Study of Multimodal Dataset for Various Forms of Distracted Driving

Guru Pavan Kumar Kamakolanu, Chandana Roopa Reddy Rajala

May 6, 2017

1 Introduction

The data consists of 68 subjects. All the subjects are subjected to a controlled experiment on a driving simulator. These subjects are volunteers that are male and female. The subjects that drove the same highway under four different conditions: No distraction, cognitive distraction, emotional distraction, and sensorimotor distraction. The experiment closed with a special driving session, where all subjects experienced a startle stimulus in the form of unintended acceleration - half of them under a mixed distraction, and the other half in the absence of a distraction. During the experimental drives key response variables and several explanatory variables were continuously recorded. The response variables included speed, acceleration, brake force, steering, and lane position signals, while the explanatory variables included perinasal EDA, palm EDA, heart rate, breathing rate, and facial expression signals; biographical and psychometric covariates were also obtained. This dataset enables multidimensional research into driving behaviors under neatly abstracted distracting stressors, which account for an increasing number of car crashes. The set can also be used in physiological channel benchmarking and multi spectral face recognition.

2 Hypothesis

2.1 Null Hypothesis:

There is no significance elevation in the physiological signal when stress is applied in different driving modes.

2.2 Alternate Hypothesis:

There is a significant elevation in the physiological signals when stress is applied in different driving modes.

3 Cleaning

3.1 Perinasal EDA Signal(PP)

Perinasal NR EDA Signals are taken from the volunteers for all the eight drives BL, PD, RD, CD, ED, MD, FD and ND. This data is already cleaned. It consists of 59 signals in PD, 59 signals in RD, 59 signals in ED, 59 signals in MD, 59 signals in CD, 58 signals in ND, 59 signals in FD and 56 signals in BL.

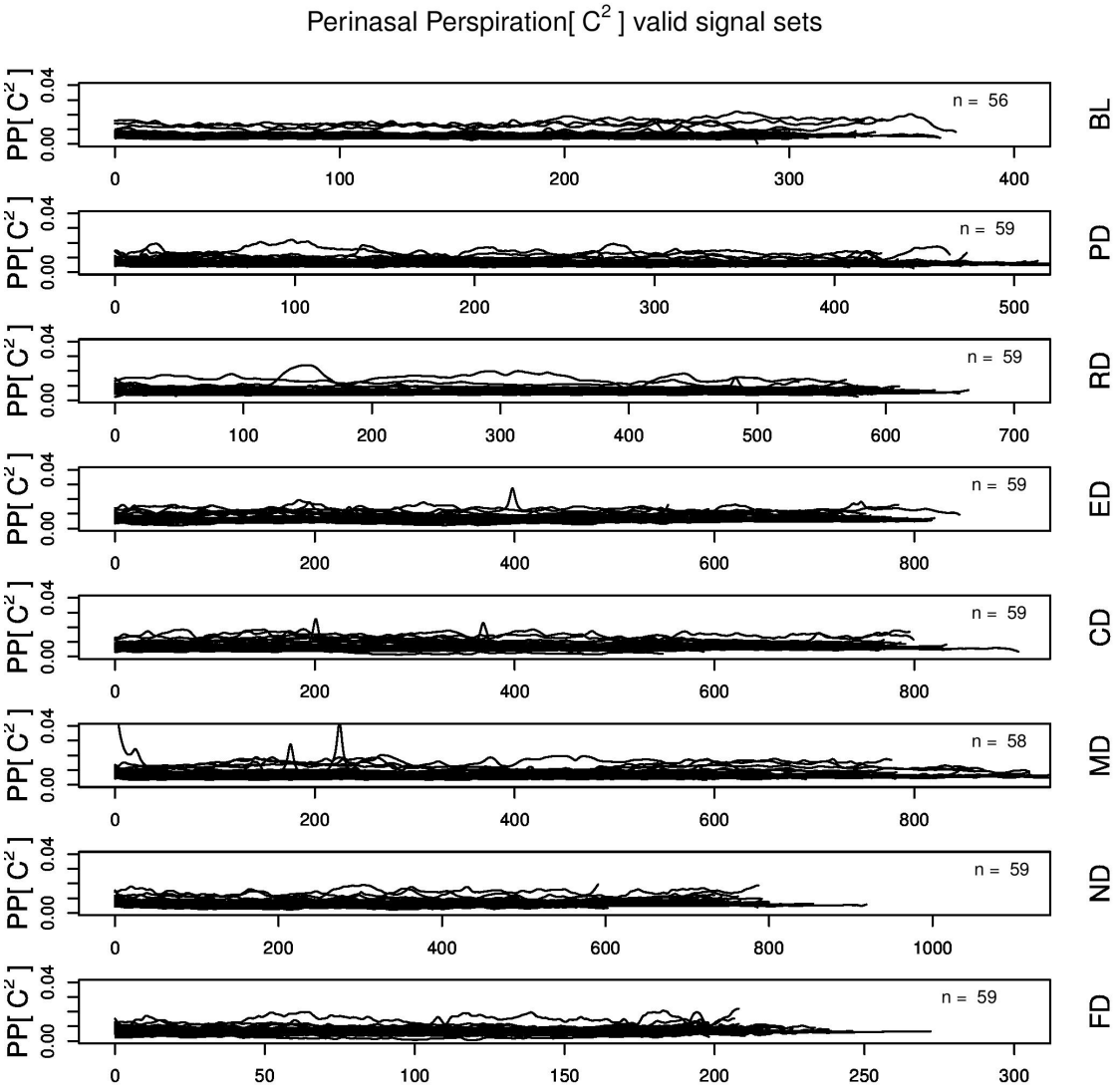

Figure 1: Perinasal NR EDA signals per drive, for all the subjects.

3.2 Breathing Rate Signal(BR)

Breathing Rate Signals are taken from the volunteers for all the seven modes ND, PD, RD, CD, ED, MD, FD. The valid range for the breathing rate is [4,70] bpm. Initially before cleaning there are 65 signals in PD, 65 signals in RD, 66 signals in ED, 62 signals in MD, 66 signals in CD, 66 signals in ND, 65 signals in FD. To make the data valid we have to eliminate the data points that are not this range i.e., [4, 70] bpm. After cleaning there are 63 signals in PD, 64 signals in RD, 62 signals in ED, 58 signals in MD, 66 signals in CD, 66 signals in ND, 63 signals in FD.

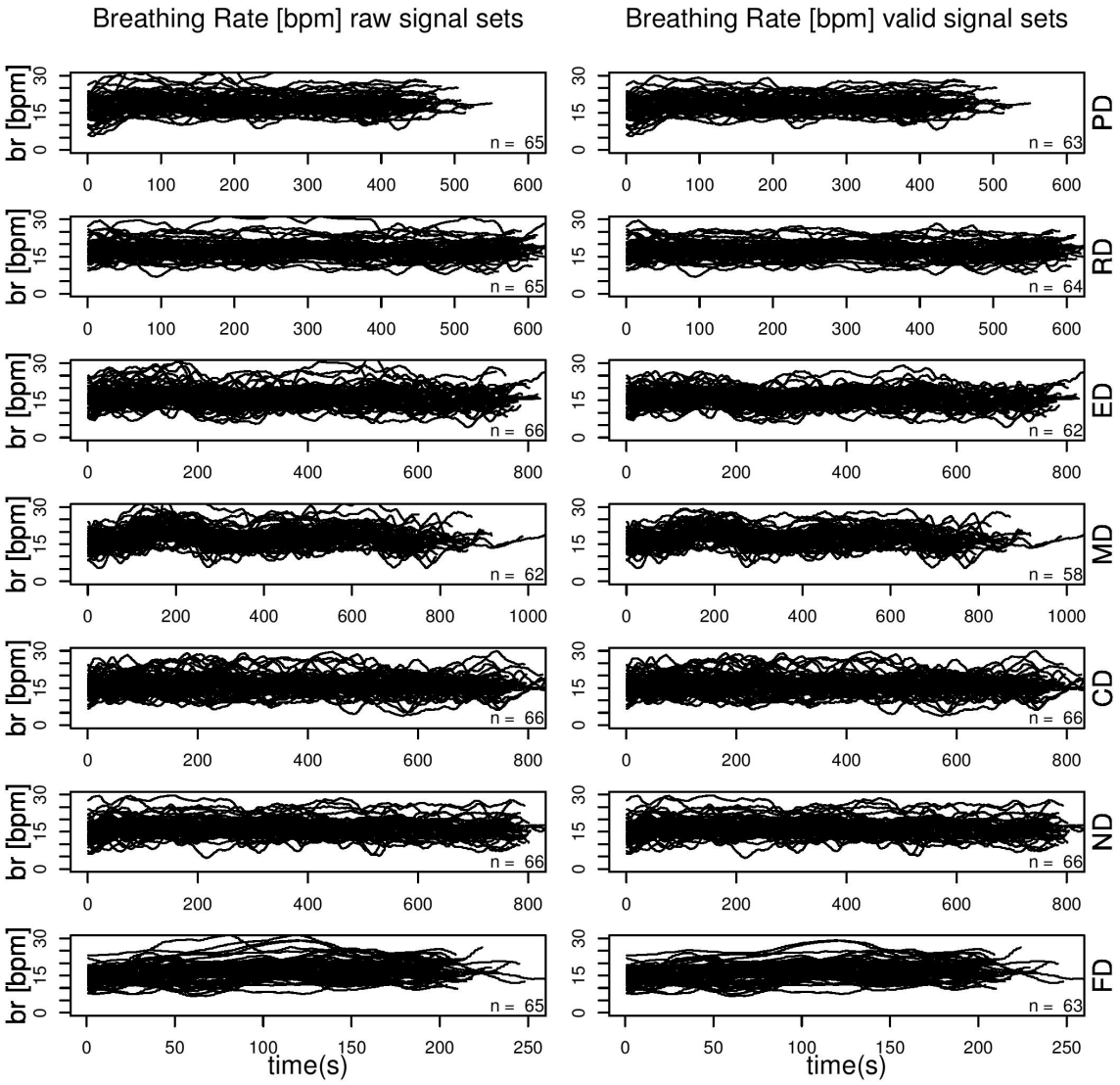

Figure 2: Breathing Rate Signals per drive, for all subjects.

3.3 Heart Rate Signal(HR)

Heart Rate Signals are taken from the volunteers for all the seven modes ND, PD, RD, CD, ED, MD, FD. The valid range for the breathing rate is [4,140] bpm. Initially before cleaning there are 65 signals in PD, 65 signals in RD, 66 signals in ED, 62 signals in MD, 66 signals in CD, 66 signals in ND, 65 signals in FD. To make the data valid we have to eliminate the data point that does not have the valid range i.e., [4, 140] bpm. After cleaning there are 58 signals in PD, 59 signals in RD, 59 signals in ED, 56 signals in MD, 58 signals in CD, 58 signals in ND, 58 signals in FD.

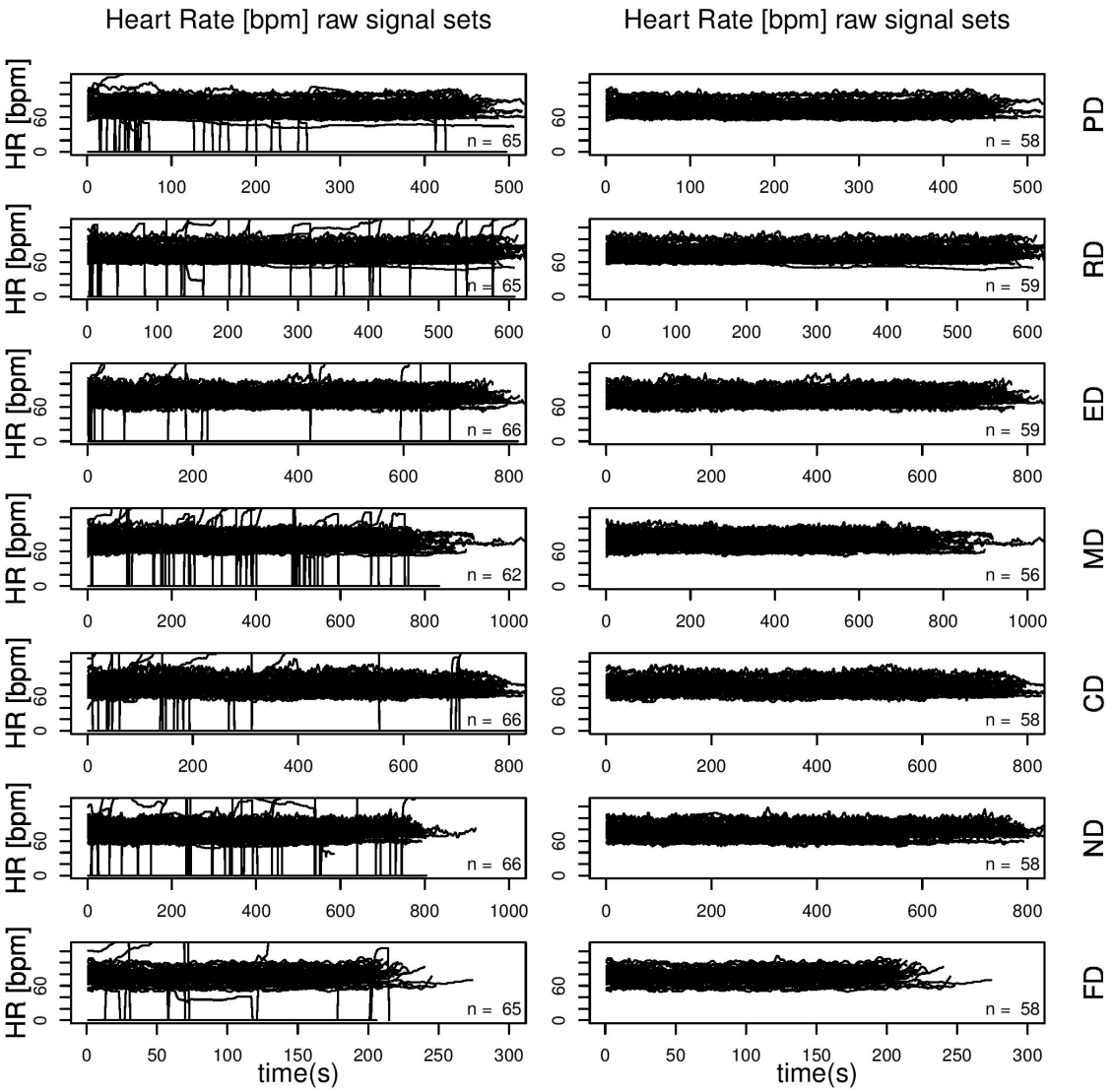

Figure 3: Heart Rate signals per drive, for all subjects, before and after noise reduction.

3.4 Palm EDA Signal(PEDA)

Palm EDA Signals are taken from the volunteers for all the seven modes ND, PD, RD, CD, ED, MD, FD. The valid range for these signals is 10 - 4,700 kOhm. Initially before cleaning there are 59 signals in PD, 59 signals in RD, 58 signals in ED, 54 signals in MD, 66 signals in CD, 59 signals in ND, 54 signals in FD. To make the data valid we have to eliminate the data point that does not have the valid range i.e., 10 – 4,700 kOhm. After cleaning there are 58 signals in PD, 56 signals in RD,54 signals in ED, 52 signals in MD, 58 signals in CD, 58 signals in ND, 52 signals in FD.

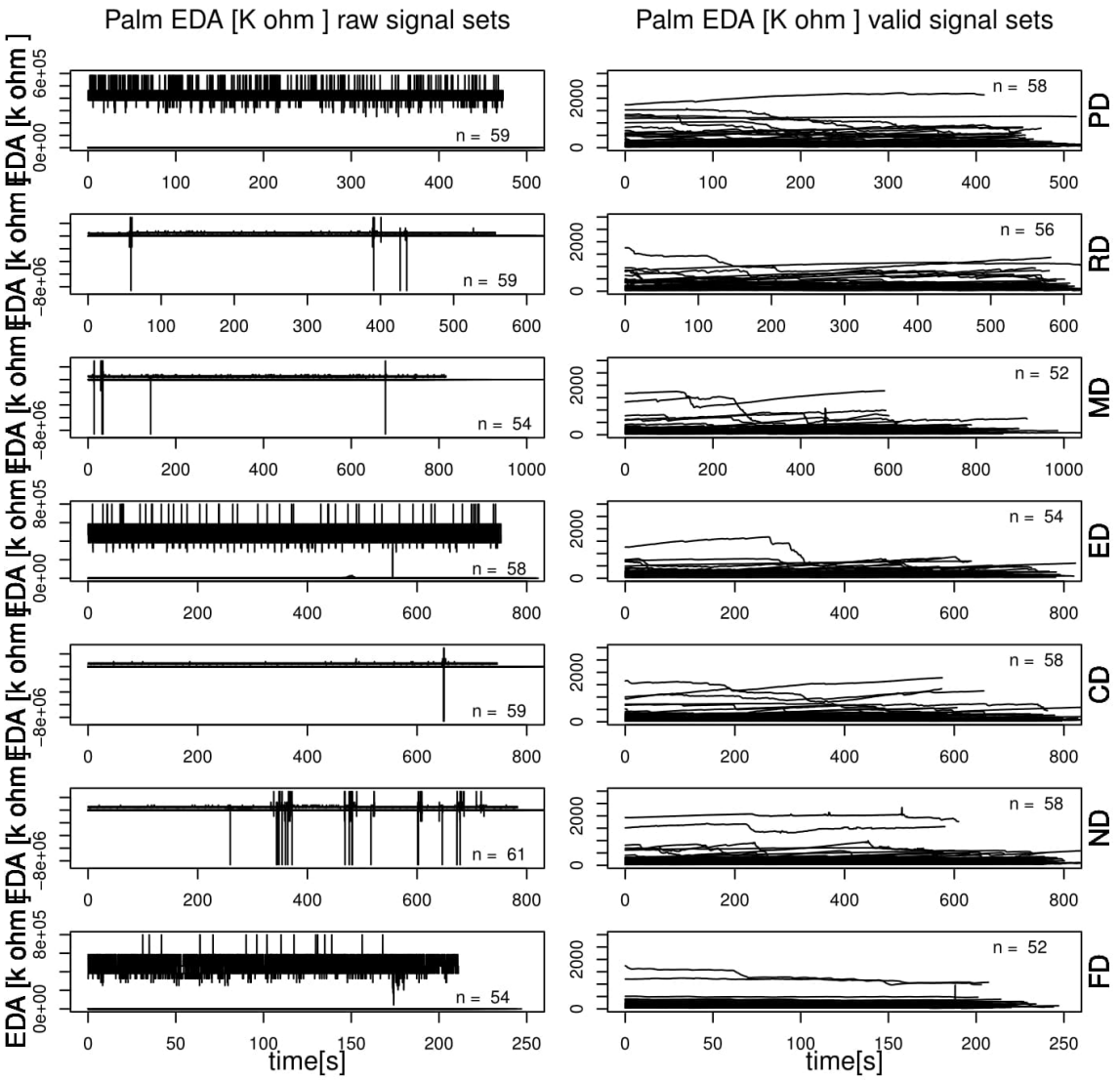

Figure 4: Palm EDA signals per drive, for all the subjects after noise reduction.

3.5 Performance Response Variable

The Response variable excel file contains 7 columns i.e. Frame, Time, Speed, signal, Acceleration signal, Brake Force signal, Steering signal and Lane Position signal. Hence, scanning each row from left to right we find the chronological rank order of the instantaneous measurement, the time [in s] .The composite measurement was taken with respect to the beginning of the session. Speed signals which has values between -0.1 and +0.1 kph are replaced with X = 0 kph, while we have considered only values which are >-0.1. In accelerometer signals, we omitted negative values and reported signals. For brake force signals, we replaced values Y > 300 N with Y = 300 N. There are 8509 signals that are greater than 300. We have replaced all of them.

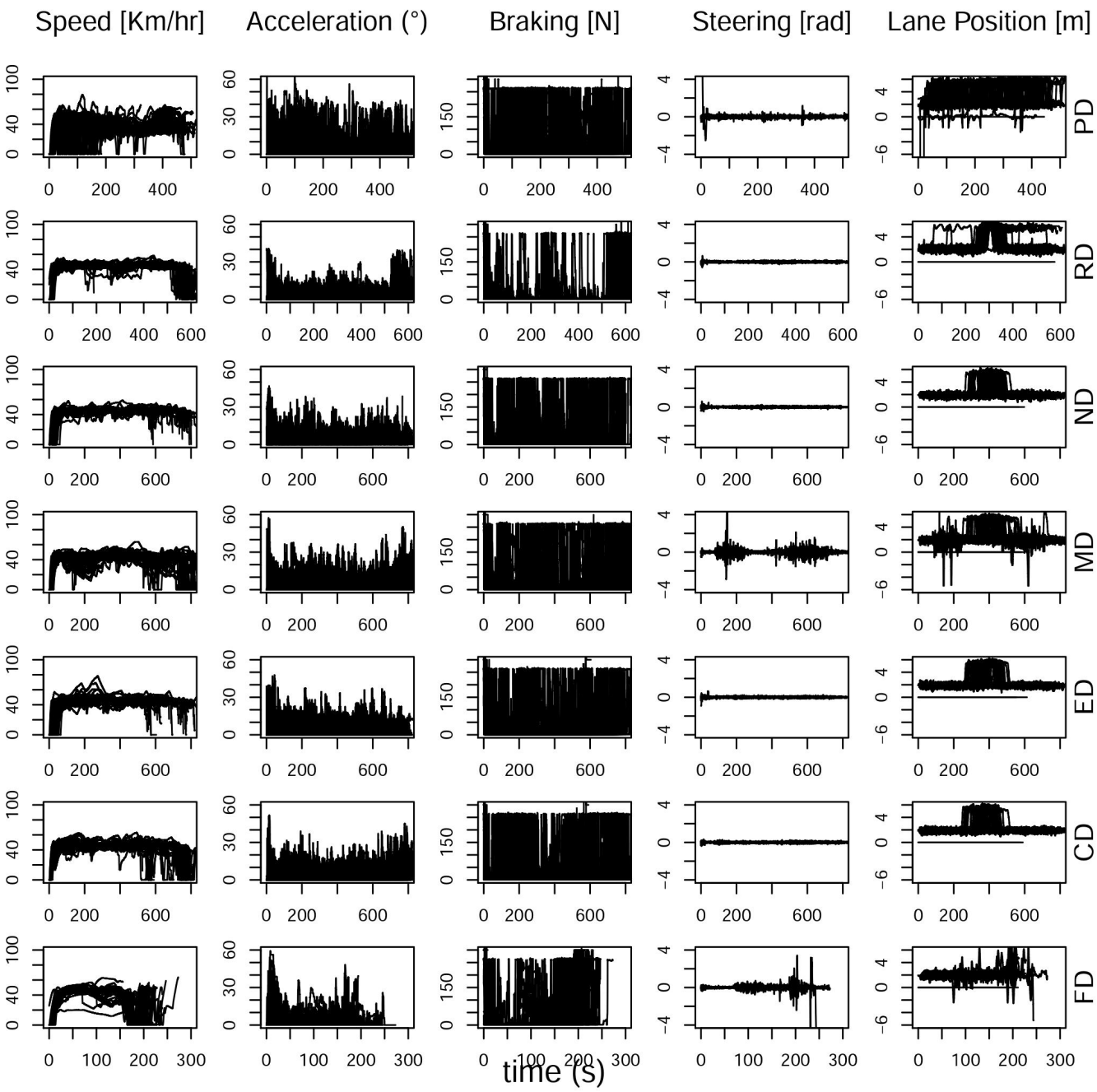

Figure 5: Cleaned Speed, Acceleration, Braking, Steering, Laneposition Plots for all the subjects.

4 Stress Validation

In this section, we check whether the stressors applied on different driving modes is producing stress or not. The different signals i.e Heart rate, Breathing rate, Persinasal EDA and Palm EDA signals can be compared for effect of stress applied in different time intervals under various distractions in different driving modes.

4.1 Process

- We need to first identify the folders with all the files i.e. Stress file, signal file in one of the driving mode under distractions folder and the same signal file in Normal driving mode without distraction.

- From the given stress files, we can consider five phases from the given time intervals.
- For each phase of time interval, mean of the respective data is calculated for each driving mode.
- QQ plot and QQ line is drawn for the above data and the data is tested for normality. From the QQ plot, if the data points are around the fringes of QQ line, then the data is said to be normalized. If any outliers are found that are not around the fringes, then the data is not normalized.
- If the data is not in a normalized distribution, we normalize the data using log transformation.
- We test this data distribution using Shapiro test to verify the normality of data.
- Then, we calculate the difference of means between a driving mode with distractions and driving mode with no distraction.
- On this calculated data, we perform t-test to check whether there is a mean difference after applying the stress.
- Box plot is plotted for this data and stars are plotted based on p-value range as per the below table.

Table 1: P Value - Stars.

P- Value	**No of stars**
$p < 0.05$	1
$p < 0.01$	2
$p < 0.001$	3

4.2 Example for one signal HR in CD mode:

- From every subject folder,We need to first identify the files i.e. Stress file, HR signal file in CD folder and the HR signal file in ND folder.
- From the given stress files, we can consider five phases from the given time intervals.
- For each phase of time interval, mean values of cleaned HR data is calculated for CD mode from all the subjects folder.
- QQ plot and QQ line is drawn for the above data and the data is tested for normality. From the QQ plot, if the data points are around the fringes of QQ line, then the data is said to be normalized. If any outliers are found that are not around the fringes, then the data is not normalized.
- If the data is not in a normalized distribution, we normalize the data using log transformation.
- We test this data distribution using Shapiro test to verify the normality of data.
- Then, we calculate the difference of means between CD and ND of Heart rate signal.
- On this calculated data, we perform t-test to check whether there is a mean difference after applying the stress.
- Box plot is plotted for this data and stars are plotted based on p-value range as per the table mentioned.
- This process is carried out for ED and MD.
- These 3 modes are plotted using box plot and compared side by side to check for the significance of stress applied.

- The above example is applied on other signals i.e. BR, PEDA and PP.

4.3 Tests Performed

4.3.1 Shapiro Test

The Shapiro–Wilk test is a test of normality in frequentist statistics.Most of the parametric tests require that the assumption of normality be met. Normality means that the distribution of the test is normally distributed (or bell-shaped) with 0 mean, with 1 standard deviation and a symmetric bell shaped curve.

QQ plot can be used as graphical method for normality test of data distribution.In this method, observed value and expected value are plotted on a graph.If the plotted value vary more from a straight line, then the data is not normally distributed. Otherwise data will be normally distributed.

4.4 T-Test

A statistically significant t-test result is one in which a difference between two groups is unlikely to have occurred because the sample happened to be atypical. Statistical significance is determined by the size of the difference between the group averages, the sample size, and the standard deviations of the groups. For practical purposes statistical significance suggests that the two larger populations from which we sample are "actually" different.

The common assumptions made when doing a t-test include those regarding the scale of measurement, random sampling, normality of data distribution, adequacy of sample size and equality of variance in standard deviation.

5 Results for Stress Validation

5.1 Heart Rate signal:

Table 2: P-Adj Values for HR

Phase/Mode	CD	ED	MD
Phase - 1	6.275511e-02	0.0639998445	1.614792e-04
Phase - 2	5.327824e-08	0.0001798596	2.615559e-11
Phase - 3	6.847083e-01	0.6298101581	3.591908e-02
Phase - 4	6.103371e-05	0.0154088981	9.690576e-10
Phase - 5	8.867425e-01	0.3369402359	4.226014e-01

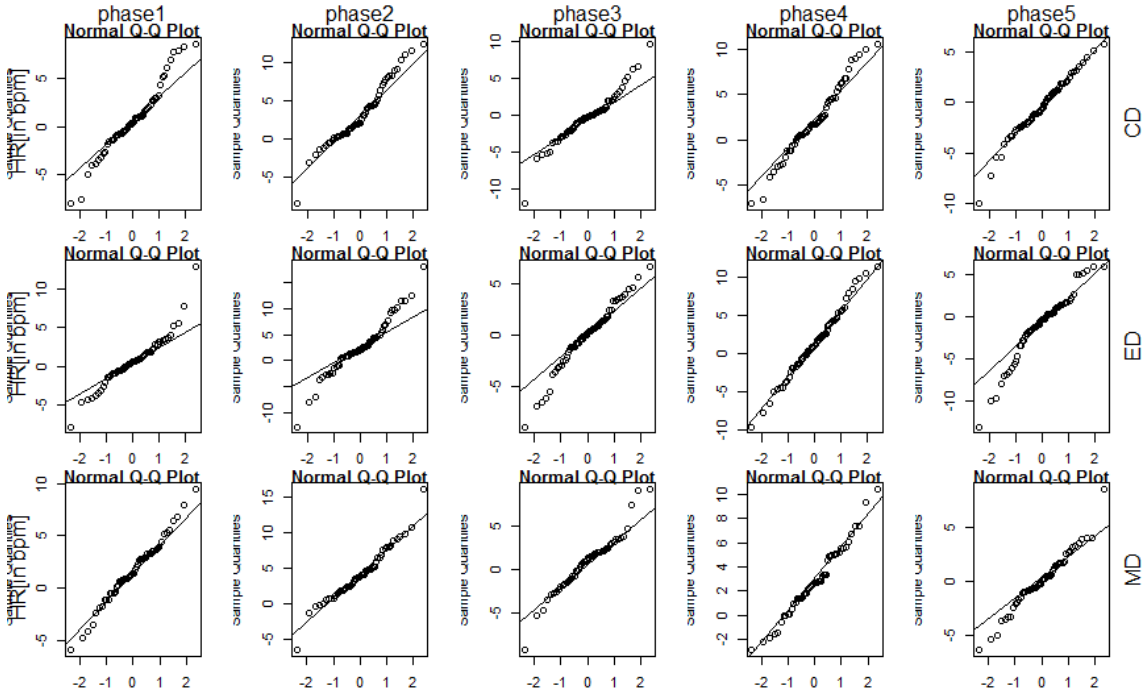

Figure 6: QQ Plot for Heart Rate.

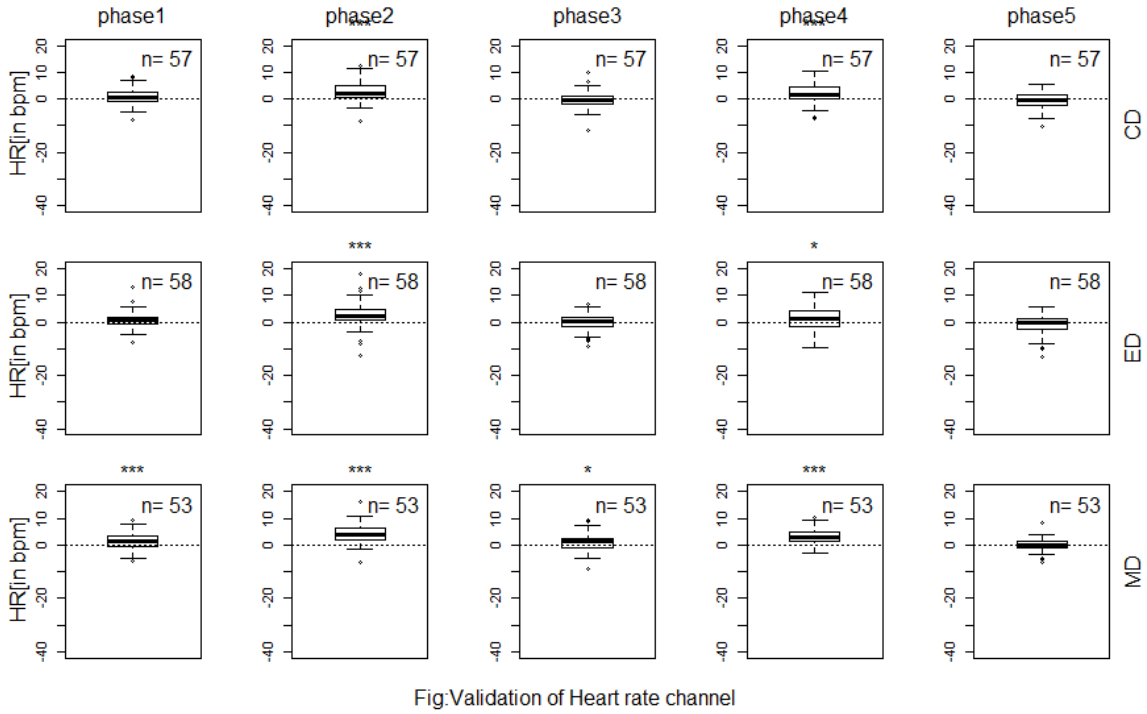

Figure 7: Box plot for Heart Rate

5.2 Breath Rate signal:

Table 3: P-Adj Values for BR

Phase/Mode	CD	ED	MD
Phase - 1	0.2378767	0.4241604	1.000000e+00
Phase - 2	1.0000000	1.0000000	1.409785e-12
Phase - 3	1.0000000	1.0000000	1.596361e-01
Phase - 4	1.0000000	1.0000000	8.347758e-14
Phase - 5	1.0000000	1.0000000	1.884806e-02

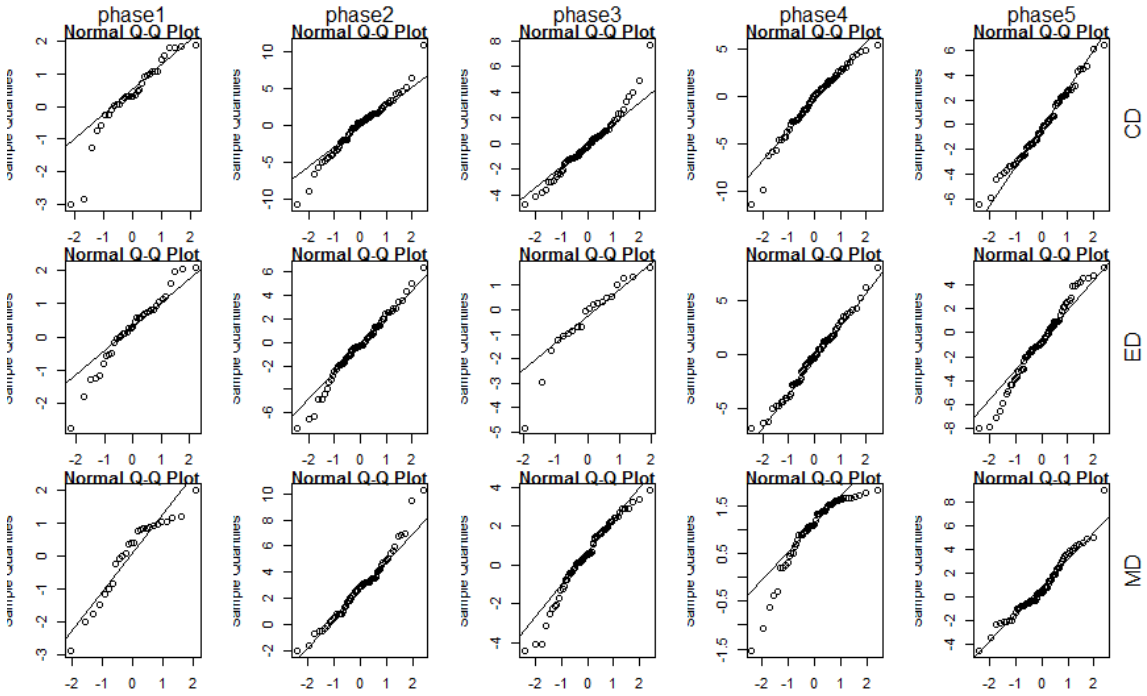

Figure 8: QQ Plot for Breath Rate.

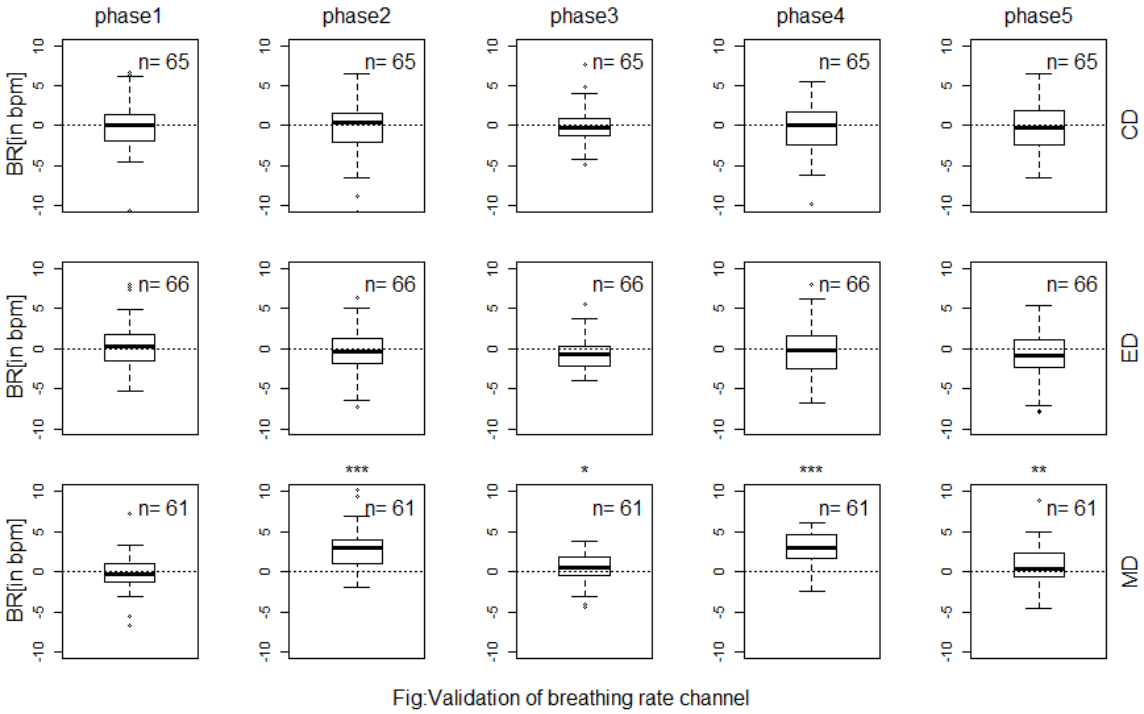

Figure 9: Box plot for Breath Rate

5.3 PP signal:

Table 4: P-Adj Values for PP

Phase/Mode	CD	ED	MD
Phase - 1	0.1078357932	5.947057e-02	1.234169e-01
Phase - 2	0.0000737499	8.886262e-06	8.929479e-05
Phase - 3	0.0051491598	9.529881e-02	1.374344e-01
Phase - 4	0.0001202307	1.417759e-04	2.813087e-03
Phase - 5	0.5906642940	3.814127e-01	6.221029e-01

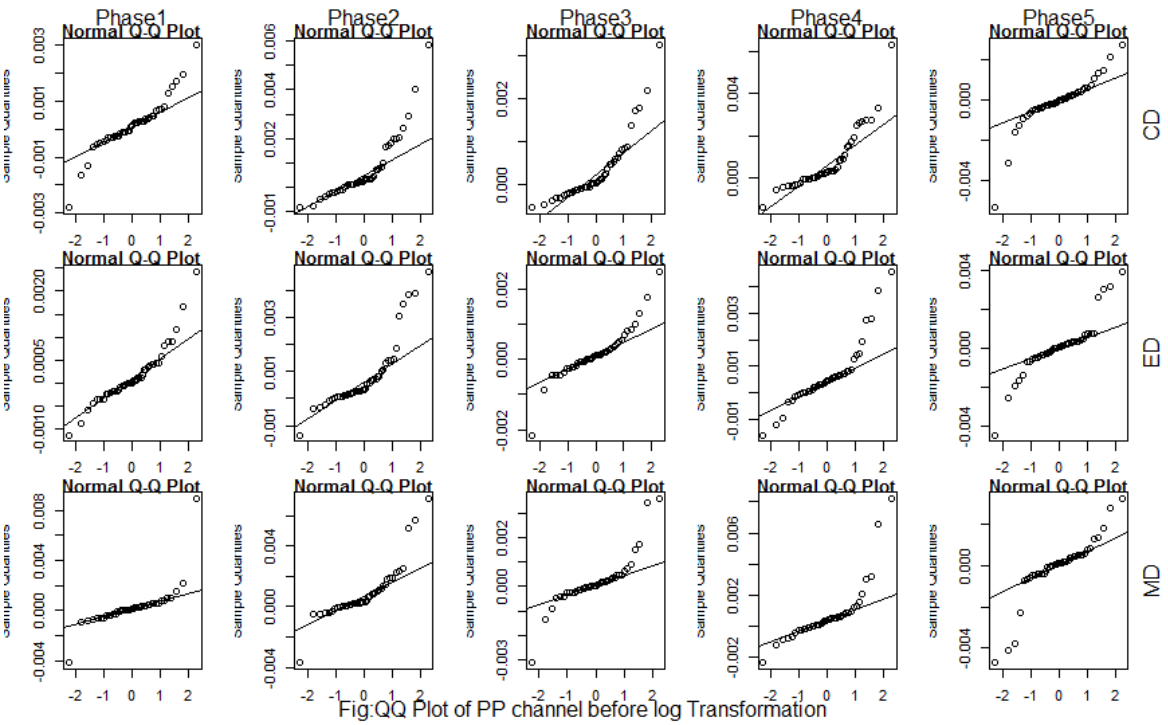

Figure 10: QQ Plot for PP signal before Normalization

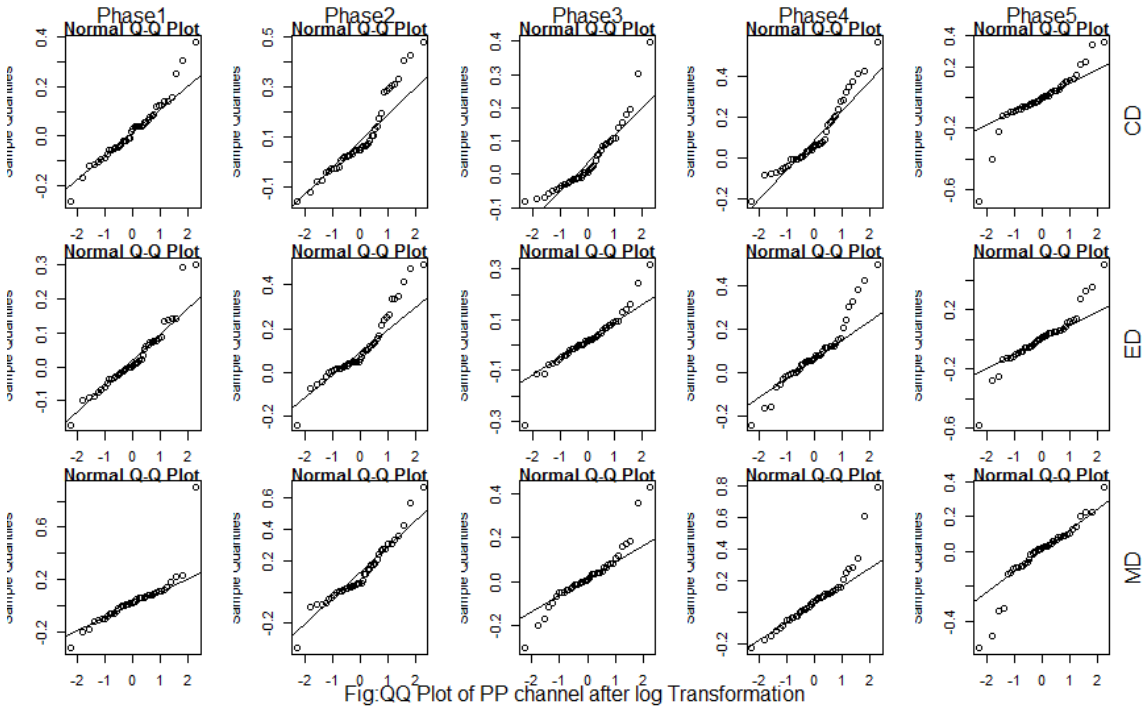

Figure 11: PP signal after Normalization

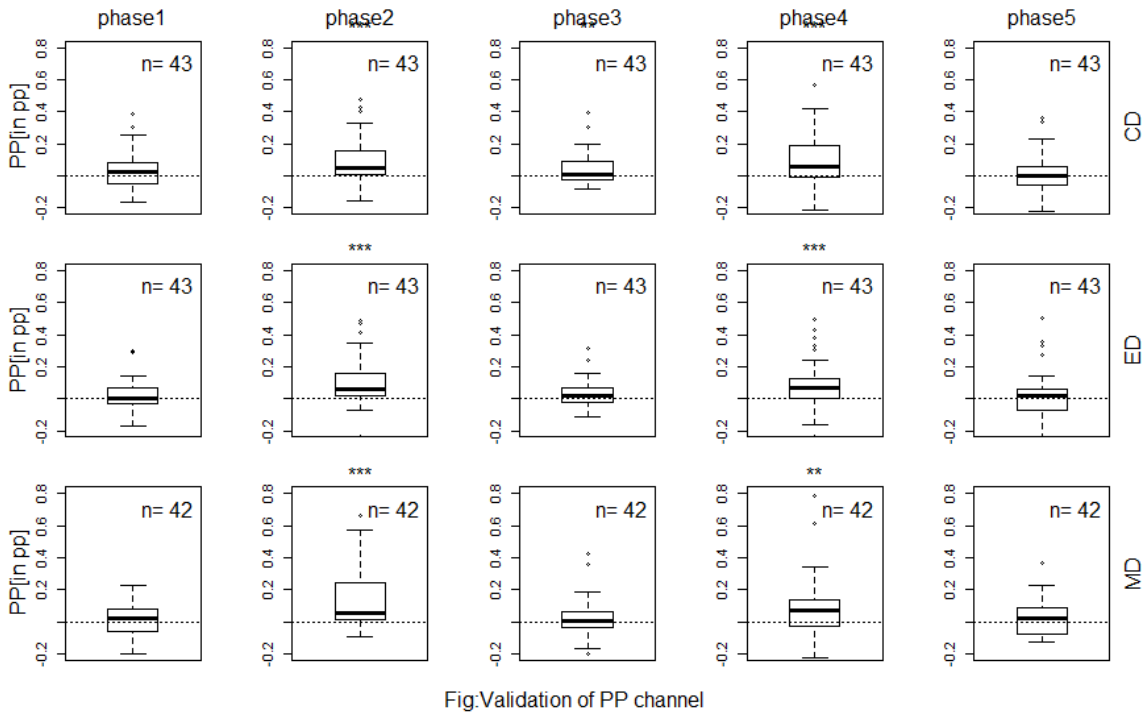

Figure 12: Box plot for PP signal

5.4 PEDA signal:

Table 5: P-Adj Values for PEDA

Phase/Mode	CD	ED	MD
Phase - 1	0.2195839	0.6841143	0.4404561
Phase - 2	0.5205684	0.9780811	0.7809731
Phase - 3	0.0876236	0.9951422	0.9146097
Phase - 4	0.9188388	0.9956869	0.9877445
Phase - 5	0.9235092	0.9882091	0.9219716

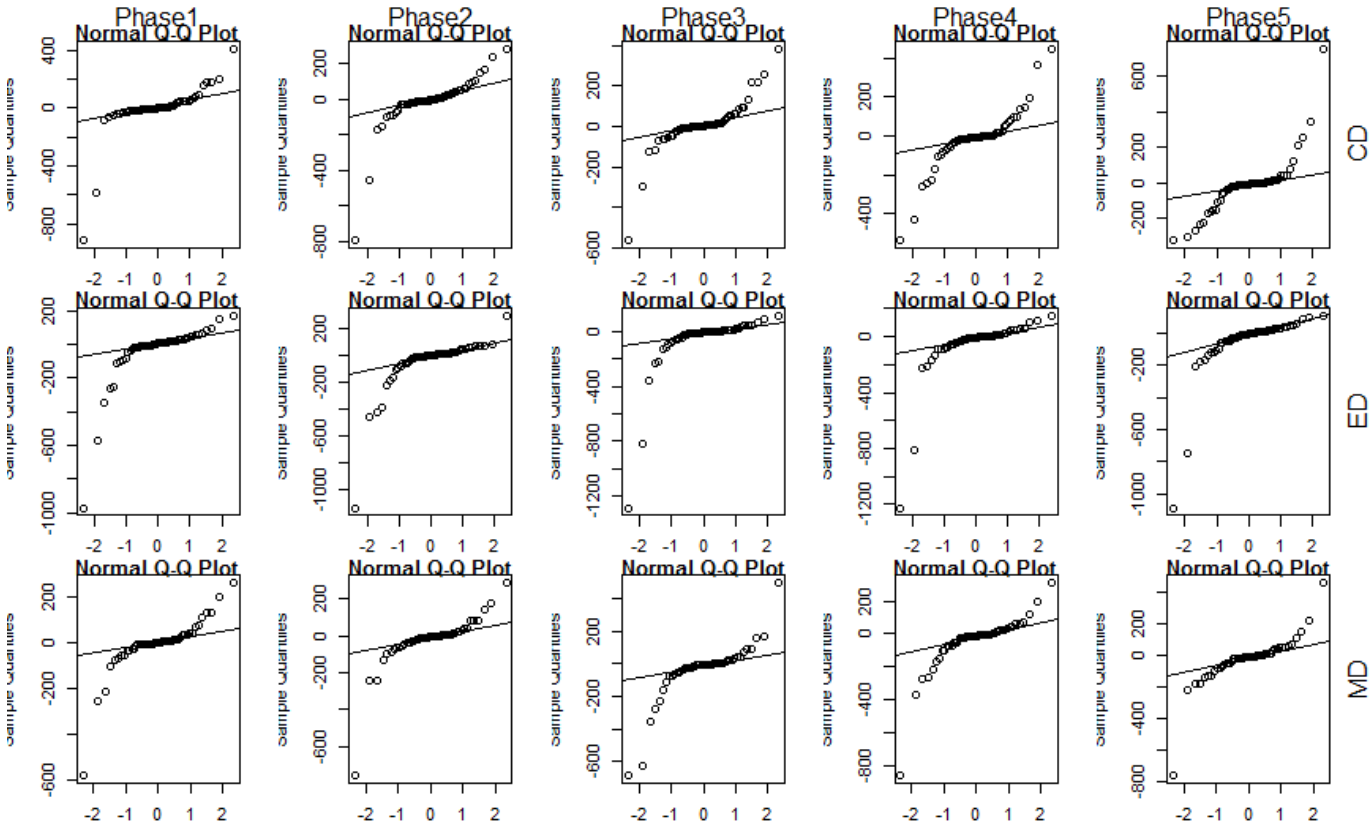

Figure 13: QQ Plot for PEDA signal before Normalization

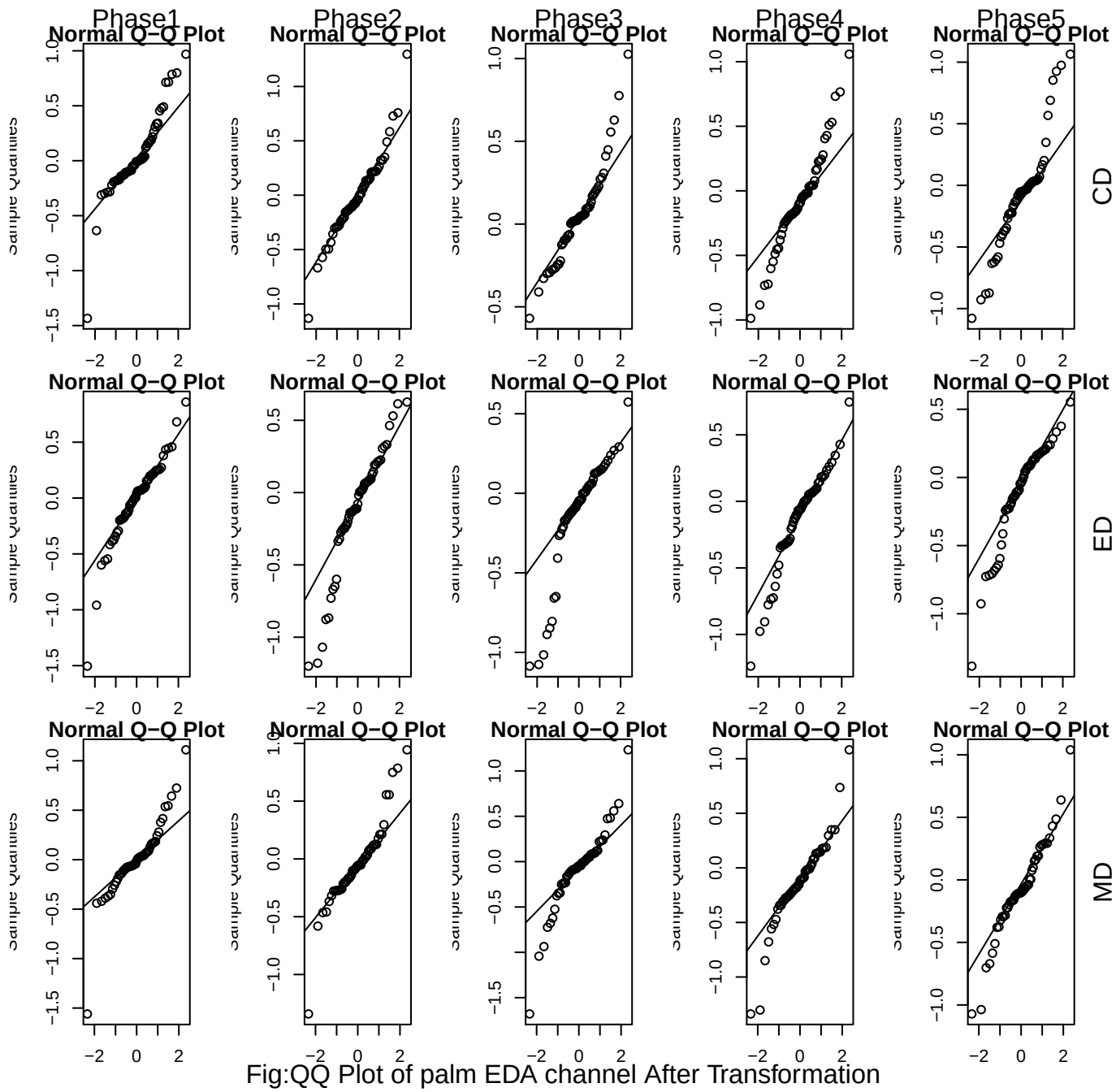

Figure 14: PEDA signal after Normalization

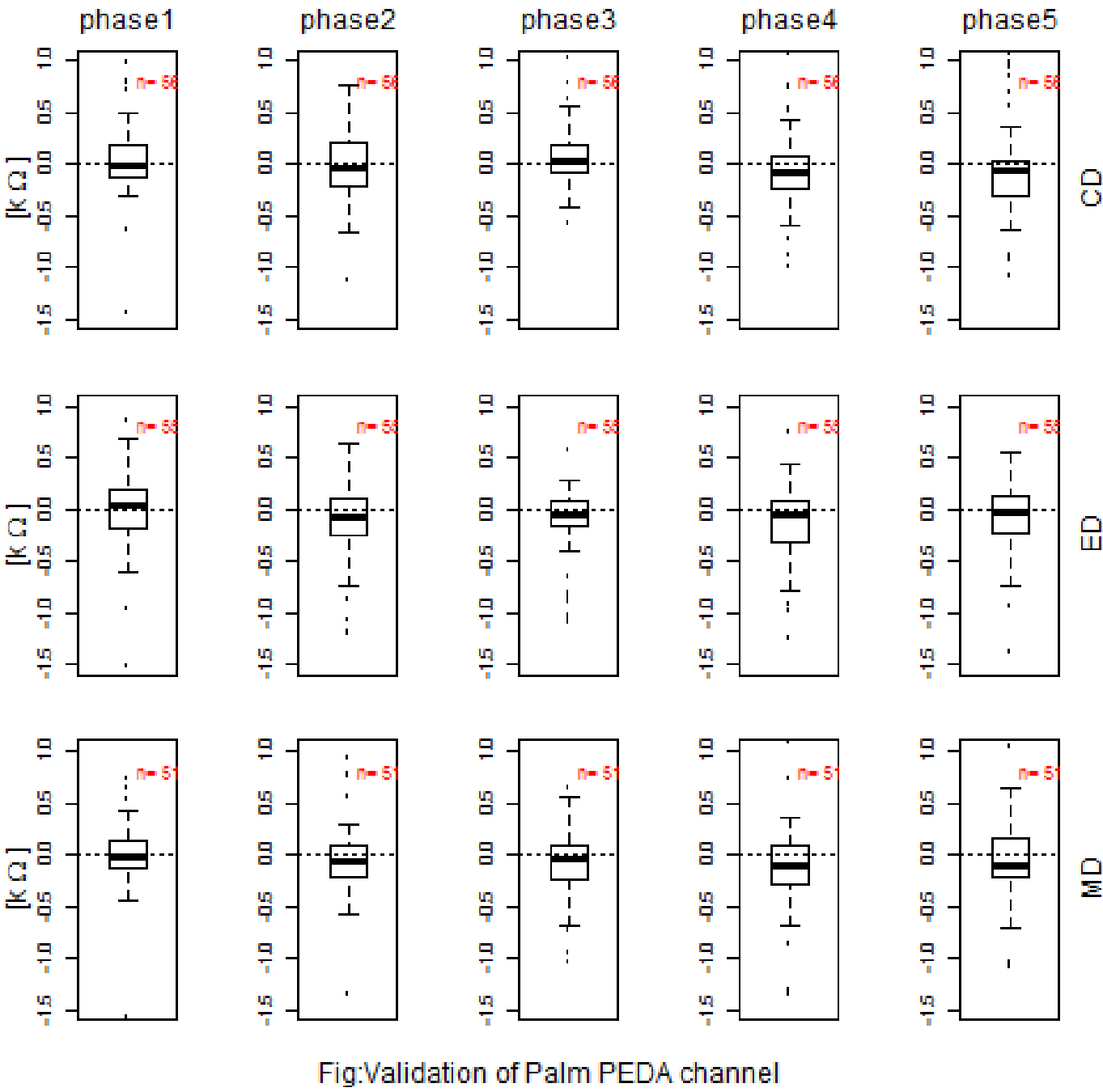

Figure 15: Box plot for PEDA signal

6 Observations

From the above plots, we can draw observations from resultant box plots on different signals of various driving modes.

- HR
  - There is a significant elevation in HR signal due to the stress applied.
  - The elevations can be observed in phase 2 of driving modes - (CD,ED,MD) and phase 4 of driving modes - (CD,MD).

  Therefore, null hypothesis is rejected.

- BR
  - There is a significant elevation in BR signal due to stress applied observed only in Sensorimotor driving mode but not in other modes.

  Therefore, null hypothesis is accepted.

- PP
  - There is a significant elevation in PP signal due to the stress applied.
  - The elevations can be observed in phase 2 and phase 4 of all the driving modes i.e. CD,ED and MD.

  Therefore, null hypothesis is rejected.

- PEDA
  - There is no significant elevation in stress in any of the driving modes.

  Therefore, null hypothesis is accepted.

7 Conclusion

- The significant elevations due to stress can be observed using HR and PP signals.
- PEDA and BR signals doesn't provide any significant elevations in effect of stressors applied.

8 Challenges

- The dataset is huge.
- We faced difficulties normalizing the PEDA and PP signals.
- Finding out the stress files and signal files in the same folder.