Semesterproject Signal processing and Analysis of human brain potentials

Julius Voggesberger

1 Introduction

This report discusses the semester project of the lecture "Signal Processing and Analysis of human brain potentials". For the project an EEG dataset is chosen, preprocessed and analysed. These steps will be discussed in the following chapters. The chosen dataset is the N170 dataset, which was created in a face perception study. In the study participants were shown different stimuli consisting of an object (face, car) and a texture (intact, scrambled) [2]. The first part of the report will cover the pre-processing and cleaning steps that were applied to the data. The second part will cover the analysis that was done on the pre-processed and cleaned data. Lastly the results of the analysis will be discussed.

2 Experimental Setup

In this chapter the used libraries and their versions will be listed. The programming language used in this project is Python (Version 3.7). For plotting visualizations the library matplotlib[1] (Version 3.3.3) is used. The library MNE[2] (Version 0.22.0) is used for everything EEG-data related, MNE-Bids[3] (Version 0.6) is used to load in the BIDS data. Further libraries that were used for different purposes are Numpy[4] (Version 1.19.4), Scikit-Learn[5] (Version 0.23.2), Pandas[6] (Version 1.2.0) and Scipy[7] (Version 1.5.4).
To ensure reproducibility of the results a fixed random state was chosen at each section where the results are effected by randomness. Furthermore the annotations for bad segments, which were found in the manual pre-processing step, are included in the repository. The annotations are needed, so that the reproducibility of the manually cleaned subjects can be ensured. The repository which contains the code can be found here: `https://git.ffhartmann.de/Julius/semesterproject_lecture_eeg`.

---

[1] `https://matplotlib.org/`
[2] `https://mne.tools/stable/index.html`
[3] `https://mne.tools/mne-bids/stable/index.html`
[4] `https://numpy.org/`
[5] `https://scikit-learn.org/stable/`
[6] `https://pandas.pydata.org/`
[7] `https://www.scipy.org/`

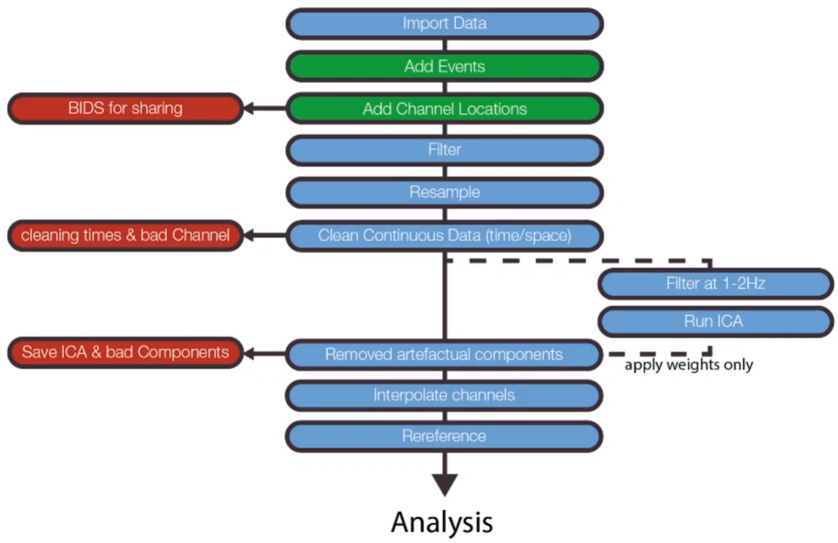

Fig. 1: The pre-processing pipeline used in the project. The picture was taken from the first lecture.

3 Pre-Processing and Cleaning

As an orientation for the pre-processing stage of the project, the pre-processing pipeline shown in the first lecture and seen in figure 1, is taken. The first three steps will not be discussed here, as the data was already available in the BIDS format. For this reason, the data only had to be imported. The subsequent pre-processing steps are discussed in the following sections of this chapter. Pre-processing and cleaning steps needed only to be done for three out of 40 subjects, which are subject 001, 003 and 014 for this report. The rest of the subjects are pre-processed and cleaned with already given pre-processing information.

3.1 Filter

The first pre-processing step, which is applied to the data, is filtering. Filtering is applied to all subjects, not only the three selected ones, as no filter information was given and it is easy to apply to all subjects. The goal for this step is to eliminate high frequencies and slow drifts in the EEG data. An example for unfiltered data, where the slow drifts and high frequency noise can clearly be seen, is shown in figure 6. As a filter for the data a bandpass filter from 0.5 to 48 Hz with a FIRWIN design was chosen. The need for a low-pass filter can be seen in figure 2, which shows the power spectral density plot for subject 014. At 60 Hz the powerline frequency, which has to be filtered out, can clearly be seen. The filtered data can than be seen in figure 3. Other low-pass filter frequencies were tested, but at above 48 Hz the 60 Hz peak still existed in many subjects. The

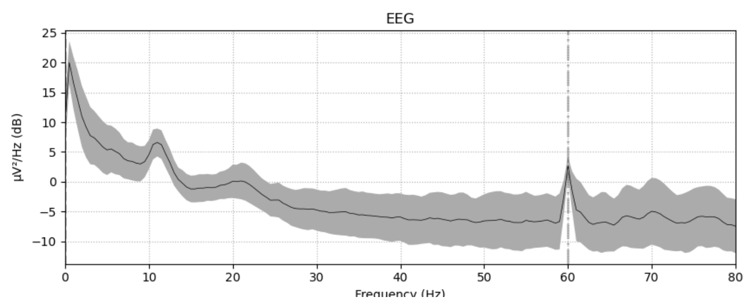

Fig. 2: Power spectral density plot of subject 014 averaged across all channels.

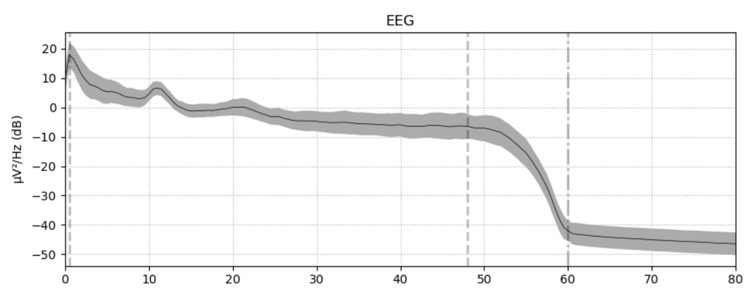

Fig. 3: Power spectral density plot of the filtered subject 014 averaged across all channels.

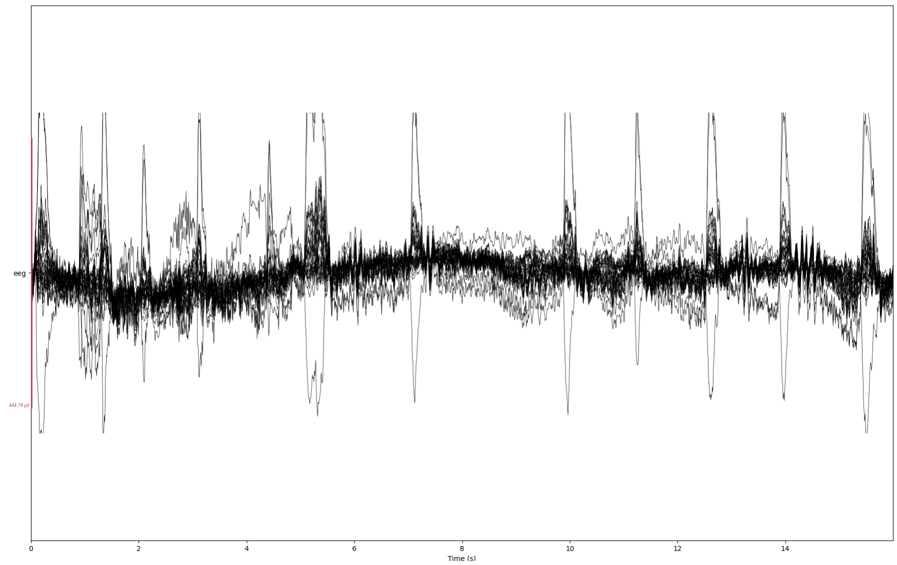

Fig. 4: Butterfly plot of all channels of subject 003 with bandpass filter from 0.1 Hz to 48 Hz and in a time window from 0 to 15 seconds.

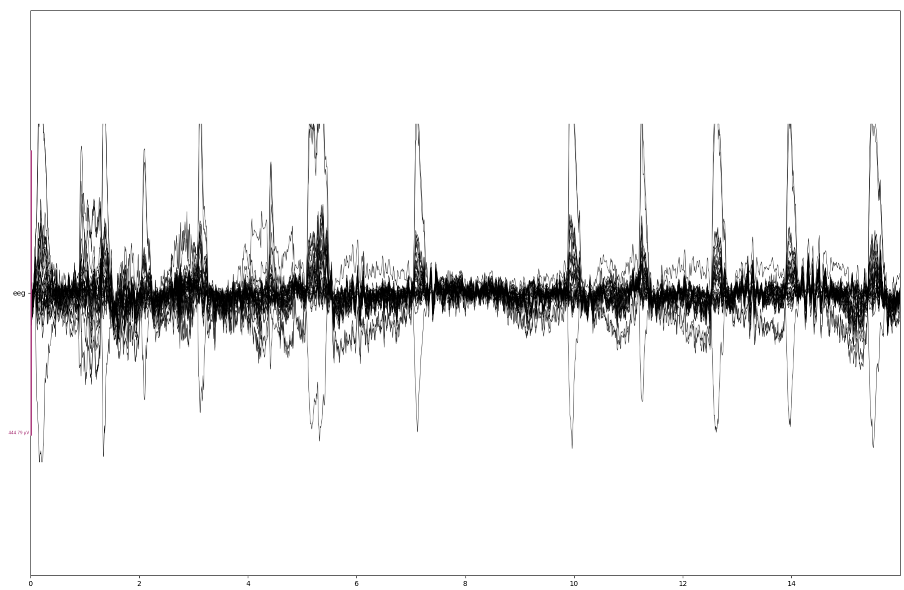

Fig. 5: Butterfly plot of all channels of subject 003 with bandpass filter from 0.5 Hz to 48 Hz and in a time window from 0 to 15 seconds.

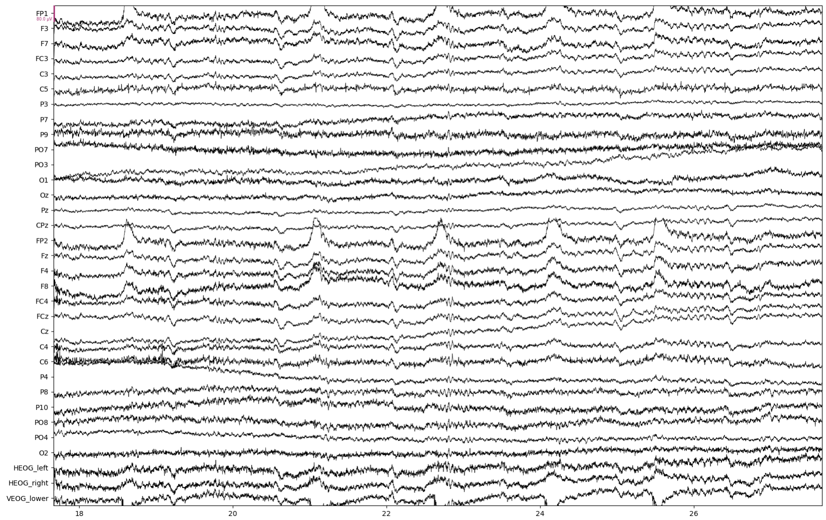

Fig. 6: All channels of the unfiltered subject 014 from around 18 to 26 seconds.

Fig. 7: All channels of the filtered subject 014 from around 18 to 26 seconds.

frequency for the high-pass part are chosen according to the current literature [1]. As such the frequency of 0.1 Hz was chosen at first. In figure 4 it can be seen, that filtering at 0.1 Hz does not remove the slow drifts. To remove the slow drifts completely a cutoff frequency of 0.5Hz was chosen, which worked well, as can be seen in figure 5. The filtered result can be seen on the example of subject 014 from around 18 to 26 seconds in figure 6 and figure 7.

3.2 Cleaning Data

This step deals with the marking of bad channels and bad segments in the data. One bad channel was found for subject 001, which is the F8-Channel. In figure 8 the bad channel and its neighbouring channels can be seen. The F8-Channel oscillates a lot and is more noisy than its neighbours. For the other subjects no bad channels were found.
Bad segments are selected conservatively. Only segments that were clearly bad, e.g. sudden spikes or distortions, are marked. One such example can be seen in figure 9. Most of the bad segments are found outside of time-intervals where an experiment was taking place, i.e. in the waiting time in between experiments. Some artefacts like blinks, etc. are not marked in this step, as they will be removed later on in the independent component analysis.

3.3 Independent Component Analysis

In this section the independent component analysis is discussed. This step is used to sort out components that are seen as bad, i.e. are representing artefacts, so that the reconstructed signal is cleaner. Before viewing the independent components (IC) the data is filtered with a lower bound of 1Hz to remove slow drifts and improve the IC analysis [6]. This filter is only relevant for this section as it is only used for the IC analysis and not on the data that is later used in the analysis. As a first step to marking ICs, the EOG-channels are used to find components, which contain eye artefacts like blinks or eye movements. An example for an eye component can be seen in figure 11. Following the selection of ICs with the help of the EOG-channels, the remaining ICs are checked manually. As for cleaning data in section 3.2, a conservative approach is chosen. ICs that are clearly brain components are identified by a peak around 10 Hz in the power spectral density plot, an activity in the range of around $-20\mu V$ to $20\mu V$, and a clear bipolar topography. An example for such an IC can be seen in figure 10. In contrast to the other described components, the brain components are not removed from the signal. ICs that are manually marked as bad are all muscle artefacts. They are selected by higher plateaus in high frequencies and poles on the border of the topography plot. One such example can be seen in figure 12. ICs that could not be clearly identified as an artefact are kept. The removed components for each subject are:

- Subject 001: 0, 1, 2, 4, 8, 14, 16, 25 - via automated EOG removal: 0, 1
- Subject 003: 0, 2 - via automated EOG removal: 0, 2

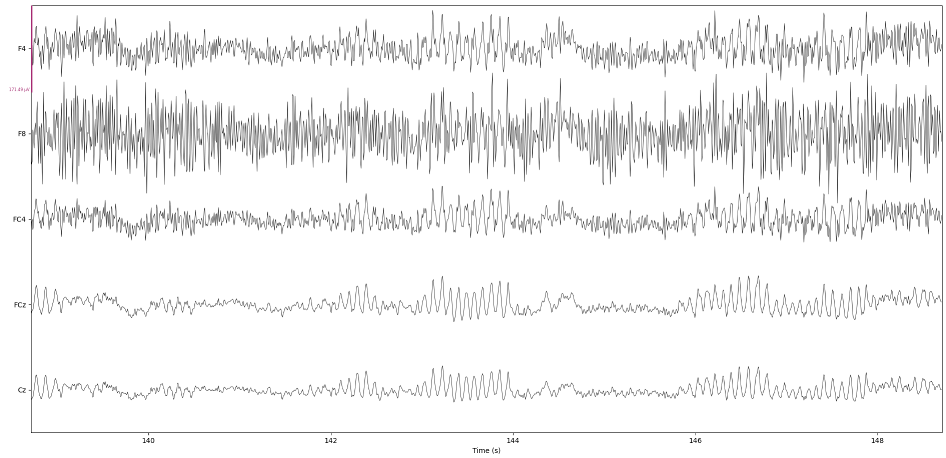

Fig. 8: Channel F8 of subject 001 and its neighbouring channels.

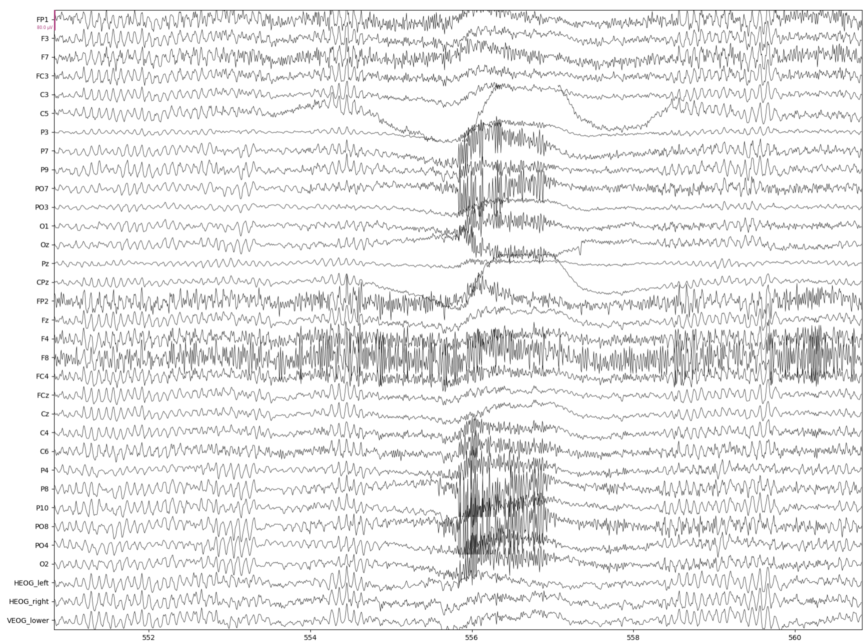

Fig. 9: A bad segment in subject 001 at around 556 seconds.

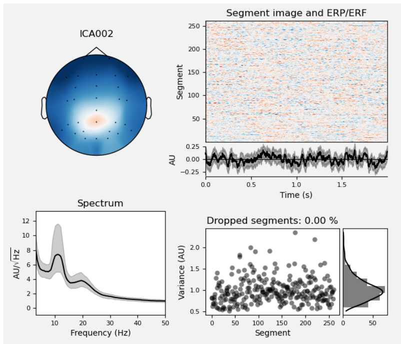

Fig. 10: A brain component.

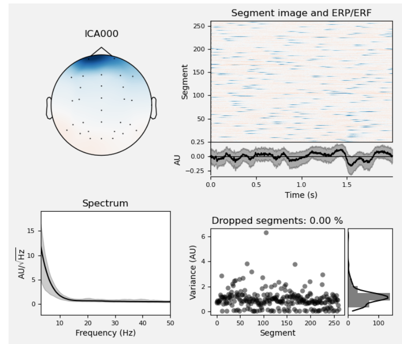

Fig. 11: An eye component.

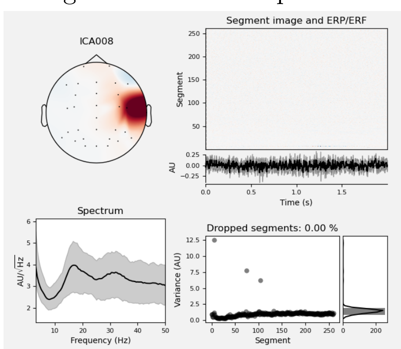

Fig. 12: A muscle component.

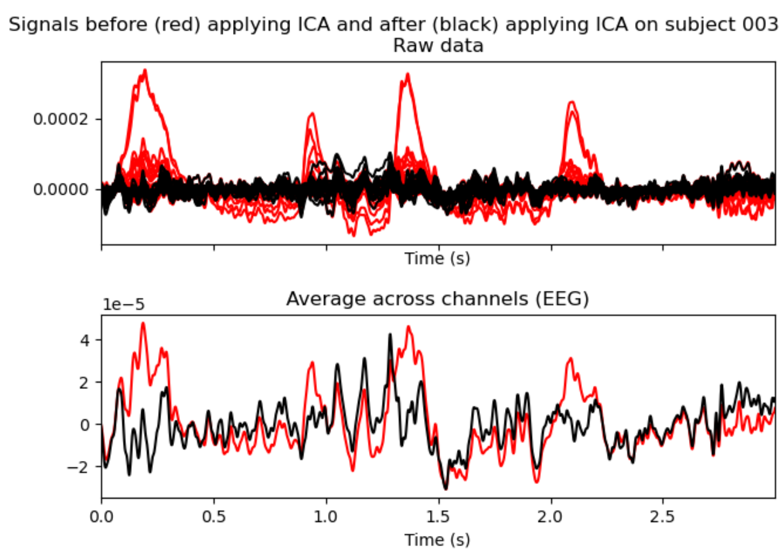

Fig. 13: Subject 003 with applied ICA.

– Subject 014: 0, 1, 9 - via automated EOG removal: 0, 1

As an example subject 003 can be seen in figure 13 with the signal before ICA and after applying ICA.

3.4 Interpolation and Re-referencing

Lastly the bad channel that was detected in subject 001 is interpolated and the data rereferenced. As the re-referencing technique, the "averaging across all channels"-technique is chosen, which works well for face recognition experiments [5]. The results of the re-referencing can be seen in figure 15.

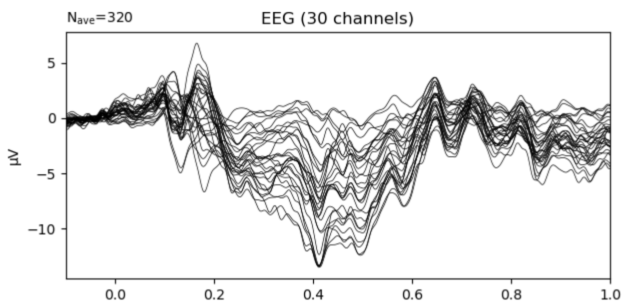

Fig. 14: Subject 014 without re-referencing.

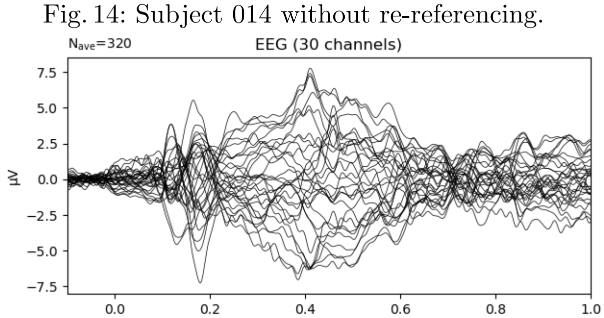

Fig. 15: Subject 014 re-referenced.

4 Analysis

In this chapter the different analysis approaches that were chosen for this report are discussed. The data is analysed with respect to the effect of faces on the N170 peak. As the experimental contrast the difference in the N170 Peak w.r.t. to different stimuli will be analysed. The different stimuli are faces against cars, which both can be either intact or scrambled. Different analysis methods were utilized to analyse the above mentioned contrast and will be presented in the following sections. These methods are the analysis of ERP peaks, decoding over time and time-frequency analysis.

4.1 ERP Peak Analysis

Using the ERP peak analysis, the ERPs at the N170 Peak will be analysed with respect to different stimuli. The analysis is done by computing the significance of peak-differences between two conditions of a contrast. The peak-difference for each subject is defined as seen in equation 1, where $peak_a$ and $peak_b$ are the N170-Peaks of the different conditions. E.g. to compute the peak-difference of intact stimuli for a given subject, $peak_a$ is the N170 peak of stimulus "face intact" and $peak_b$ is the N170 peak of stimulus "car intact" of the subject.

$$\text{Peak-difference d for Subject i: } d_i = peak_a - peak_b \tag{1}$$

Before starting with the analysis of the ERP-components, the channels as well as the time window for the N170 peaks were chosen. The time window in which the N170 peak is searched is between 130 ms and 200 ms, which is often chosen in the literature [4,3]. The channels chosen for this analysis part are the electrodes P7, P8, PO7, PO8, as often used in the literature [4]. For each of these channels the grand average over all subjects and epochs is computed, as can be seen in figure 16. In the grand average plots the N170 peak can be clearly seen between 150 ms and 200 ms, which lies within the proposed time window. Furthermore the differences between the stimuli can be seen. The peaks of the intact face and car stimuli are more pronounced, then the peaks of the scrambled stimuli, which show a smaller dip at the N170 peak. Furthermore the P7- and P8-Channel show the peak more clearly then the PO7- and PO8-Channel.
In the next step the peak differences are tested for significance via a one-sample t-tests against a t-distribution of mean 0. A difference is seen as significant, if its p-value is smaller then the chosen significance level of 0.05. The null hypothesis always has the form: "There is no difference between the peak of stimulus a and the peak of stimulus b", while the alternative hypothesis has the form: "There is a difference between the peak of stimulus a and b". The tested conditions can be seen in Table 1. As a first test the N170 peaks of all faces where compared against the peaks of all cars, regardless of the stimulus condition. To compare the peak-difference feature the mean of the peaks face intact and face scrambled where taken as $peak_a$ in equation 1. This was done analogously for car and $peak_b$. While the resulting p-value was quite low for the P7 (0.0012) and P8 (0.0009) channels it was not significant for the PO7 (0.0540) and PO8 (0.1903) channels. This is in accordance to what was observed in figure 16. As such the difference of all faces against all cars can not be seen as significant. The second hypothesis that was tested compares the intact stimuli. All channels have p-values that are smaller than the significance-level, so it can be said that the peak-difference between intact faces and cars is significant. This seems to confirm that the N170 peak for intact face stimuli is significantly more distinct then the N170 peak for cars. The same test was also done for the scrambled stimuli. For this test the

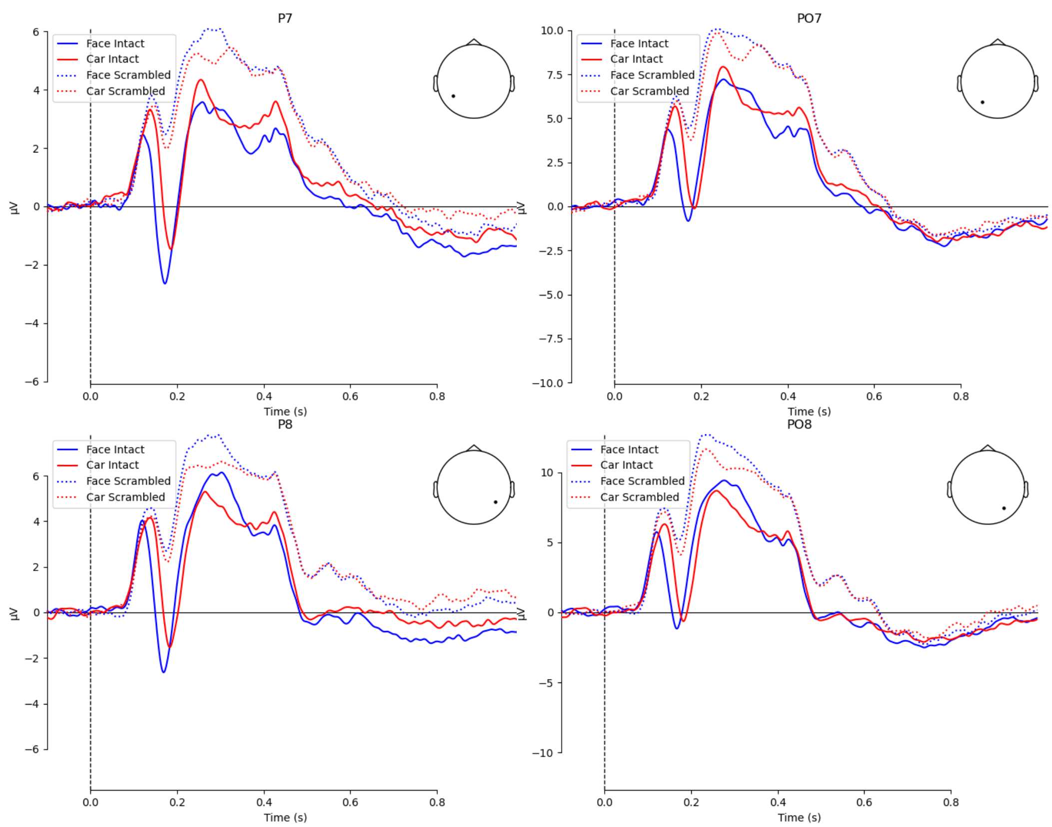

Fig. 16: grand average diagrams over all subjects, for the Channels P7, PO7, P8, PO8.

Analysed Effect Stimulus(Condition)	Channel	p-Value
Faces(All)-Cars(All)	P7	0.0012
	PO7	0.0540
	P8	0.0009
	PO8	0.1903
Faces(Intact)-Cars(Intact)	P7	$2.2e^{-06}$
	PO7	0.0003
	P8	$6.4e^{-06}$
	PO8	0.0017
Faces(Scrambled)-Cars(Scrambled)	P7	0.0551
	PO7	0.0117
	P8	0.0014
	PO8	0.0001
Faces(Intact)-All other	P7	$6.9e^{-12}$
	PO7	$7.8e^{-12}$
	P8	$1.0e^{-12}$
	PO8	$4.5e^{-10}$

Table 1: P-Values of different ERP peak analysis t-tests.

alternative hypothesis could not be seen as significant, as the channel P7 has a p-value (0.0551) higher than the significance-level, while the p-values of the other channels are smaller then the significance level. This result is in line with the observations of the grand averages in figure 16, as the N170 peaks for scrambled stimuli are less distinct then the peaks for intact stimuli. Lastly the N170 peaks for intact faces are compared against all other stimuli pairs. The alternative hypothesis for this test can be clearly seen as significant, as all p-values for all channels are smaller than the significance-level.

4.2 Time-Frequency Analysis

In this section time-frequency analysis is discussed, to investigate the main experimental contrast and its underlying oscillations. As the effect of interest, which will be evaluated using time-frequency analysis, the contrast of the intact face stimulus against all other stimuli is chosen. The time-frequency analysis is used on each subject separately by applying the short time Fourier transform (stft) with Morlet wavelets on the induced data of each condition of the contrasts. The analysed frequency spectrum starts at 0.1 Hz and ends at 50 Hz. The lower bound is chosen, so that delta bands are included into the analysis and the upper limit is chosen in accordance to the upper cutoff frequency of the bandpass filter from section 3.1. For the number of cycles used for the Morlet wavelets a different number of cycles is used for each frequency. The number of cycles are scaled according to frequency, by $frequency/2$. This allowed for a good trade-off between time and frequency scaling. Furthermore the analysis is done for two different scalings of the frequency spectrum. One is the logarithmic scaling, which allows for more accuracy in the lower frequencies. The second scaling used is the linear

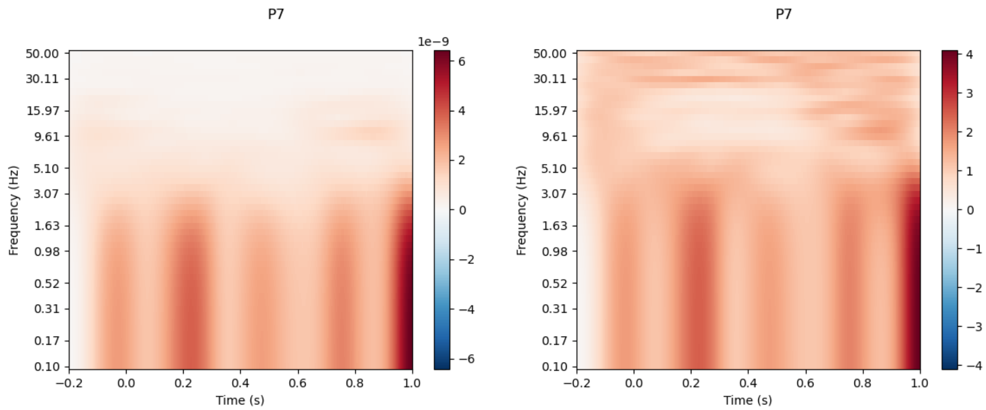

Fig. 17: Stft transformed induced logarithmic scaled data of subject 014 of channel P7 of condition intact faces without a baseline on the left and stft transformed data with baseline on the right. A clear shift of the values can be seen. As a result the patterns for higher frequencies are more pronounced.

scaling, so that all frequencies are viewed on the same scale. Epoching is done for a time interval starting at 200 ms before stimulus onset and ending 1000 ms after stimulus onset. As a baseline for baseline-correcting the stft data, the interval until the stimulus onset is chosen, i.e. from -200 ms to 0 ms. The result of applying the baseline can be seen for the example of subject 014 in figure 17 for the logarithmic scaling. Figures for the linear scaling can be found in the appendix. The stft was then computed for each subject. To test for significant differences between two conditions a cluster-permutation test is used on the grand averages of the stft data over all subjects. Using the grand averages allowed to compute the differences in conditions over all subjects. The grand average of one condition of the contrast is then compared to the grand average of the other condition of the contrast with a cluster-permutation test and tested for significance. An example for the grand averages of the contrast intact faces against rest can be seen in figure 18 for the logarithmic scaling. As the power is averaged over all subjects, the visual representation is not as clear for the grand averages, as it was for the single subjects. When viewing the grand averages of the oscillation bands of both conditions in figure 19 clear differences between both conditions can be seen. The delta, theta and gamma bands exhibit the largest differences between both conditions, as different poles and intensities can be seen in them. The alpha and beta bands are mostly similar except some smaller differences in the power. Using the cluster permutation test, the conditions over all channels, time and frequencies are compared against each other. The result of the cluster permutation test for both scalings can be seen in figure 20. For the analysed contrast a significant difference between both conditions is found. Looking at the logarithmic scaling this difference is driven by an effect between 0 ms and 200 ms and 0.1 Hz to around 10 Hz. As such the observed effect seems to be

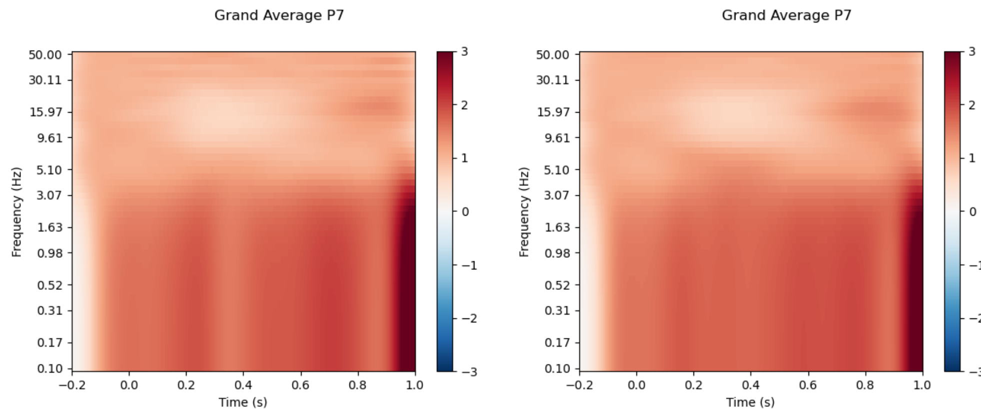

Fig. 18: Grand average over subjects of stft transformed induced logarithmic scaled data of channel P7. On the left the grand average for the condition intact faces is seen, while on the right the condition of all other stimuli is seen.

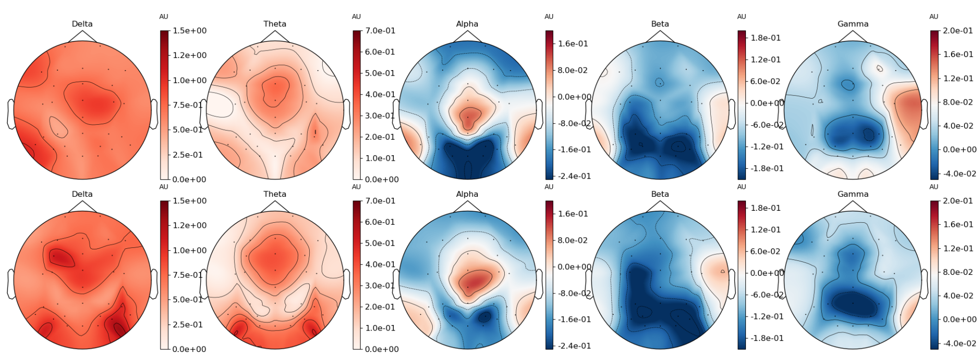

Fig. 19: Topology plots of the grand averages over all subjects of the oscillation bands using the logarithmic scaling. The plots where done from 130 ms to 200 ms. The upper row is for the condition of intact faces, while the lower row is for the condition including all other stimuli.

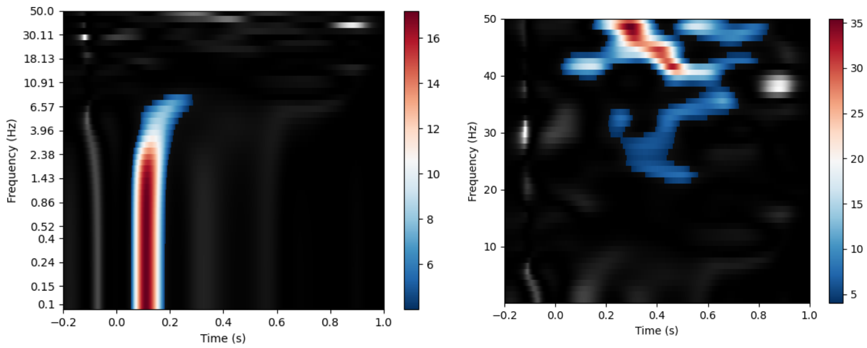

Fig. 20: Cluster permutation test of intact faces against all other stimuli. The background in grey is the result of the whole cluster-permutation test and the coloured segments are the clusters with p-value <= 0.05. On the left the result of the logarithmic scaling can be seen, on the right the result of the linear scaling can be seen.

influenced by delta and theta bands, as the significant difference seems to be at the N170 peak. If the result for the linear scaling is viewed the significant difference between both conditions is more stretched over time. The difference is driven by an effect between 0 ms and 800 ms and between 20 Hz and 50 Hz. As the significant cluster is not located at the N170 peak it is difficult to say what oscillation bands influence the difference between conditions. But looking at the marked cluster, the frequencies between 40 Hz and 45 Hz may influence the effect, which would indicate the gamma band.

4.3 Decoding Analysis

The decoding analysis is used to analyse the main contrast of the experiment over time, by using classifiers to distinguish the conditions of the contrast at each time point. The contrasts that are being analysed are all faces against all cars and intact faces against all other stimuli. As the classifier for the decoding the logistic regression is chosen, as it is fast to compute and works well. The classifier is used to compute the accuracy of each time point for a binary classification problem. To ensure that the classifier wont overfit too much, cross validation was used, by separating the data into 10 portions. The classes used for classification are chosen to reflect the analysed contrast, i.e. for all faces against all cars one class is the condition of all face stimuli, while the second class is the condition of all car stimuli. As this is done for each subject and at each time point the results are 40 accuracy values per time point. The mean over all subjects at each time point is then computed and can be seen in figure 21. In both figures a clear peak in accuracy can be seen between 150 ms and 200 ms. At that time interval the decoder could distinguish both classes more clearly. An important aspect when analysing the images in figure 21 is that for the contrast intact

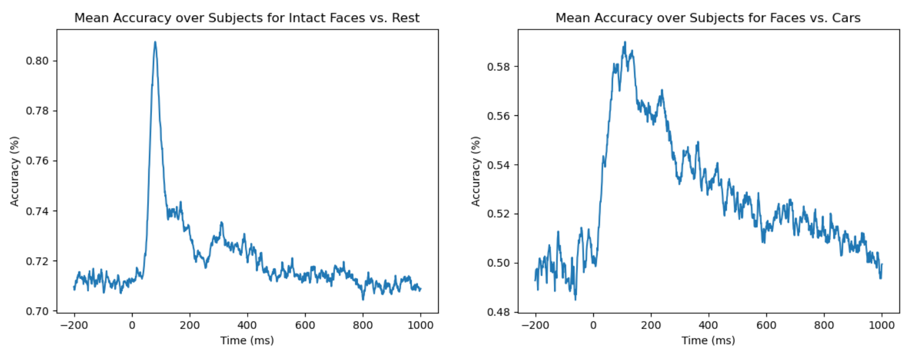

Fig. 21: Mean accuracy over subjects

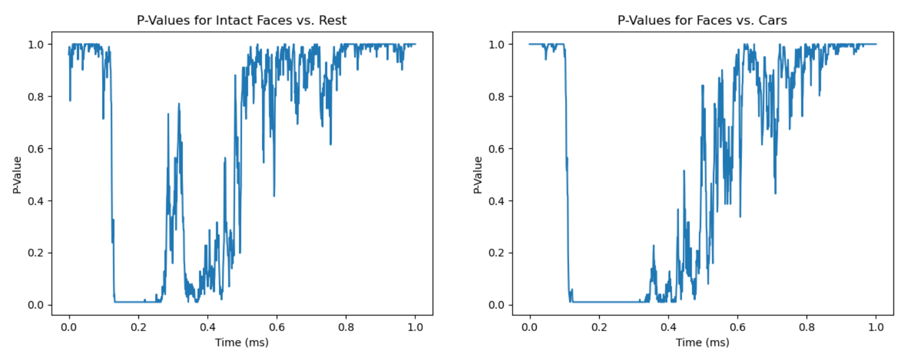

Fig. 22: P-Values for the permutation test on decoding results.

faces against all other stimuli, the accuracy for guessing the correct class is at around 70%. This is the case, as 3/4th of the data are part of the non intact face stimuli. The accuracy for guessing the correct class can be seen in both cases in the time interval from -200 ms to 0 ms, as the classifier can not distinguish clearly between the different classes before the stimulus onset. For the contrast of intact faces against all other stimuli the classifier could distinguish the classes with an accuracy of more then 80% at the time interval from 130 ms to 200 ms, then for the other contrast, which has an accuracy of more then 58% at the same time interval. To determine if the observed peak in the figures 21 is significantly different to chance, a permutation test against the accuracy for the baseline is done, with 1000 permutations. A simple permutation test on the results is done, as a permutation test for each time point classification needs an enormous amount of runtime. Using a permutation test on the results helps to bypass that. The baseline is chosen as the time interval from -200 ms to the stimulus onset at 0 ms. The results for the permutation test can be seen in figure 22. For the contrast intact faces against all other stimuli the results are significant for the time-interval 120 ms to 260 ms. This interval includes the N170 peak, which shows that around the N170 peak the condition information becomes available and improves the ability of the classifier to distinguish the two conditions of the contrast. Looking at the contrast of all faces against all cars the time-interval in which the accuracy is significant is even more apparent. The time-interval starts at around 125 ms and ends at around 340 ms. This includes the N170 peak as well and also shows that the information needed by the classifier to distinguish the two classes becomes available at the N170 peak. Interestingly this information persists even long after the N170 peak. To summarize, this shows that for both conditions a clear distinction between the classes can be made at the N170 peak.

5 Discussion

Summarizing the results of the analysis in section 4, the effect of faces on the N170 peak is significantly different then the effect of other stimuli on the N170 peak. This could be already seen in section 4.1 where the comparison of the N170 peak of intact faces stimuli against some other stimuli is always significant. Using scrambled faces as a stimuli did on the other hand yield no significant difference in the N170 peak to another stimuli. As a result, it can be only said that intact faces as a stimuli have a significant different effect on the N170 peak than other stimuli and as such this was chosen as the main contrast of this experiment. This was further analysed in section 4.2, where time-frequency analysis was used to analyse the influence of the underlying oscillation bands on this main contrast. Using a cluster-permutation test to analyse the influence of the oscillation bands yielded that the delta, theta and gamma bands seem to influence the contrast. Lastly the effect is analysed by decoding the contrast over time, by using logistic regression. This shows a clear difference between the stimuli at the N170 peak. By using a permutation test this difference is found to be significant. Surprisingly a significant difference could also be found at the

N170 peak for the contrast of all face stimuli against all car stimuli. This stimuli was found to be not significant in the ERP-analysis but a significant difference could now be found in the decoding analysis. To summarize, it can be said that the N170 peak shows a more significant effect, when intact faces are used as a stimulus, then when stimuli that are not faces or have a different texture are used.

References

1. Acunzo, D.J., MacKenzie, G., van Rossum, M.C.: Systematic biases in early erp and erf components as a result of high-pass filtering. Journal of neuroscience methods **209**(1), 212–218 (2012)
2. Kappenman, E.S., Farrens, J.L., Zhang, W., Stewart, A.X., Luck, S.J.: Erp core: An open resource for human event-related potential research. NeuroImage **225**, 117465 (2021). https://doi.org/https://doi.org/10.1016/j.neuroimage.2020.117465, `https://www.sciencedirect.com/science/article/pii/S1053811920309502`
3. Rossion, B., Caharel, S.: Erp evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. Vision research **51**(12), 1297–1311 (2011)
4. Rossion, B., Jacques, C.: Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? ten lessons on the n170. Neuroimage **39**(4), 1959–1979 (2008)
5. Wang, Y., Huang, H., Yang, H., Xu, J., Mo, S., Lai, H., Wu, T., Zhang, J.: Influence of eeg references on n170 component in human facial recognition. Frontiers in neuroscience **13**, 705 (2019)
6. Winkler, I., Debener, S., Müller, K.R., Tangermann, M.: On the influence of high-pass filtering on ica-based artifact reduction in eeg-erp. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 4101–4105. IEEE (2015)

6 Appendix

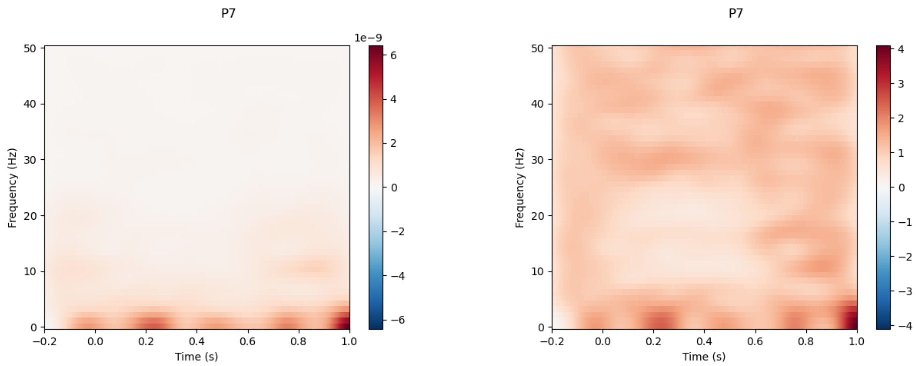

Fig. 23: Stft transformed induced linear scaled data of subject 014 of channel P7 of condition intact faces without a baseline on the left and stft transformed data with baseline on the right.

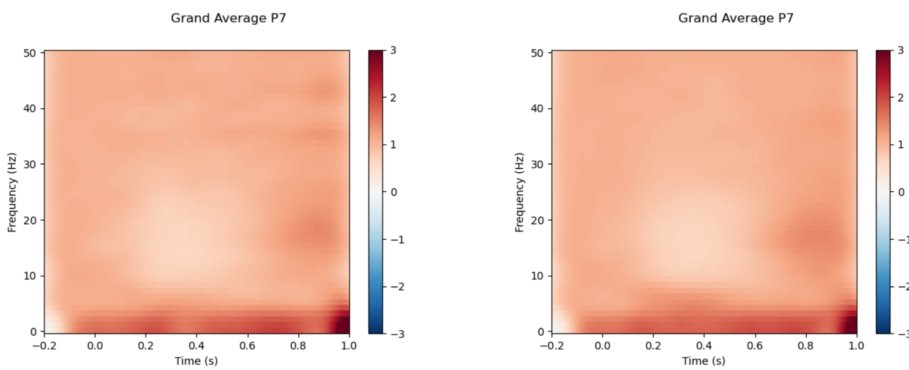

Fig. 24: Grand average over subjects of stft transformed induced linear scaled data of channel P7. On the left the grand average for the condition intact faces is seen, while on the right the condition of all other stimuli is seen.

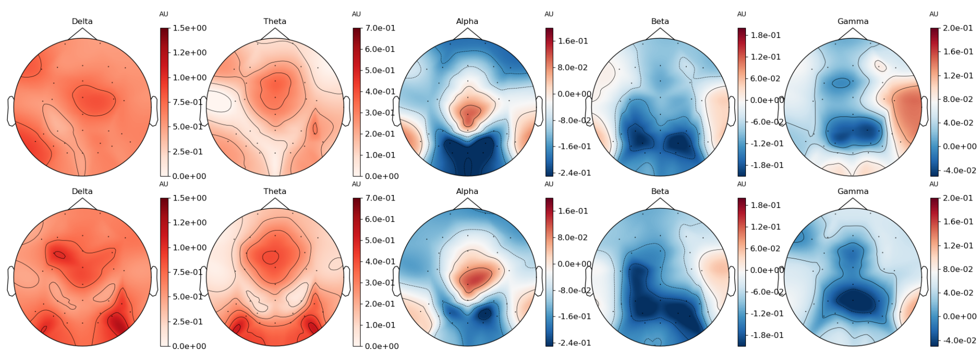

Fig. 25: Topology plots of the grand averages over all subjects of the oscillation bands using the linear scaling. The plots where done from 130 ms to 200 ms. The upper row is for the condition of intact faces, while the lower row is for the condition including all other stimuli.