

Semesterproject EEG  
Analysis of N170 dataset  
SS 2021

Jingxi Zhang  
3126431

## Content of zip:

- Generating the paths with generatePaths.py
- Notebooks for cleaning: Cleanings-sub-{id}.ipynb
- Notebooks for ERP:
  - o ERP-Peak-Extraction.ipynb contains explanation for reading raw/precomputed data and extracting peaks
  - o ERP-EvokedPlotting.ipynb contains a sanitycheck of a subject with confidence intervals
  - o Ttest.ipynb contains one sampled ttests for extracted ERP peak values and cluster permutation test for extracted values
- Execution files:
  - o ERP\_Sanitycheck.py -- ERP-EvokedPlotting.ipynb same content, just running for each subject generating plots for sanity checks subjectwise
  - o ERP\_Extraction.py – ERP-Peak-Extraction.ipynb same content, just running over all subjects and writing peaks, times and baseline to csv file.
- Notebooks for time frequency analysis:
  - o Tf\_singleSubject.ipynb: my thought process on time frequency analysis
  - o Tf\_test.ipynb: running permutation test on between conditions on time frequency
- Execution files:
  - o Tf\_allsubjects.py – tf\_singleSubject.ipynb same content just for multiple(all subject together)
  - o Tf\_subjectwise.py – tf\_singleSubject.ipynb same content just for each subject
  - o Tf\_test\_subjectwise – tf\_test.ipynb same content just running permutation\_cluster\_test for each subject
- Notebooks for source estimate (lags out computer completely, for some parts it was not possible for me to run):
  - o source-Estimate.ipynb contains the thought process and an example
  - o sourceEstimate-CPTTest.ipynb contains a cluster permutation test.
- Execution files:
  - o Source\_Estimate.py
  - o Source\_Estimate\_subjectwise.py
  - o SourceEstimate-CPTTest.py
  - o SourceEstimate\_CPTTest\_Subjectwise.py

## Introduction:

In this semester project I will analyse the N170 dataset. The reference to the dataset can be found at <https://osf.io/pfde9/>. In this analysis I will go into cleaning with filtering, re-referencing and ICA. Then I will do an ERP Peak analysis and filter out one peak per subject and condition. For this dataset there are two stimuli face and car images and a scrambled version of each. Overall there are four conditions that can be compared and analysed.

For the final I will discuss Source estimation of the signal and do a time frequency analysis. In these parts I focused on analysing the relation between intact face and intact car. As well as intact and scrambled. Combining all conditions resulted in too long computation times and too large use of computational power especially for source estimates.

The documentation of my choices and thoughts are intermingled with code in jupyter notebooks for an easier understanding. Interpretations are documented here in this document. The required packages are also listed on github. I provide the citations as direct links to the pages/papers for convenience. You do not need to scroll all the way to the reference and open it from there. But I will add the reference at the end anyway as an overview. I will not add the mne documentation as a reference as it is a basic tool being used all the time in this semester project.

The conditions were as mentioned as above intact face, intact car and scrambled face, scrambled car according to “Face perception task used to elicit the N170. On each trial, an image of a face, car, scrambled face, or scrambled car was presented in the center of the screen, and participants indicated whether a given stimulus was an “object” (face or car) or a “texture” (scrambled face or scrambled car).” (ERP CORE Manuscript)

The samples were from “ 40 participants (25 female, 15 male; Mean years of age = 21.5, SD = 2.87, Range 18–30; 38 right handed) from the University of California, Davis community. Each participant had native English competence and normal colour perception, normal or corrected-to-normal vision, and no history of neurological injury or disease (as indicated by self-report). Participants received monetary compensation at a rate of \$10/hour.” (ERP CORE Manuscript)

The main electrode site is the PO8 electrode according to ERP CORE and the event is a stimulus locked one with epoch window of -100 to 800 ms and a baseline period of -100 to 0 ms.

Recorded electrodes were from “30 scalp electrodes, mounted in an elastic cap and placed according to the International 10/20 System (FP1, F3, F7, FC3, C3, C5, P3, P7, P9, PO7, PO3, O1, Oz, Pz, CPz, FP2, Fz, F4, F8, FC4, FCz, Cz, C4, C6, P4, P8, P10, PO8, PO4, O2” (ERP CORE)

According to ERP CORE “all signals were low-pass filtered using a fifth order sinc filter with a half-power cutoff at 204.8 Hz and then digitized at 1024 Hz with 24 bits of resolution. The signals were recorded in single-ended mode (i.e., measuring the voltage between the active and ground electrodes without the use of a reference), and referencing was performed offline”

I followed their guideline and values as they did some testing and showing results with them.

## Preprocessing:

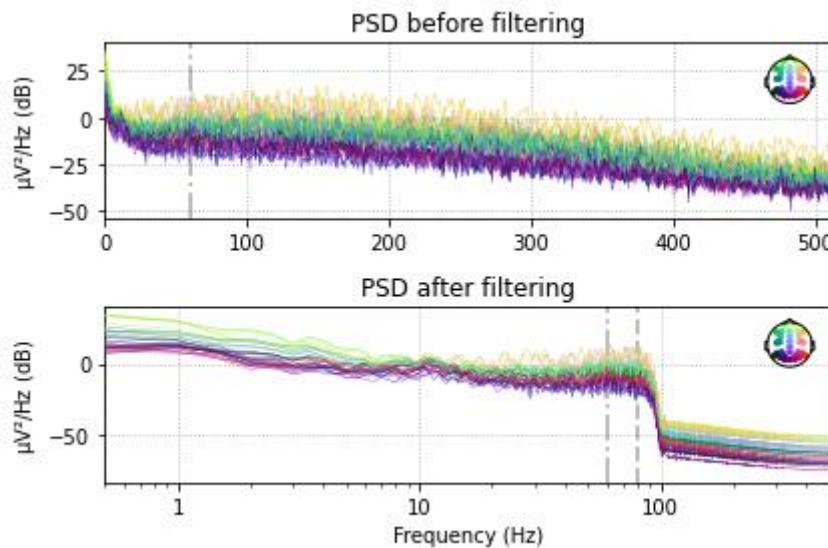
For my EEG semester project I used the N170 dataset and cleaned subjects 6, 8 and 13. The cleaning steps and results can be found in the notebookfiles (Cleanings-sub-{subjectid}.ipynb).

When reading the raw data mne version 0.23.0 did not recognize HEOG\_left, HEOG\_right and VEOG\_lower as eog channels which had to be set manually. This might have some consequences for later results. As for me I could not fix this problem as I later needed a function from the 0.23 version. Maybe this will be fixed in a later version of mne.

## Filtering:

First I did a filtering of the raw data. The data was filtered with 0.5hz highpass and 80hz lowpass. To check the results I plotted the power spectra for reference. These seem right to me compared with the references from mne documentation

[https://mne.tools/dev/auto\\_tutorials/preprocessing/30\\_filtering\\_resampling.html](https://mne.tools/dev/auto_tutorials/preprocessing/30_filtering_resampling.html) and our results from exercise 2 about filtering.



Removing the spiking of the PSD

## Re-referencing:

EEG signals were referenced to the average of all 33 sites according to ERP CORE.

At this point I did a little sanity check by plotting raw, filtered and re-referenced data to view the effect. This is as expected less noisy and less occlusions. But eye blinks and noise from muscles are remaining. These I then cleaned with the ICA in the following.

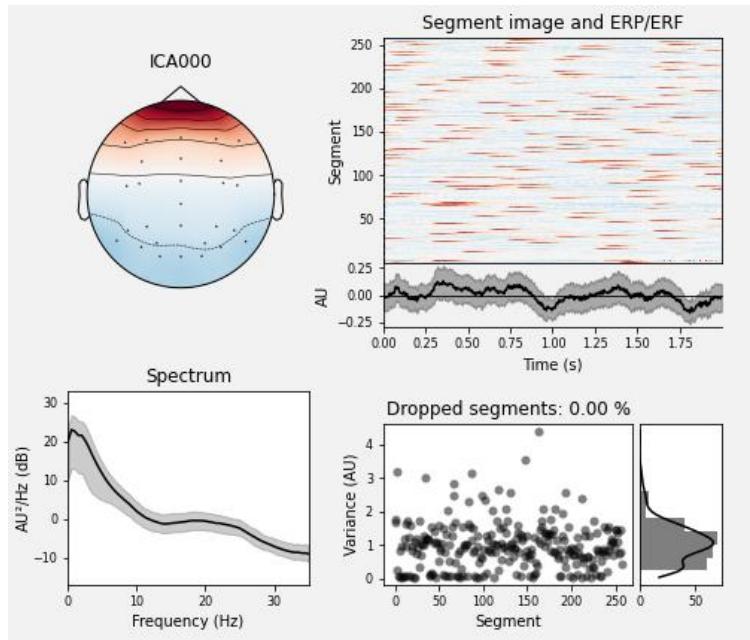
## ICA

The full ICA is in the appendix or in folder img/sub-{subjectid}/. Following methods were applicable fastica, picard and infomax. I decided on fastica which is more dependant than picard and faster than infomax.

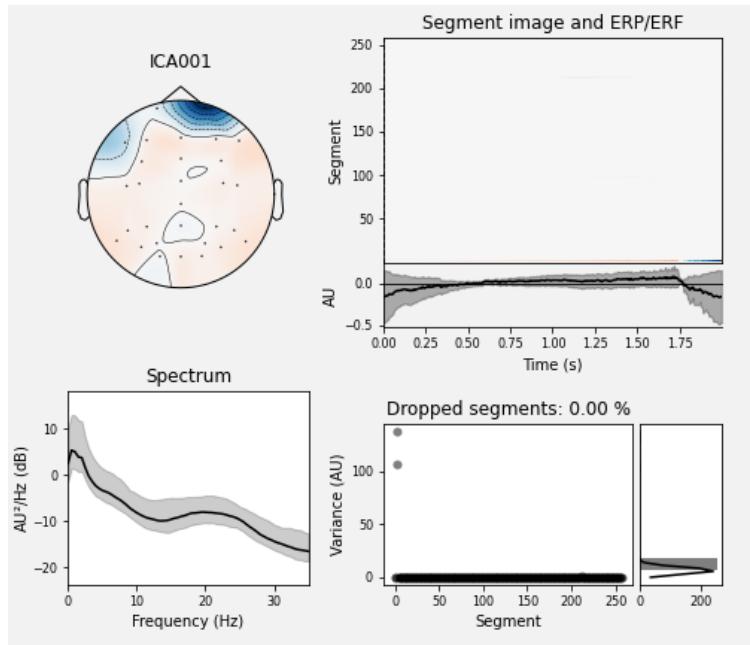
(<https://www.iis.org/CDs2017/CD2017Spring/papers/ZA832BA.pdf>) shows some differences between the algorithms to perform ica. This is also my reason for fastica.

ICA decomposition of subject 6. I removed the following components:

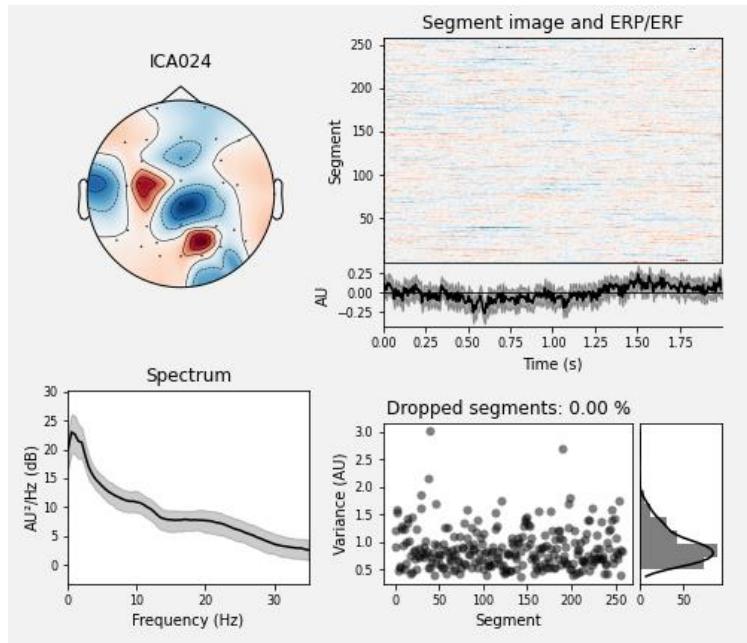
- Component 0: This is an obvious eye blink.



- Component 1: is some noise, ERP also seems odd.

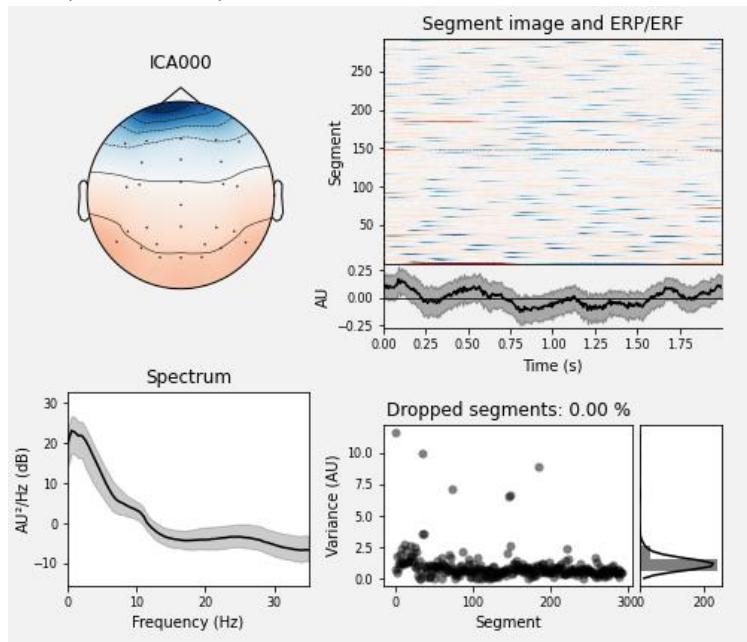


- Component 24: is very noisy, might be a muscle



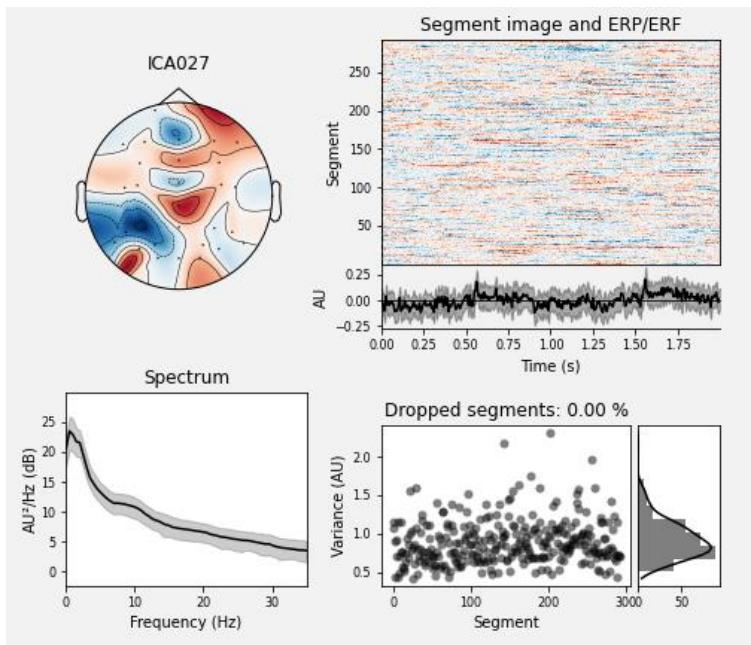
ICA decomposition of subject 13: I removed the following components:

- Component 0: eye blink



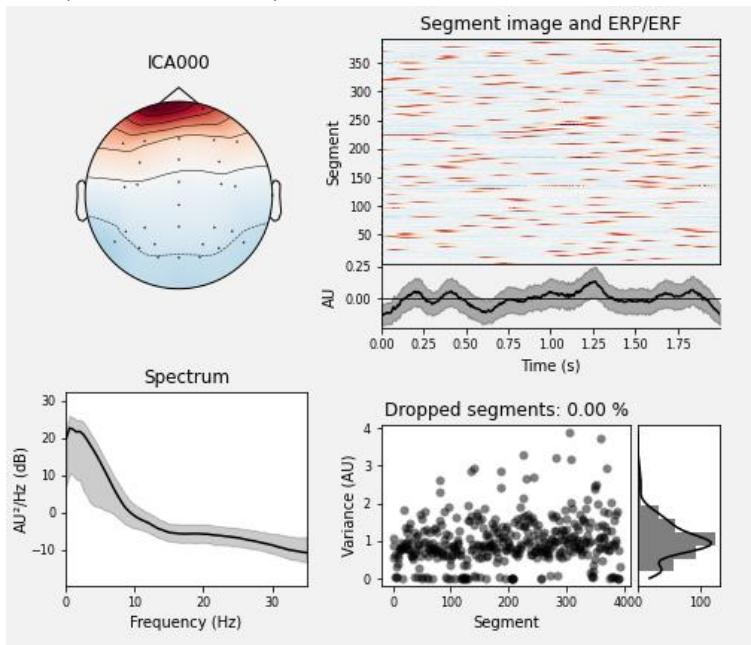
- Component 27: seems to me to be a muscle movement on left side of head, but could also be some erp signal judging from the spectrum. As I cannot be fully sure on this

one I did not remove this one.

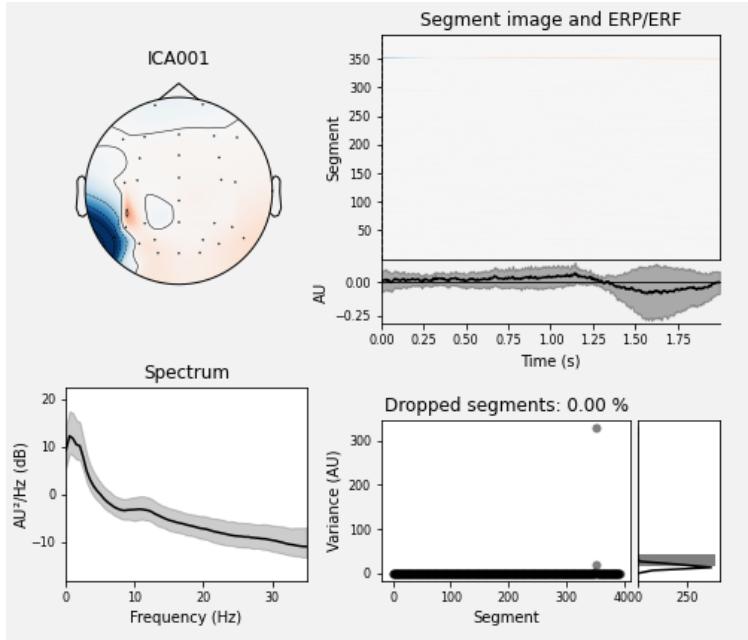


ICA decomposition of subject 8: I removed the following components:

- Component 0 is an eye blink



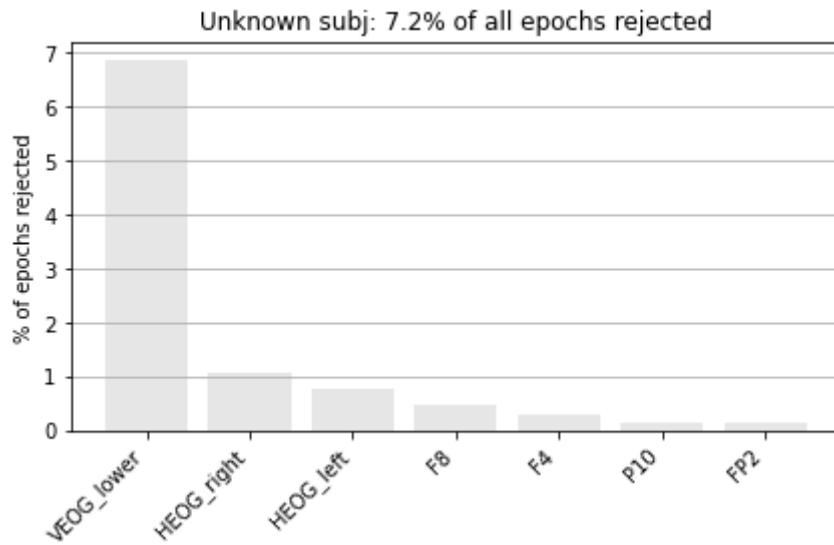
- Component 1: other, no ERP signal after stimulus and only small ups and downs



After applying ICA I checked the plots of raw, filtered, re-referenced and post ica. Each plot showed significant improvements for noise and after removal of eyeblinks with ica cleaner data. Similar to mne documentation of ica ([https://mne.tools/dev/auto\\_tutorials/preprocessing/40\\_artifact\\_correction\\_ica.html#sphx-glr-auto-tutorials-preprocessing-40-artifact-correction-ica-py](https://mne.tools/dev/auto_tutorials/preprocessing/40_artifact_correction_ica.html#sphx-glr-auto-tutorials-preprocessing-40-artifact-correction-ica-py)) the artifacts of eyeblinks are repaired after the application of the ica.

## Data Cleaning:

The rejection of bad segments is hard to reproduce if done by eye. So I decided to use rejection criteria on epoched data which were defined for each subject/dataset individually. As each persons ERP is different the rejection criteria had to be tuned for each subject. I did the cleaning on the stimulus epochs as the response epochs have no duration and are always filtered by using rejection criteria.



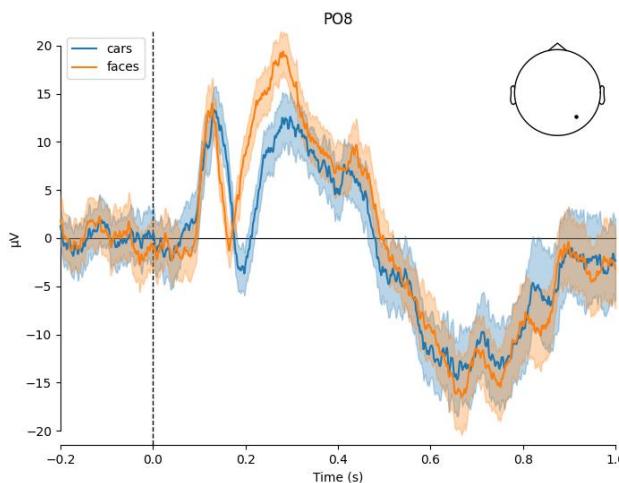
Example drop log I plotted for orientation. Notice that mostly the EOG channels cause dropped epochs and loss of data for some reason.

Here I used a sanity check to fine tune the rejection criteria. I viewed a plot of the epochs first. Then I did a rejection for 200e-6 for both eeg and eog. The result I viewed with the drop log. If I found that too many epochs were dropped then I tuned towards a less strict value. With eg.80% dropped I would lose too many epochs with valuable data. For subject 7 I tried cleaning, there were two cases. Either all epochs are dropped because of channel FP1 or almost none at all. This could be because the data was too clean before dropping or the channel FP1 was a bad electrode.

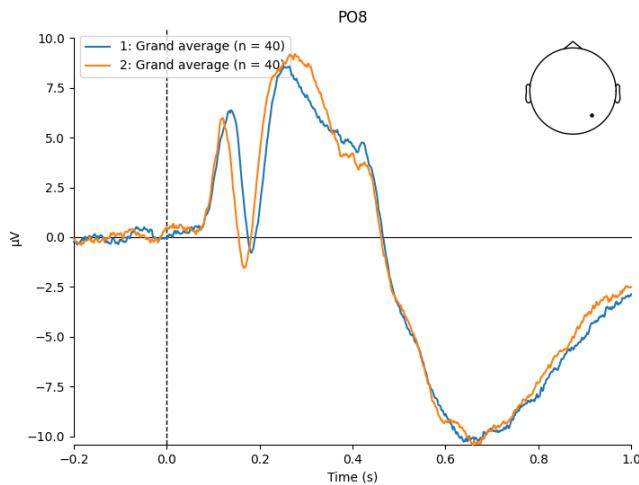
#### Autoreject:

As I did not know if this machine learning pipeline was allowed, I did not want to present this as main solution for my data cleaning part. I used this as a comparison and sanity check. For that I compared on the epoched data on the PO8 channel (see literature ERP CORE).

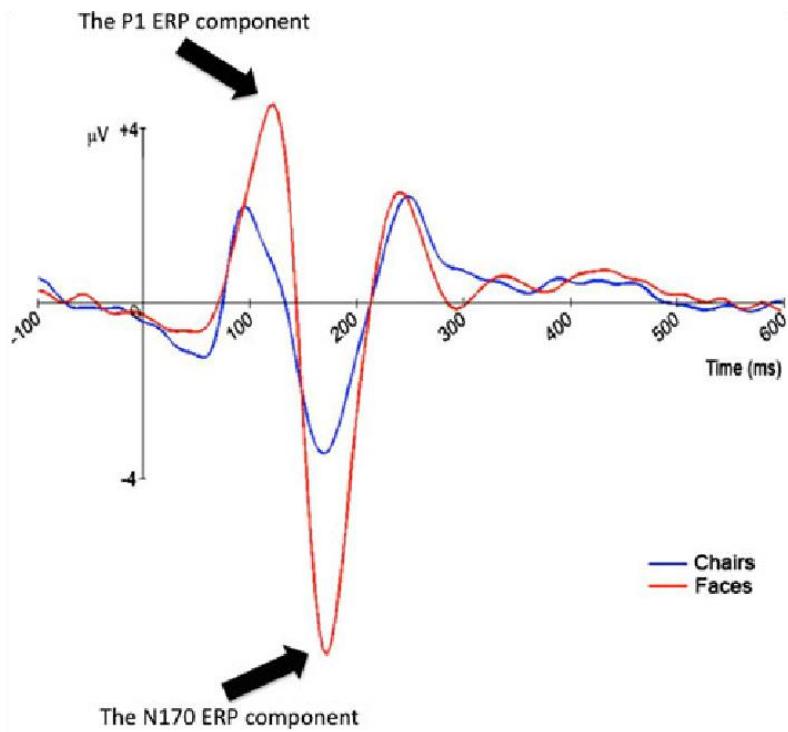
ERP of subject 37 on channel PO8:



Averaged ERP showing PO8 channel:



This seems correct to me in comparison to literature N170 ERPs. For example the following image of an N170 ERP. Note that their P1 and N170 seems to be much more extreme for faces than ours. For that I ran some tests mentioned later on in the ERP Peak Analysis part.



source: [https://www.researchgate.net/figure/The-P1-and-N170-ERP-components-The-graph-shows-the-grand-average-ERP-responses-from-ten-subjects-faces-in-red-and-chairs-in-blue-fig1\\_49833438](https://www.researchgate.net/figure/The-P1-and-N170-ERP-components-The-graph-shows-the-grand-average-ERP-responses-from-ten-subjects-faces-in-red-and-chairs-in-blue-fig1_49833438) The P1 and N170 ERP components. The graph shows the grand average ERP responses from ten subjects to faces (in red) and to chairs (in blue). Note that the peaks at 100 ms and at 170 ms are larger for faces. Three channels have been averaged in each subject (P8, PO8, and P10). (see references: 1.)

The rejected channels/segments are documented in “out\_data/” under subject id with manual(with reject criteria) or autoreject. In those files the segments are documented with the onset, duration and a description as BAD\_ .

### **ERP Peak Analysis:**

Thoughtprocess to readData.py which wraps the used functionality of ccs\_eeg\_semesterproject.py are explained in ERP-Peak-Extraction.ipynb .

### **Preprocessing&Cleaning:**

Here I used the provided pre-processed and cleaned data and loaded with ccs\_eeg\_semesterproject.py. These functions I wrapped in readData.py which I continued to use in the later parts of my semester project.

### **Mapping to dictionary:**

I changed the stimulus dictionary to the following {faces, cars, scrambled\_faces, scrambled\_cars} to be easier to understand. For each condition I created an evoked by using the average. From the averaged evoked I used get\_peak() to retrieve peak amplitude and timing. Here I noticed that the method gets the global peak, so I had to crop the data. As recommended in ERP CORE I used the following interval 110ms to 180ms for intact stimuli and 110ms to 250ms for scrambled ones as they tend to be more delayed. This resulted in more sensible peaks.

The resulting table looks like the following one. Values are imaginary here. For exact ones please look at erpData.csv.

<b>subjectid</b>	<b>PO8</b>	<b>time</b>	<b>baseline</b>	<b>stimulus</b>	<b>condition</b>
1	5.7	0.14	-3	face	intact
1	8.5	0.16	1.3	car	intact
1	-5.3	0.18	-1.2	face	scrambled
1	-6.28	0.19	-1.6	car	scrambled
2					

### **Remarkable points of the ERP:**

This part I used the grand average for each condition over all subjects. For each condition there is one evoked. These evoked can be compared among each other.

Calling get\_peak() the resulting timing was around 151ms for the grand averaged evoked.

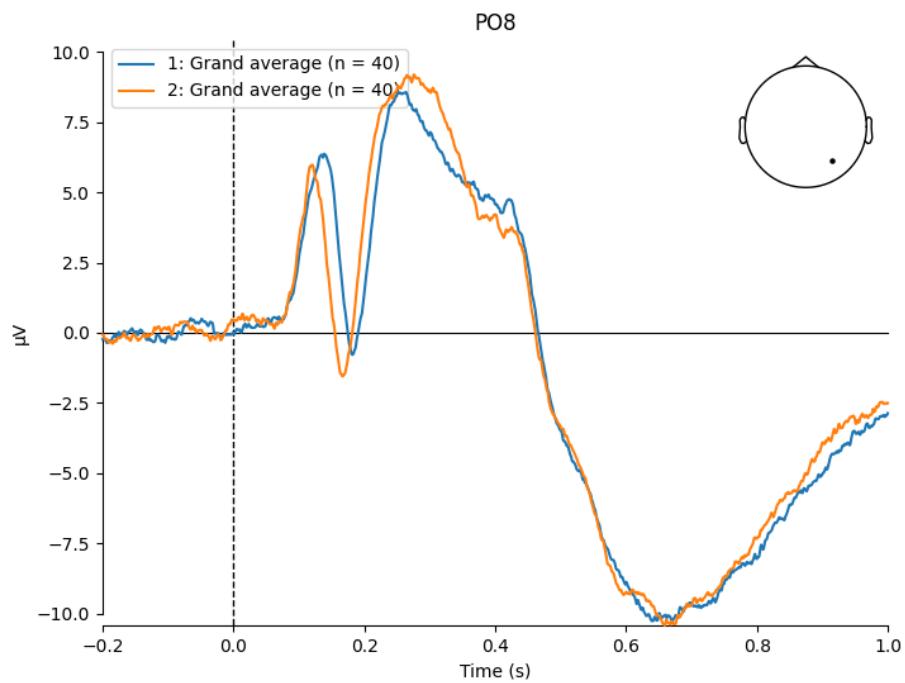
For one of these I compared according to ERP CORE face and car stimuli waveforms.

### **For Comparison I also used the PO8 channel:**

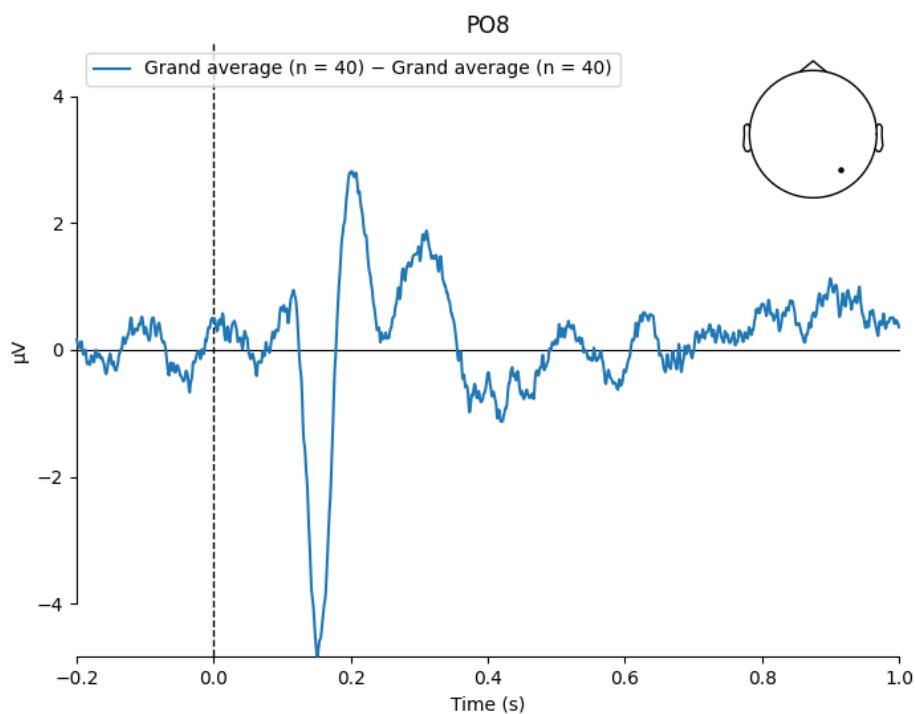
ERP of intact stimuli

Orange line: stimuli faces

Blue Line: stimuli cars



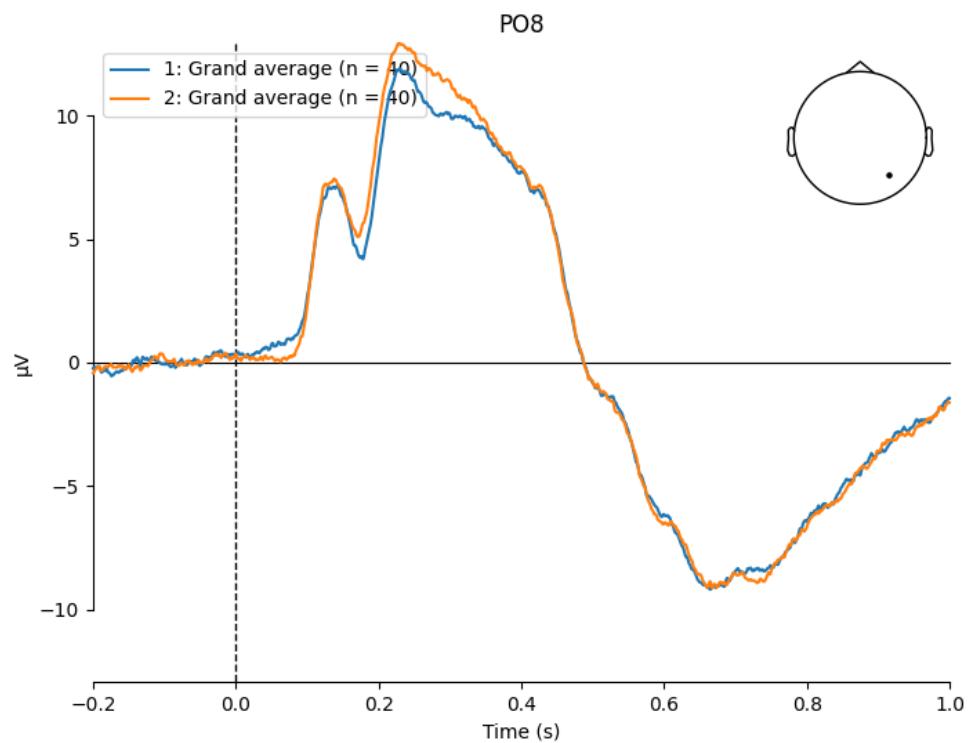
Difference of ERPFace – ERPCar



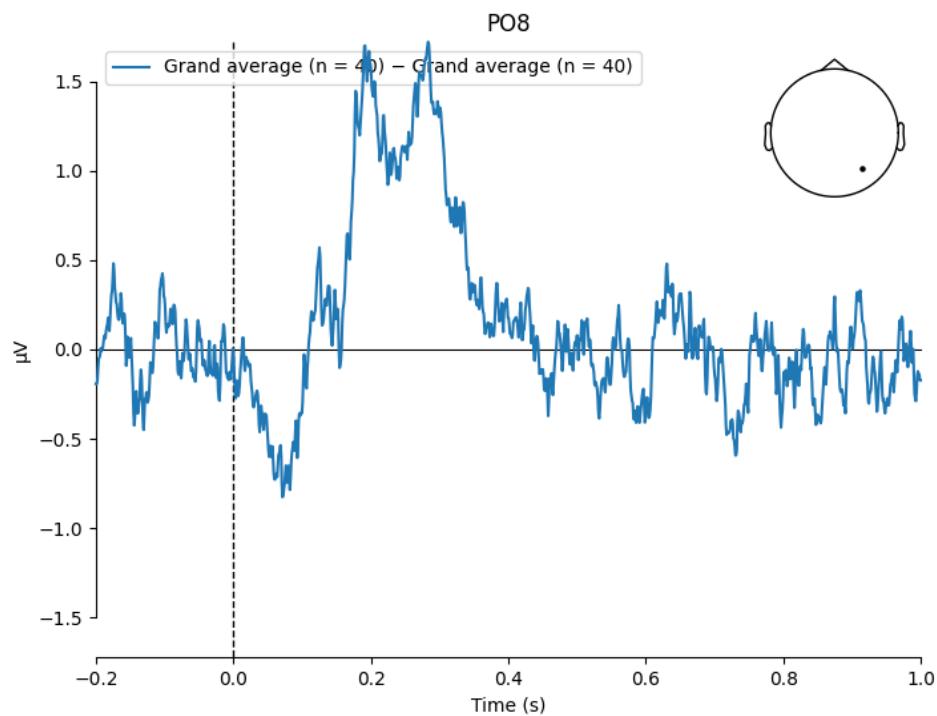
ERP of scrambled stimuli

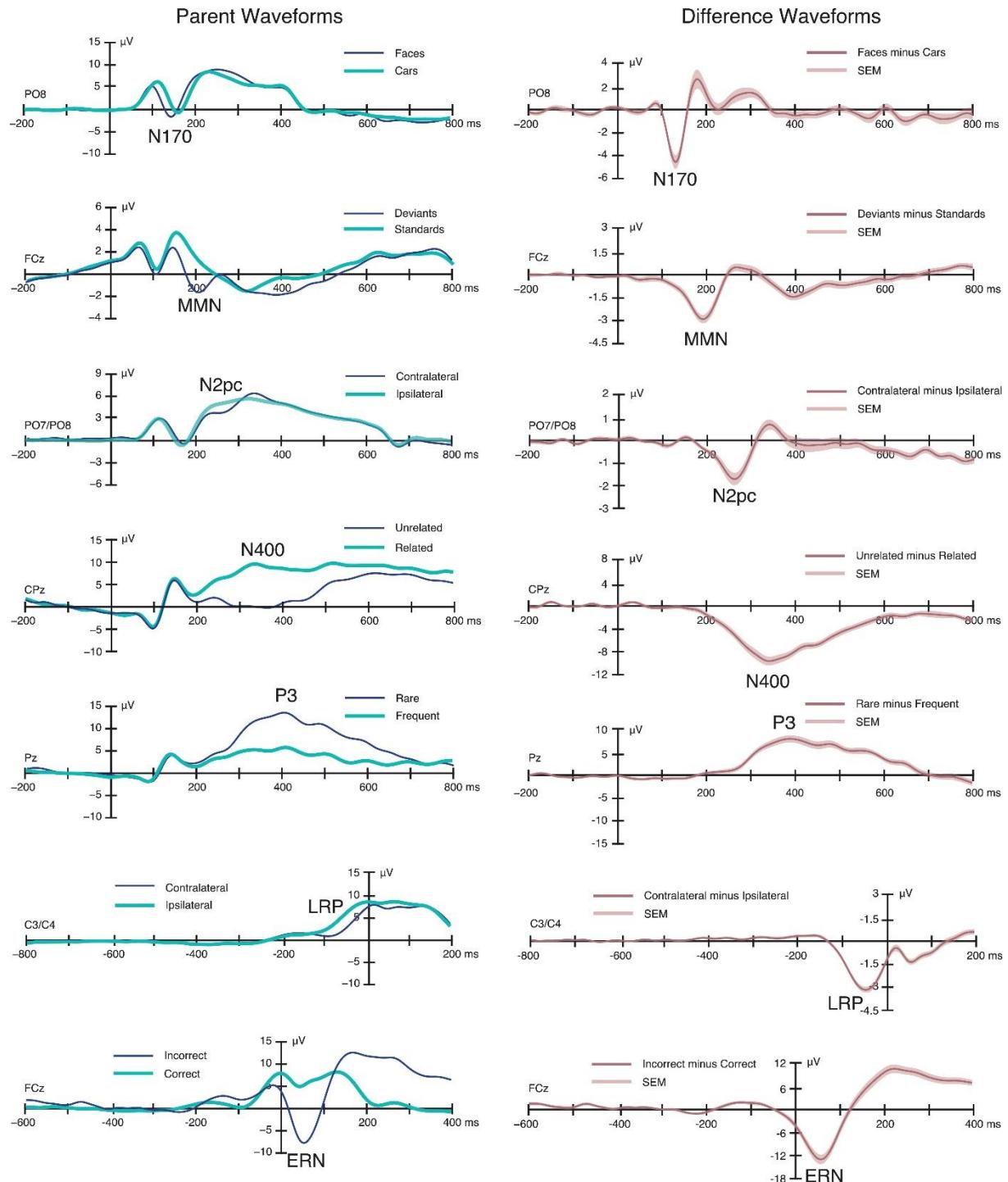
Orange: face

Blue: car



Difference of ERP\_scrambled\_face – ERP\_scrambled\_cars

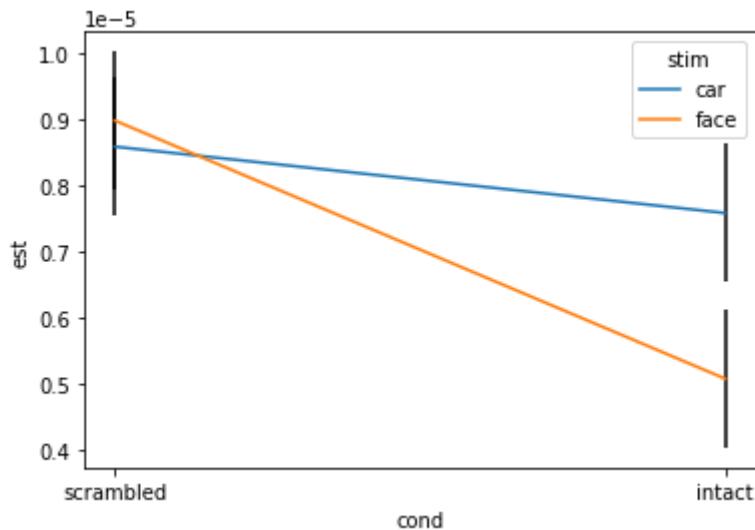




Unsurprisingly the N170 waves were similar, and the face peaks were larger and earlier than the peaks for cars. I also did a comparison between scrambled faces and scrambled cars. Here the difference was less significant and the N170 peak for cars was larger than for faces.

### Linear model to confirm the obeservations:

The linear model shows me, that scrambled differences are not significant where the confidence intervals overlap while the intact ones do not overlap. For full documentation please view the ttest.ipynb file.



In the following part I also ran some one sampled t-tests between the condition to confirm this plot.

#### Extraction of Peaks:

##### Statistically Testing Peaks:

For ERP Peak analysis I used all data and statistically checked them for relations and their significance with multiple one-sided t-tests.

I calculated the mean(face) – mean(car) for the one sampled t-tests.

##### Test results:

- One sampled t-test for comparison of conditions face – car: t-value: -2.34, p-value: 0.024
- One sampled permutation cluster test for comparison of conditions face-car: one cluster with p-value: 0.22  
⇒ Significant difference between the conditions face and car
  
- Comparing between intact faces and intact cars proved to be significant with a p-Value of 0.0065
- Comparing between scrambled faces and scrambled cars showed to be non significant with a p-value of 0.23 > 0.05
- Obviously comparing between intact and scrambled proved to be significant with 0.00005

The test results also confirm that the difference between the scrambled conditions is not significant. Whereas the difference between faces and cars is significant. Obviously, the difference between intact and scrambled turned out to be significant.

Using a one sampled permutation cluster test on the data also results in similar values and prove further the results above.

## Time frequency analysis

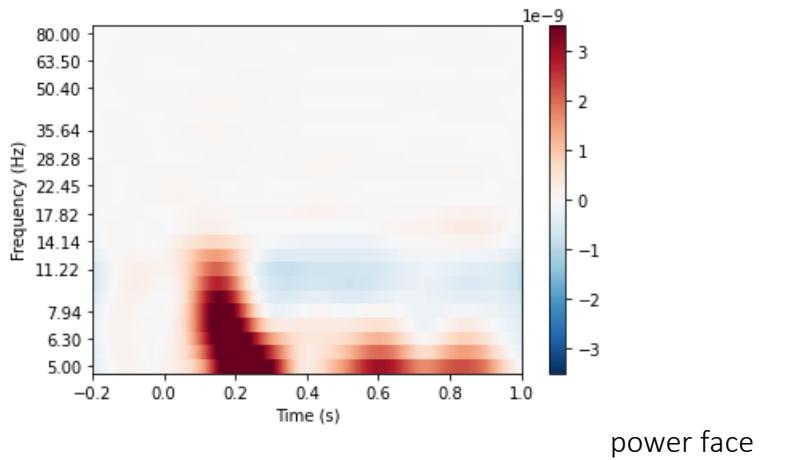
I ran a grand average over each condition for all subjects. This results in 4 values. One for intact faces, one for intact cars, one for scrambled faces and one for scrambled cars.

From the epoched data I got induced powers for each condition and subject and ran a grand average for each condition. This also results in 4 values.

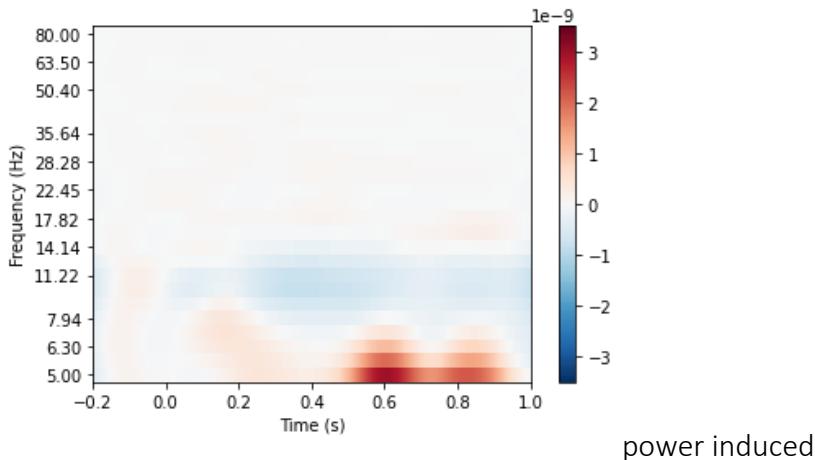
Overall there are 2 values per condition. One for the power and one for the induced power.

With these values I got the difference of power for the condition face.

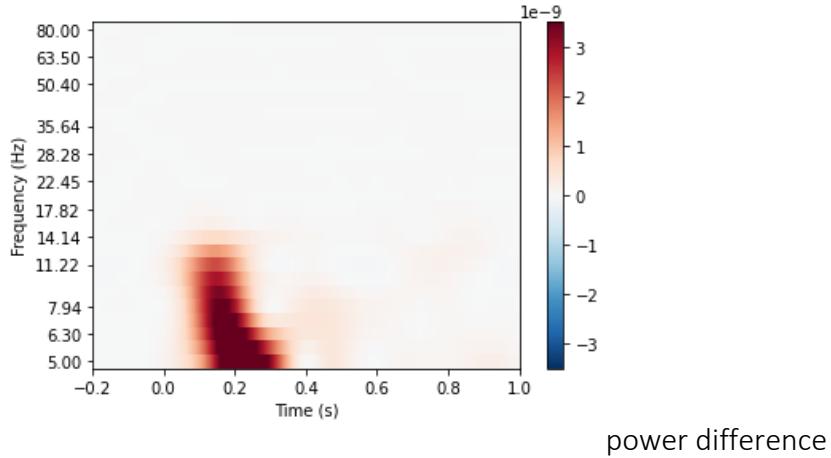
This is an example on subject 7. The main time of the onset of higher amount of frequencies is around 150-200ms. According to ERP CORE the values should lie between 110ms and 180ms. This seems acceptable as this is just one of 40 subject. The full time frequency plots are in the Tf\_singleSubject.ipynb notebook.



power face



power induced

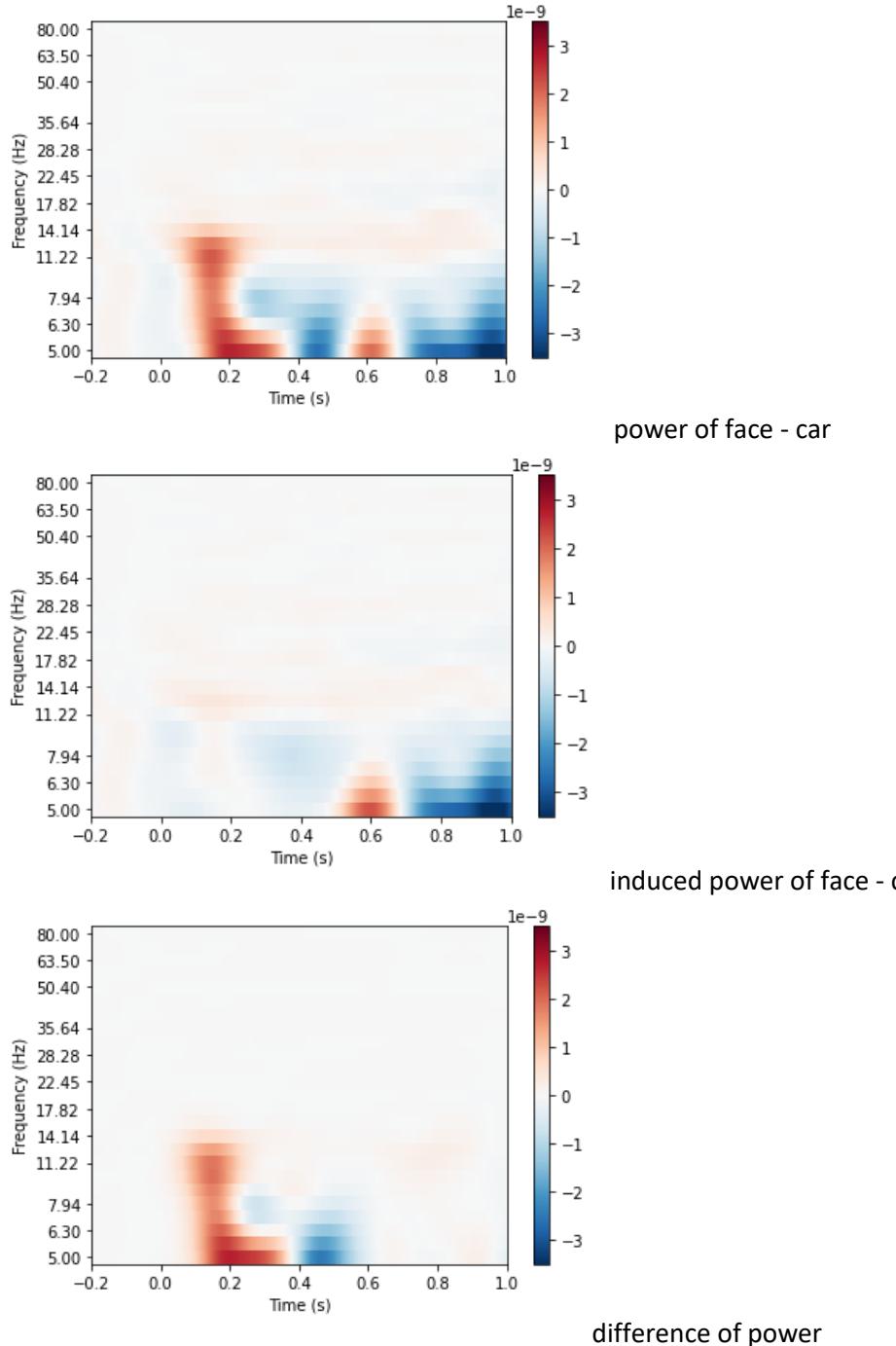


The time frequency plot for the power of the condition face shows that around 200ms after stimulus presentation there are high power at lower frequencies.

The difference plot confirms that there are many lower frequencies around 200ms. This shows that there is a high response from the brain after the presentation of the face stimulus. The same goes for the car stimulus.

For the full 40 subjects I was not able to run that many datasets at once as my computational power was limited. I ran some visualisations for 20 subjects.

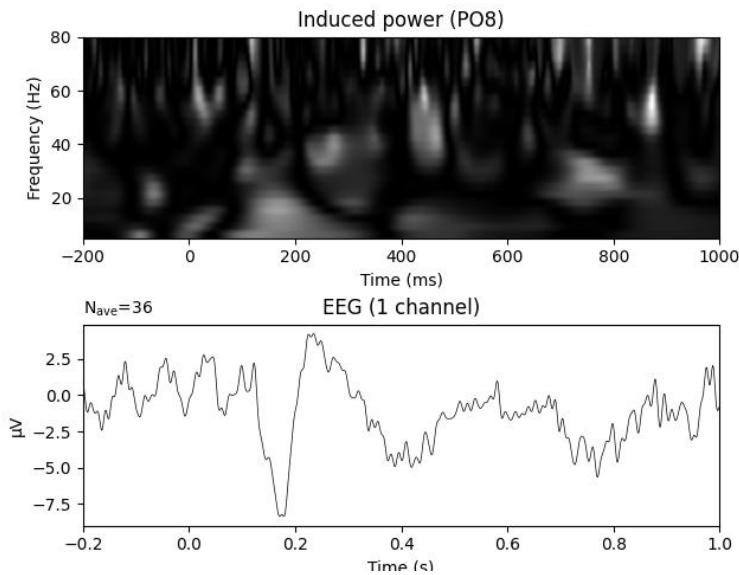
For condition face – car I took the combined evoked of them with weights 1 and -1. Same for their induced spectra. From here I calculated the power difference.



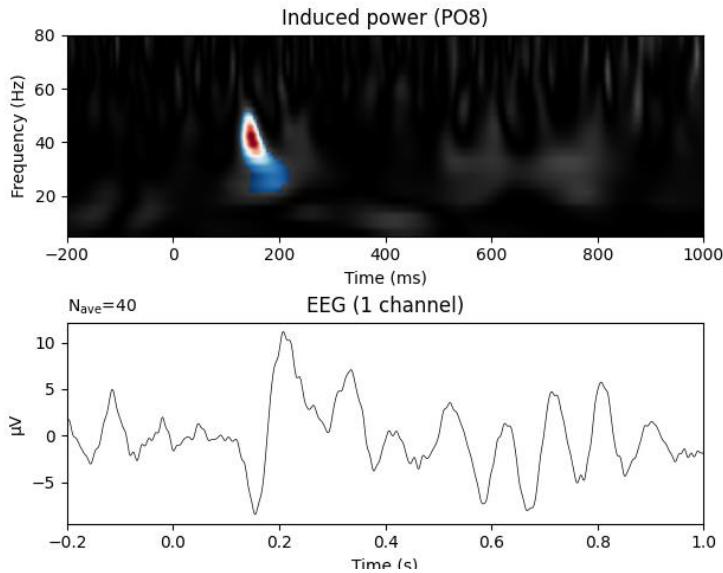
The plots show an increased amount of low frequencies around 200ms. This seems to make sense. As the main difference should be around 170ms. There are some higher frequencies in the difference as expected. This shows that the response for the face stimulus is stronger than for the car stimulus.

I used a one sampled permutation cluster test to show the significance of the results. Following plots show the significance. Most subjects showed a significant difference at around 170ms. This is as expected from the N170 dataset.

For example subject 40 did not show a significant difference



Comparison to subject 38 with the following result.



I also ran if for each subject which can be found in the img folder.

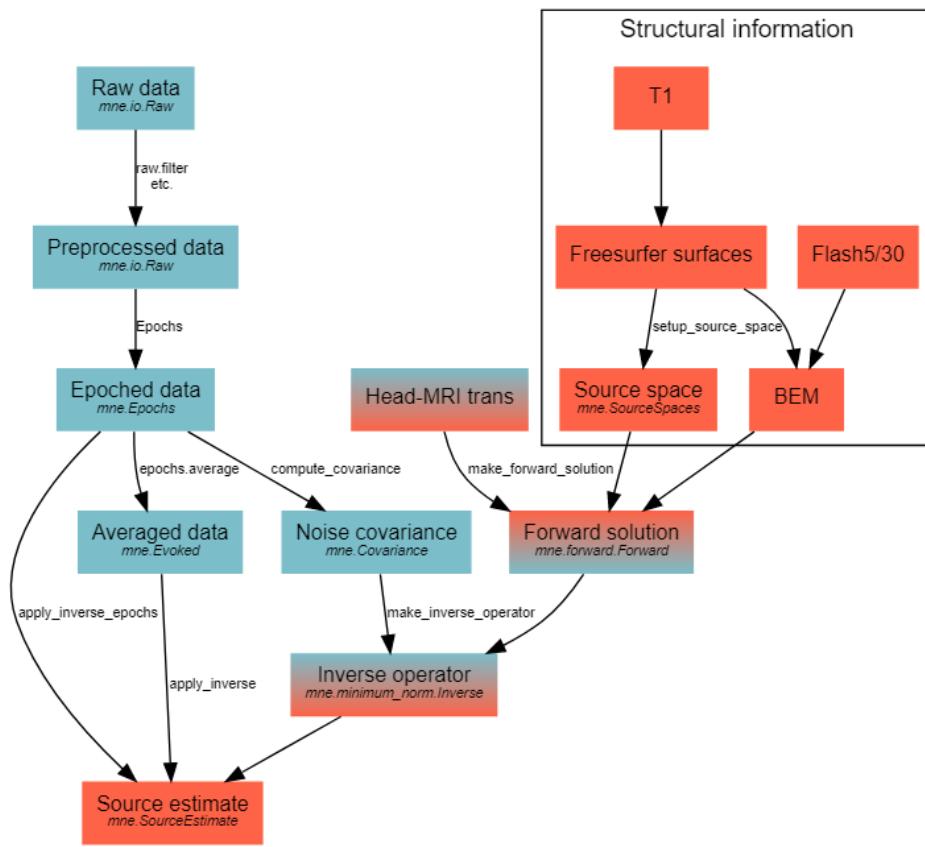
There were some subjects that showed no power for faces or an earlier car recognition than for faces. Overall in the grand average this did not affect the outcome. These subjects require further investigation.

### Source estimation for N170

I ran this part on two different PCs one with more power than the other one. My biggest problem in this part was the installation of pyvista/pysurfer+mayavi. My recommendation after days of debugging is using pysurfer+mayavi only on Linux/Unix and pyvista on Windows systems. The dependencies are too restrictive and installation fails in most cases.

Now for the setup:

For source estimation I used the commonly used pipeline:



I used an already-segment "default" MRI 'fsaverage'. As we do not have individual headmodels I used the model and assumed that all vertices and dipoles are at the same position for all subjects. This of course should not be done in practice.

Results for a grand average for faces and for cars showed a difference in the timing of the peaks for each condition. That is already shown in the tfr. The source of the signal is from the visual cortex of the human brain. Most times the right half activates earlier and more intense.

Timings of the peaks can also be found from the time frequency analysis and from ERP Peak analysis.

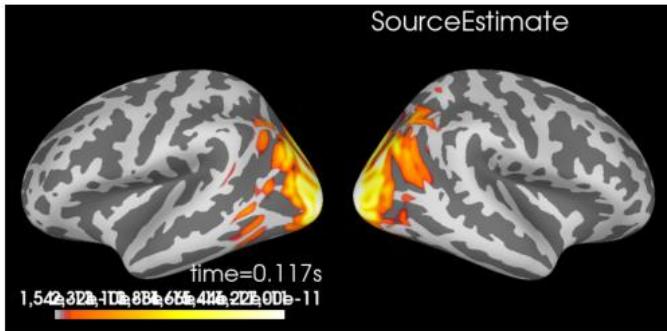
For a difference of the face and car stimuli the difference is between the peak of face and car recognition. This is as expected like in other literature

( <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6637847/> ). The results seem similar as the left cortex seem to have less activity than the right one.

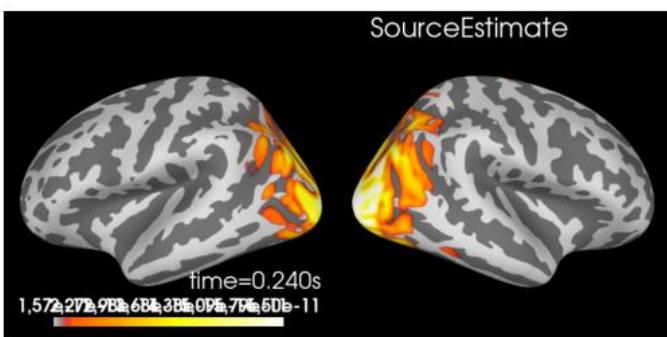
My results might differ from literature as my computer could only handle data of 20 subjects for source estimation at once. I also ran the full pipeline subjectwise. The results can be found in the img folder.

Following are reference images from running the source estimations:

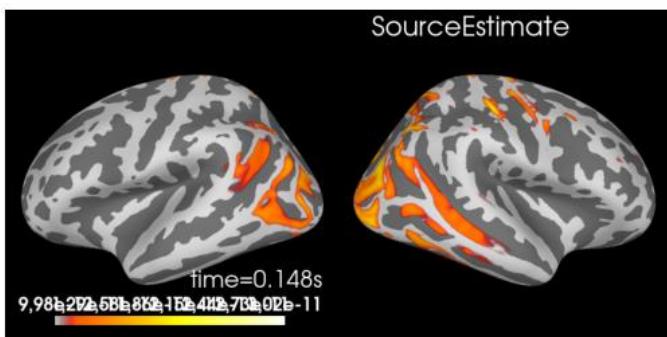
Grand average for face stimulus



Grand average for car stimulus



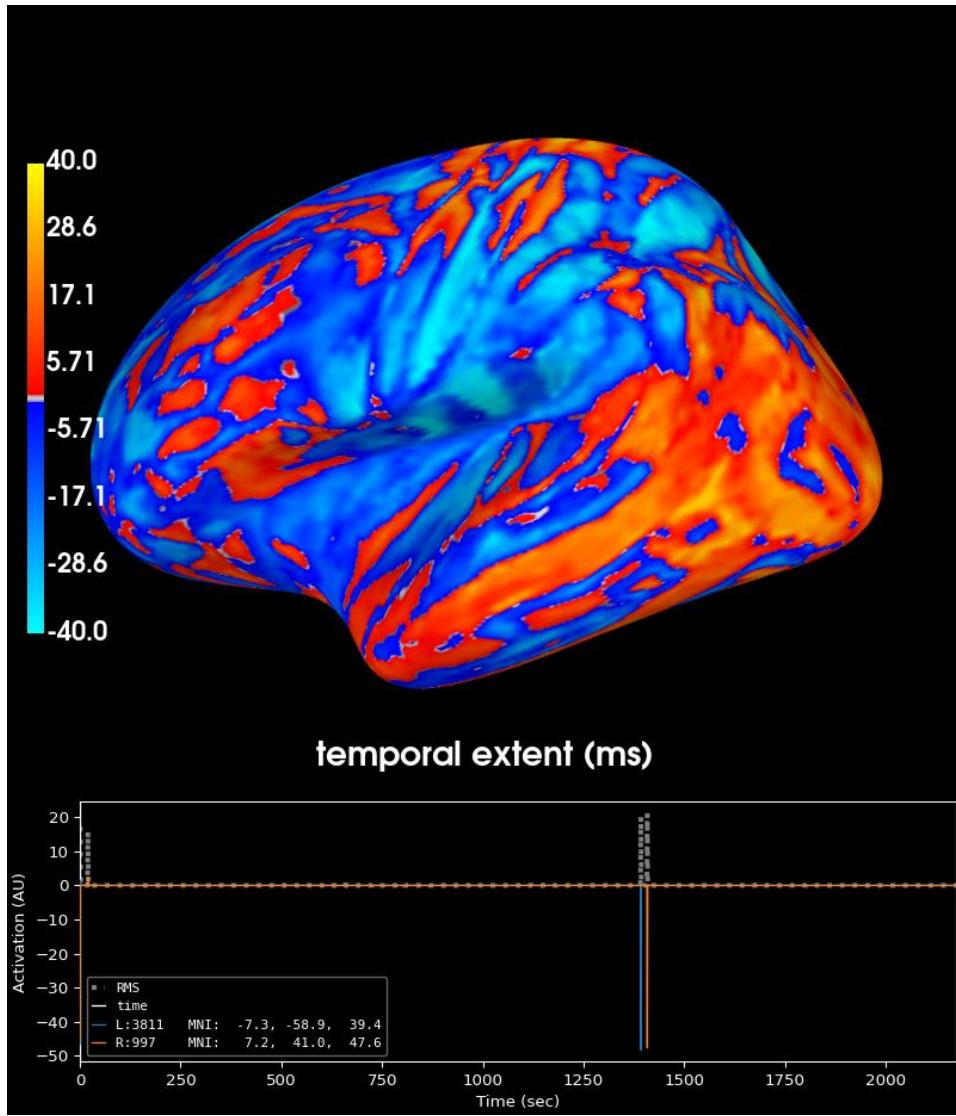
Comparison between face and car stimulus. The area with color is the difference of face – car



There are also some interesting subjects that show less/no peak for a face stimulus. Some subjects recognized cars faster / more intense than a face stimulus. One might speculate that those subjects are some car experts or have some disability in recognizing faces.

Statistics for source estimate:

Statistical Testing according to [https://mne.tools/dev/auto\\_tutorials/stats-source-space/20\\_cluster\\_1samp\\_spatiotemporal.html](https://mne.tools/dev/auto_tutorials/stats-source-space/20_cluster_1samp_spatiotemporal.html) with a one sampled spatio temporal cluster test. Red colored areas are more related with the face stimulus and blue ones are more related with car stimulus. The visual cortex is more related with the face recognition.



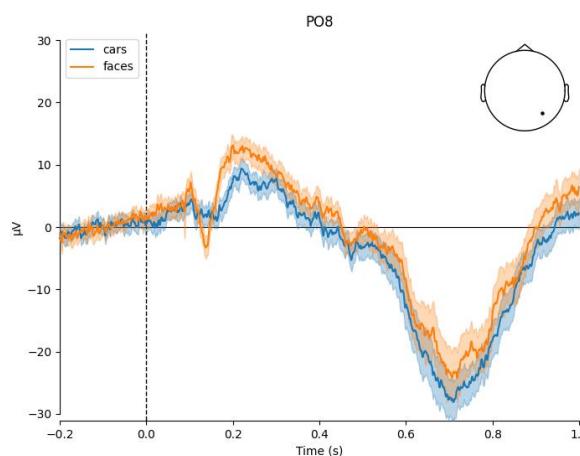
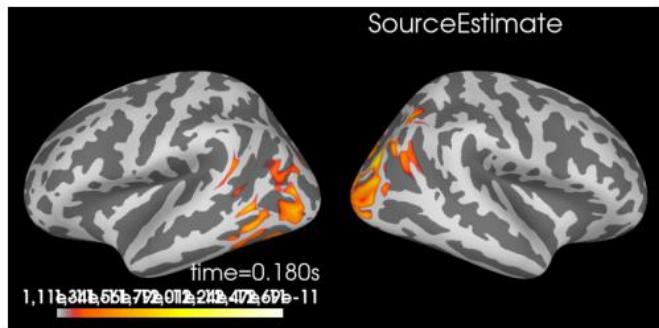
Here the result may differ and interpretations are limited. My PC can only stand 10 samples for this statistical analysis and is at its limits. The amount of data for cluster permutation is far too large if I used all datasets. For more accurate results and better interpretations more computational power or different approaches would be needed.

## Results and discussion:

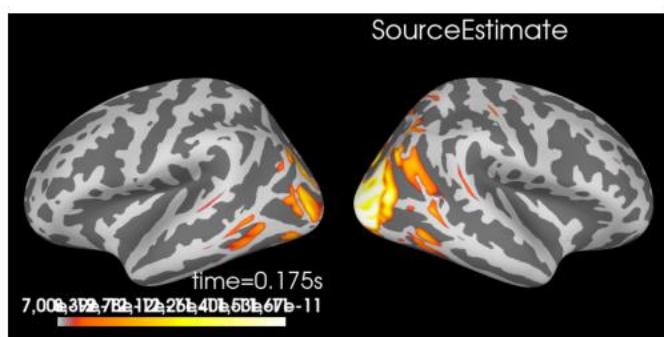
Some subjects require further investigation. Some show noisy data and some show no certain peak. There also are some unclear subjects, that turn out to be good after comparing the conditions. In the following I will show up some examples (subjects 4, 16 and 20).

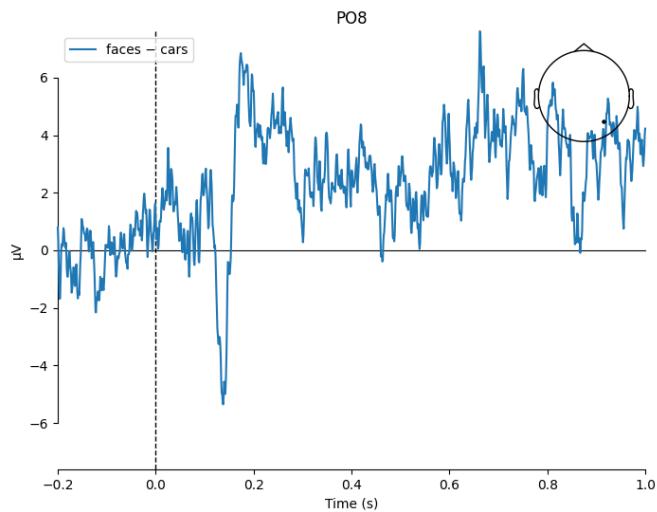
Subject 4 for example shows almost no activity for face stimulus and noisy data even after cleaning and ICA.

The peak for stimulus face is very small:

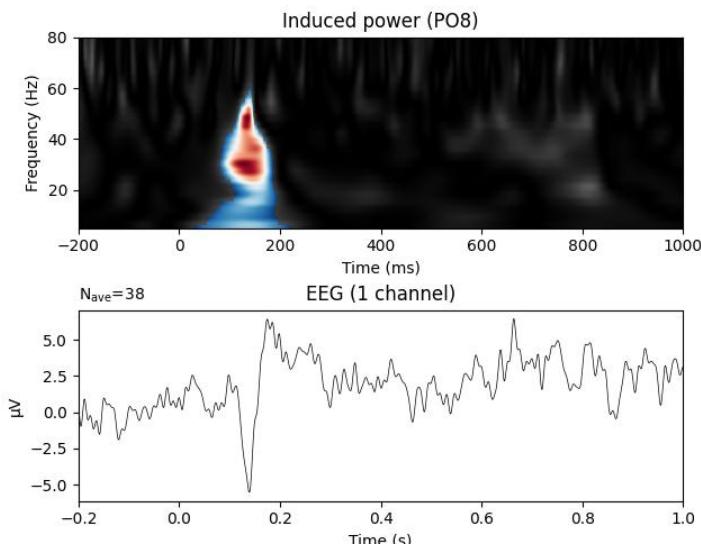


For comparison face-car the peak is almost like out of literature:

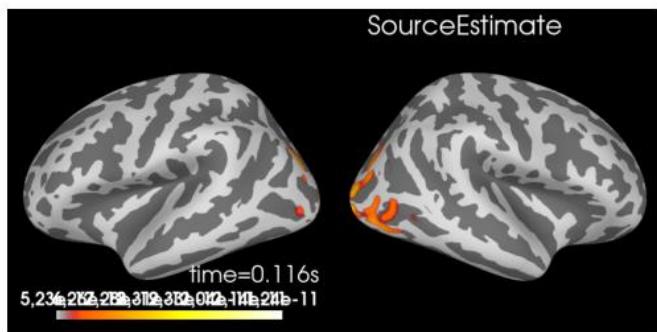




The statistical test for time frequency also shows significant difference between face and car stimulus.

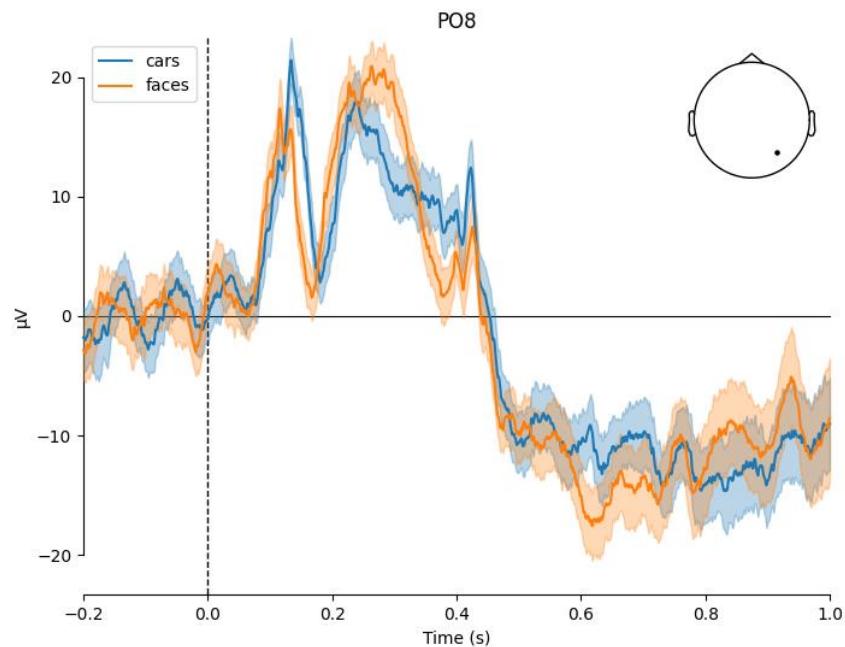


On the other side there are subjects like subject 16:

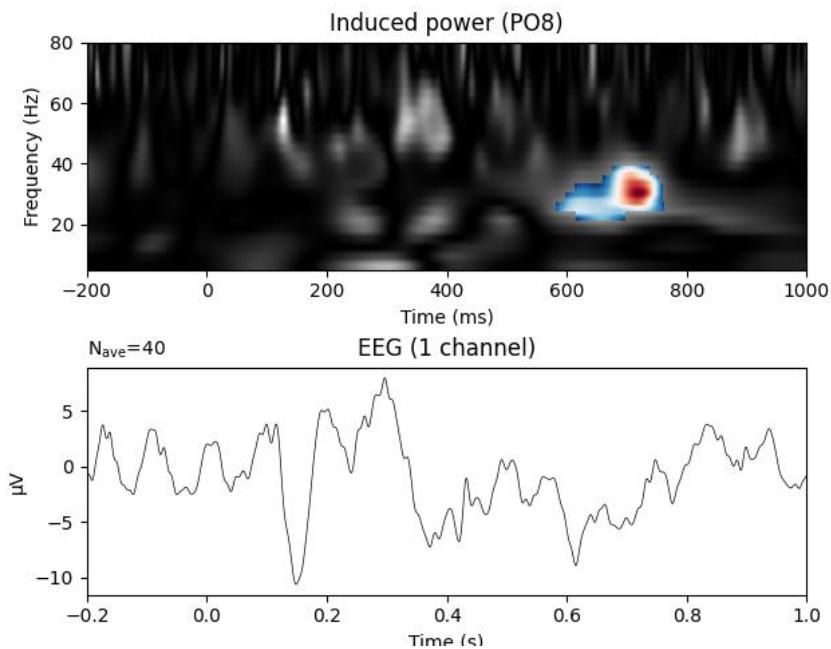


Where the face recognition shows almost no significant signal. The ERP show that the brain is constantly reacting to the stimulus after onset. This can influence the analysis as the expected peak is lower than peaks following the time span. Also interesting is that the peak for car

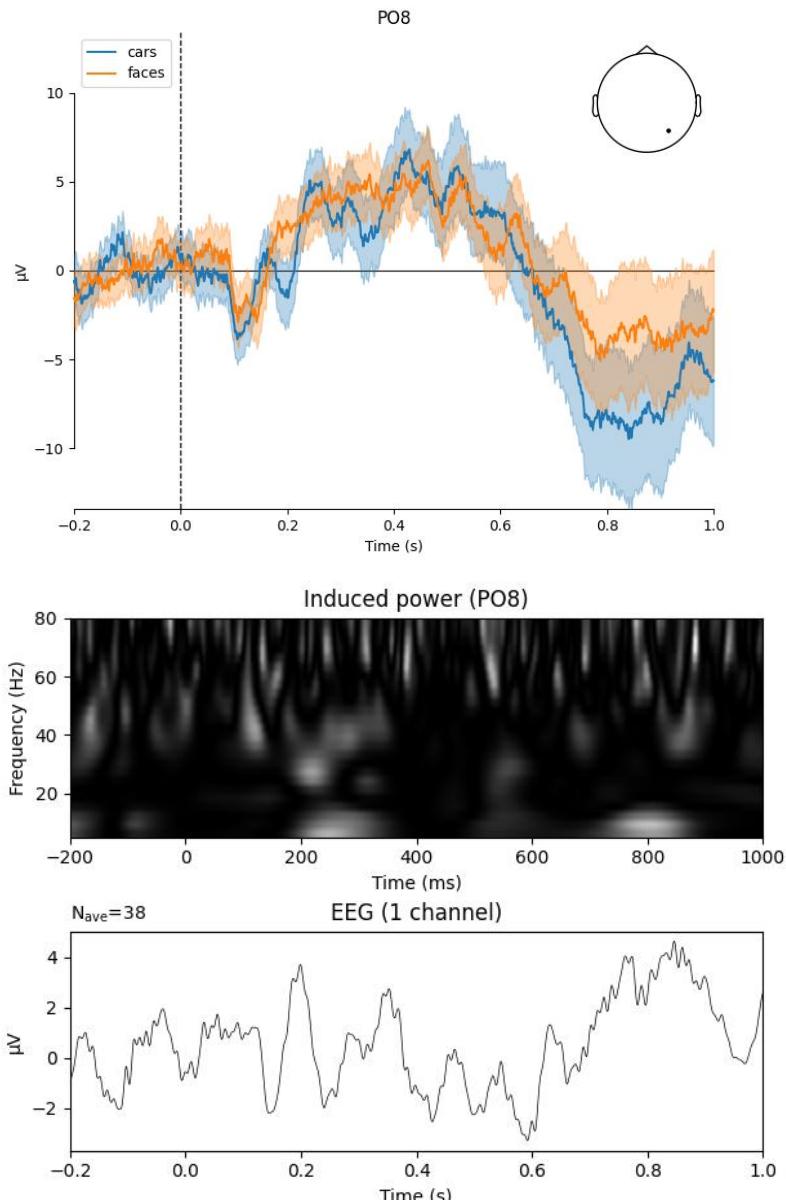
stimulus is more intense than for face. This might be a person who cannot recognize faces that well.



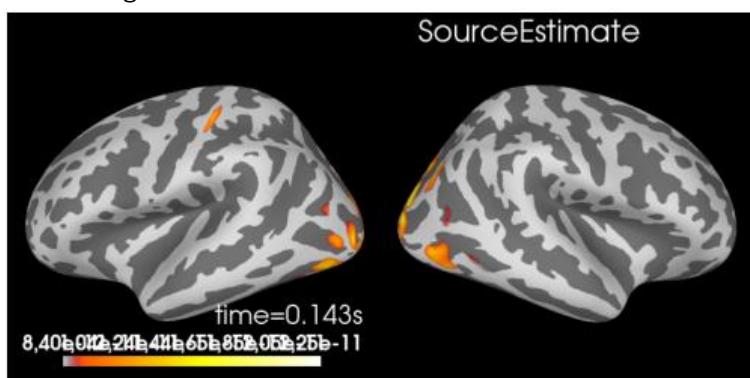
Analysis with time frequency also seems odd. There is a significant difference between face and car, but the timing is outside of our expectation. There is a negative peak around 180ms, but is not as significant as expected. There might be some need to investigate this subject.



Subject 20 also shows noisy data after cleaning:

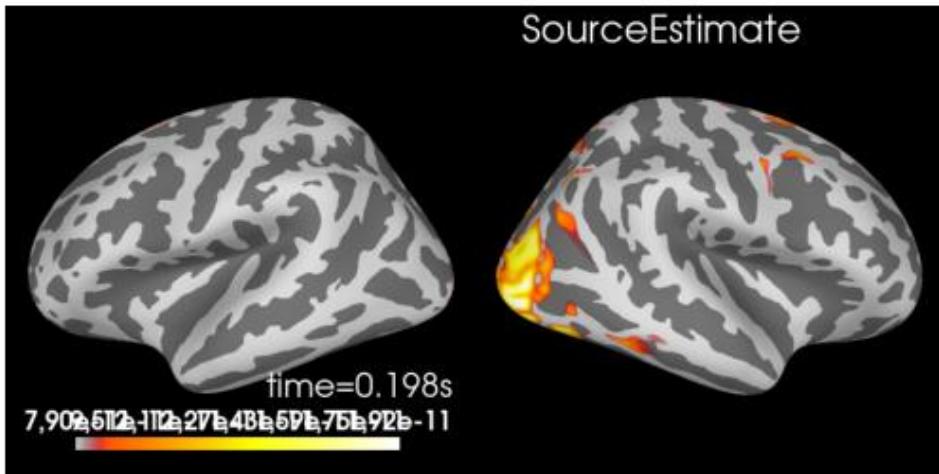


Here no significant difference between the conditions can be found.

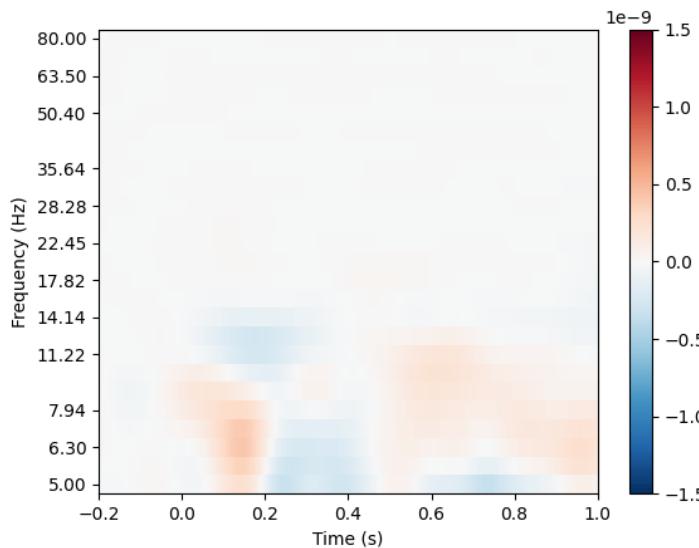


After source estimation there is almost no significant source.

Comparison between the conditions only shows a small area in the visual cortex activating.



The difference in the time frequency is also unclear. There is no certain increase in frequencies.



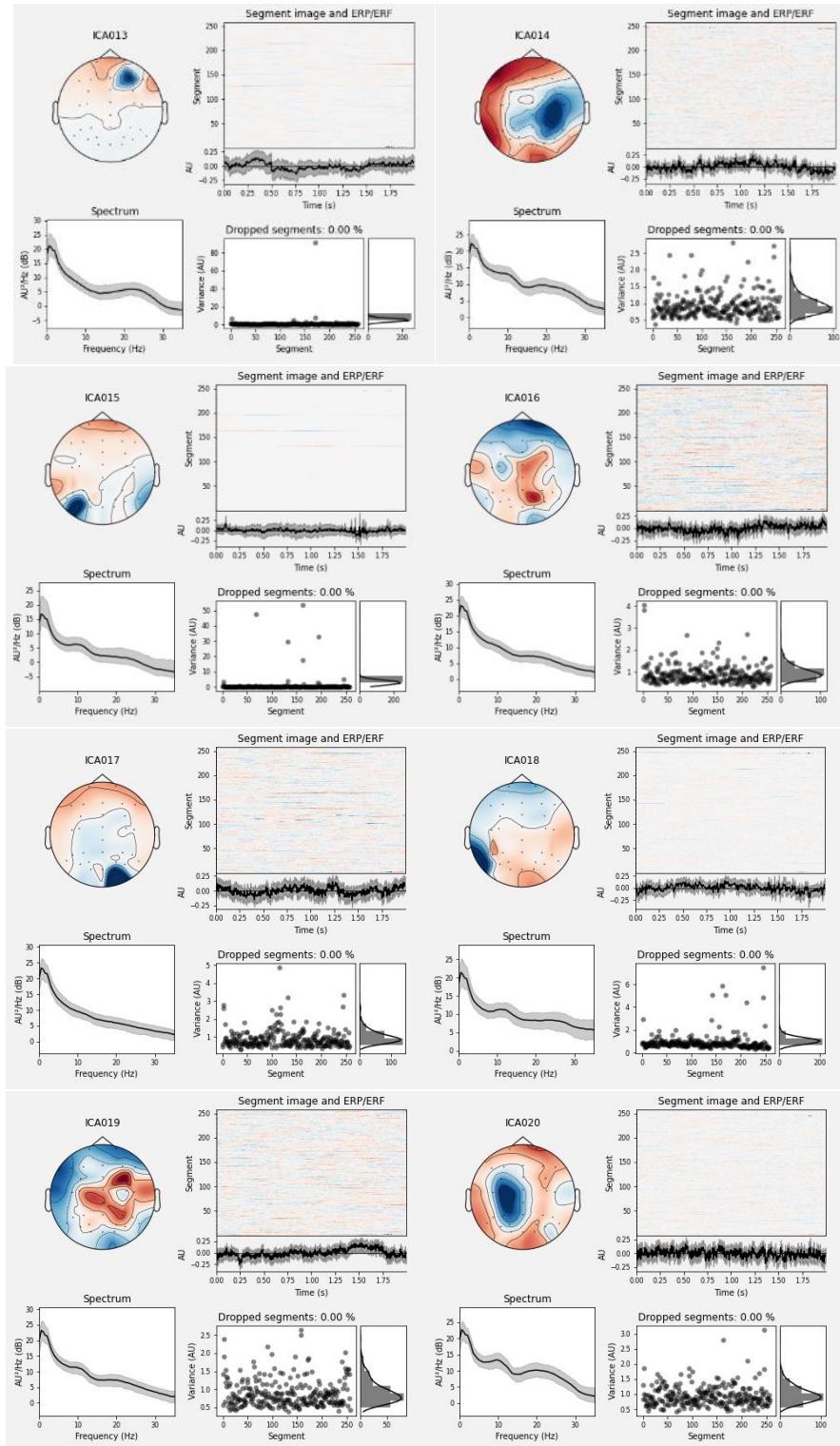
Subject 21 is also an example for noisy data where the ERP shows very large confidence intervals and a N170 peak which is very small in comparison to literature.

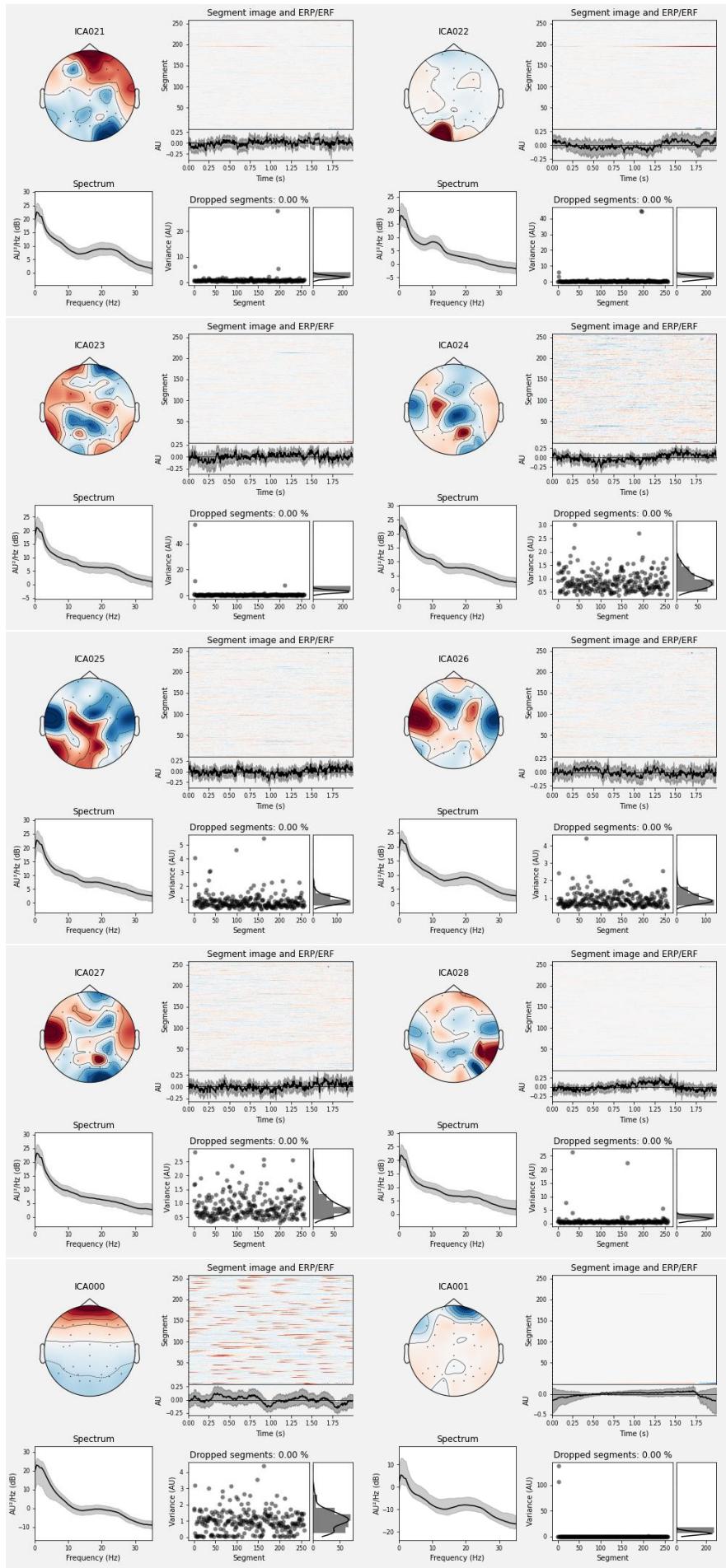
With the full samplesize of 40 subjects covering all kinds of people, that recognize cars faster than faces and people that have difficulty recognize faces, one might try to generalize this on the general. As shown in the t-tests on the ERP peaks there is a significant difference recognizing faces and cars.

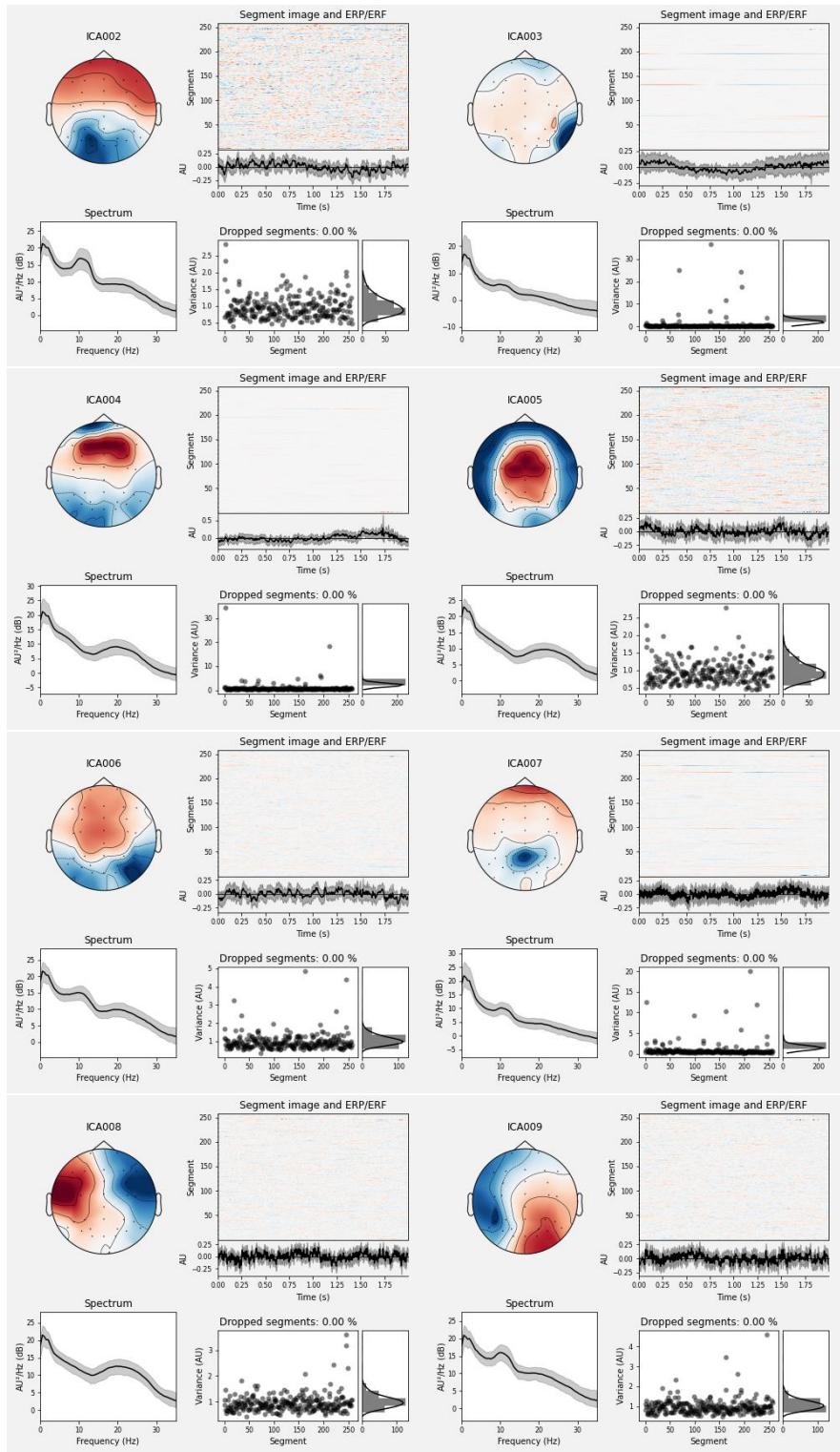
This obviously is not the same for every person. On closer inspection there are subjects that have difficulties. I would recommend collecting more data and analyzing the outcome again to get a better sample size and more reliable results from it. The small sample size of 40 subjects can contain subjects that differ from the group and affect the results by a significant amount.

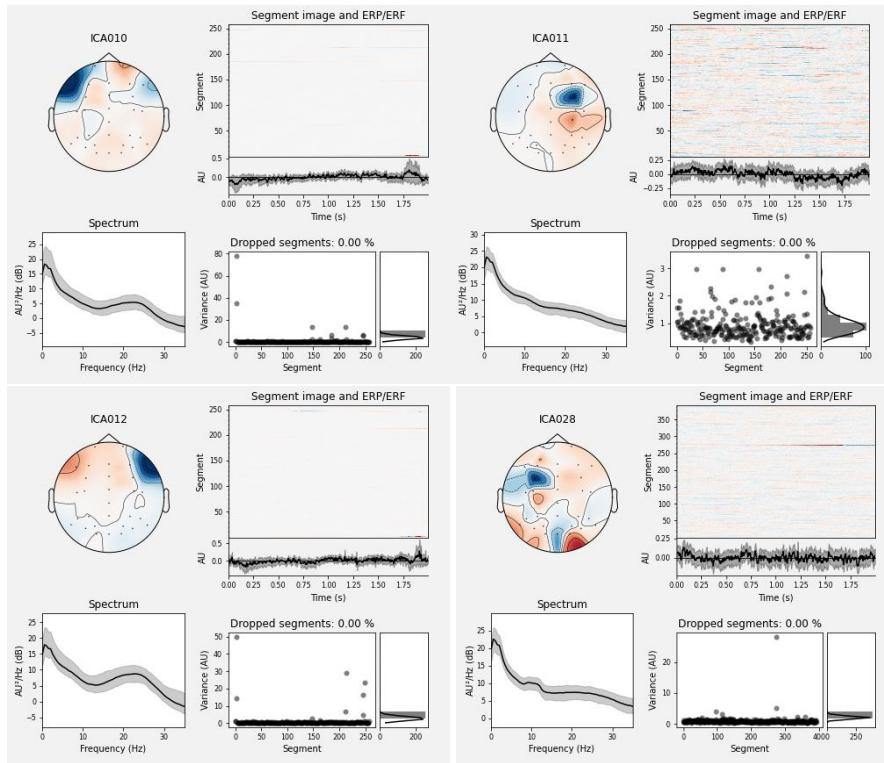
## Appendix

### ICA: subject 6:

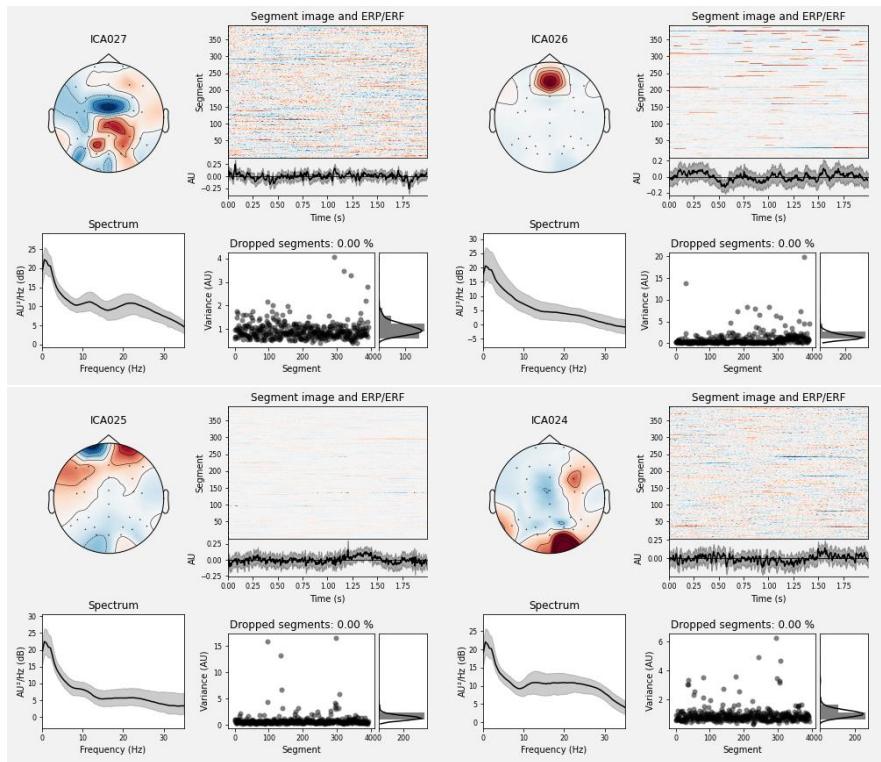


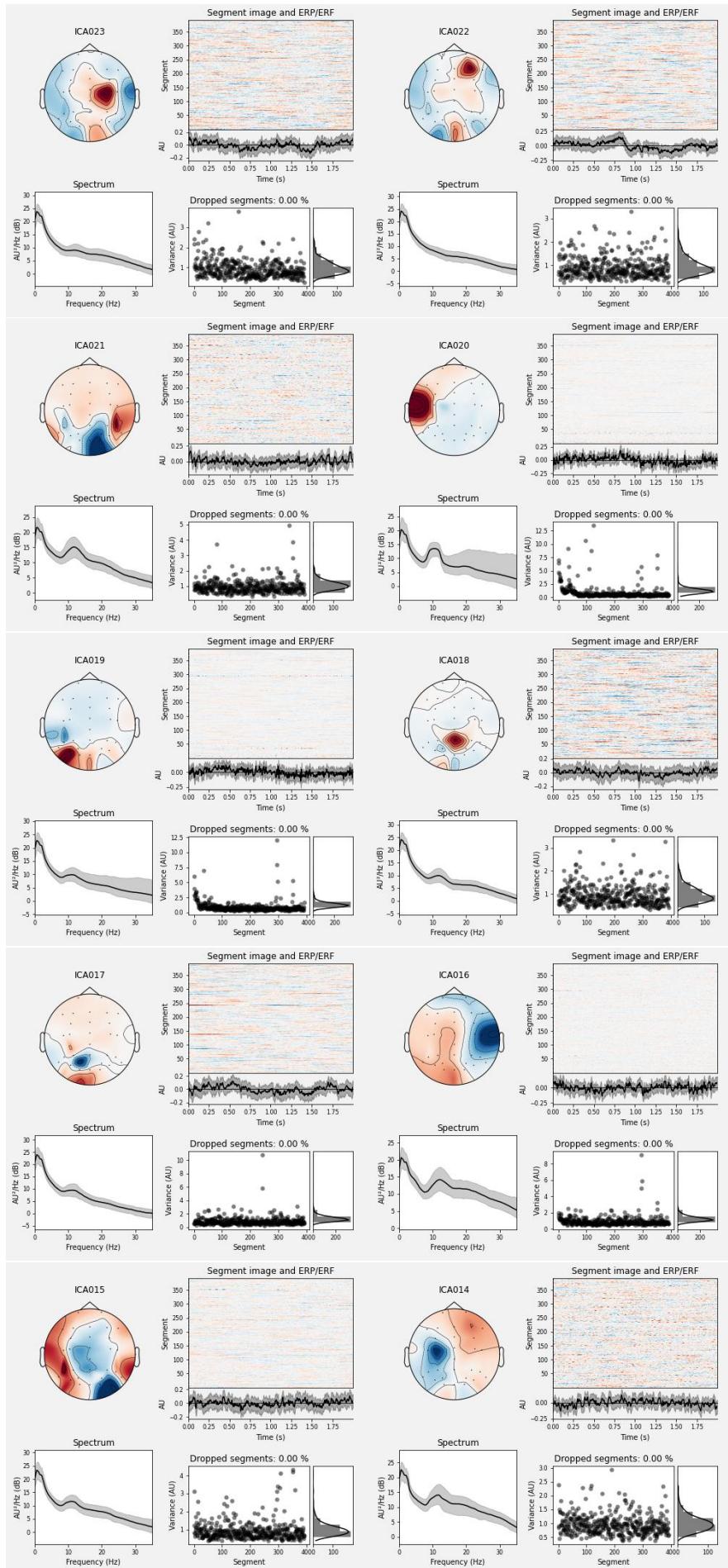


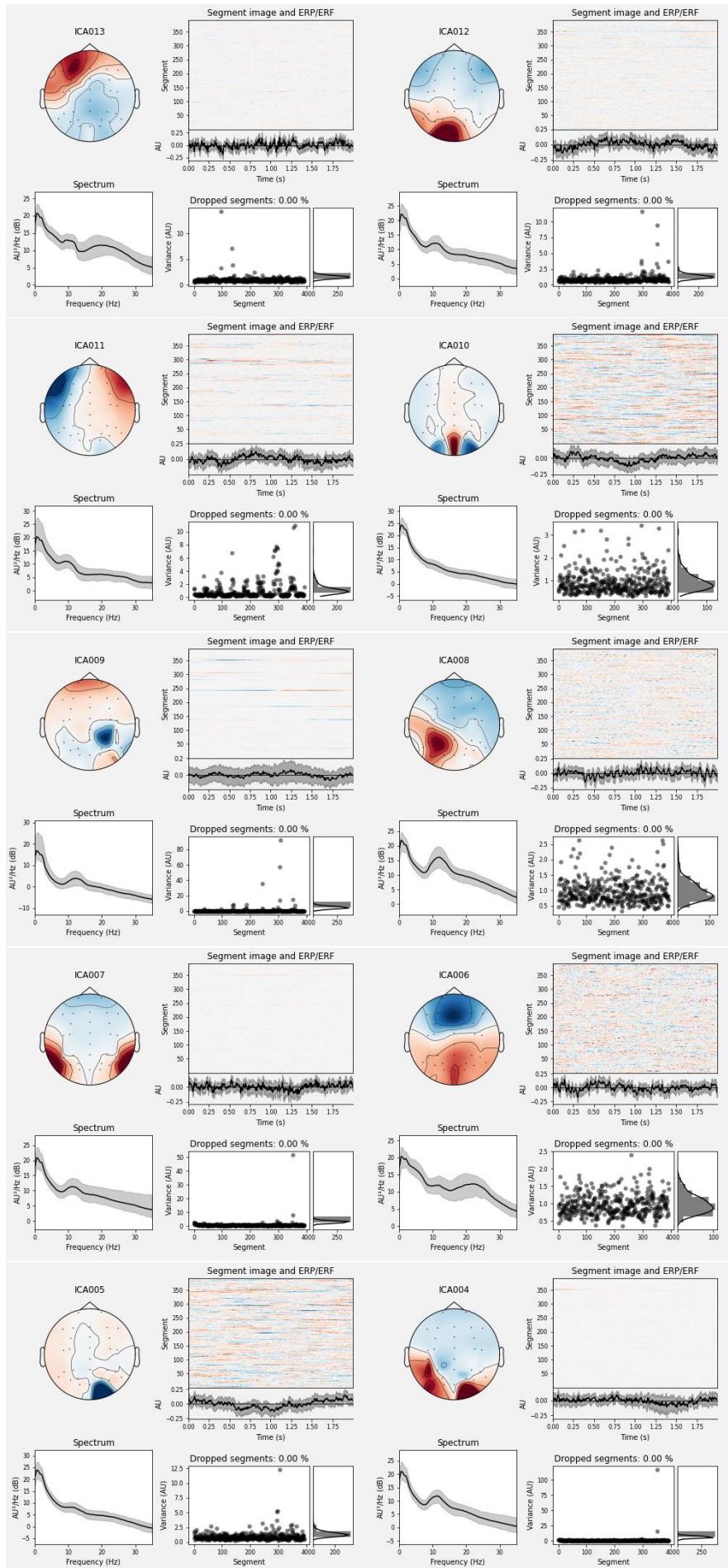


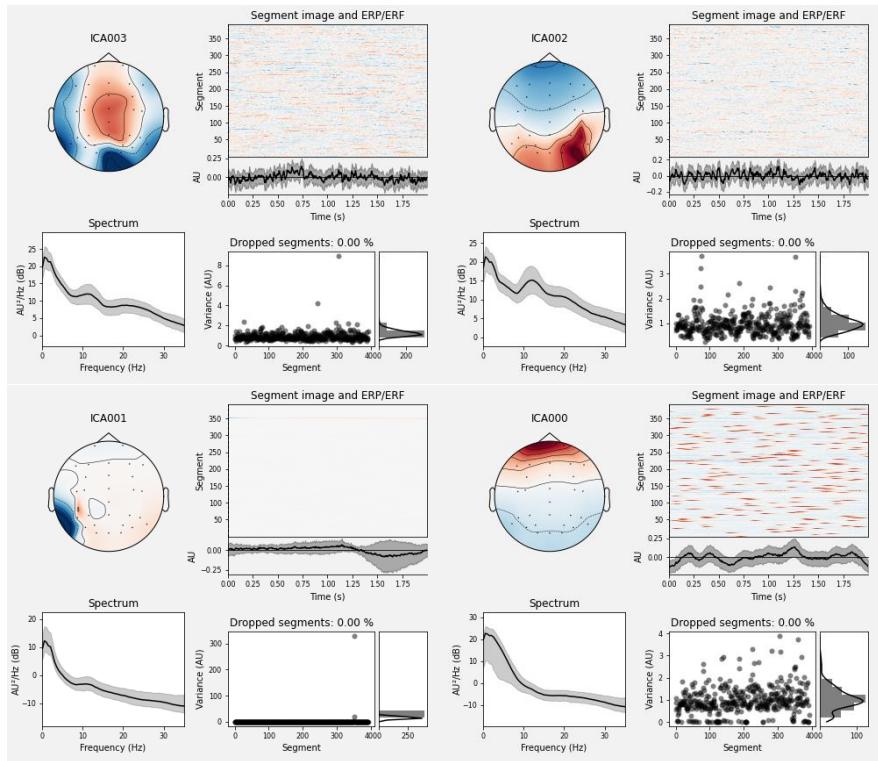


## Subject 8:

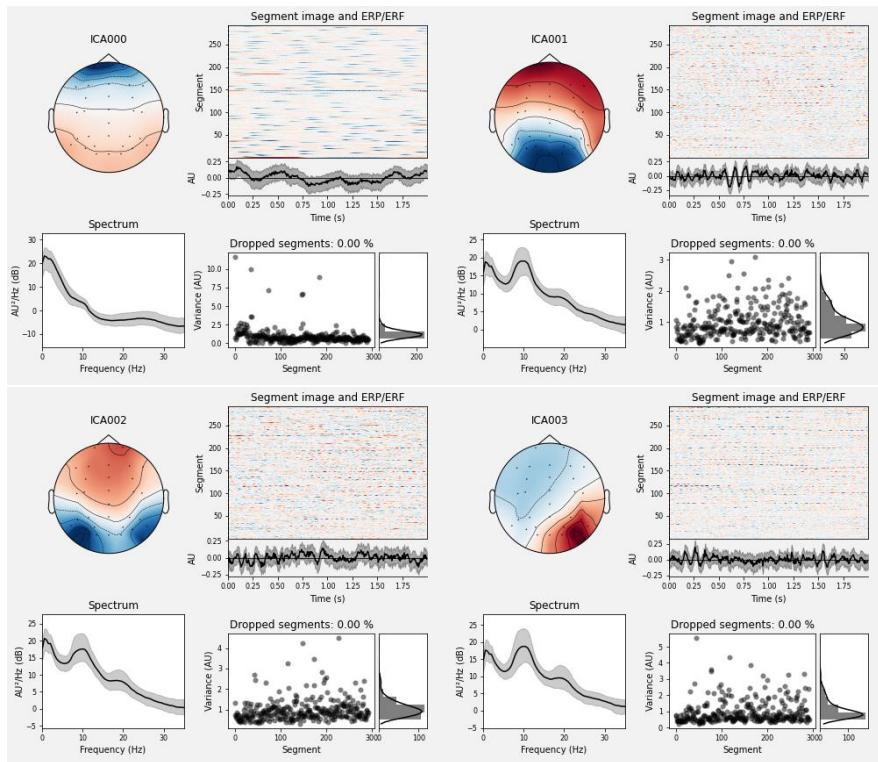


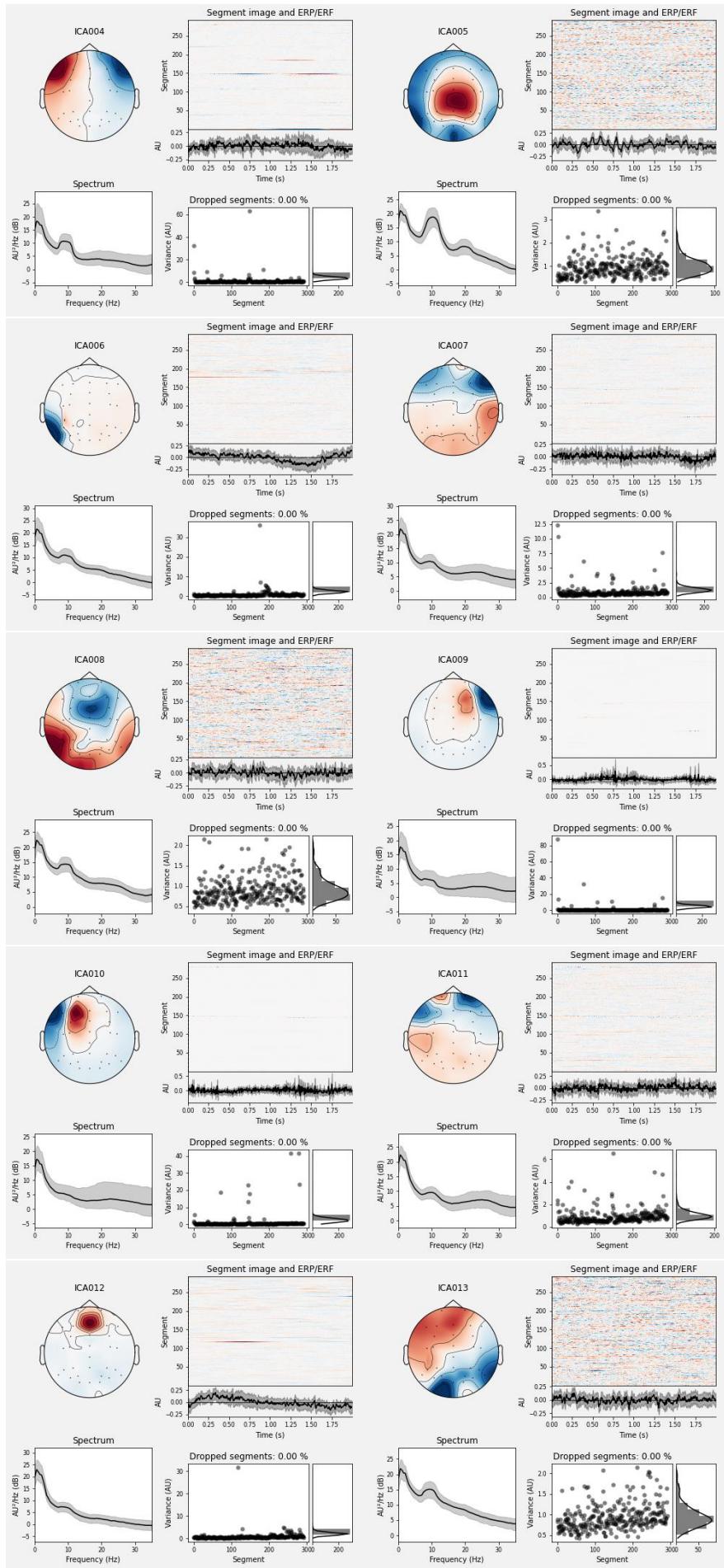


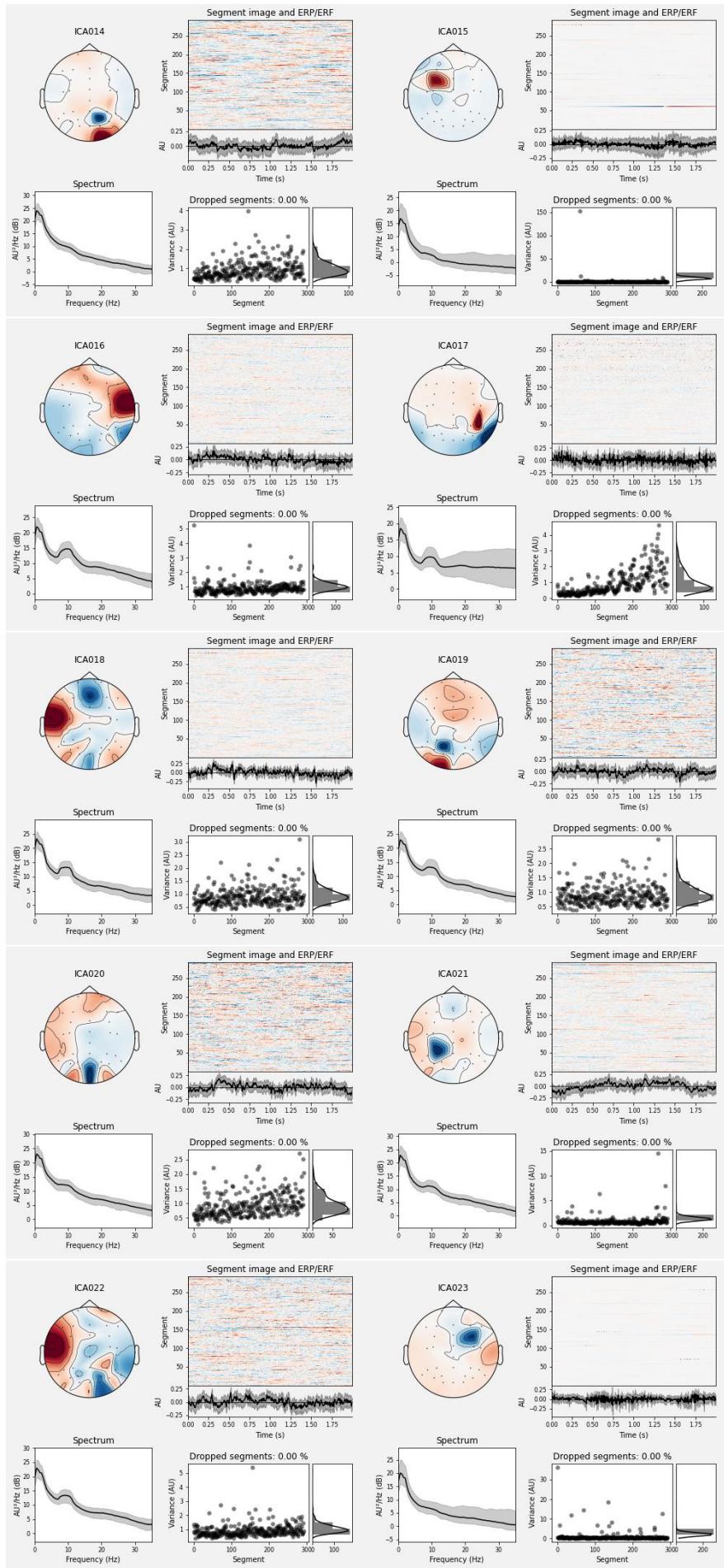


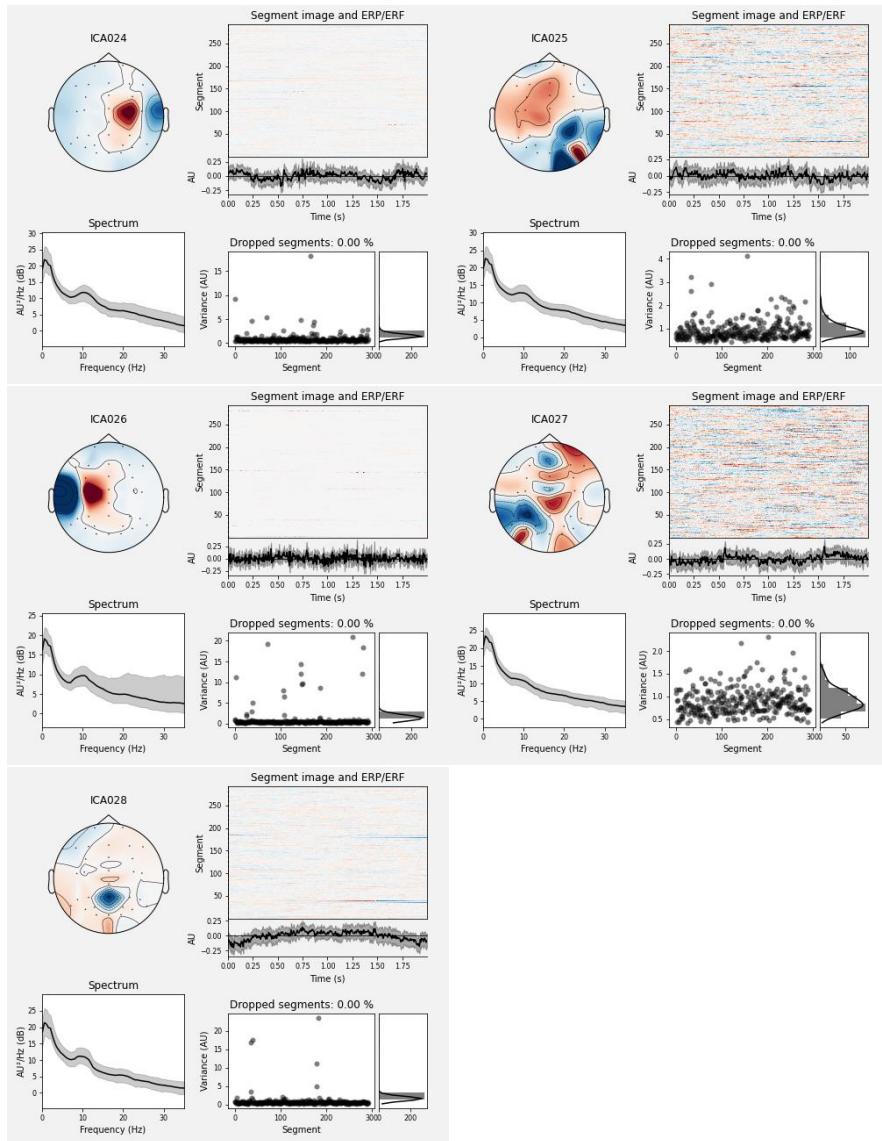


Subject 13:











References:

1. Pitcher, David & Walsh, Vincent & Duchaine, Brad. (2011). The role of the occipital face area in the cortical face perception network. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale.* 209. 481-93. 10.1007/s00221-011-2579-1.
2. Emily S. Kappenman, Jaclyn L. Farrens, Wendy Zhang, Andrew X. Stewart, Steven J. Luck, ERP CORE: An open resource for human event-related potential research, *NeuroImage*, Volume 225, 2021, 117465, ISSN 1053-8119, <https://doi.org/10.1016/j.neuroimage.2020.117465>.
3. Guillermo SAHONERO-ALVAREZ, Humberto CALDERON. A Comparison of SOBI, FastICA, JADE and Infomax Algorithms. Laboratorio de Ingeniería en Computación, Universidad Católica Boliviana “San Pablo”