Project 2

Voeurn Yong Ann
y.voeurn@innopolis.university

1. Task description

In this report, we present a data analysis of EEG data of Mona Lisa portraits using a linear and nonlinear method. EEG was recorded at a sampling rate 250 Hz on 5 different participants with 31 electrodes and 3 referents.

Data set description:
Mona Liza portraits color images with different brightness values I were captured during 60 second time intervals with 10 different values of I from 0.1-1.0 with a step of 0.1. The values of I = 0 and I = 1 correspond, respectively, to 0% and 100% of natural pixels' luminance of the picture.

2. Linear Method

In this section, Pearson correlation was used as a linear method to analyse the EEG data. Pearson correlation [1] is the bivariate correlation which is a measure of linear correlation between two sets of data and defined as following:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

In this particular task, it is used to estimate the connections between different EEG channels in α(8-12 Hz) and β (15-30 Hz) frequency regions for different values of brightness intensity I in between 0.1 to 1.0. Then finding an optimal value of I for each subject characterized by maximal connectivity.

2.1 Create two graphs α(8-12 Hz) and β for five users

In the implementation (see) α(8-12 Hz) and β (15-30 Hz) is generated with Scipy library *scipy.signal.firwin* and convulse with each channel of the given dataset. Then calculating means performance Pearson correlation coefficient in each regions. Consequently, we get the result as shown in Fig.8.

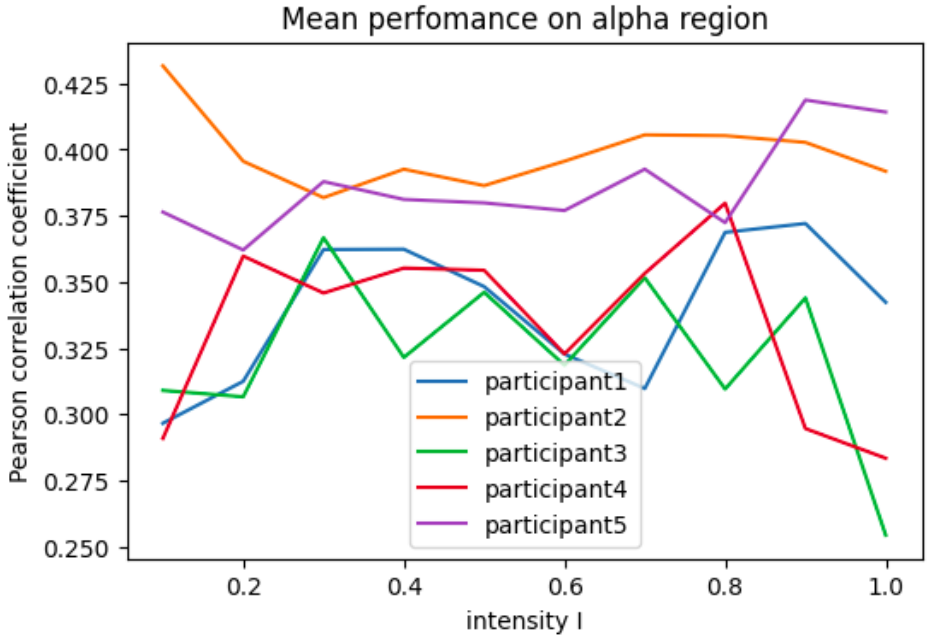

Figure 1: Mean of Pearson correlation coefficient of region α(8-12 Hz)

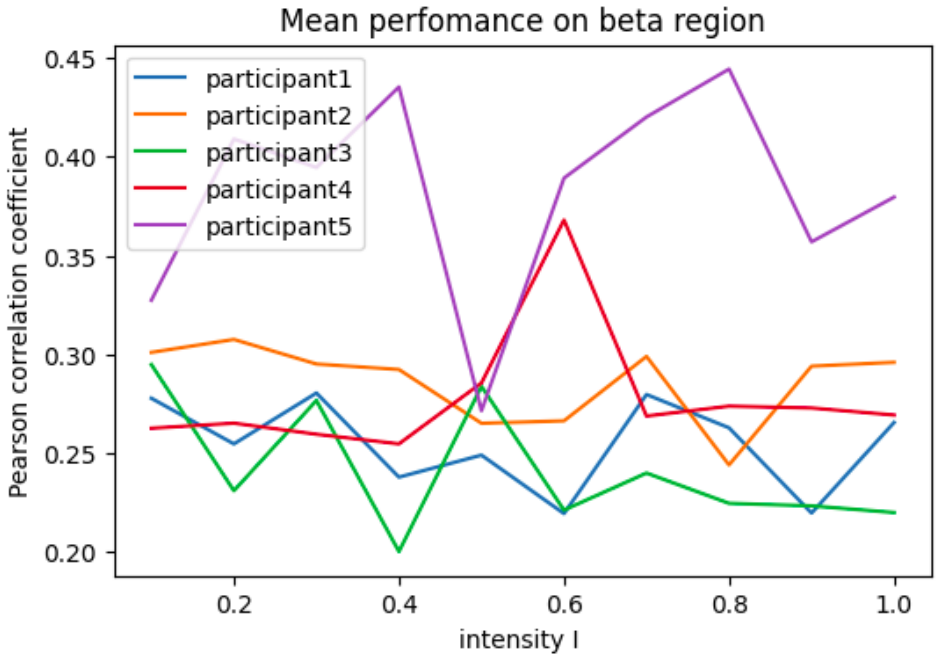

Figure 2: Mean of Pearson correlation coefficient of region β (15-30 Hz)

The graphs (Figs.1 and 2) show that the maximal connectivity of alpha regions at intensity of each person respectively 1 to 5, is 0.9, 0.1, 0.3, 0.8, and 0.9. Another region β the highest peaks of correlation of people 1 to 5 are at the following sequences 0.3, 0.2, 0.1,

0.6 and 0.8. Consequently, we can conclude that different participants have the maximal connectivity at different intensities.

2.2 Maximum co-efficient value matrix for I

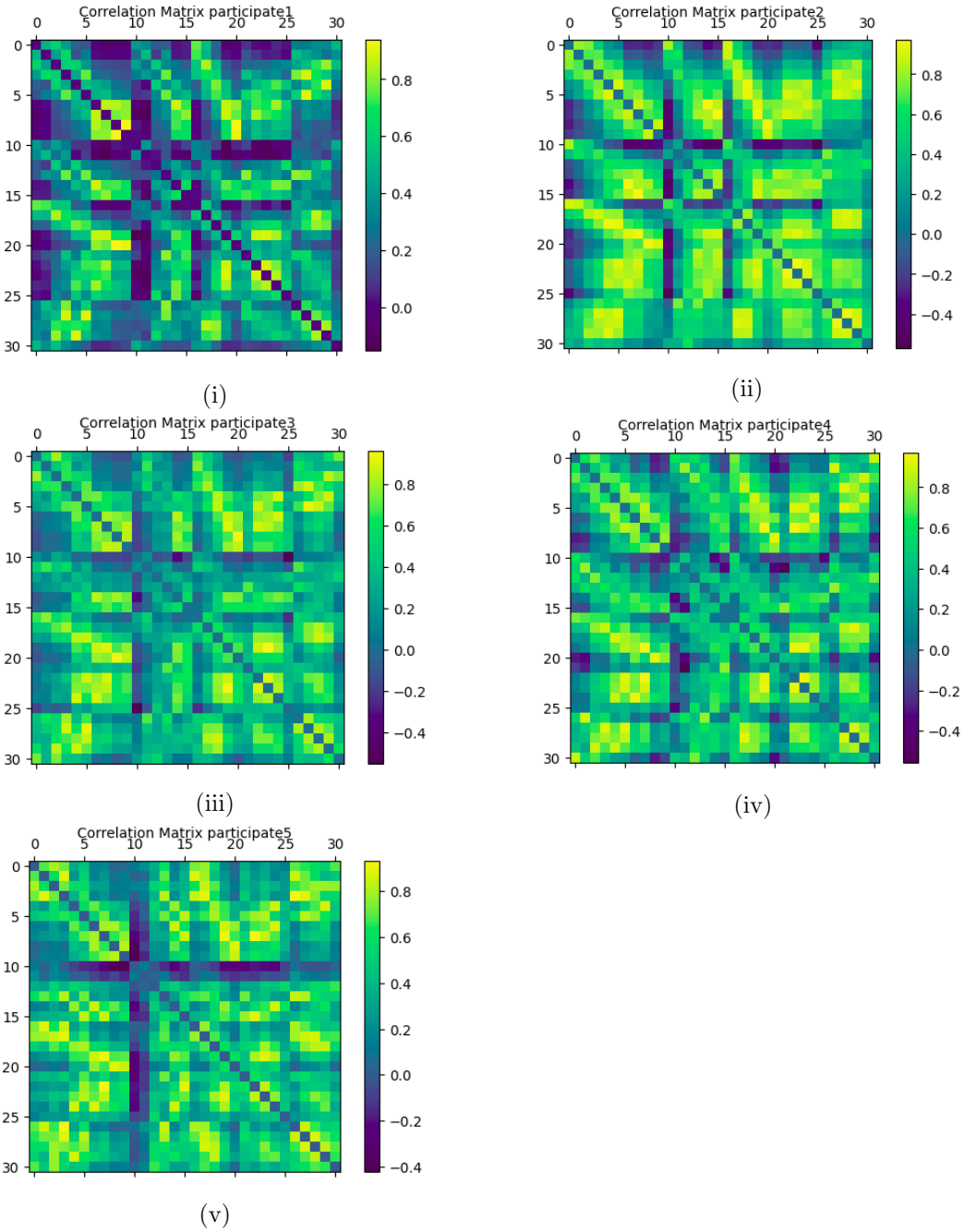

Figure 3: α(8-12 Hz): co-efficient value matrix for I [31*31 channel]

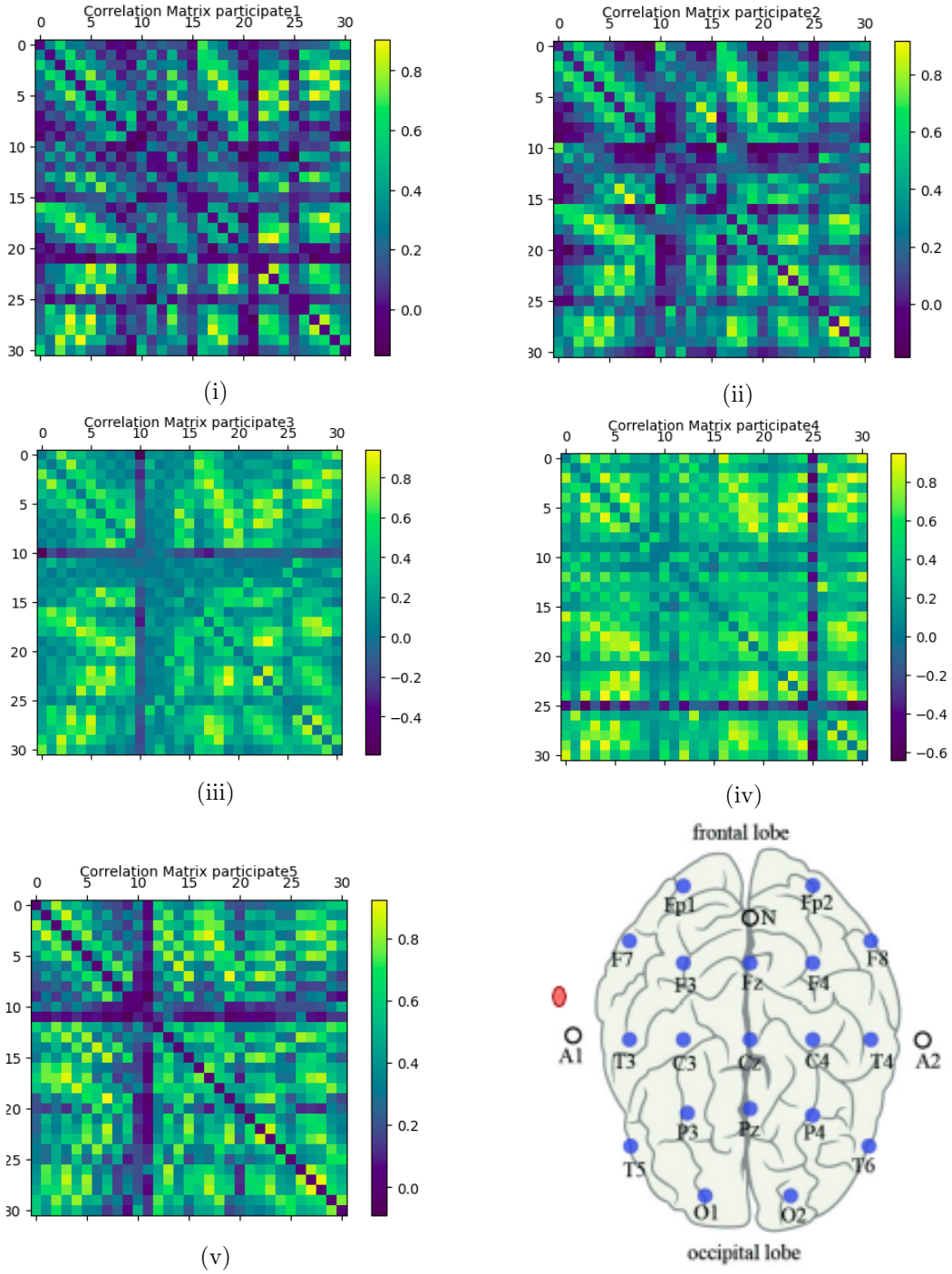

Figure 4: β (15-30 Hz) regions:Maximum co-efficient value matrix for I [31*31 channel] ;Each electrode channel number 1-31 corresponds to location as shown in (vi) with the following names: 1. O2 12. T5 23. FC3 2. O1 13. T4 24. Fcz 3. P4 14. T3 25. FC4 4. P3 15. F8 26.FT8 5. C4 16. F7 27. TP7 6. C3 17. Oz 28. CP3 7. F4 18. Pz 29. Cpz 8. F3 19. Cz 30.CP4 9. Fp2 20. Fz 31. TP8 10. Fp1 21. Fpz 11. T6 22. FT7

3. None-linear method

In this section, Feed-forward artificial neural network was used to define the connectivity by feed the data of each channel as input and output channel .

After that we quantify the degree of functional dependence, we use a metric called R^2-score [2] (coefficient of determination), which evaluates "goodness of fit" of the original data collected from a response system and is defined as:

$$R^2 = 1 - \frac{\sum_{d=1}^{D}\sum_{i=1}^{N}(y_d(t_i) - y^{'}(t_i))^2}{\sum_{d=1}^{D}\sum_{i=1}^{N}(y_d(t_i) - \bar{y}(t_i))^2} \tag{2}$$

In the first trial, we train the model and calculate R^2 as shown in Fig.5.

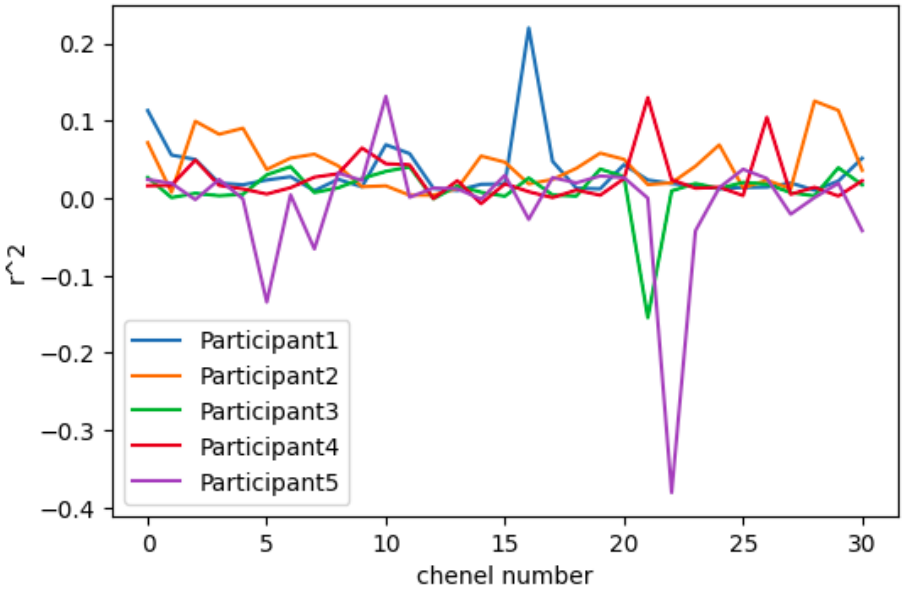

Figure 5: R^2 scores in each channel

3.1 Time-dependence

In this task, we investigate on the time-dependence of the coupling strength, finding an optimal value of the time window for calculation of the coupling strength.

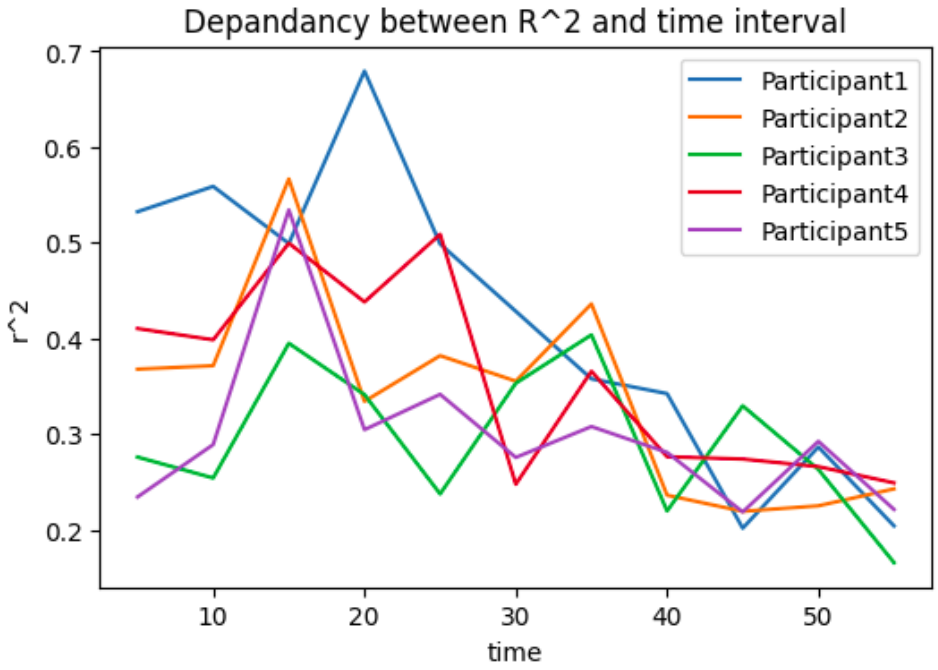

Figure 6: R^2 scores in each time interval

First, we created time window 5s, 10s, 20s, ...,60s. Then the model is trained and calculate R^2 in each window time interval. The result are shown in fig.6. These results are summarise in the table below.

Time interval of maximum connectivity of each participant:

Participant	Time interval	channel number(name)
Participant 1	20s	16(F7)
Participant 2	15s	28(CP3-A1)
Participant 3	35s	6(C3-A1)
Participant 4	24s	21(Fpz-A2)
Participant 5	15s	10(Fp1-A1)

4. Statistical test

In this task, Non-parametric test are used for checking the hypothesis of achieving maximal connectivity in the found range of the optimal brightness intensity.

Chi-squared test

The Chi-Squared test [3] is a statistical hypothesis test that assumes (the null hypothesis) that the observed frequencies for a categorical variable match the expected frequencies for the categorical variable.

In our particular case, we are checking whether the range of optimal brightness intensity that has been found could be accepted or rejected.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

In the implementation, "scipy.stats.chisquare" was used to calculated a one-way chi-square test and got the following result:

```
participant 1
max intensity = 0.9
stat=180.964, p=1.000

participant 2
max intensity = 0.1
stat=289.277, p=1.000

participant 3
max intensity = 0.3000
stat=222.654, p=1.000

participant 4
max intensity = 0.8
stat=250.622, p=1.000

participant 5
max intensity = 0.9
stat=189.570, p=1.000
```

(i) χ^2_α

```
participant 1
max intensity = 0.3000
stat=210.123, p=1.000

participant 2
max intensity = 0.2
stat=197.746, p=1.000

participant 3
max intensity = 0.1
stat=218.707, p=1.000

participant 4
max intensity = 0.6000
stat=233.323, p=1.000

participant 5
max intensity = 0.8
stat=118.464, p=1.000
```

(ii) χ^2_β

Figure 7: The result of Chi-Squared test

Interpretation:

- H0: the two samples are independent.

- H1: there is a dependency between the samples.

If $p \leq 0.05$ independent (fail to reject H0), otherwise dependent (reject H0).

Base on the result we can see that the frequencies for a categorical variable does not match the expected frequencies for the categorical variable.

T-test

T-test [4] is a statistical hypothesis test whether the means of two independent samples are significantly different.

$$t = \frac{m - \mu}{s/\sqrt{n}} \tag{4}$$

Assumption: H0: the means of the samples are equal. H1: the means of the samples are unequal.

If $p \leq 0.05$, so probably has the same distribution; otherwise it has different.

In the implementation, "scipy.stats.chisquare" was used to calculated a one-way chi-square test and got the following result:

```
participant 1
stat=7.710, p=0.000
Probably different distributions


participant 2
stat=6.423, p=0.000
Probably different distributions


participant 3
stat=3.757, p=0.000
Probably different distributions


participant 4
stat=-5.023, p=0.000
Probably different distributions


participant 5
stat=-2.112, p=0.035
Probably different distributions
```

Figure 8: The result of T-test

Full implementation of this paper can be found within this Colab's Link

References

[1] Wikipedia, "Pearson correlation coefficient — Wikipedia, the free encyclopedia." http://en.wikipedia.org/w/index.php?title=Pearson%20correlation%20coefficient&oldid=1008691934, 2021. [Online; accessed 27-February-2021].

[2] N. Frolov, V. Maksimenko, A. Lüttjohann, A. Koronovskii, and A. Hramov, "Feed-forward artificial neural network provides data-driven inference of functional connectivity," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 9, p. 091101, 2019.

[3] Wikipedia, "Chi-squared test — Wikipedia, the free encyclopedia." http://en.wikipedia.org/w/index.php?title=Chi-squared%20test&oldid=1008237521, 2021. [Online; accessed 03-March-2021].

[4] Wikipedia, "Student's t-test — Wikipedia, the free encyclopedia." http://en.wikipedia.org/w/index.php?title=Student's%20t-test&oldid=1009373024, 2021. [Online; accessed 03-March-2021].