

Research Placement Report:  
Developing Pre-processing pipeline for Resting State MEG data

Supervisor: Prof. Roni Tibon

PSGY4011- Research Placement

Anwesha Chakraborty

4<sup>th</sup> May 2023

## Abstract

This article provides an overview of the use of multi-layer network analysis and resting-state magnetoencephalography (MEG) to research functional brain connectivity, with a focus on aging-related changes. Recently, the approach of multi-layer network analysis has been used to explore the age-related changes in resting-state MEG data, and it has been demonstrated that a multi-modal approach can successfully capture the complete impact of ageing on brain connection. The main goal of this research is to improve our understanding of age-related changes in brain connectivity by leveraging the NIMH open-source database to apply the multi-layer network technique to both fMRI and MEG data. The placement project contributes to the development of the resting-state MEG data pre-processing pipeline, which includes several processes such artifact detection and correction, source localisation, and spectral analysis. The successful implementation of the pre-processing pipeline will enable the researchers to conduct a rigorous analysis of the MEG data and facilitate the exploration of age-related changes in brain connectivity using the multi-layer network approach. The results of this study could provide insights into the relationship between age-related changes in brain connectivity and cognitive decline. Lastly, the report gives a thorough explanation of the pre-processing pipeline, followed by a paragraph of reflection on my experience throughout the placement project.

Keywords: MEG, pre-processing, multi-layer network, artifacts, ICA.

## Introduction

Resting-state magnetoencephalography (MEG) is a non-invasive technique used to study functional brain connectivity. It involves recording magnetic fields produced by the brain while a person is not performing a specific task, i.e., they are at rest state. This method can help reveal how different regions of the brain are functionally connected to each other, allowing researchers to study spontaneous neural activity and connectivity patterns. Some characteristics of the MEG make it a highly attractive alternative for connectivity analyses in the human brain. These include high temporal resolution, high spatial resolution and lastly, direct measurement of neural activity (Brookes et al., 2016). In general, the resting state MEG data is analysed through the following steps: pre-processing, source reconstruction, functional connectivity analyses and multi-layer network analyses. The choice of the analysis method depends entirely on the nature of data and the research question being asked.

In this project, we will utilize the multi-layer network analysis method to study the functional connectivity between different frequency bands in the brain (alpha, beta, gamma). This method is based on the idea that different frequency bands reflect different types of neural activity and functional connectivity. By analyzing these frequency bands separately, we can gain insights into the organization and dynamics of the brain's functional networks. In this approach, the brain is modelled as a network of nodes (brain regions) connected by edges (functional connectivity between regions). After dividing the data into frequency bands, a network is constructed using a connectivity metric (eg. Coherence or phase locking value) to measure the strength of connectivity between pairs of nodes. Once the networks have been constructed for each frequency band, they can be analysed separately or combined into a single multi-layer network (Rubinov & Sporns, 2010). In simpler terms, the multi-layer network can be described as a 'network of networks' consisting of individual network layers all of which are interconnected (Madke et al., 2018). Such a framework enables integration of information across different networks. Unlike the traditional connectivity analysis methods, the multi-layer network approach is more comprehensive and considers the dynamic nature of brain connectivity, allowing for the identification of multiple patterns of connectivity that can change over time (Domenico, 2017). This shows that the multi-layer network approach is a powerful tool for studying the functional connectivity of the brain using resting-state MEG data. It has the potential to advance our understanding of the interactions between different brain regions and their role in cognitive and behavioral processes.

Resting-state MEG and multi-layer network analysis have become popular tools for studying brain connectivity. Many studies have utilized this approach to investigate the effect of aging on large-scale brain connectivity. For instance, aging is seen to be associated with reduced connectivity within and between brain regions (Betzel et al., 2014), as well as changes in network topology (Cao et al, 2014; Geerligs et al., 2015). Some studies have also used this approach to look into the connectivity within and between frequencies among schizophrenic patients (Brookes et al., 2016). To further capture the complex patterns of brain connectivity that underlie aging-related changes, recent studies have employed a whole-brain, multi-modal approach. For example, Varangis et al. (2019) utilized a combination of MEG, EEG, and fMRI data to investigate the effect of aging on functional connectivity and found that the multi-layer network approach would aid in capturing the full effect of ageing on brain connectivity. This lays an emphasis on the importance of using a whole-brain,

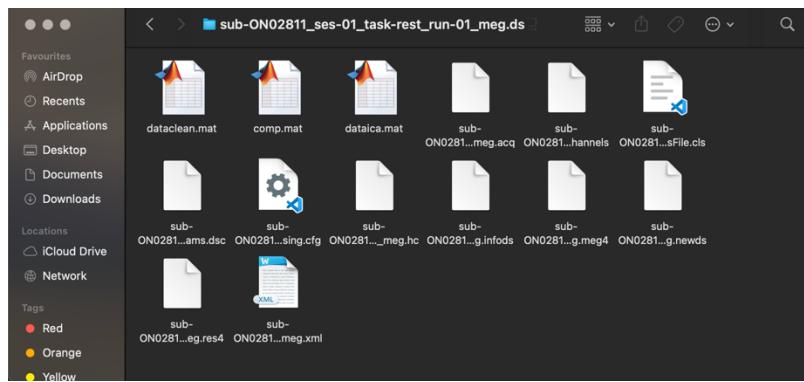
multi-technique approach in resting-state MEG research to better understand the complex patterns of brain connectivity in healthy as well as clinical populations.

The aim of the present study is to investigate the changes in resting state MEG data that occur with age. The study seeks to apply the multi-layer network approach to both MEG and fMRI resting state data in order to explore changes in brain connectivity across the lifespan. Data for the study will be obtained from the NIMH open-source database (Nugent et al., 2002). After the data undergoes pre-processing, connectivity analyses will be run with the help of toolboxes (Fieldtrip). The results of the study could help future research in exploring the relationship between age-related changes in brain connectivity and cognitive decline and may also reveal the extent to which other variables such as gender, lifestyle, and sleep affect these changes over time. The placement project plays a crucial role in achieving the long-term goal of this study. Specifically, the focus of the placement project is to develop a pre-processing pipeline for resting state MEG data, which is an essential step in the overall analysis. This pipeline will involve a range of procedures like artifact detection and correction, source localization, and spectral analysis, among others. The successful implementation of the pre-processing pipeline will enable the researchers to conduct a rigorous analysis of the MEG data, thereby facilitating the exploration of age-related changes in brain connectivity using the multi-layer network approach.

Pre-processing of neuroimaging data involves a series of procedures to prepare raw data for subsequent analysis. It is an essential step as it minimizes the influence of various errors that can arise during data acquisition and maximizes the quality of the final data (Newman, 2020). Failure to eliminate these sources of noise from the dataset makes it unlikely for researchers to arrive at valid interpretations. The pre-processing pipeline encompasses of the script that can be applied to the dataset. In the case of MEG, the most common artifacts are physiological in nature. They might include eye movements, blinks, heartbeat, breathing etc. These disturbances are also known as artifacts. The following section will outline the pre-processing steps taken for this project.

### **Placement Project- Methodology**

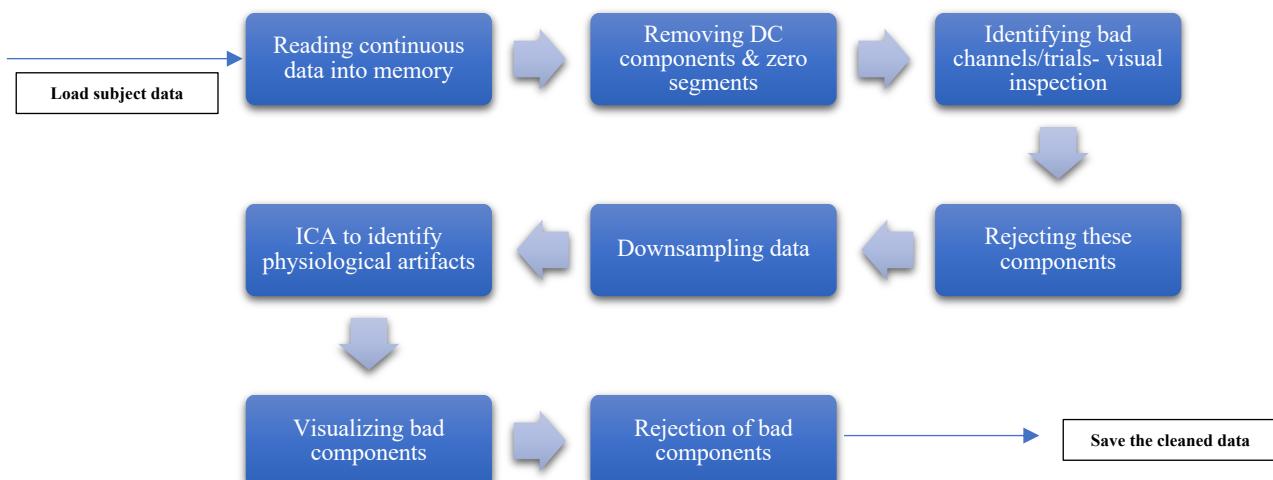
Before starting with the pre-processing pipeline, the first step was to download the data from the NIMH open-source database. The subjects were chosen such that both resting state MEG and fMRI data were available. The data was then subsequently organized into separate directories for each subject. This is an important step as it lays the foundation for subsequent analysis and makes it accessible for MATLAB to locate the files quickly. By organizing the data systematically and consistently, the pre-processing can be performed more accurately and executed with efficiency.



*Image 1: This is an example of the data organization for sub-ON02811. The .mat files are also saved in the respective subject folders after running it through the pre-processing pipeline. Doing so, makes it easier to locate the cleaned data for subsequent analyses.*

In this project, a subject pool consisting of eight datasets was selected from the NIMH: Intramural Healthy Volunteer open-source database, containing both fMRI and resting-state MEG data. The database is composed of healthy research participants curated by the National Institute of Mental Health (Nugent et al., 2002). The availability of both fMRI and MEG data served as the basis for the subject pool's selection criteria. The sample was divided equally between older (49-65 years) and younger (21-30 years) populations to investigate the effect of age on functional brain connectivity. Even though the sample size is small, it was deemed to be appropriate for the early phase of this study. The detailed list of the subject pool is attached in the appendix.

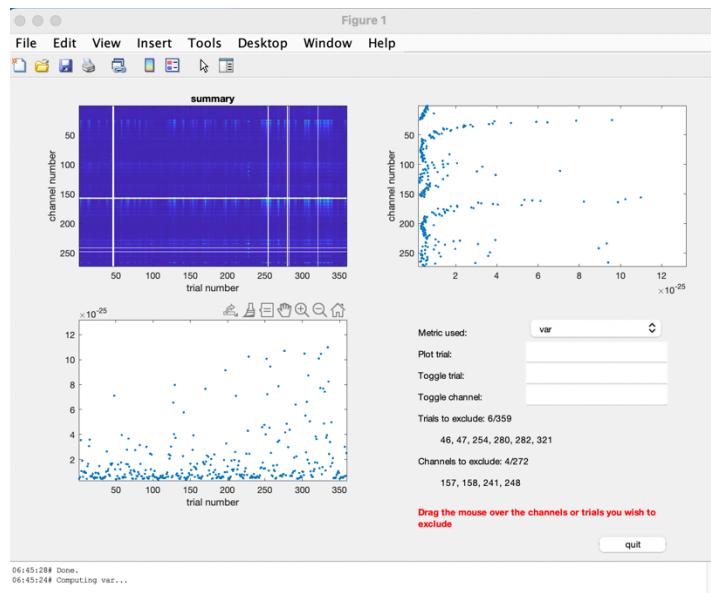
During the initial meetings with my supervisor Prof. Roni Tibon, we looked at the pre-processing tutorial on the FieldTrip website. FieldTrip is a MATLAB software toolbox for MEG and EEG analysis developed by the members of the Donders Institute of Brain, Cognition and Behaviour at Nijmegen, Netherlands (Oostenveld et al., 2011). For practice, I used sample EEG and MEG data sets to gain familiarity with the script and its functionality. Understanding this would eventually help me modify the script and make changes as per the needs of the project. Following two weeks of background reading and acclimation to the script, I was able to gain a rudimentary understanding of the steps involved in pre-processing of MEG data. Using these tutorials as a guide, the pre-processing pipeline for the project was developed. The pipeline comprised of reading the continuous data, rejecting the bad trials and channels based on visual inspection, using ICA to identify artifacts and finally visualizing and rejecting the bad components.



*Figure 1: Overview of components in the pre-processing pipeline*

First the location of the dataset as well as the toolbox (Fieldtrip) is entered into the MATLAB editor. Using *ft\_preprocessing*, the continuous MEG data is loaded. For MATLAB to identify the relevant dataset, the researcher would have to enter the location of the subject specific folder in *cfg.dataset = ''*. The function *ft\_redfintrial* segments the data into epochs. In this pipeline, the epoch is defined as 2 seconds length, similar to the tutorials. An epoch refers to a segment of data that is time-locked to a particular event or stimulus. The present project considers resting state MEG data- which is a continuous recording of brain activity in the absence of a specific task or external stimulation and segmenting into epochs may not be necessary. However, this enables the software to segment the data into numerous trials instead of considering it as a single epoch. Doing so makes the analysis of the data more accurate. The DC components and zero segments are then removed from the data sample. DC components are a part of the signal that represents its average or constant value (offset). Removing DC components are necessary to eliminate any baseline shift or drift, which can interfere with subsequent analysis. On the other hand, zero segment is the part of the signal whose amplitude is zero or within a certain threshold. These segments should be rejected as they are classified as ‘noise’ and do not add any value to the data.

After removing the DC components and zero segments, the pre-processed data is visually inspected to identify bad trials and channels. Data segments or sensors that have been tampered with by artifacts, noise, or other kinds of interference are referred to as "bad trials" or "bad channels" and are therefore useless. Their removal is achieved by plotting a scatter plot of the channels and trials, enabling easy identification of outliers. The bad trials or channels are then excluded from the dataset. To aid in the visualization of the location of different nodes in the measurement tool, the function *cfg.layout* is used to provide MATLAB with information about the version of the MEG tool used, which in this case is ‘CTF275.lay’. This information is crucial in the accurate representation of the spatial distribution of the nodes in the measurement tool. The *cfg.method* command is used to switch between three modes - summary, trial, and channel as illustrated below. All of these modes serve the same purpose, i.e., identifying bad trials or channels. After identification, the identified components can be selected for exclusion from the dataset. With the help of *ft\_rejectvisual*, ‘Dataclean’ then stores the revised number of trials and channels post the visual inspection. This is used in the following steps of the pre-processing pipeline. It is also important to keep track of all the bad trials and channels that have been excluded from the data. This ensures transparency and reproducibility of the analysis. Additionally, it serves as a record of the decisions made during the analysis, which can help identify potential sources of error. For this project, I have noted the bad trials and channels on Visual Studio Code (VSCode)- a snippet of which is attached in the appendix.



*Image 2:* This data is from subON88614 (older population) and it shows when cfg.method='summary'. The bad trials and channels can be visually identified by their outlier position in the scatter plots. On selection, they are excluded from the data set.



*Image 3:* These are the alternative view modes for subON88614 (older population) as well. On the left is cfg.method='trial'. While the right one is cfg.method ='channel'. These give a good overview to quickly spot any bad trials/ channels. On identification, they have to be clicked on to exclude them in from dataset.

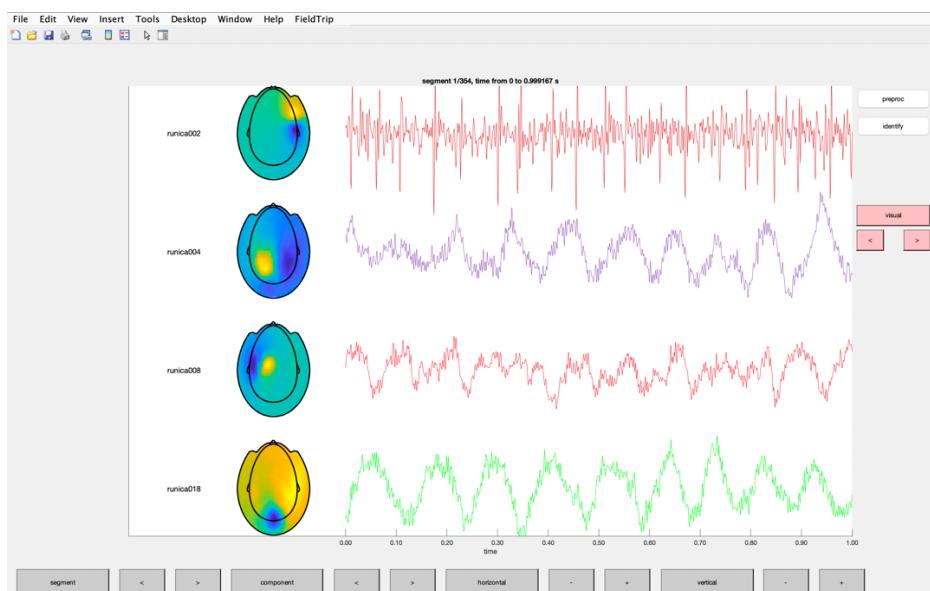
Next, the data undergoes the process of downsampling. This is done before independent component analysis (ICA) with the aim of reducing the computational cost. To ensure accurate measurement of the signals, MEG data is often obtained at extremely high sampling rates. As a result, it produces extremely vast datasets that are challenging to process. Downampling this data reduces the size of the data set making it easier for ICA or further analyses. This is carried out with the help of *ft\_resampled* and is stored as ‘datads’. I found the original sampling rate for all the datasets to be 1200 Hz. After the downampling process, it is reduced to 100 Hz.

Once the data is downsampled, the ICA can be performed using the *ft\_componentanalysis* function. ICA is a powerful technique used to separate independent components from multi-channel MEG data. Since the MEG data is a mixture of signals from different neural sources, and ICA is utilised to distinguish these sources from each other. The above function is able to estimate the unmixing weights (coefficients that are used to linearly combine the measured MEG signals to obtain the independent components) along with the component time course of the separate components. The underlying brain activity that produced the MEG signals is reflected in these time courses. After the ICA is completed, the independent components are inspected visually with the aim of rejecting any artifacts or noise components. This is possible through the

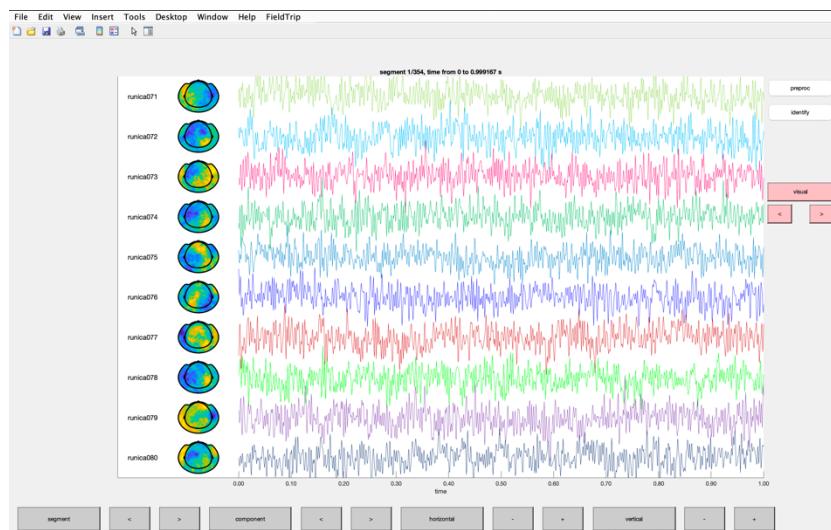
function `ft_databrowser`. As discussed earlier, the main sources of artifacts in MEG data are physiological in nature- eye movements, blinks, heartbeat, breathing etc. On running these set of commands, we get the entire list of components along with their location and time course. After the identification of bad components, they can be written as an array of numbers defining ‘badcomp’. Once ‘badcomp’ is defined, running these commands will display only these components. This can be used to confirm if the identified components are infact, artifacts and should be rejected. This rejection of bad components is done using the function `ft_rejectcomponent`. The data obtained after this step is deemed to be clean and should be saved in the respective subject directory (as shown in *image 1*). There are yet more steps to be added to the pre-processing pipeline and revisions may be required however, this concludes the initial stage.



*Image 4: This is the display when `ft_databrowser` is done for subON03748 (younger population) and no badcomps have been defined. We can scroll through the segments and components to identify bad components.*



*Image 5: From the initial display, I identified components 2,4,8 and 18 as artifacts. I then defined these as ‘badcomp’ and ran `ft_databrowser` again to confirm. This view shows that component 2 is the heartbeat owing to the periodic spikes while the others represent eye movements or blinks. These components are then removed using `ft_rejectcomponent`.*



*Image 6: This is also from subON03748 and represents normal MEG activity is displayed. The localisation of each component can also be seen in the topograph on the left.*

## Reflection

During my placement project, I had the opportunity to develop a pre-processing pipeline and apply it to the resting state MEG data, obtained from the NIMH open-source database. This was a challenging task as I needed to have a deeper understanding of the underlying principles of MEG data acquisition and processing, as well as proficiency in MATLAB coding and data analysis. Nevertheless, this experience provided a glimpse into the importance of attention to detail in pre-processing pipelines, the significance of collaboration and communication in research & the challenges of conducting neuroimaging research. It also helped me appreciate the practical applications of the MATLAB fundamentals I had learnt earlier in the course.

One of the key insights I gained was the importance of attention to detail in developing a pre-processing pipeline. Given the highly complex nature of MEG data, even minor inaccuracies or inconsistencies in the pre-processing pipeline can lead to significant errors in subsequent analyses. As such, I needed to be highly vigilant in ensuring that each step of the pipeline was running accurately, and artifacts were identified and rejected during visual inspection of the data. This was solidifying the reliability of the results. Another valuable lesson I learned was the crucial role of collaboration and communication in research. In developing the pre-processing pipeline and its application, I worked closely with my supervisor, regularly discussing my progress, seeking feedback, and sharing ideas. This collaborative approach was relatively new to me, but it was instrumental in ensuring the accuracy and efficacy of the pipeline. Additionally, it facilitated a greater understanding of the underlying research goals and methodology and allowed me to gain a deeper interest and appreciation for the broader context of the project. Throughout the duration of the placement project, I applied the pre-processing pipeline on several subjects and had to verify the accurate rejection of artifacts. I also spent time fixing bugs and errors in the MATLAB code, which improved my troubleshooting skills. This process allowed me to hone my time management skills and problem-solving skills, which are essential in research.

Overall, I found the experience of developing a pre-processing pipeline for resting state MEG data to be highly rewarding and enlightening. It provided me with a deeper understanding of the intricacies of MEG data processing, as well as a greater appreciation for the importance of attention to detail and collaboration in

research. Moving forward, I look forward to continuing to apply these insights in my research endeavours, and to further developing my skills and expertise in the field of neuroimaging.

## References

- Betzel RF, Byrge L, He Y, Goñi J, Zuo XN, Sporns O. Changes in structural and functional connectivity among resting-state networks across the human lifespan. *Neuroimage*. 2014 Nov 15;102 Pt 2:345-57. doi: 10.1016/j.neuroimage.2014.07.067. Epub 2014 Aug 7. PMID: 25109530.
- Brookes, M. J., Tewarie, P., Hunt, B., Robson, S. E., Gascoyne, L. E., Liddle, E. B., Liddle, P. F., & Morris, P. S. (2016). A multi-layer network approach to MEG connectivity analysis. *NeuroImage*, 132, 425–438. <https://doi.org/10.1016/j.neuroimage.2016.02.045>
- Varangis, E., Habeck, C. G., Razlighi, Q. R., & Stern, Y. (2019). *The Effect of Aging on Resting State Connectivity of Predefined Networks in the Brain*. *Frontiers in Aging Neuroscience*, 11. doi:10.3389/fnagi.2019.00234
- Geerligs L, Saliasi E, Maurits NM, Renken RJ, Lorist MM. Brain mechanisms underlying the effects of aging on different aspects of selective attention. *Neuroimage*. 2014 May 1;91:52-62. doi: 10.1016/j.neuroimage.2014.01.029. Epub 2014 Jan 25. PMID: 24473095.
- Nugent, A. C., Thomas, A. G., Mahoney, M., Gibbons, A., Smith J., Charles, A., Shaw, J. S., Stout, J. D., Namyst, A. M., Basavaraj, A., Earl, E., Riddle, T., Snow, J., Japee, S., Pavletic, A., Sinclair, S., Roopchansingh, V., Bandettini, P. A. & Chung J (2022). *The National Institute of Mental Health (NIMH) Intramural Healthy Volunteer Dataset*. OpenNeuro. Retrieved May 4 2023, from <https://openneuro.org/datasets/ds004215/versions/1.0.0>
- Rubinov M, Sporns O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*;52(3):1059-69. doi: 10.1016/j.neuroimage.2009.10.003.
- Madke, K., Meier, J., Brookes, M. J., O'Dea, R. D., Mieghem, P. V., Stan, C. J., Hillebrand, A & Tewarie, P (2018). Comparing multilayer brain networks between groups: Introducing graph metrics and recommendations. *NeuroImage*, 166, 371–384. doi:10.1016/j.neuroimage.2017.
- De Domenico M. Multilayer modeling and analysis of human brain networks. *Gigascience*. 2017 May 1;6(5):1-8. doi: 10.1093/gigascience/gix004. PMID: 28327916; PMCID: PMC5437946.
- Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M. & He, Y. (2014). Topological organization of the human brain functional connectome across the lifespan. *Developmental Cognitive Neuroscience*, 7, 76–93. doi:10.1016/j.dcn.2013.11.004
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, M-J. (2011). [FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data](#). Computational Intelligence and Neuroscience, 2011:156869.
- Newman, A.J (2020) Artifacts in EEG Data: *Data Science for Psychology and Neuroscience — in Python*. [https://neuraldatascience.io/7-eeg/erp\\_artifacts.html](https://neuraldatascience.io/7-eeg/erp_artifacts.html)

## Appendix

### 1. Study Sample

1	sub-ON02747	Female	54
2	sub-ON02811	Female	49
3	sub-ON62003	Male	64
4	sub-ON88614	Female	49
5	sub-ON03748	Male	22
6	sub-ON05311	Female	29
7	sub-ON08792	Female	23
8	sub-ON52083	female	21

Table 1: List of subjects whose data underwent the pre-processing pipeline for this project. Each subject has a unique ID. Their gender and age are also provided.

### 2. Snippet of VSCode-

```

40
41      NIMH subON88614
42 353 trial; 272 channels
43 new trial no.= 347; new channel no.= 268
44
45 --Trials to exclude (6): 46, 47, 254, 280, 282, 321
46 -- Channels to exclude (4): 157, 158, 241, 248
47 | MRF12, MRF13, MRT31, MRT41
48
49 badcomp= [2 4 8 9 11]
50
51      NIMH subON03748
52 353 trial; 272 channels
53 new trial no.= 347; new channel no.= 268
54
55 --Trials to exclude (6): 115, 116, 117, 187, 197, 198
56 --Channels to exclude (4): MLT31, MLT41, MRT31, MRT41
57
58 badcomp= [2 4 8 18]
59

```

Image 7: This is a snippet from my VSCode for subON88614 (older population) and subON03748 (younger population). Here, I kept track of the bad trials and channels that were excluded as well as the badcomp that were identified after ICA.