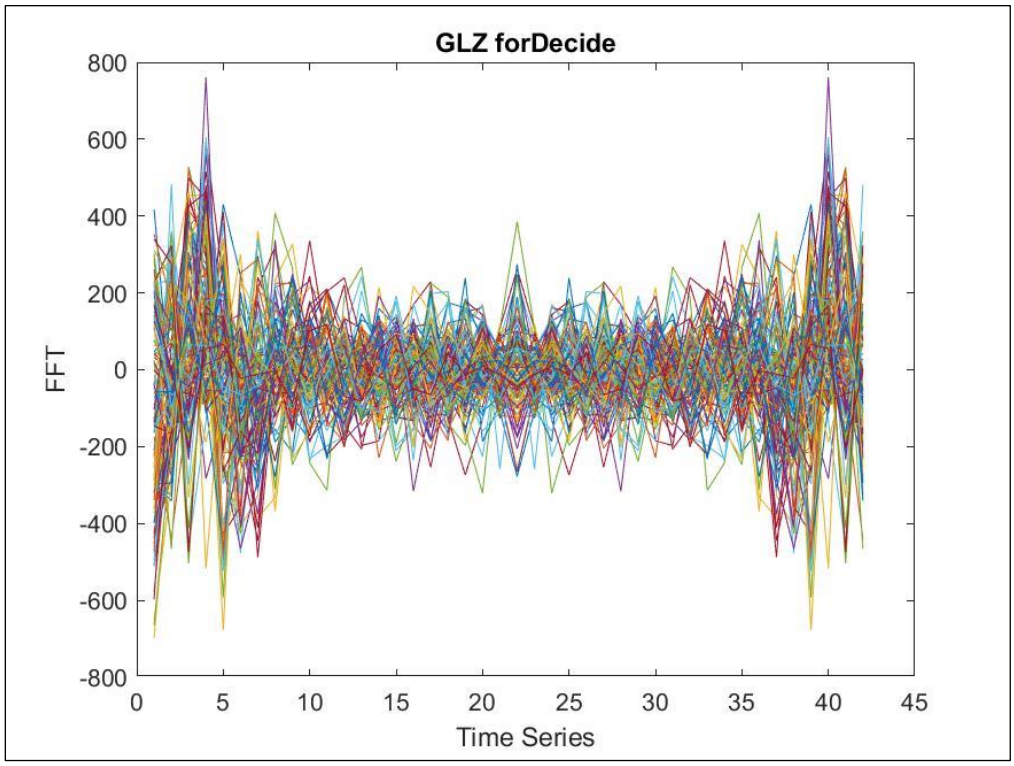
Below are the graphs plotted for DWT against time series:



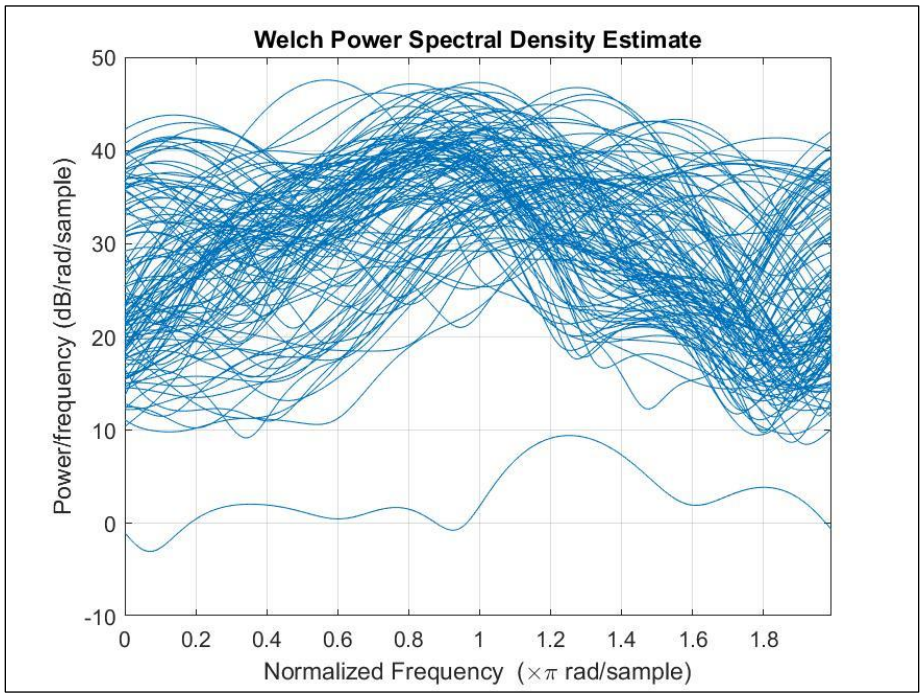


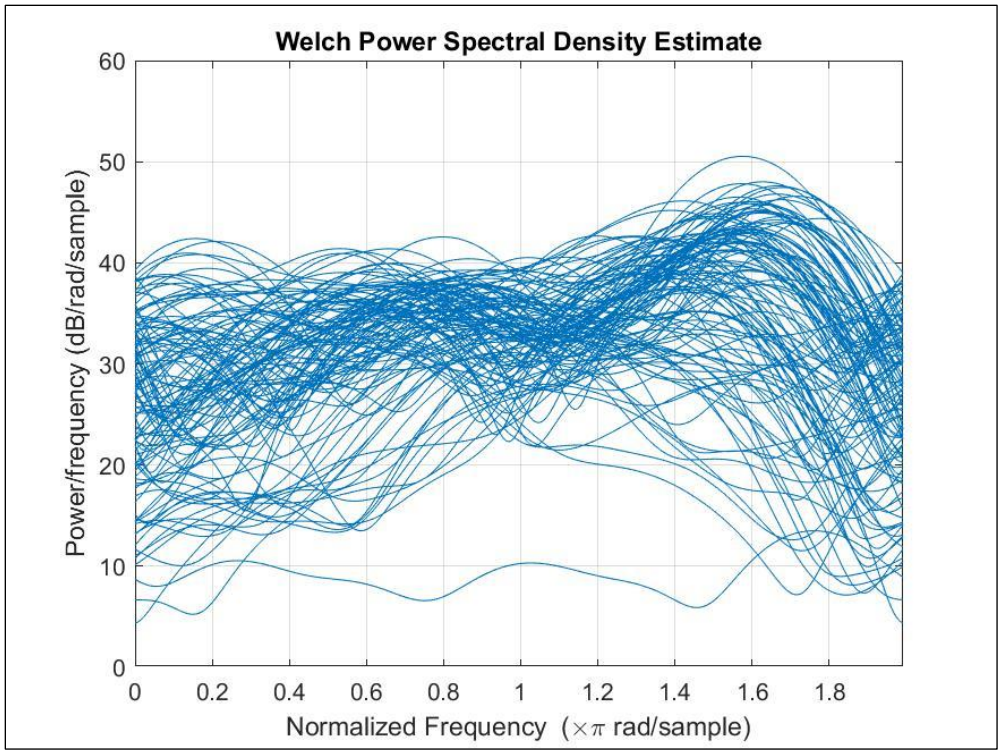

Below are graphs plotted for FFT against time series:



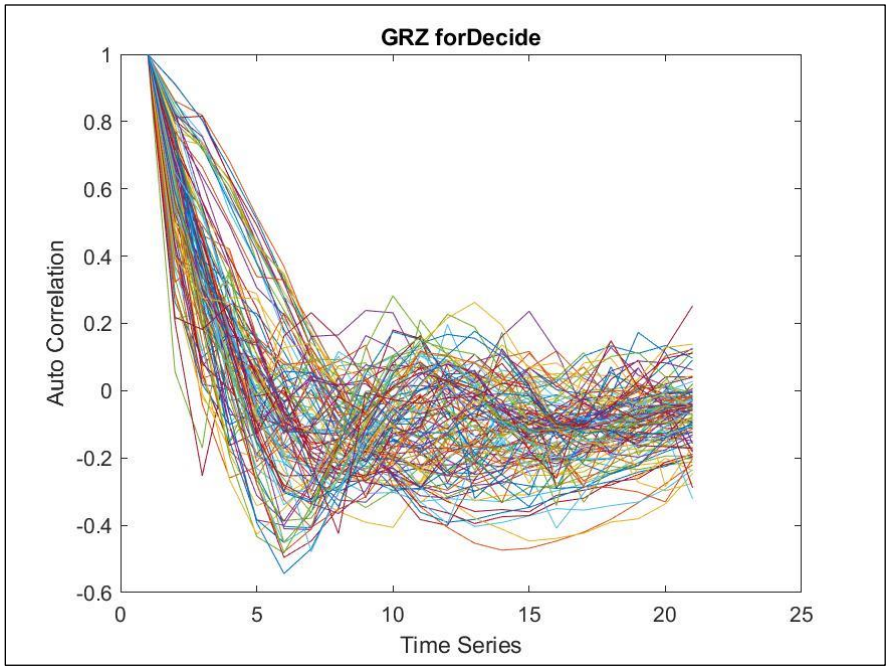


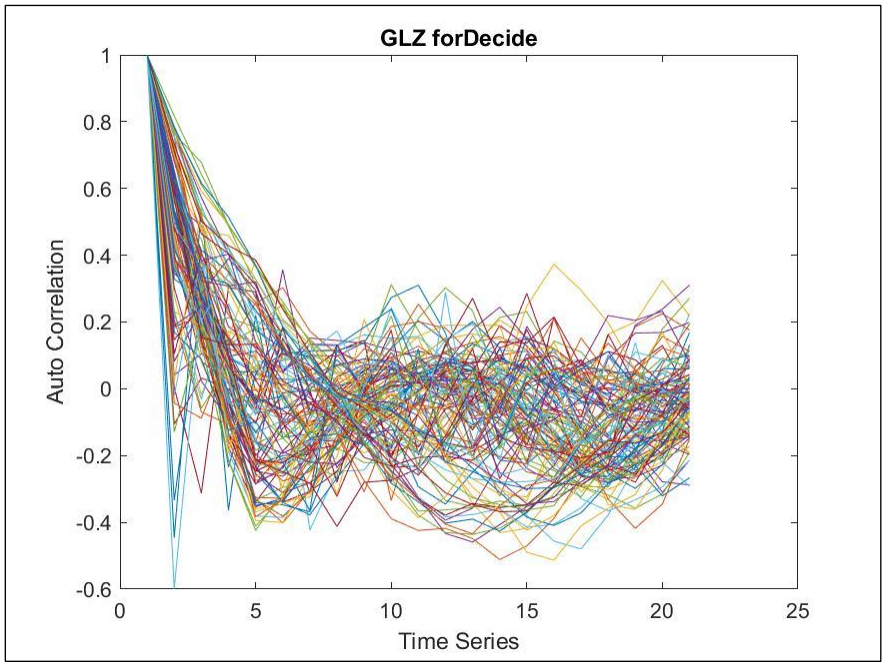

Below are graphs plotted for PSD against time series:



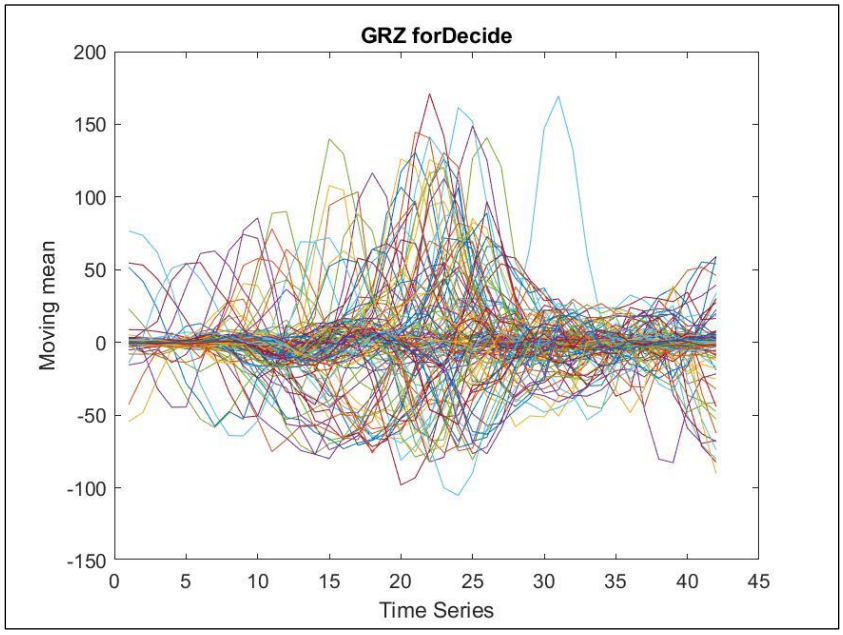

Welch Power Spectral Density Estimate
Power/frequency (dB/rad/sample)
Normalized Frequency (×π rad/sample)
0
10
20
30
40
50
60
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8

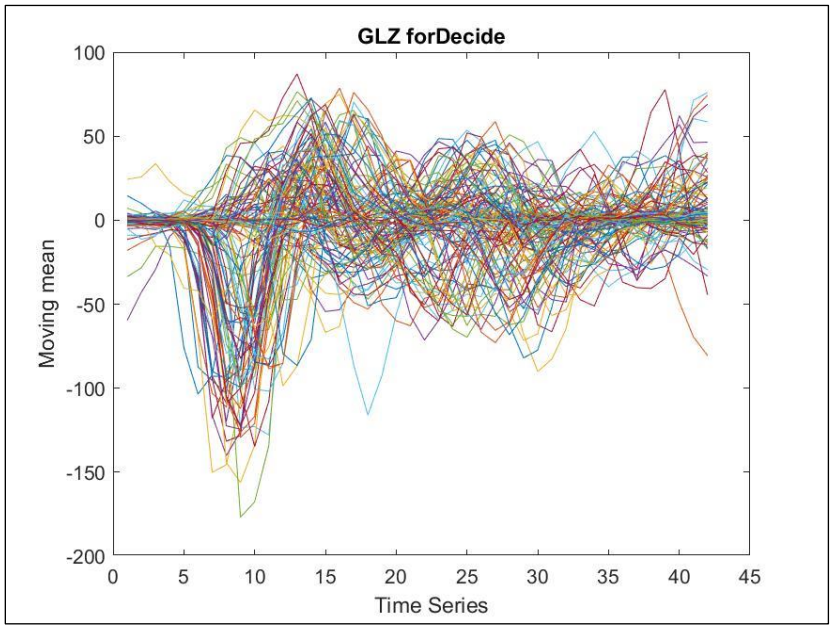
Below are graphs plotted for Auto Correlation against time series:



Below are graphs plotted for Moving Mean against time series:



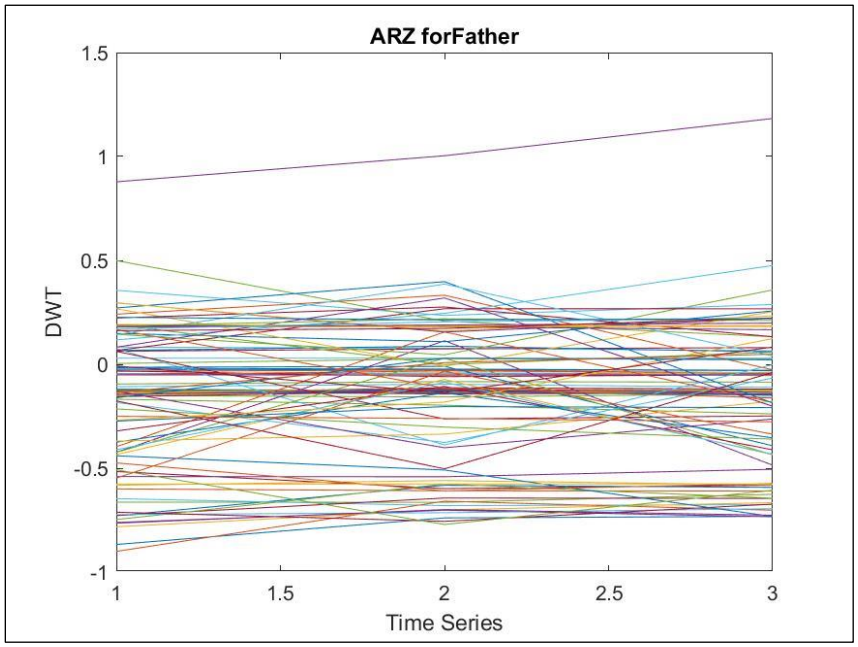

e) Analysing all the graphs plotted for all the features for the gesture “Decide”, we found that the above given graphs have shown the maximum variation i.e. the graphs for Gyroscope along z axis for both right and left hands when compared to the graphs of remaining features. The remaining features have shown overlapping curves which shows, minimum or no variation. So, our initution is mostly correct.

### 4.1.7. FATHER

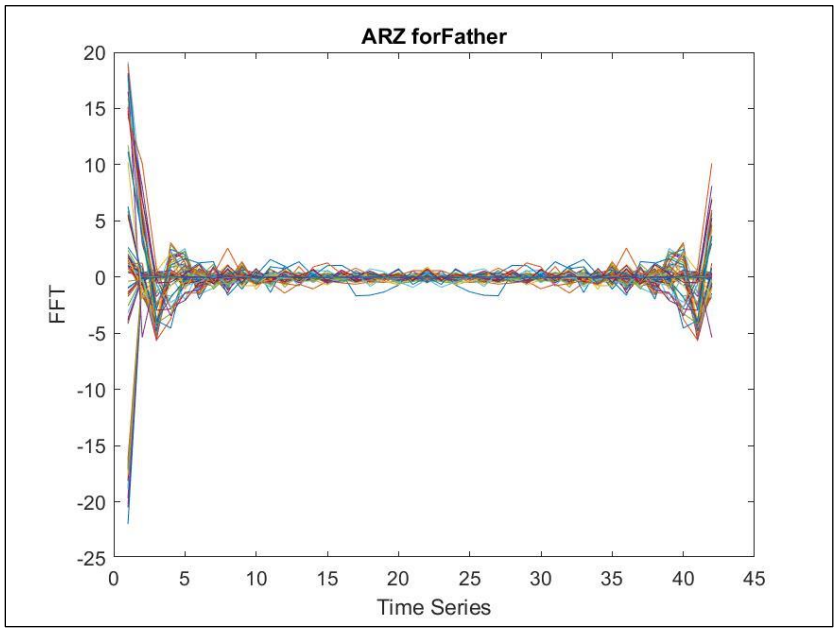
b) Our intiution of the most useful feature for identifying the gesture Father is Accelerometer, as the gesture has only the movement of right hand in a linear pattern.

d) The feature showing maximum variation in the graphs is Accelerometer along z axis.

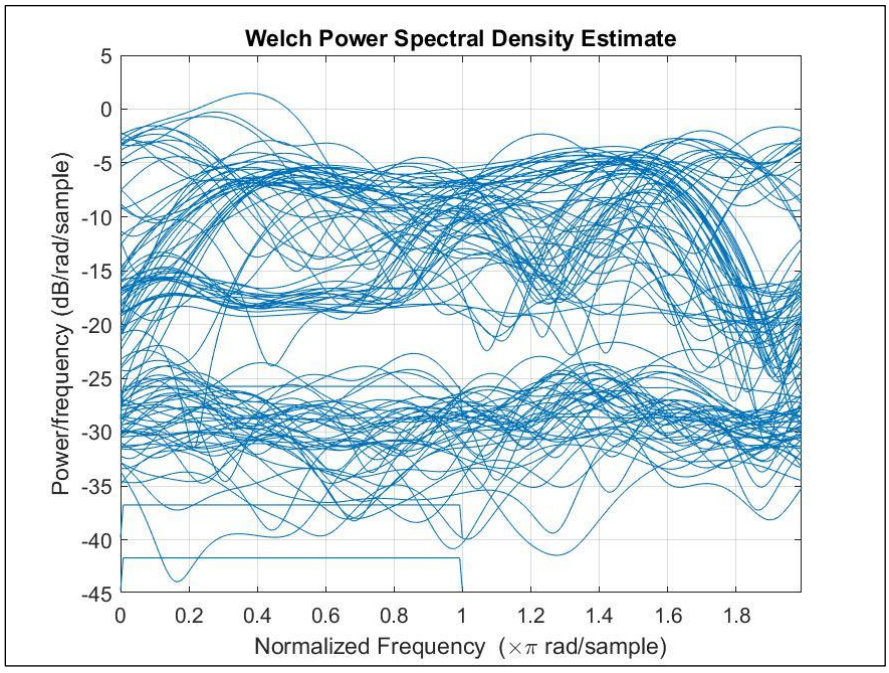
Below are graphs plotted for Discrete Wavelet Transform against time series:



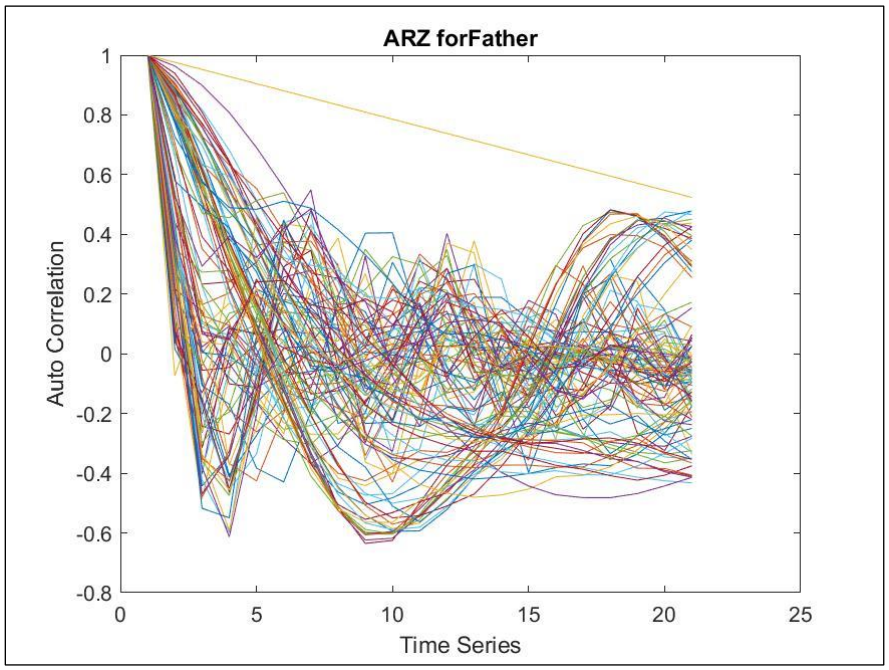

Below are graphs plotted for Fast Fourier Transform against time series:



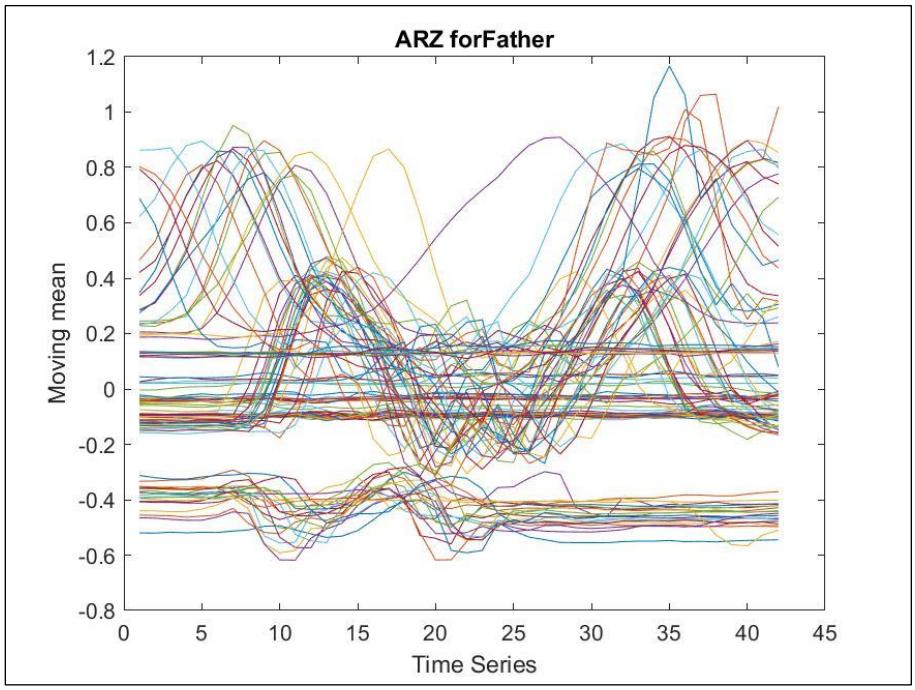
Below are graphs plotted for Power Spectral Density against time series:




Below are graphs plotted for Auto correlation against time series:



Below are graphs plotted for Moving Mean against time series:



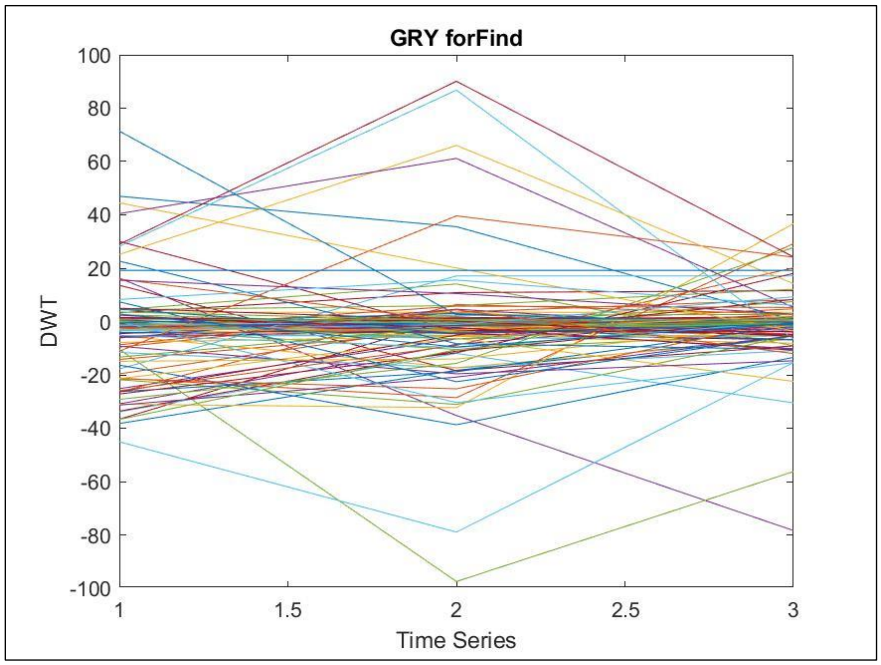

e) All the graphs have been carefully analysed to compare the intiution and the results found by ploting the graphs using MATLAB code. And we found that, our intiution is partially correct as, the maximum variation is found in 4 out 5 techniques used. In the graph plotted using Fast Fourier Transform, we have not seen much variation.

### 4.1.8. FIND

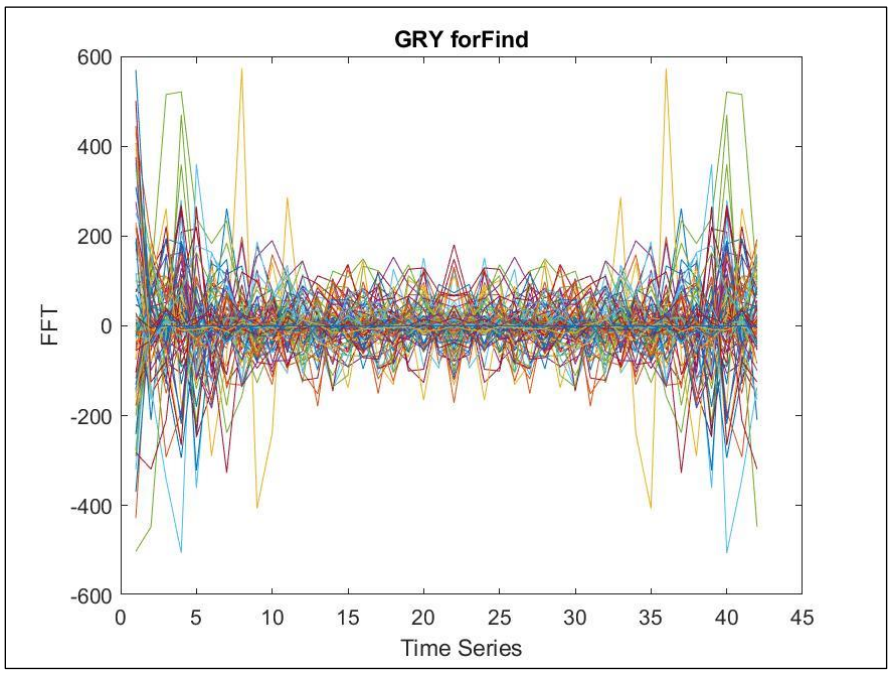
b) The useful feature which might show the maxium variation among the actions is Gyroscope, because the movement of right hand shows, an angular motion seen in the gesture “Find” from downward to upward.

d) The feature showing maximum variation in the graphs is Gyroscope along Y axis.

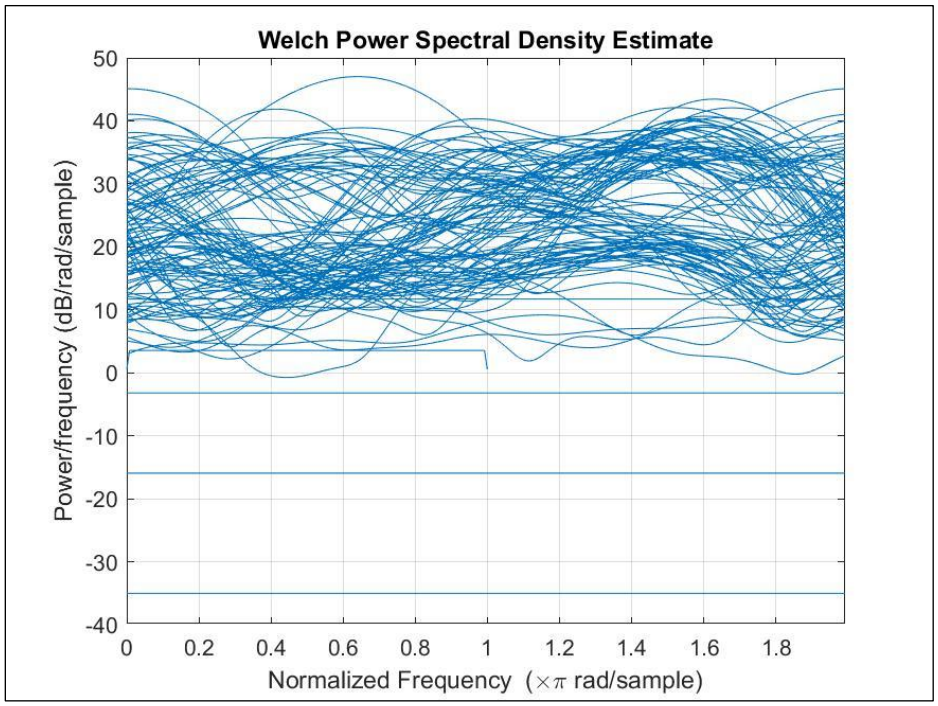
Below is the graph plotted for Discrete Wavelet Transform against time series:



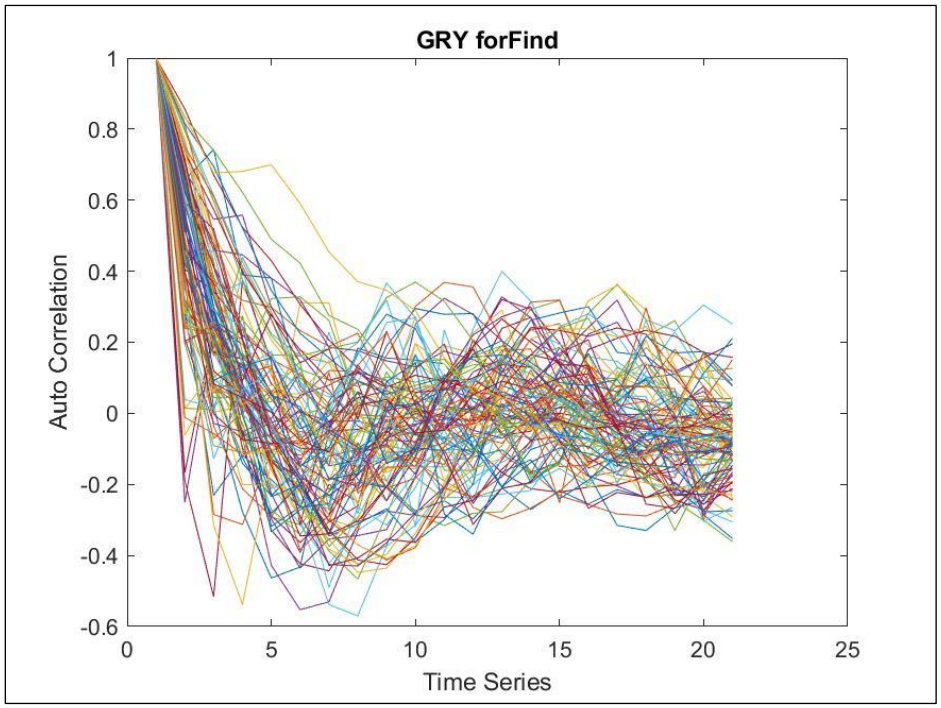

Below is the graph plotted for Fast Fourier Transform against time series:



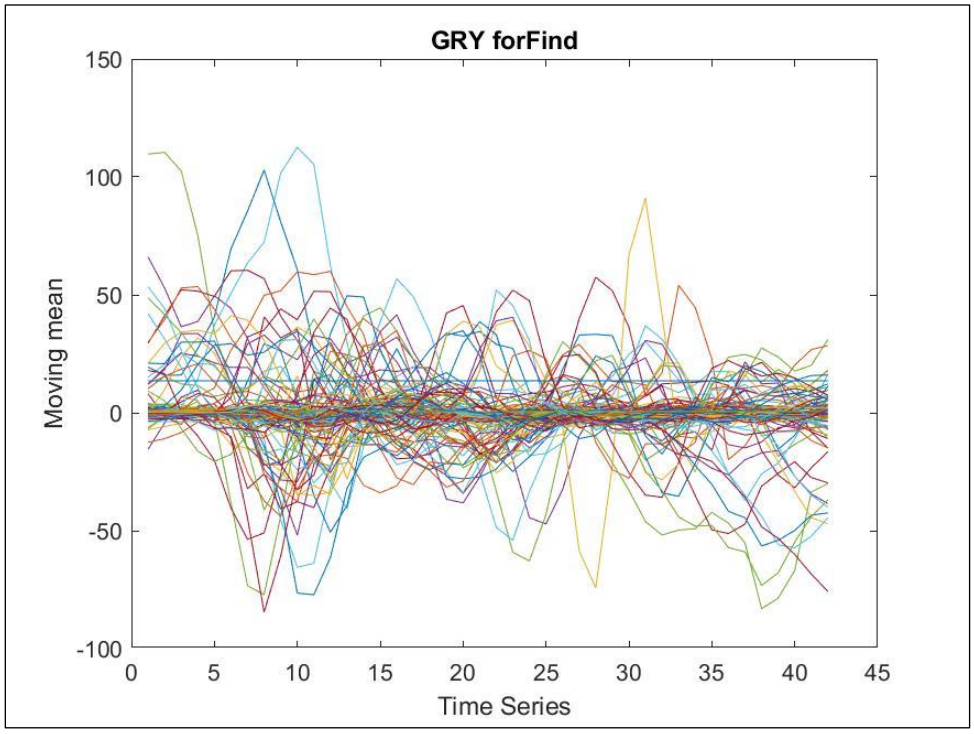
Below is the graph plotted for Power Spectral Density against time series:




Below is the graph plotted for Auto correlation against time series:



Below is the graph plotted for Moving Mean against time series:



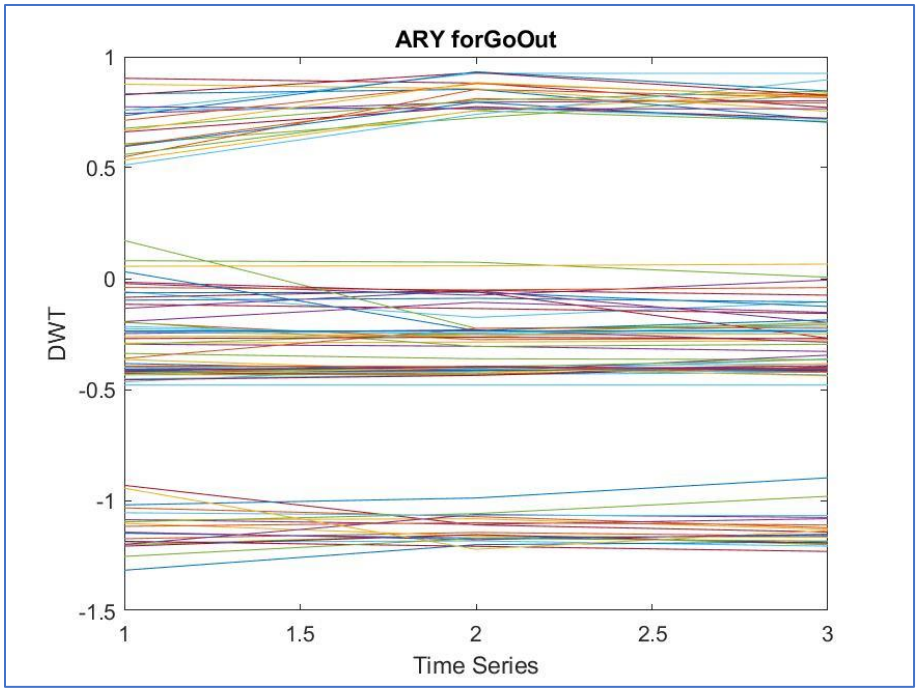

e) As we compare all the graphs with the above graphs generated, the above graphs have shown the maximum variation in the curves plotted, hence our intiution is mostly correct. The angular velocity of the hand movement is more dominant when compared to other features, hence Gyroscope is the useful feature.

### 4.1.9. Go out

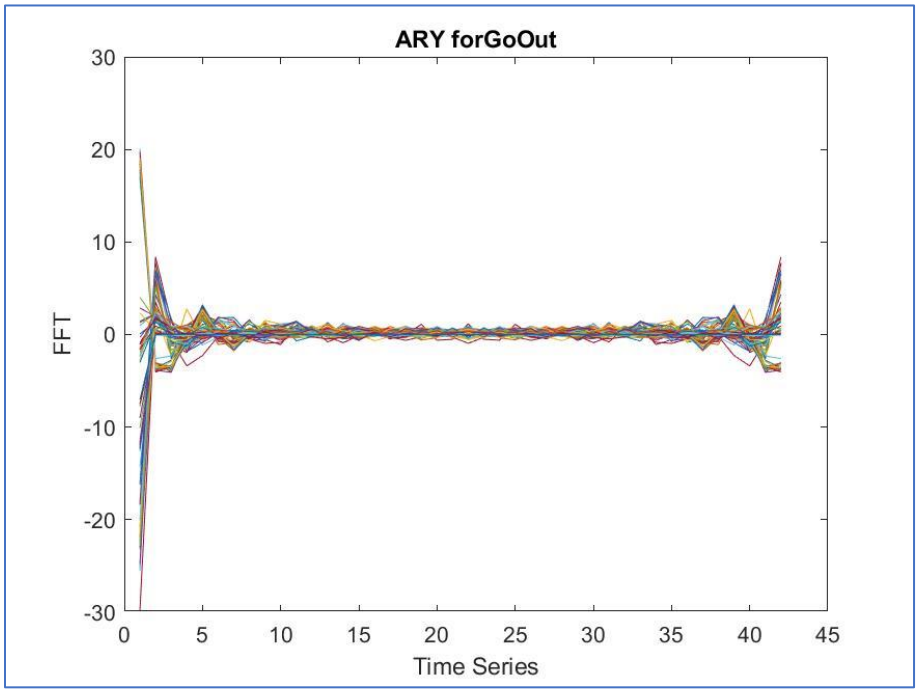
b) The hand movement is going in a linear direction for the gesture “Go out”, hence our intution is the accelerometer can show maximum variation in the graphs.

d) The feature showing maximum variation in the graphs is along Accelerometer Y axis using Right Hand

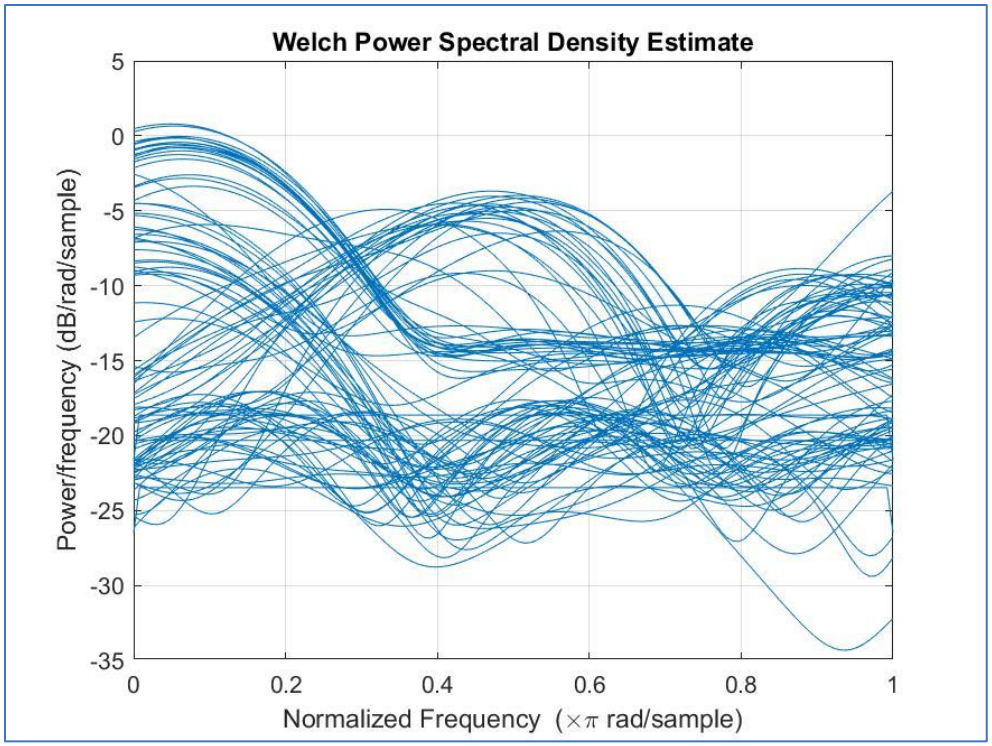
Below are graphs plotted for Discrete Wavelet Transform against time series:



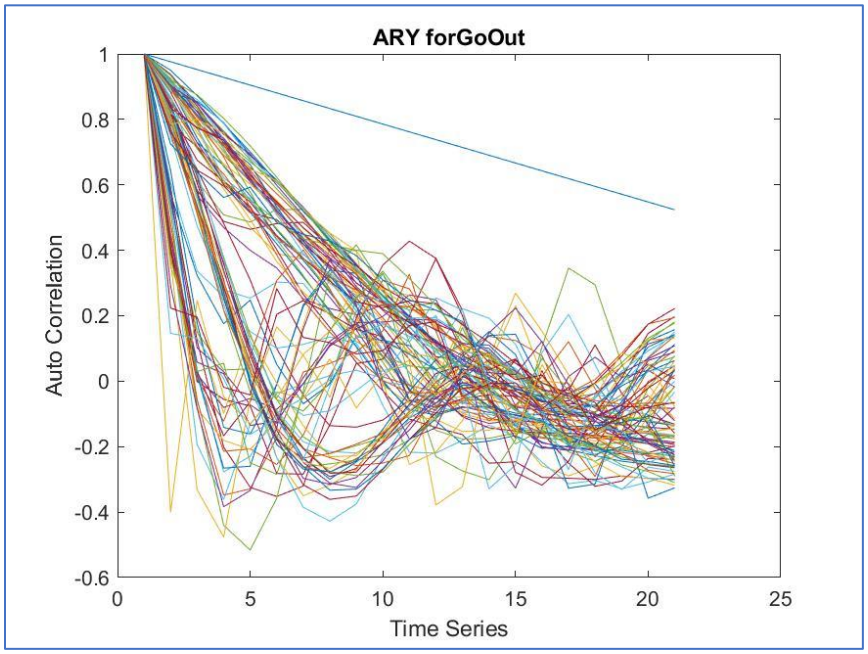

Below are graphs plotted for Fast Fourier Transform against time series:



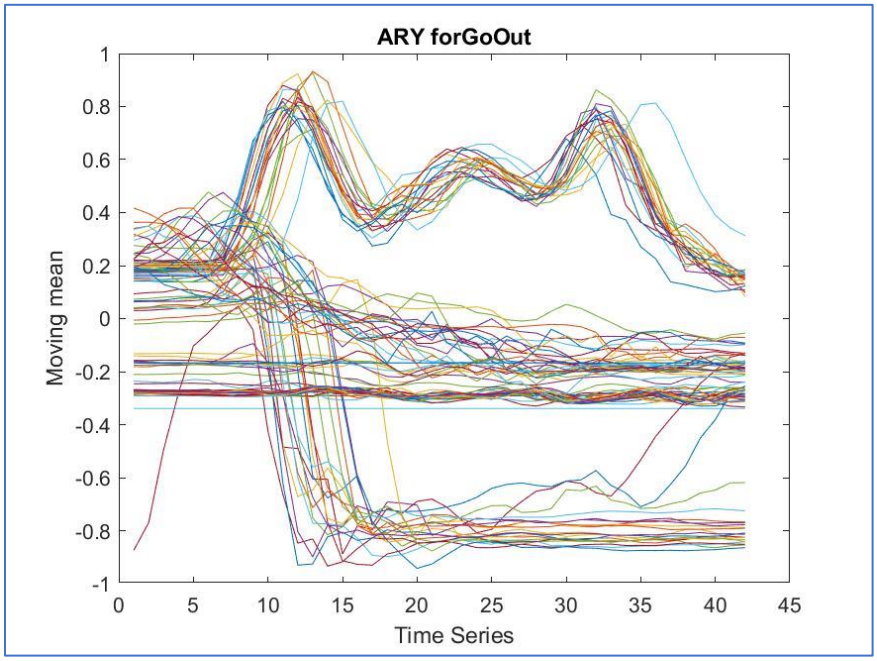

Below are graphs plotted for Power Spectral Density against time series:


Welch Power Spectral Density Estimate
Power/frequency (dB/rad/sample)
5
0
-5
-10
-15
-20
-25
-30
-35
0
0.2
0.4
0.6
0.8
1
Normalized Frequency (×π rad/sample)

Below are graphs plotted for Auto correlation against time series:




Below are graphs plotted for Moving Mean against time series:



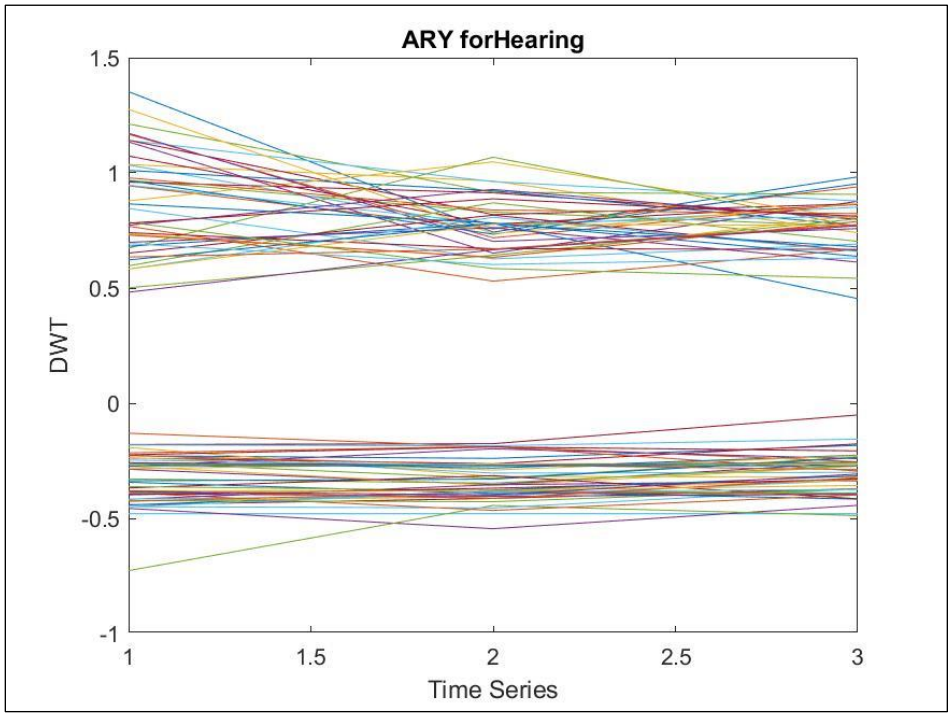

e) We have considered all the graphs generated and found that the graphs shown above are showing the maximum variation, hence our intution is mostly correct. And we recommend considering Accelerometer sensor data as the useful data for identifying gesture “Go out” effortlessly.

## 4.1.10.Hearing

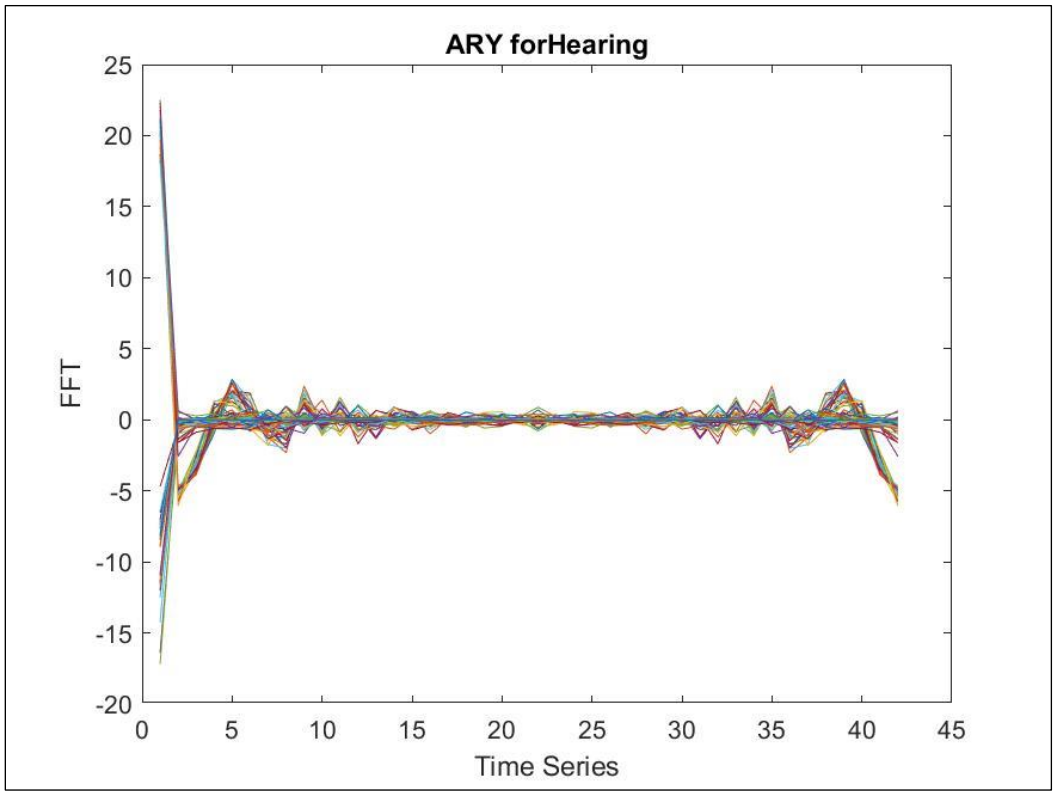
b) The Hearing gesture involves the movement of right-hand index finger up and down in front of the lips. As per our Intuition, the Accelerometer should identify the variation in the movement of the right hand. Accelerometer along Y-axis should identify the difference better than other sensors since there is no angle change for gyroscope or Orientation sensors to pick up the signal.

d)

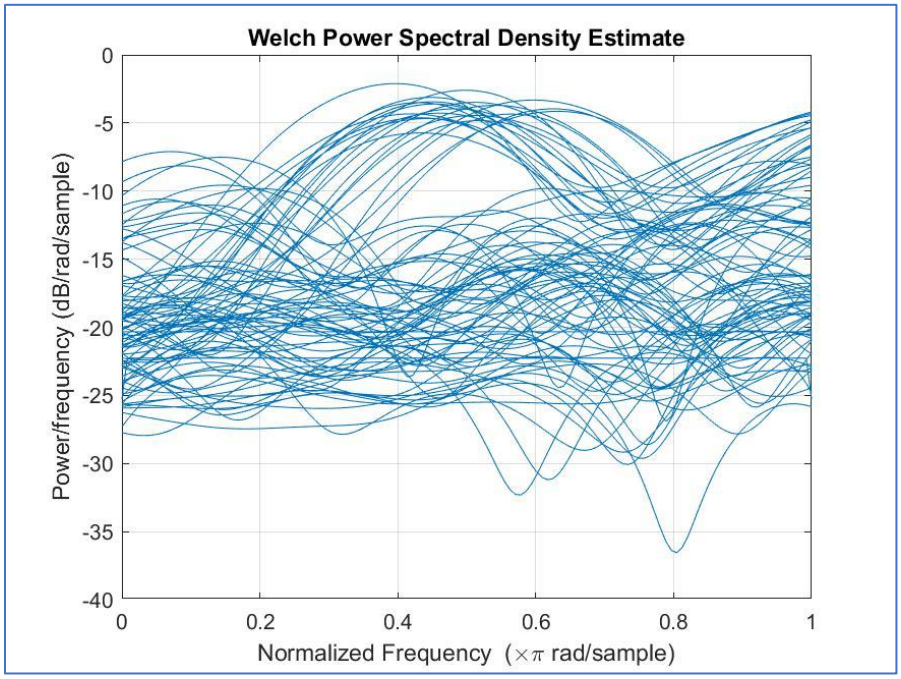
Below are graphs plotted for Discrete Wavelet Transform against time series:



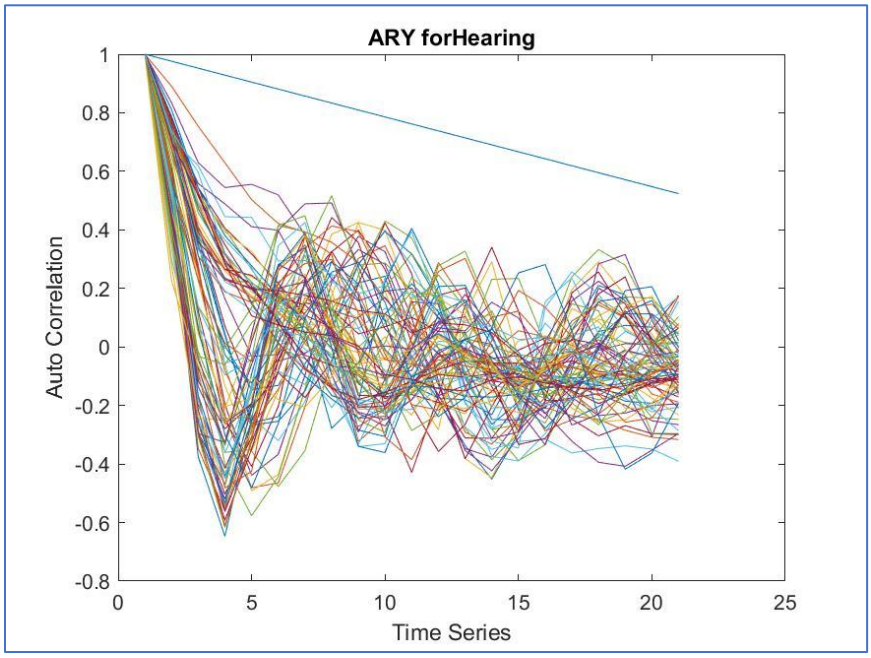
Below are graphs plotted for Fast Fourier Transform against time series:



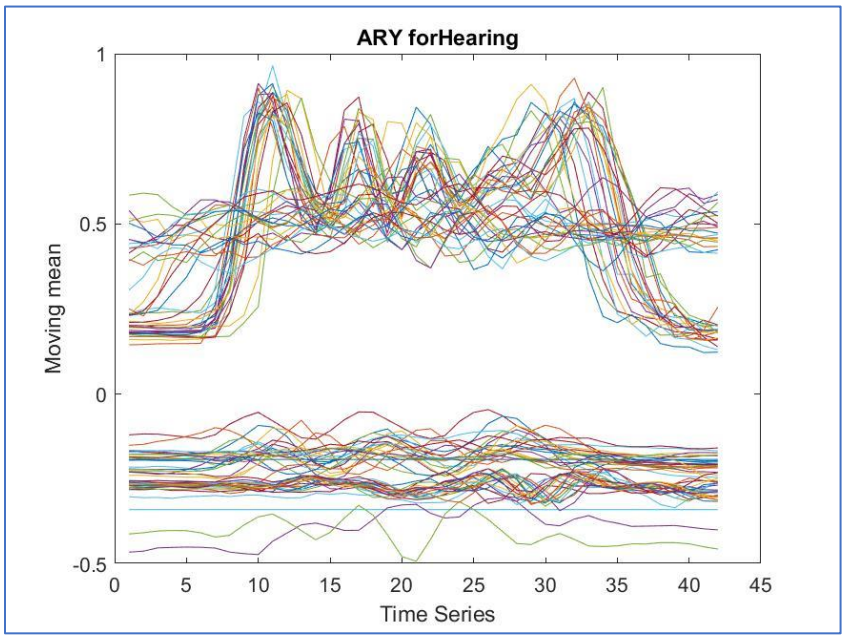

Below are graphs plotted for Power Spectral Density against time series:



Below are graphs plotted for Auto correlation against time series:




Below are graphs plotted for Moving Mean against time series:




e) he data from the graph shows that our intuition was correct, and Accelerometer along Y-axis did pick up the difference in movement of the right-hand index finger. As for the other sensors that deal with angle there was no major difference as most of the graphs indicate overlapping.

## 4.2 Feature Selection

### 4.2.1 Arranging 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 eigen vector then the new feature matrix can be obtained by simply multiplying the eigen vector with the old feature matrix. (You might need ten matrices corresponding to the ten classes)

Write your logic of feature matrix arrangement.

We have created a TemporaryFeatureMatrix which stores all the actions of single feature in every gesture. We have stacked all the features of a single Gesture in FinalFeatureMatrix by running the loop 34 times. The TemporaryFeatureMatrix will be having dimensions of 120*42 ( where 120 =6 Persons*20 Actions of each Feature). The FinalFeatureMatrix will be having dimensions of 4080*42(where 4080=6Persons*20Actions*34Features)

### 4.2.2 Execution of PCA

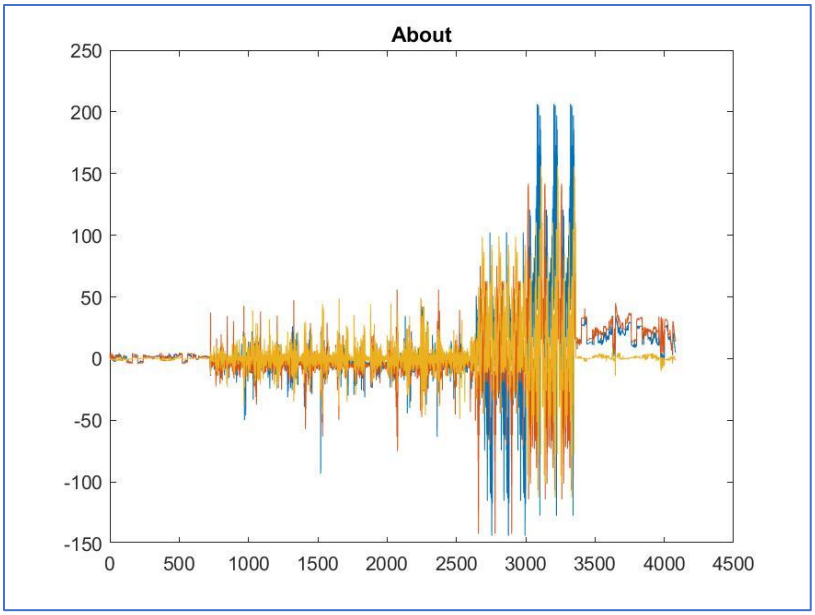
1. Use Matlab's PCA function to run PCA on your feature matrix. Show all the eigen vectors in a plot.
2. We have applied an inbuilt PCA library function `princomp()` which takes an input of FinalFeatureMatrix(4080*42) and returns Coefficient and latent matrices.
3. Coefficient is a *p*-by-*p(i.e 42*42* matrix), each column containing coefficients for one principal component. The columns are in order of decreasing component variance.
4. `Latent` is a vector containing the eigenvalues of the covariance matrix of FinalFeatureMatrix.

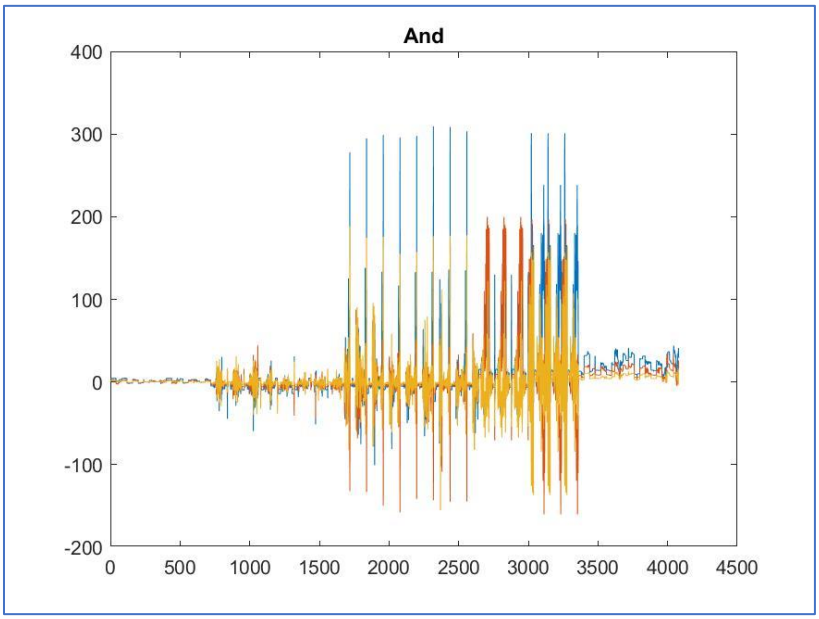
### 4.2.3 Discussion of Eigen Vectors

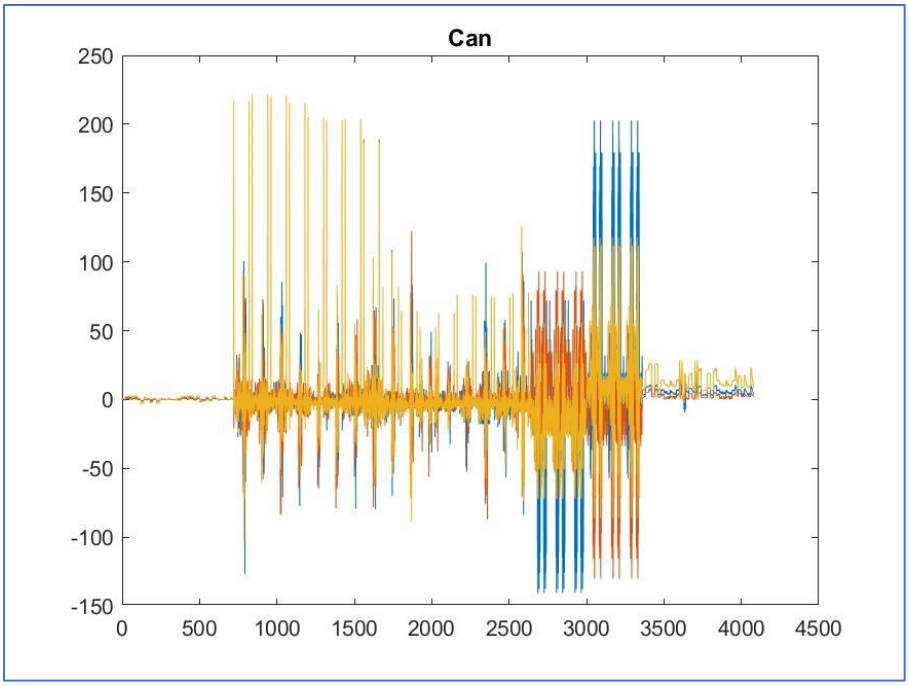
1. Latent matrix obtained contains the eigenvalues of the covariance matrix. The values in the latent will be in decreasing order of variance. Similarly, The columns in Coefficient are in order of decreasing component variance. Since the first 3 columns are prominent we consider coefficient matrix with dimensions 42*3
2. Write an explanation on the reason why the eigen vectors turned out the way they did.

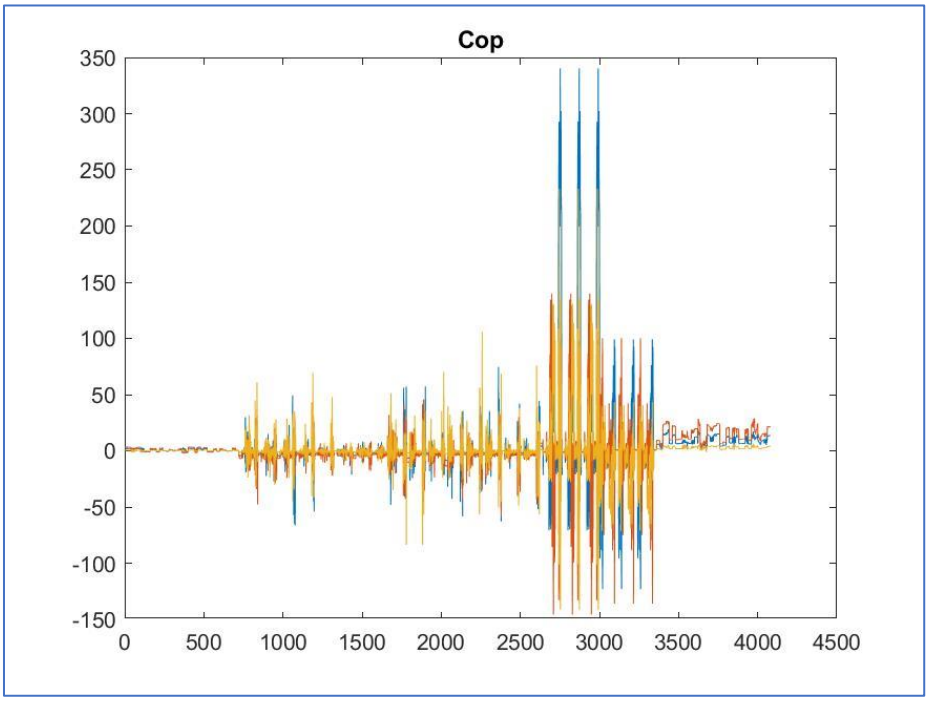
### 4.2.4 Results of PCA

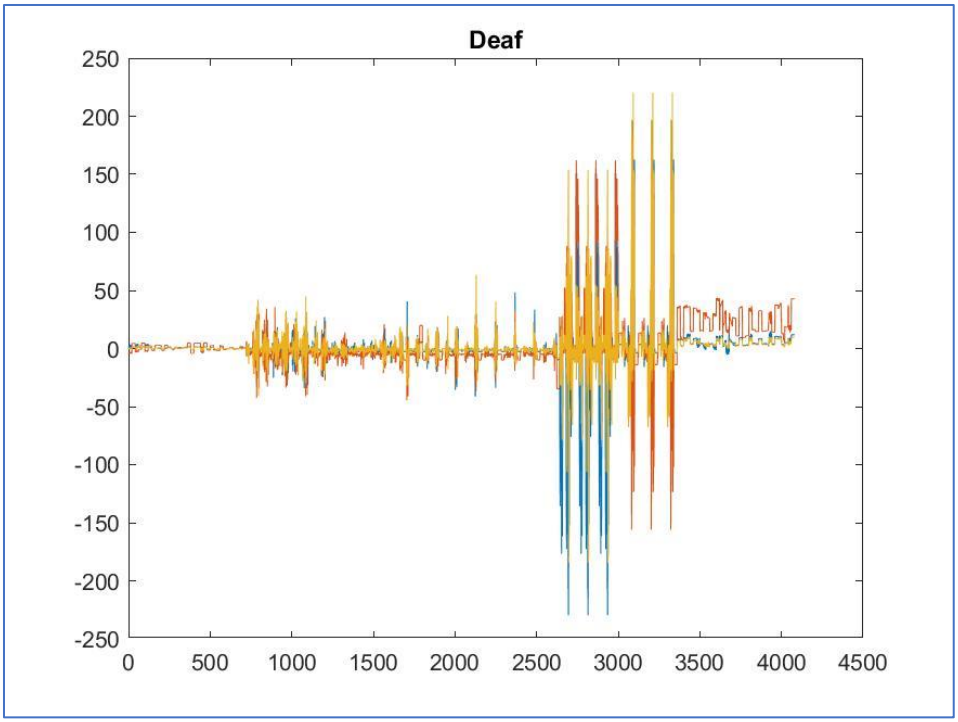
1. Results of PCA. Create the new feature matrix. Again generate 10 matrices. For multiple actions you can choose to overlap the plots.
2. The FinalFeatureMatrix (4080*3) is multiplied with the reduced coefficient matrix (42*3) obtained in step 4 which gives us a NewFinalFeatureMatrix of dimensions 4080*3,Where the 3 columns are the prominent features.We have created 10 NewFinalFeatureMatrix for all the gestures and plotted as shown below

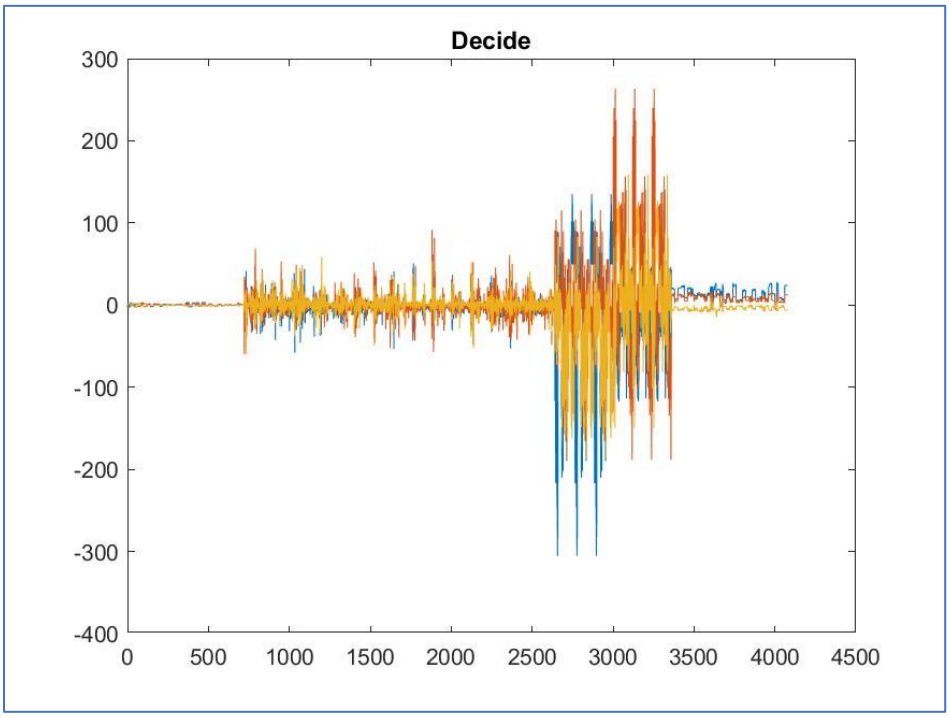

About
250
200
150
100
50
0
-50
-100
-150
0
500
1000
1500
2000
2500
3000
3500
4000
4500

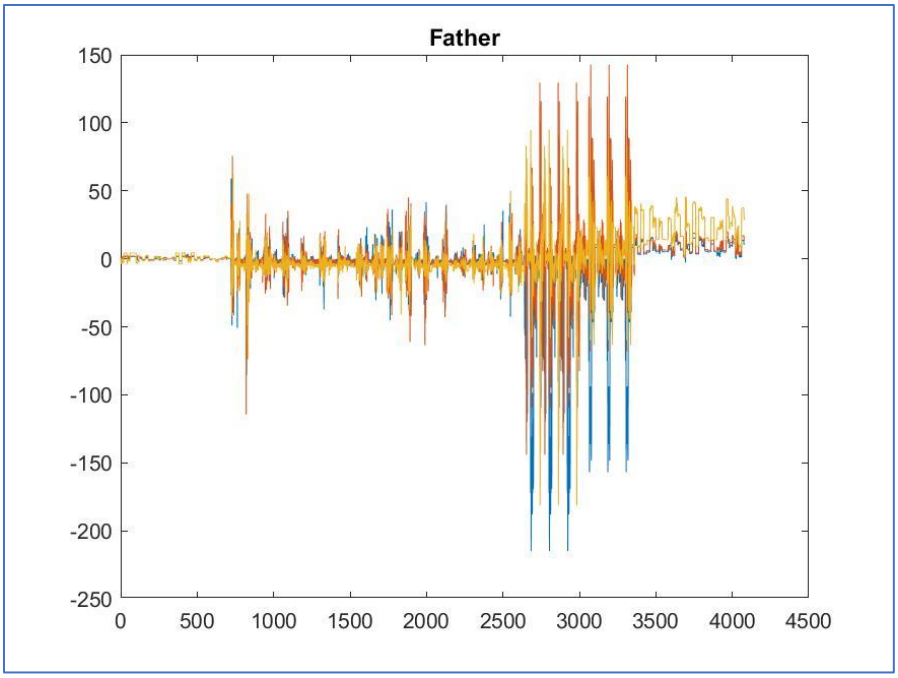

And
400
300
200
100
0
-100
-200
0
500
1000
1500
2000
2500
3000
3500
4000
4500

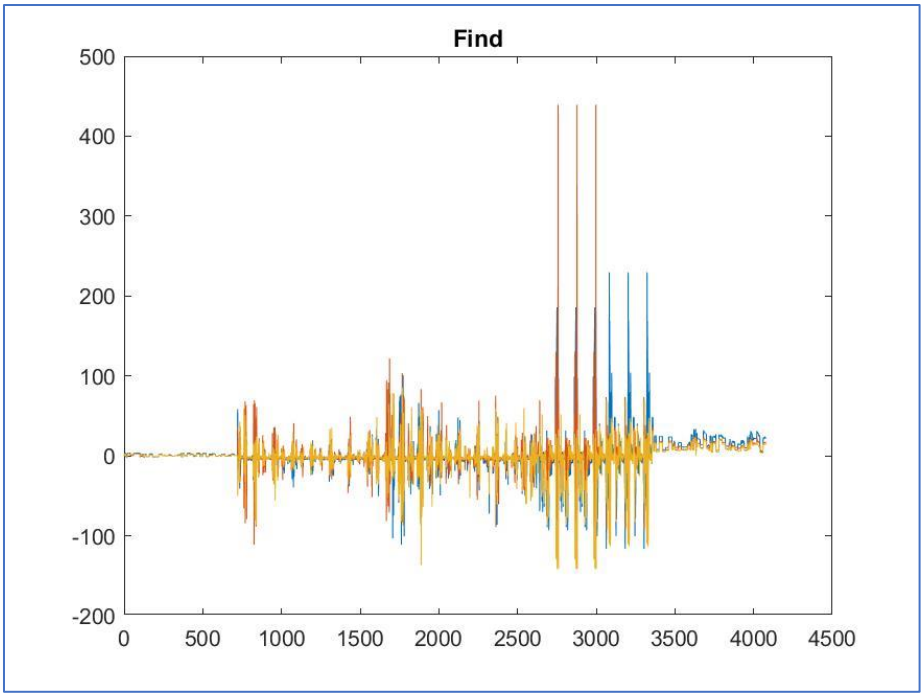

Can
250
200
150
100
50
0
-50
-100
-150
0
500
1000
1500
2000
2500
3000
3500
4000
4500

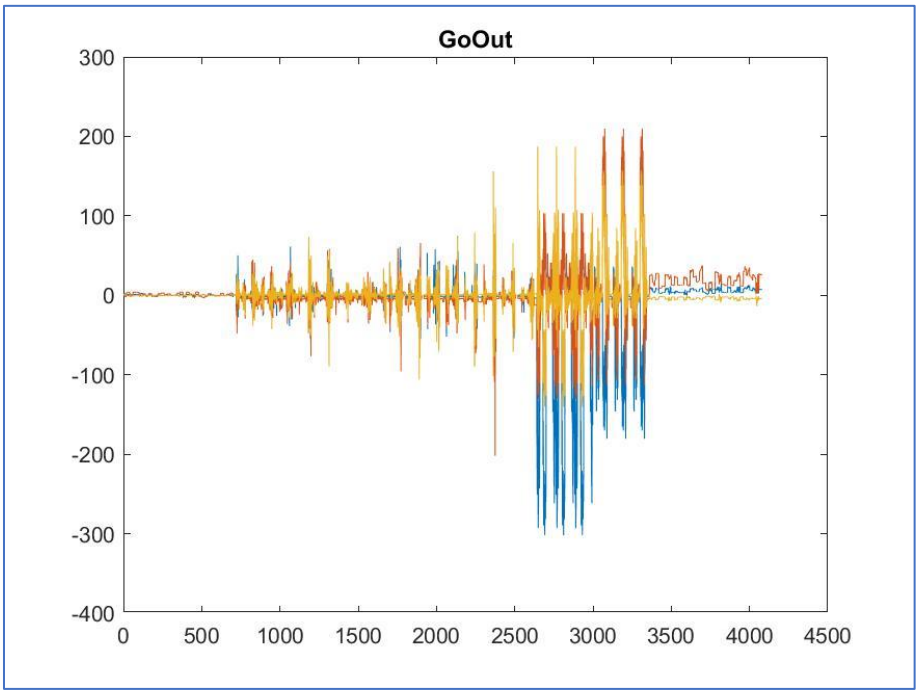

Cop
350
300
250
200
150
100
50
0
-50
-100
-150
0
500
1000
1500
2000
2500
3000
3500
4000
4500

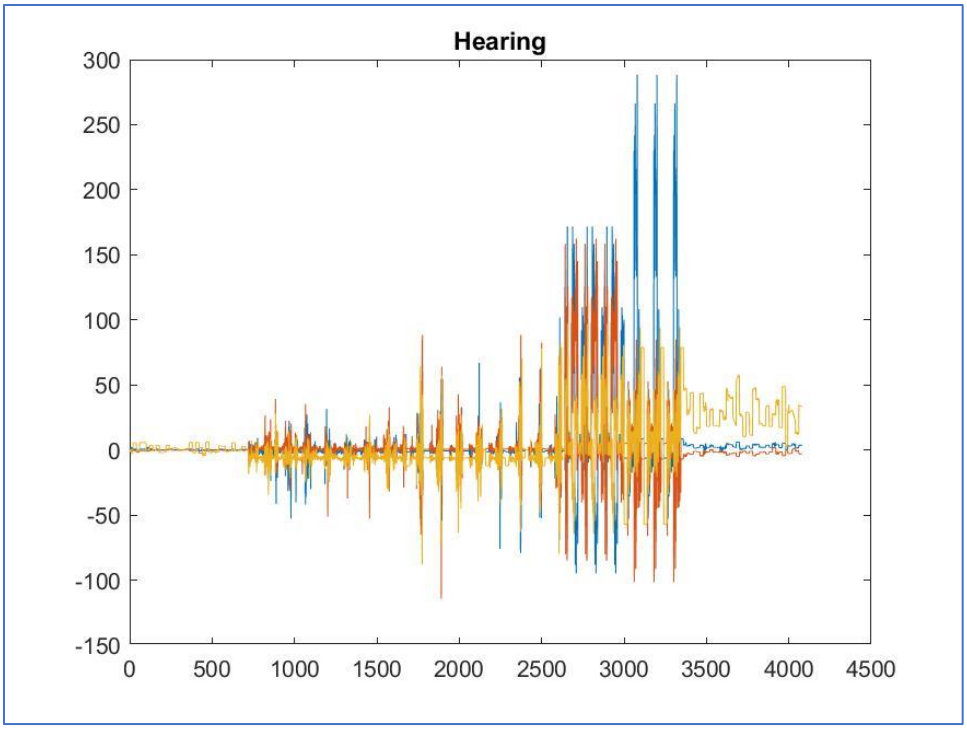

Deaf
250
200
150
100
50
0
-50
-100
-150
-200
-250
0
500
1000
1500
2000
2500
3000
3500
4000
4500


Decide
300
200
100
0
-100
-200
-300
-400
0
500
1000
1500
2000
2500
3000
3500
4000
4500


Father
150
100
50
0
-50
-100
-150
-200
-250
0
500
1000
1500
2000
2500
3000
3500
4000
4500


Find
500
400
300
200
100
0
-100
-200
0
500
1000
1500
2000
2500
3000
3500
4000
4500


GoOut
300
200
100
0
-100
-200
-300
-400
0
500
1000
1500
2000
2500
3000
3500
4000
4500




### 4.2.5 Conclusion on PCA

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 was very helpful in reducing the dimensions of dataset. With such a huge dataset containing many dimensions, PCA helped us to find eigen vector matrix which in turn helped us to find the prominent features among the entire dataset.