The reliability of an evoked brain response measured with electroencephalography (EEG)

Annabel Hazewinkel (4478223), Julia van der Elst (4307240),
Michiel Tjeerdsma (4438485), Thomas van Hengel (4235959)

***Abstract*—Electroencephalography (EEG) is a widely used method to measure brain activity. Since the brain is a complex system, the recorded signals contain a lot of noise. By applying perturbations, an evoked brain response can be measured. The signal to noise ratio (SNR) is a measure of the quality of this response in which noise is present. The reliability of this response can be determined by calculating the intraclass correlation coefficient (ICC). The acquired ICCs show a good level of significance, which is in line with the literature.**

I. Introduction

The brain is a complicated system. Herculano-Houzel (2009) concludes that a piece of brain tissue the size of a grain of sand contains 100,000 neurons and one trillion synapses that all ”talk” to each other. That is why it is very important that the working of it is well understood. A technique that contributes to this knowledge is electroencephalography, also known as EEG.

Electroencephalographic measurements record the electrical activity of the brain and are often used in medical and research areas. Those measurements can contribute to the investigation into the working of the brain and what happens in case the brain gets damaged. EEG measures potential differences caused by the ion flow in the nerve cells of the brain. These potential differences are measured by electrodes. A cap containing electrodes is placed on the head of the participant. Since the measured differences are in microVolts (μV), an amplifier is used to enhance the signal for display.

According to Rana (2017) sixteen channels are used for a scalp EEG when doing a medical analysis. However, when doing research, more channels are of need to precisely map the brain activity. Therefore, in this research, a cap containing 128 electrodes will be used.

When measuring brain activity it is of high importance for the researcher that he or she considers the reliability of what is measured. A few number of studies can be found when researching the reliability of this measurement technique. However not with the specifications of this research. For example, a study by Salinsky et al. (1991) and a study by Moezzi (2018) examine the reliability of EEG frequencies in rest state. Most studies regarding this subject are done in rest state: the participant is not receiving any stimulus. For this research, a perturbation is used to study the evoked brain response, an innovative concept in nowadays research. By applying the same perturbation and then studying the evoked brain response, the reliability can be checked of this response.

Studying this response will be done by using the signal to noise ratio (SNR). A lot of noise in the signal is present by nature during EEG measurements. The signals from the brain are really small. Since these signals are strongly amplified, the noise will be amplified as well. This noise is caused by several circumstances such as blinking with the eyes or a distraction. This may reduce the reliability of the response that is analysed. The SNR is a good index to indicate the quality of a signal in which a disturbing noise is present.

The reliability of EEG will be determined by using intraclass correlation coefficient (ICC). In order to be able to draw conclusions regarding evoked brain response, it is important to consider the test-retest reliability of the EEG measurements in order to obtain valuable data. The ICC is an often used index to determine the test-retest reliability. The ICC indicates the reproducibility of a certain output. In this research, the analysed output will be the SNR. Per participant the SNR of the active tasks will be analysed and via the ICC the correlation of those outputs will be determined. An ICC can vary between 0 and 1, respectively from lowest to highest reliability.
this research has been conducted as part of the Bachelor End Project for Mechanical Engineering students at the TU Delft in the third year of their Bachelor studies.

II. Research question

What is the reliability of an evoked brain response measured with electroencephalography (EEG), when applying perturbations to the wrist?

III. Hypothesis

EEG is a commonly used method to measure brain activity. This implies that the reliability of this testing method should not be too low, otherwise the measurements done so far, would be of no use. The SNR is a commonly used output for determining the quality of the signal relative to the noise. Also within neuroscience, the SNR is used to describe the activity of the brain as stated by Schultz (2007). This implies that obtaining an ICC by using SNR would be a justified method. A study by Rogers et al. (2016) uses the ICC to determine the reliability of EEG. However, this study differs in procedure and amount of participants. This results in an ICC between 0.57 and 0.85. Nevertheless, the hypothesis of this research states that the ICC will resemble these results since the same method for obtaining the ICC is used.

IV. Methodology

Participants
The data will be acquired by measuring brain activity of ten participants. The participants are all right handed, healthy and between the age of 18 and 30.

Testequipment
The tests will be executed at the TU Delft in a special isolated chamber, to cancel out distractions. Each participant will wear an EEG cap with 128 electrodes as shown in Figure 1.A. These electrodes make contact with the skin. A gel will be applied between skin and electrode to lower the electrical resistance. An impedance check (Appendix A Figure 1) is made to check if the impedances are low enough. The measurements can start when all electrodes are connected properly. For the tests, a machine called the Wristalyzer will be used. This haptic robot, shown in Fig. 1.C, inflicts small vibrations to the right hand of the participant. The muscle activity of the right forearm will also be monitored by 4

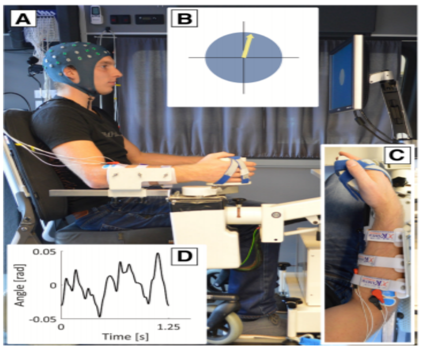

Fig. 1: A: Experimental setup, B: What the participant sees on the TV screen, the circle and crosslines are always visible. The arrow only during the force task. C: The right forearm of the participant is strapped into the Wristalyzer, with the 4 EMG electrodes. D: one repetition of the signal the haptic robot follows. Reprinted from Vlaar et al. (2017). Reprinted with permission.

electromyography (EMG) electrodes. EMG records the electrical activity produced by muscles. This data will not be used for this research. The signals will be filtered with special software. EEGlab and Fieldtrip are specialized MATLAB Toolboxes to analyse the signals of EEG.

Procedure

The participant will carry out three different tasks twice on two different days. This research only takes into account the two series of day one. The other data is used for a different research. The first task is a resting state. The participant will not actively do anything and is trying to relax. This task will be carried out 5 times, each lasting a minute. The second task involves the Wristalyzer.The Wristalyzer is strapped to the right forearm and hand. The forearm is fixed and the hand will be moved by the robot, as seen in Fig. 1C. The robot inflicts small vibrations, along a specific perturbation signal, which repeats itself during the measurement. The left part of the brain will be activated due the fact that the Wristalyzer will move the right arm (Cunnington, 2016). The participant should not exert any force on the Wristalyzer. This test will be carried out 12 times each lasting 50 seconds. The last task is very similar to the second task. Now the participant is asked to exert a certain amount of force on the Wristalyzer. This force is measured and displayed on a screen for the participant to see. Now the participant can adjust the amount of force he/she is applying to reach the desired amount of force. This test will be carried out 24 times and will last 25 seconds each. The force needed for this task may be tiring. To make sure fatique is not an influencing factor, the entire test is divided in more repetitions in comparison to the relax task.

The tasks consist of multiple repetitions. Each repetition is also divided into shorter parts called epochs. Epochs have a duration of 1.25 seconds. EEG is a noisy signal, so a large amount of epochs is chosen to be measured. An overview of the measurements and the amount of epochs per task are shown in Fig. 2.

Data cleaning

For the analysis only the relax and force task will be analysed. The first task, the resting state, will not be used for this analysis. The data retrieved by executing this task is used for other research.

Raw data

During the measurements the potential differences caused by the ion flow in the nerve cells of the brain are measured by the electrodes and plotted against time in a graph. The output of the EMG electrodes (muscle activity), the input and output signal of the haptic robot are also plotted in this graph, all shown in Appendix A in Figure 2. At the start of each measurement a trigger is added to the graph and embedded in the data. These triggers are added to be able to retrieve the starting point of each measurement later.

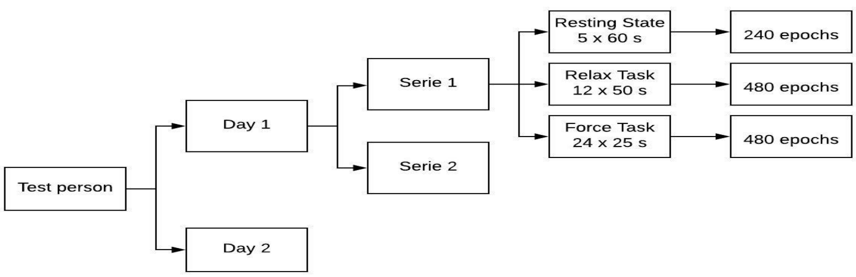

Fig. 2: Structure of the measurements

Cleaning
In order to clean the data properly, a number of steps have to be taken, shown in Fig. 3.

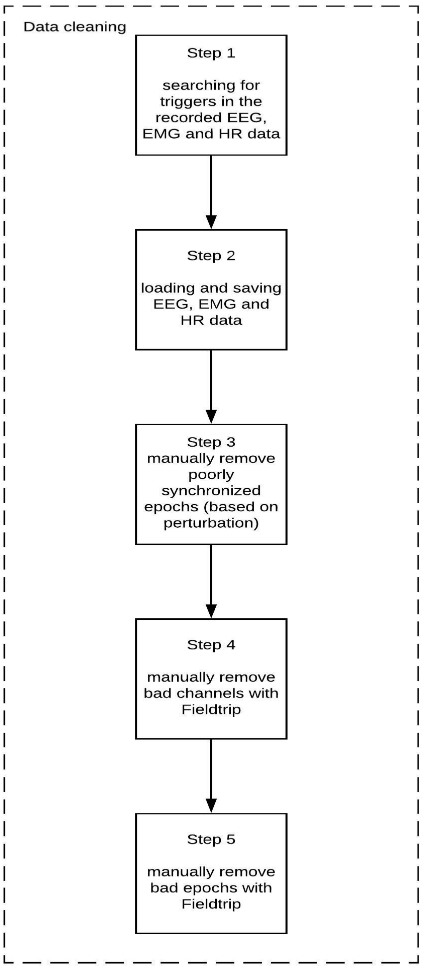

Fig. 3: Steps data cleaning

Firstly, the triggers embedded in the data are extracted. Secondly, the data is saved in a new file format for later use. Thirdly, the epochs of the haptic robot channel are checked for errors. The data in each epoch should all be distributed along a specific curve, namely the input signal of the robot. However, sometimes during an epoch, the output signal of the haptic robot is not the same as the input signal of the robot due to a mistake in the alignment of the epochs. This is caused by a miscommunication between the sent triggers and the robot. These misaligned epochs are selected and stored automatically. The faulty epochs are now known and will not be used further on. An overview is seen in Fig. 4.

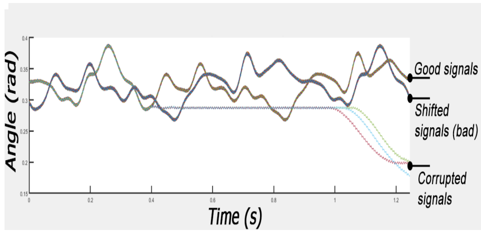

Fig. 4: Step 3 epochs of the Haptic Robot displaying a good signal, a normal signal but shifted and a corrupted signal

The channels of the EEG are also a possible source for errors. Some channels show a noisy signal or no signal and should therefore be removed in the results. Therefore the fourth step, shown in Fig. 5, consists of manually removing bad channels with Fieldtrip. FieldTrip is a MATLAB tool that shows all the epochs of one channel in one figure. This is done by the same operator. The operator can mark a channel as ‘good’ or ‘bad’ and move on to the next channel. Channels that are marked as ‘bad’ will be filtered out when all the channels are evaluated.

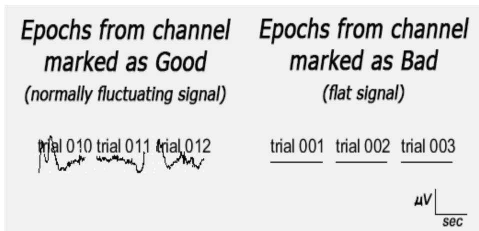

Fig. 5: Step 4 bad and good channels

Finally a measurement can be distorted due to muscle activity or blinking. Both are clearly visible, shown in Fig. 6. Again FieldTrip is used to plot the graph of all remained channels against time. The operator can now select parts of the data that are to be taken out. FieldTrip deletes this data.

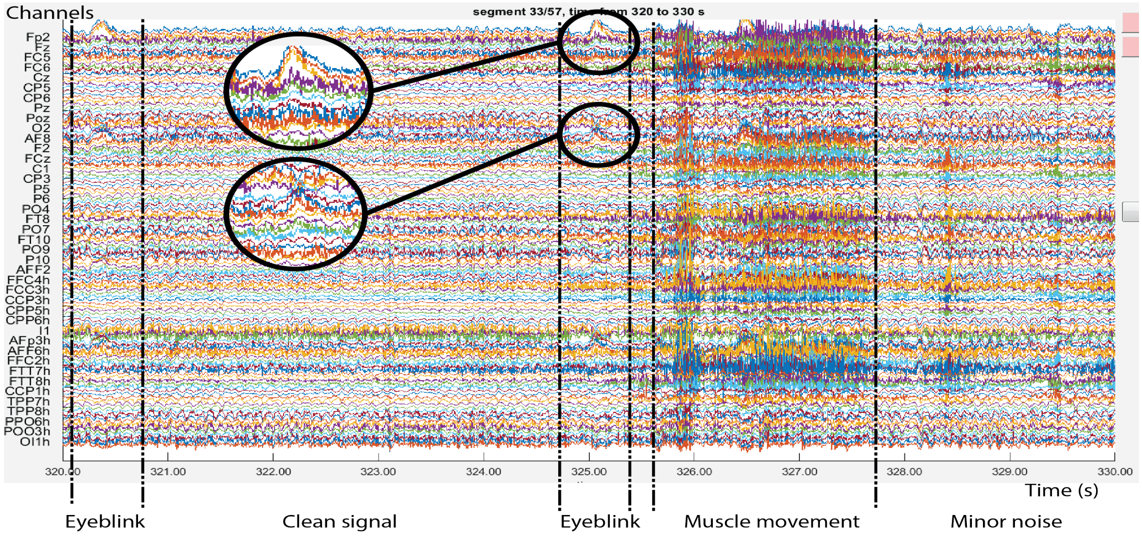

Fig. 6: Step 5 part of EEG measurement showing different kinds of noise; eyeblinks, muscle movement and minor noise.

Data analysis

After cleaning the data, it is ready to be analysed. Firstly, the SNR will be calculated. This will be the major indicator for the test-retest reliability. Secondly, the topographic maps will be analysed in order to select the channels that highly respond to the evoked brain response. Thirdly, the ICC will be calculated. An overview of the analysis is shown in Fig. 7.

SNR

Firstly, in order to calculate the SNR, all the epochs of one channel will be placed in a single epoch time span of 1,25 seconds. This is done by calculating the mean of all epochs, shown in formula (1). (Vlaar et al. ,2017)

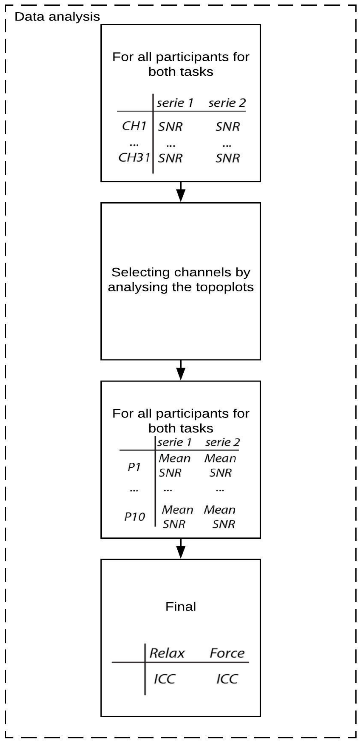

Fig. 7: Steps data analysis

The SNR is a ratio of the power of the steady-state response E and the variance σ^2, seen in formula (2). Where $\hat{x}(k)$ is the average of the recorded signal per epoch, x is the recorded signal, k is a step in epoch p. An epoch is made out of N time steps of k. In this research N is 2560 time steps. Each k has a signal value $x[p]$. P is the amount of epochs, which is 480 per task. Each channel will receive its own SNR.

(1) $$\hat{x}(k)=\frac{1}{P}\sum_{p=1}^{P} x^p(k)$$

(2) $$SNR=\frac{\hat{E}_x}{\hat{\sigma}_x^2}=\frac{\sum_{k=1}^{N} x^p(k)^2}{\sum_{k=1}^{N}\frac{1}{P-1}\sum_{p=1}^{P}(x^p(k)-\hat{x}(k))^2}$$

After these calculations there will be an SNR for every channel for every task per serie. These values can be plotted in a topographic map to provide a clear view of the location of the calculated SNR values.

Topographic maps
Fig. 8 shows an example of a topographic map. In this figure the differences in SNR values are clearly visible. Blue areas display low values and yellow areas show high values. According to Vlaar et al. (2017), there was a fixed set of channels to be analysed, located at the left brain half. However in this research, the topographic maps were analysed in order to determine precisely the channels for which the evoked brain response is active. The mean of the SNR values are to be calculated, in order to correlate them later on into the ICC. Only the values located in the cluster with high values are useful, this was were the evoked brain response occurred. A selection of channels must be made in order to enclose the cluster with high values. The SNR of the selected channels will be averaged to a mean SNR per participant per serie per task.

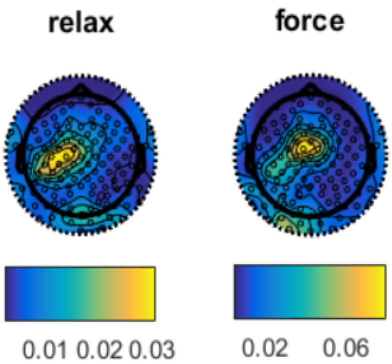

Fig. 8: Example of a topographic map

ICC
The last step in the data analysis is the calculation of the intraclass correlation coefficient (ICC). Ten different forms of ICCs can be chosen. The form applicable for this research is:

Two-way mixed effects, absolute agreement, single raters/measurements.

To determine to this form, the flow diagram of "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research " by Koo & Li (2016) is used.

This form is calculated by formula (3) where MS_R is the mean square for rows, MS_E mean square for error and MS_C mean square for columns. k is the number of series per participant and n is the number of participants. The program SPSS, statistical computer software, is used in order to obtain our results.

(3) $$ICC=\frac{MS_R-MS_E}{MS_R+(k+1)MS_E+\frac{k}{n}(MS_C-MS_E)}$$

The ICC will be calculated with the mean of the SNRs of the selected channels for every participant. In the end, one ICC for the relax task and one for the force task, reflecting the reliability, will be obtained.
An ICC can vary between 0 and 1, respectively from lowest to highest reliability and divided into significance levels. The significance level is considered poor for ICC < 0,40, fair for 0,40 < ICC < 0,60, good for 0,60 < ICC < 0,75 and perfect for 0,75 < ICC < 1,00. (Li et al., 2015)

V. Results

As mentioned above, in order to be able to answer the research question, the SNR and the ICC needed to be calculated.

SNR
Per participant four SNRs were calculated, one per serie per task. All SNRs of the relax and the force task are plotted in order to check the correlation regarding the optimal ICC of one. These graphs are shown Appendix B in Figure 1 and Figure 2, respectively. If all points would lie on this line, evoked brain response test would be completely replicable. The deviation from the point to the line shows the degree of unreliability per person.
One SNR, the rightmost data point, shown in Appendix B Figure 1 has a substantial deviation from the ideal curve. The corresponding topographic maps, shown in Appendix C Figure 4 01, show indeed a normal plot for serie 1, but a different plot for the second serie. Here the cluster of high SNR values is located more towards the front.

All the SNRs of all the participants are shown in the appendix B in Table I.

Topographic maps
Differences in SNR values are visualized in the topographic maps shown in Appendix C in Figure 1 and Figure 2. A cluster of high SNR values is clearly visible. This area is, as expected, located on the left hand side of the brain. The exact location of the cluster of high SNR values however, is never the same per participant.
The selection is made so that the high value cluster of all topographic maps shown in Appendix C Figure 1 and Figure 2 is enclosed. This selection is shown in Appendix C Figure 3 and Figure 4.

ICC
After the SNRs were obtained, the ICC could be calculated, again using scripts programmed in MATLAB. As mentioned earlier, one ICC per task was obtained. The results are shown in Table I.

ICC	
Relax Task	Force Task
0,61	0,73

TABLE I: Intraclass correlation coefficient (ICC)

Calculating the ICC values gives an answer to the research question. The values of the relax task and the force task are 0,61 and 0,73 respectively. This gives for both tasks a ’good’ significance level.

VI. Discussion

The goal of this research was to calculate the reliability of EEG. This was done by using the signal to noise ratio (SNR) and the intraclass correlation coefficient (ICC). As a result, the ICC indicated that there is a good level of significance for both tasks. A couple of factors might have influenced the result. Starting with the amount of participants. A group of ten participants is rather small. A larger group will increase the reliability of the final calculated ICC values. The measurements themselves might have been influenced beforehand, during or after the measurements, during the filtering. This may have led to less representable results.

Before measurements
Starting with discussing the influencing factors beforehand. The factor that influenced the measurements the most is the lowering of the impedances by applying gel between the eletrodes and the skin. For some of the participants the available caps were too large. Making it more difficult to lower the impedance for certain electrodes, where in some cases the impedance did not get low enough. A smaller cap would have been a better fit and would have made it easier to make contact with the scalp. Also the amount of hair could influence the lowering of the impedance, since thick or curly hair makes it harder to make contact with the scalp.

Some electrodes were not working properly and are therefore removed from the dataset. These defects can have multiple causes: defect electrodes, a faulty cable or not thoroughly cleaning of the cap after an earlier measurement.

During measurements
The EEG signal of all electrodes was sometimes influenced by one of the electrodes during the measurements. These bad electrodes had to be removed halfway. This is due to impedance variance which might happen if the cap moves a bit. For the first few measurements, the ground sensor was not connecting correctly, this caused noise.

After measurements
Swallowing, eye-blinking and other movements of the head disturb the signal. This noise had to be filtered out. However in parts of the measurements, muscle activity dominated the recording. A lot of the data had to be removed. Resulting in a smaller set of useful data than expected.
The removing of epochs and channels is done manually by one operator. By automating the process, a more consistent, time efficient and reliable result could be obtained.

Furthermore, the effect of removing epochs and channels must be considered. A faulty channel can show a flat or a heavily fluctuating line. Now, these channels are removed in the analysis. A removed channel can be one of the selected 40 channels used for reliability analysis. If this was the case, the value of this SNR was replaced by a mean of the other channels in the selection. Adding this mean SNR value won't influence the ICC since the ICC is calculated with the mean of all SNR's. However if the bad channels would not be removed in the first place, the SNR of this channel will be calculated and used in the calculation of the mean of all SNR's. This will influence the ICC. The ICC would be lower in the case of not removing any channel. With removing epochs, the amount of data is lowered. A conclusion is then based upon less data. It is not clear how the removal of epochs influenced the data. The final value is probably not influenced much since the distribution of faulty epochs in a channel is random.

Some topographic maps show an odd result. The mean SNR values of the tasks corresponding with these bad topographic maps are currently still used in the calculation of the ICC. Removing these bad tasks will increase the final ICC value, but decrease the reliability of the ICC value since less values are used for calculation.

VII. Conclusion

The significance level for both tasks is defined as 'good' since the ICC of the relax and force task are respectively 0,61 and 0,73. The research question, 'What is the reliability of an evoked brain response measured with electroencephalography (EEG), when applying perturbations to the wrist?', has been answered. The hypothesis stated that the expected ICC value should be between 0,57 and 0,85. The values of both tasks are as expected.

Acknowledgements
We would like to acknowledge Joost van Kordelaar for his knowledge during the meetings.

Ethics approval and consent to participate
The study has been approved by the Ethics Committee of the TU Delft. All participants gave written informed consent prior the experiments.

References

Cunnington, R. (2016, September 28). How our brain controls movement and makes new connections when parts are damaged.
Retrieved October 19, 2018, from http://theconversation.com/how-our-brain-controls-movement-and-makes-new-connections-when-parts-are-damaged-63520

Herculano-Houzel, S. (2009, 9 november). The Human Brain in Numbers: A Linearly Scaled-up Primate Brain.
Retrieved October 17, 2018, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2776484/

Koo, T. K., Li, M. Y. (2016, June 1). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.
Retrieved October 19, 2018, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/

Li, L., Zeng, L., Lin, Z., Cazzell, M., Liu, H. (2015). Tutorial on use of intraclass correlation coefficients for assessing intertest reliability and its application in functional near-infrared spectroscopy–based brain imaging. Journal of Biomedical Optics, 20(5), 050801. https://doi.org/10.1117/1.jbo.20.5.050801

Moezzi, B. (2018, 5 augustus). Test-retest reliability of functional brain network characteristics using resting-state EEG and graph theory. Geraadpleegd op 21 december 2018, van https://www.biorxiv.org/content/early/2018/08/05/385302

Oliveira, A. S., Schlink, B. R., David Hairston, W., König, P., Ferris, D. P. (1970, January 1). Proposing Metrics for Benchmarking Novel EEG Technologies Towards Real-World Measurements.
Retrieved November 29, 2018, from https://www.frontiersin.org/articles/10.3389/fnhum.2016.00188/full

Rana, A. Q. (2017). Basics of Electroencephalography. In A. Q. Rana (Red.), Neurophysiology in Clinical Practice (pp. 7–183). Cham, Switzerland: Springer International Publishing.

Rogers, J. M., Johnstone, S. J., Aminov, A., Donnelly, J., Wilson, P. H. (2016). Test-retest reliability of a single-channel, wireless EEG system.
Retrieved October 17, 2018,
from https://ro.uow.edu.au/cgi/viewcontent.cgi?referer=
https://www.google.nl/httpsredir=1article=3410context=sspapers

Salinsky, M. C., Oken, B. S., Morehead, L. (1991). Test-retest reliability in EEG frequency analysis. In ORG. Elsevier (Ed.), Electroencephalography and Clinical Neurophysiology (pp. 382–392).
Retrieved October 17, 2018, from https://www.sciencedirect.com/science/article/pii/001346949190203G

Schultz, S. (2007, 24 juni). Signal-to-noise ratio in neuroscience - Scholarpedia.
Retrieved October 17, 2018, from http://www.scholarpedia.org/article/Signal-to-noise-ratio-in-neuroscience

Vlaar, M. P., Solis-Escalante, T., Dewald, J. P. A., Van Wegen, E. E. H., Schouten, A. C., Kwakkel, G., Van der Helm, F. C. T. (2017). Quantification of task-dependent cortical activation evoked by robotic continuous wrist joint manipulation in chronic hemiparetic stroke. Journal of NeuroEngineering and Rehabilitation, 14(1), 6.
Retrieved from https://jneuroengrehab.biomedcentral.com/track/pdf/10.1186/s12984-017-0240-3

Appendix A

Impedence check and graph of EEG, robot and EMG signals

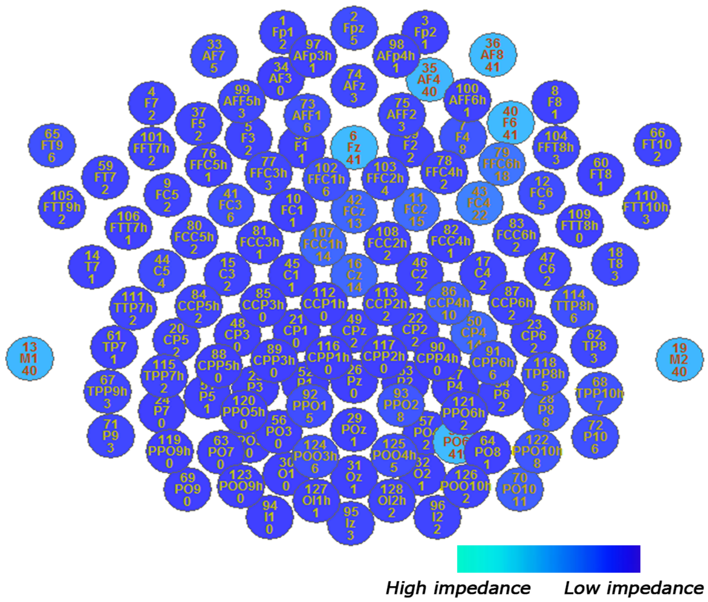

Figure 1: Impedence check

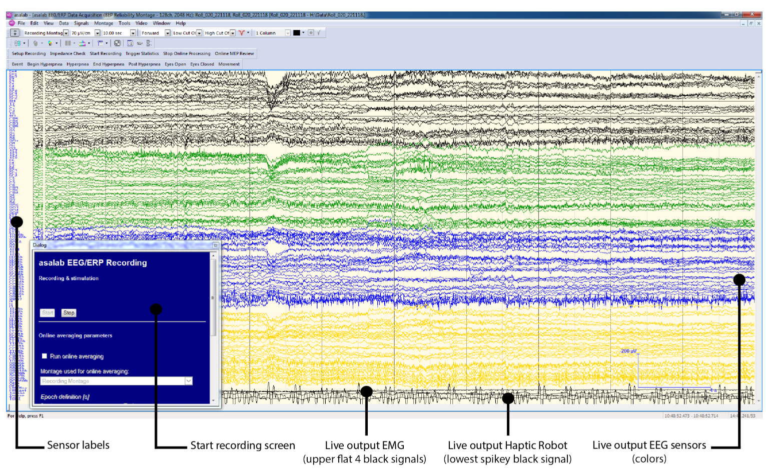

Figure 2: Graph of EEG, robot and EMG signals

Appendix B

Signal to noise ratio

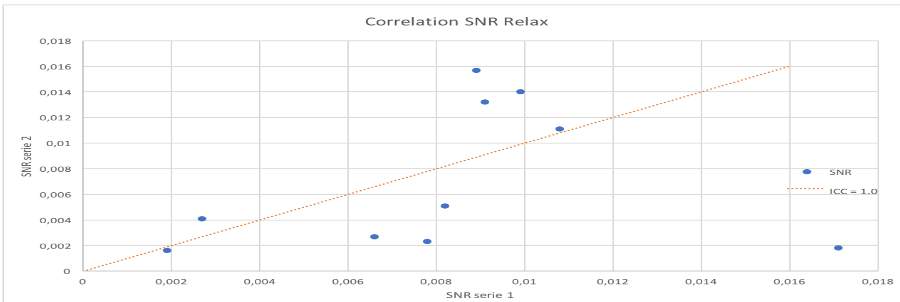

Figure 1: The correlation of the SNR of the relax task

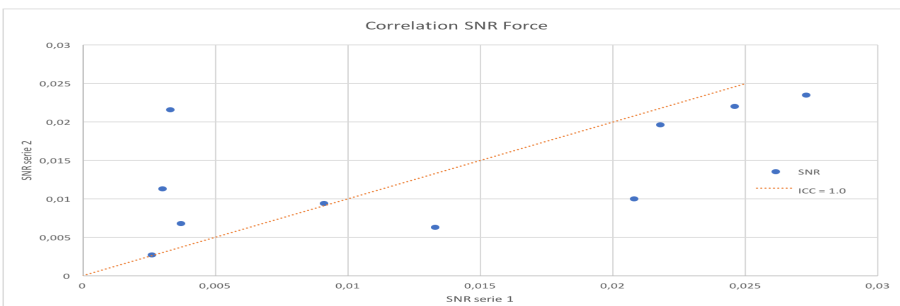

Figure 2: The correlation of the SNR of the force task

SNR		
Task and participant	Serie 1	Serie 2
Relax1	0,0171	0,0018
Force1	0,0246	0,0220
Relax2	0,0027	0,0041
Force2	0,0026	0,0027
Relax3	0,0066	0,0027
Force3	0,0037	0,0068
Relax4	0,0019	0,0016
Force4	0,0033	0,0216
Relax5	0,0108	0,0111
Force5	0,0030	0,0113
Relax6	0,0091	0,0132
Force6	0,0091	0,0094
Relax7	0,0078	0,0023
Force7	0,0273	0,0235
Relax8	0,0089	0,0157
Force8	0,0218	0,0169
Relax9	0,0099	0,0140
Force9	0,0208	0,0100
Relax10	0,0082	0,0051
Force10	0,0133	0,0063

Table 1: Signal to noise ratio

Appendix C

Topoplots

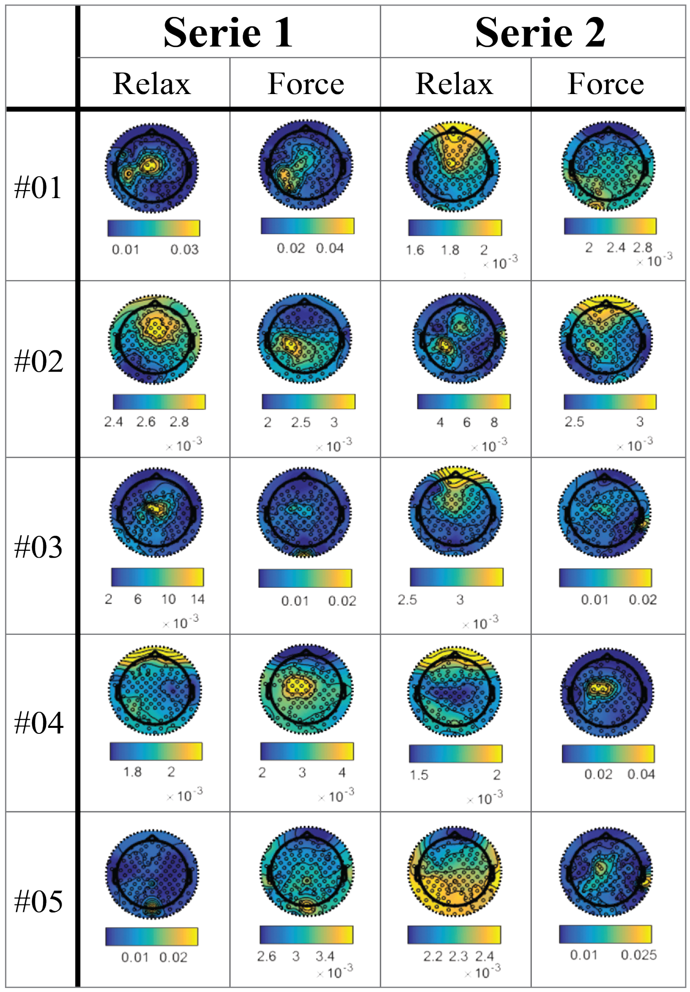

Figure 1: Topoplots of participants 1 to 5

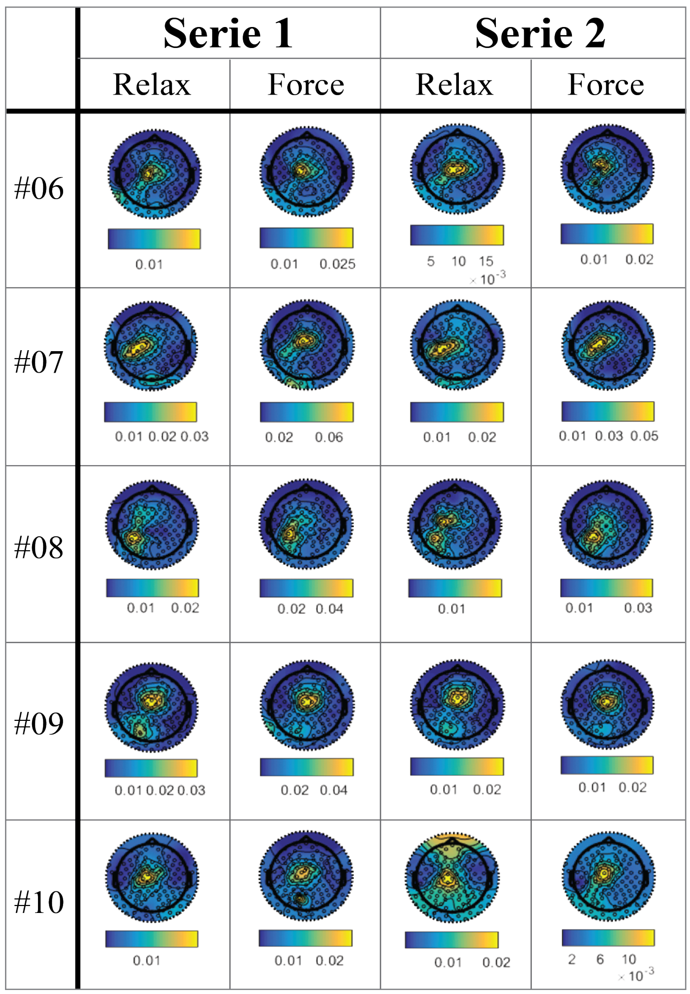

Figure 2: Topoplots of participants 6 to 10

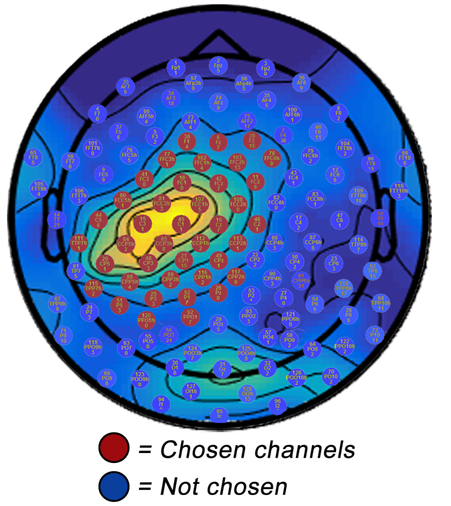

Figure 3: Selected channels from topoplot

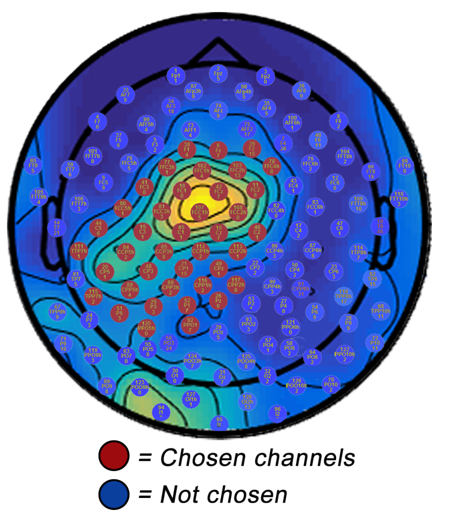

Figure 4: Selected channels from topoplot