EEG Pre-Processing Using Independent Component Analysis

Arda Ertanhan
ICT for Internet and Multimedia, 2071519

***Abstract*—This paper uses Independent Component Analysis (ICA), to identify and remove artifacts of EEG signals and proceeds to extract some relevant information. Initially, the data was loaded into a Python environment. Then, the necessary electrode locations were processed in accordance with the dataset instructions given to us. In the next stage, artifact removal was performed and various brain waves were identified and various features of time-domain and frequency-domain were determined. These identified features were interpreted. At the same time, these features were introduced for various uses, such as neurophysiology.**

INTRODUCTION

THE ELECTRIC signal produced by brain cells working together, or more precisely, the time course of extracellular field potentials produced by their synchronous action, is recorded by the electroencephalogram (EEG) [1]. Conventional EEG configurations record activity at many scalp locations. The international 10–20 system is typically used to set electrodes for recordings of about 21 channels, 10–10 for recordings of between 64 and 85 channels, or 10–5 for high-density caps of more than 300 channels [2]. It is well-known in the scientific community and is frequently used by physicians, neurologists, and researchers to track and assess brain activity as well as to diagnose and treat mental health conditions [3]. EEG preprocessing, which is essentially a series of signal processing operations that come before the major analyses of EEG data, typically entails any required digital signal processing operations to refine the raw EEG signals in order to preserve only the signals of brain activity for further analysis [4]. Because of its capacity to remove artifacts from the signal, Independent Component Analysis (ICA) is frequently utilized during the signal preprocessing phase of EEG analysis. It converts a collection of mixed signals into a collection of distinct components [5].

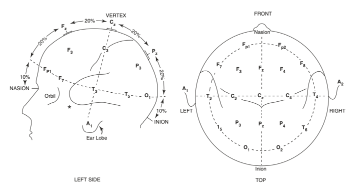

Figure 1: 10-20 system that used for electrode placement internationally [6].

PROCESSING PIPELINE

- In the project, a suitable dataset was first found on the Kaggle website. This dataset is actually a raw EEG dataset created by Garches Hospital Sleep Studies Lab (France, Paris area) and used for sleep spindles detection. This data set was then loaded into the Jupyter Notebook environment.
- After the data set was loaded, the electrodes that were the source of the EEG were identified and artifact removal was performed with ICA.
- After artifact removal, various features were identified and these features were extracted from the EEG signal.

DATASET DESCRIPTION

The dataset used in the project was taken from the Kaggle website. This dataset is actually based on EEG analysis in sleep pattern detection, and in this context, the concept of ”sleep spindles”. During non-REM (rapid eye movement) sleep, sleep spindles—transient electroencephalogram waves with a frequency of roughly 12–15 Hz—occur. It has been proposed that neuronal plasticity and memory consolidation are mediated by sleep spindles. Spindle-related spike trains that are experimentally activated have been shown to cause long-term potentiation, which is a fundamental neuronal mechanism involved in learning and plasticity. Human training on declarative and procedural learning tasks has been linked to increases in spindle density during the next sleep cycle, with the magnitude of these increases corresponding to the magnitude of the performance improvements that occur during the night [7]. There are two files on this Kaggle page; The first one is a raw EEG named ”extrait_wSleepPage01.csv”, which shows 30 minutes of sleep data, and the ”spindles.csv” file, which contains the expected pattern and allows us to detect sleep spindle based on this data set. Expected pattern file was not used in this project.

```
Sleep Page Data:
         Date  HH  MM      SS    EOG Left   EEG C3-A1     EEG O1-A1  \
0  08/03/2016   1  27      22     0,78125  5,56640625  -12,40234375
1  08/03/2016   1  27  22,005    1,953125  4,78515625  -11,23046875
2  08/03/2016   1  27   22,01   -1,67E-14    4,296875  -11,23046875
3  08/03/2016   1  27  22,015     0,78125  4,78515625  -12,01171875
4  08/03/2016   1  27   22,02  0,29296875   4,1015625  -13,18359375

     EEG C4-A1    EEG O2-A1
0    0,9765625  13,76953125
1  -0,09765625  17,08984375
2    0,9765625     17,96875
3   6,34765625   20,1171875
4   8,10546875   17,3828125
```

Figure 2: Display of several columns of the EEG data in the used dataset according to the electrodes, dates and times from which they were obtained.

ARTIFACT REMOVAL USING ICA

Data Upload

Before starting to process the data, some basic Python libraries such as ”pandas”, ”numpy”, ”matplotlib” and a python package such as MNE, which allows us to explore, visualize, and analyze MEG, EEG, sEEG, ECoG, NIRS human neurophysiological data, were implemented. Also by using this package we defined our ICA object. Then the data was loaded into the program. The time intervals during which the data were collected and the electrode names from which they were collected (’EEG C3-A1’, ’EEG O1-A1’, ’EEG C4-A1’, ’EEG O2-A1’) were defined within the project and prepared for use in signal analysis.

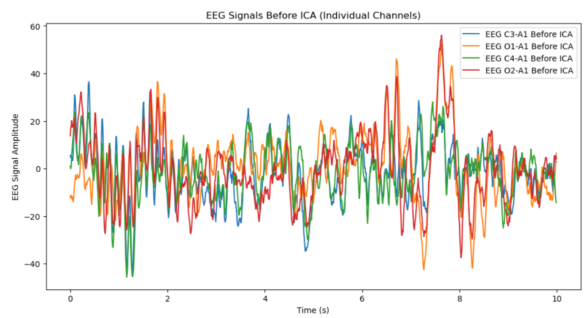

Figure 3: Amplitude-Time(s) graph of the EEG signals we received from the electrodes before ICA was applied.

Defining ICA Object

For artifact removal, we first defined our ICA variables. The variables used in the project are (All variables are implemented according to the MNE package):

1) n_components=4: Number of principal components that are passed to the ICA algorithm during fitting. Since there are 4 electrodes in our data set, we defined 4 components.
2) random_state=97: A seed for the NumPy random number generator (RNG). When this seed is run, most likely produce different output every time. It is necessary to achieve reproducible results.
3) max_iter=800: Maximum number of iterations during fit.

After defining the ICA variables, high-pass filter is applied to signal with a lower cutoff frequency defined as 1.0 Hz. It is frequently preferable to high-pass filter the data in EEG projects in order to eliminate linear trends. For high-quality ICA decompositions, it is also advised to high-pass filter the data at 1 Hz [8]. Following this, ICA was fitted to the data and components that were thought to be artifacts were removed. Typical EEG artifacts result from either external technical sources or the subject's non-neural physiological activity. Physiological artifact causes include eye blinks, eye movements, muscle activity around the head, heartbeat, and pulse. Technical artifact sources include swaying wires in the earth's magnetic field, line humming, power supply, or transformers [9]. While performing artifact removal in this project, we looked at our raw EEG data and the signals from the electrodes. In our amplitude-time graph, we took care to eliminate the parts that show electrical potential sharply increase and increase in high frequency activity/tense brief contractions through extended periods in stable ongoing EEG signals. The purpose of eliminating these was to eliminate possible eye blinks, eye movements and muscle contractions in the EEG signal [10]. As a result of these, it was concluded that the data in the EEG O1-A1 and EEG O2-A1 electrodes may contain possible artifacts and that these should be removed before feature extraction. Then, ICA was applied to the data, excluding the identified artifact components, and the cleaned data was prepared for feature extraction.

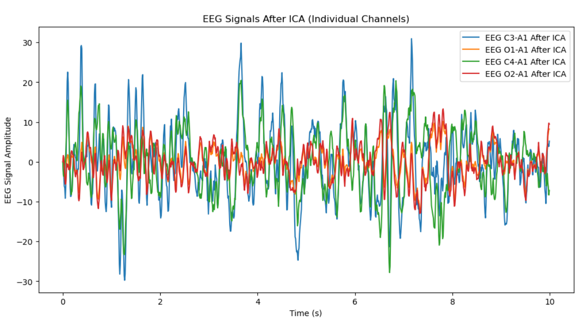

Figure 4: Amplitude-Time(s) graph of the EEG signals we received from the electrodes after ICA was applied.

FEATURE EXTRACTION

Since we would be working with specific time intervals, we decided on an epoch before we began feature extraction. We identified a few brain waves and their frequencies in parallel with the features we planned to record after establishing the epoch. Basically brain waves are oscillating electrical voltages in the brain. A scalp-position electrode EEG signal is composed of many waves, each with distinct properties [11]. There are different types of brain waves and each brain wave has a different meaning (as seen in the figure below). In this project, delta, alpha, beta and theta signals were used and features were determined depending on them.

Frequency band	Frequency	Brain states
Gamma (γ)	>35 Hz	Concentration
Beta (β)	12–35 Hz	Anxiety dominant, active, external attention, relaxed
Alpha (α)	8–12 Hz	Very relaxed, passive attention
Theta (θ)	4–8 Hz	Deeply relaxed, inward focused
Delta (δ)	0.5–4 Hz	Sleep

Figure 5: Features of Five Brain Waves [11].

Features

In this project, time-domain and frequency-domain features were obtained.

The mean power and standard deviation of the brain waves defined in the project are defined as time domain features, that is, statistical features. Mean power for each brain wave was found by averaging over epochs defined and filtered into specific frequency bands. Standard deviation was found using the "std" function from Python's "numpy" library.

The signal's power spectral density (PSD) serves as the foundation for determining the frequency domain properties. It can be calculated with several parametric and non-parametric methods. In this project, PSD was calculated with the Welch method, which is a non-parametric method[12].

A nonlinear technique for summarizing signal power irregularity across measured frequencies is spectral entropy, or SE [13]. It is computed in this project by evaluating the power ratios in particular frequency ranges . The project's spectral entropy formula is the negative sum of the power ratios inside each frequency band divided by the logarithm (base 2) of those ratios added together for all frequency bands.

RESULTS

Although EEG data interpretation is a job that would be done by a neurophysiologist or an EEG specialist, I can make a cursory and non-detailed analysis using the resources I have collected for the project and these as a reference by looking at our features' graphs. Our time-domain features show that the delta wave is more prevalent in the epochs, which makes sense given that the person or people whose data set we use is derived are asleep. The standard deviation shows that it decreases, climbs, and then decreases once more. From this, we can conclude that distinct signal features are seen in different epochs and that variability in the amplitudes of different brain waves increases at particular epoch intervals and subsequently diminishes. For spectral entropy, we can say that there are similar ups and downs with standard deviation, and for the graph, we can say that there is a more diverse distribution of frequencies among the signals over time at certain epoch intervals, and then it decreases and increases again. The PSD graphic shows how several frequency bands dynamically vary. Deep relaxation is suggested by delta's early dominance. Subsequent increases in theta may then be indicative of dream experiences, while alpha is an indication of relaxation. Beta peaks may indicate periods of hyper vigilance or activation.

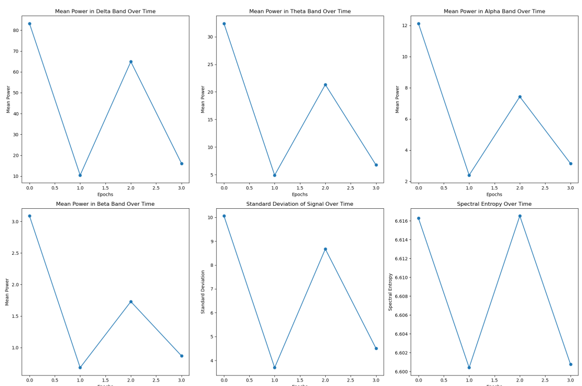

Figure 6: Time domain features and spectral entropy graph.

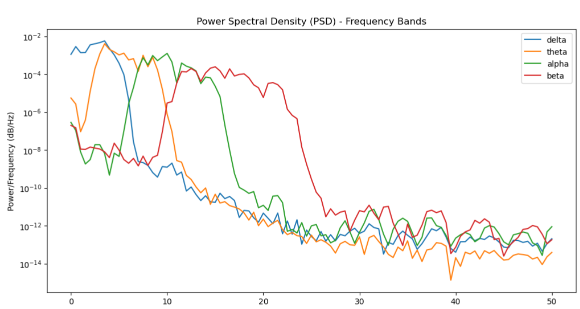

Figure 7: Time domain features and spectral entropy graph.

CONCLUSION

The main purpose of this project is to perform Independent Component Analysis using the MNE library and to monitor the performance of this method on a randomly selected data set suitable for the project and to examine the brain waves of a person or persons who are asleep and experiencing sleep spindles. As a result of the project, it was observed that ICA was able to successfully eliminate the channels containing artifacts, the brain waves I defined were extracted from the signal data from the electrodes presented in the data set, and the defined features could be obtained after artifact removal. The most important thing that limited me in this project was my doubts about the comments to be made after feature extraction and which channels should be interpreted. This problem could have been reduced to a more specific problem with a labeled machine learning set, or the data could have been interpreted better with the help of an expert. This project may help future researchers use ICA in EEG preprocessing, observing brain waves during sleep spindles, or using ICA in sleep spindles datasets. At the same time, this project can be improved by using different features and different preprocessing methods and comparing ICA with them.

REFERENCES

1. Blinowska, K. J., & Durka, P. (2006). Electroencephalography (EEG). Wiley Encyclopedia of Biomedical Engineering. https://doi.org/10.1002/9780471740360.ebs0418
2. Scrivener, C. L., & Reader, A. T. (2022). Variability of EEG electrode positions and their underlying brain regions: visualizing gel artifacts from a simultaneous EEG-fMRI dataset. Brain and Behavior, 12(2). https://doi.org/10.1002/brb3.2476
3. Zakeri, Z. (2016). Optimised Use of Independent Component Analysis for EEG Signal Processing [PhD Thesis]. The University of Birmingham.
4. Kim, SP. (2018). Preprocessing of EEG. In: Im, CH. (eds) Computational EEG Analysis. Biological and Medical Physics, Biomedical Engineering. Springer, Singapore. https://doi.org/10.1007/978-981-13-0908-3_2
5. Rejer, I., & Gorski, P. (2015). Benefits of ICA in the Case of a Few Channel EEG. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2015.7320110

6. Sazgar, M., Young, M.G. (2019). Overview of EEG, Electrode Placement, and Montages. In: Absolute Epilepsy and EEG Rotation Review. Springer, Cham. https://doi.org/10.1007/978-3-030-03511-2_5
7. Aeschbach, D., & Santhi, N. (2013). Attention and memory changes. In Elsevier eBooks (pp. 217–224). https://doi.org/10.1016/b978-0-12-378610-4.00043-7
8. Klug, M., & Gramann, K. (2020). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. European Journal of Neuroscience, 54(12), 8406–8420. https://doi.org/10.1111/ejn.14992
9. Winkler, I., Haufe, S. & Tangermann, M. Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. Behav Brain Funct 7, 30 (2011). https://doi.org/10.1186/1744-9081-7-30
10. Newman, A. J. (n.d.). Artifacts in EEG Data. Neural Data Science in Python. Retrieved January 20, 2024, from https://neuraldatascience.io/7-eeg/erp_artifacts.html
11. Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). Technological Basics of EEG Recording and Operation of Apparatus. Introduction to EEG- and Speech-Based Emotion Recognition, 19–50. doi:10.1016/b978-0-12-804490-2.00002-6
12. Stancin I, Cifrek M, Jovic A. A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems. Sensors (Basel). 2021 May 30;21(11):3786. doi: 10.3390/s21113786. PMID: 34070732; PMCID: PMC8198610.
13. Helakari H, Kananen J, Huotari N, Raitamaa L, Tuovinen T, Borchardt V, Rasila A, Raatikainen V, Starck T, Hautaniemi T, Myllylä T, Tervonen O, Rytky S, Keinänen T, Korhonen V, Kiviniemi V, Ansakorpi H. Spectral entropy indicates electrophysiological and hemodynamic changes in drug-resistant epilepsy - A multimodal MREG study. Neuroimage Clin. 2019;22:101763. doi: 10.1016/j.nicl.2019.101763. Epub 2019 Mar 12. PMID: 30927607; PMCID: PMC6444290.