**EEG Data Acquisition and Analysis**

**Abstract**

The electroencephalogram is a recording of the brain electrical activity from the scalp: the recorded time-series reflect the activity of the brain cortex. In the present work an EEG acquisition system is used to collect data from seven subjects with age ranging from 20 to 30 years. Three different scenarios were presented to each subject by means of a VR device. The collected EEG signals were filtered and processed in order to perform a Power Spectrum Analysis (PSA) and a Connectome Analysis.

**Experimental Setup**

The acquisition setup consists of an OpenBCI electrode-cap with 19 scalp electrodes connected to a Cython board featuring a signal isolation amplifier, an analog-to-digital converter and a wireless transmission module. For the purpose of the data acquisition we used the “referential montage” and only 16 electrodes to record the signals. Moreover, an Oculus Rift VR device was employed to provide three different visual stimuli to each subject. The EEG signals were collected by means of the OpenBCI software which also served as a platform to test the correct functionality of the employed electrodes via impedance test.

**Methods**

The electrode-cap was placed on the head of each subject and a high-conductive gel was applied between each electrode and the subject’s scalp, in order to increase electrical conductivity. Then, the VR device was set on the subject’s head. The correct functionality of each sensor was verified by means of an impedance test, which was repeated every time that the experimental conditions changed.

The first phase of the acquisition procedure consisted of a “rest phase”, in which the subject was asked to keep his eyes closed for a three minutes acquisition time span. The “rest phase” was performed to acquire a minimal brain activity reference.

In the second phase, a visual stimulus, consisting of a snowy landscape, was shown to the subject for seven minutes.

In the last phase, an urban environment was presented to the subject for the total duration of three minutes.

The entire procedure was repeated for each one of the seven subjects.

**Data Analysis**

The raw data were pre-processed and cleaned with the EEGLAB software, a Matlab plugin that was used to filter the time series and perform the Independent Component Analysis (ICA).

The snow samples were divided into two three-minutes time spans, removing the first and final 30 seconds: this procedure was performed to check if the signal was stable during the whole observation.

The filtering was performed applying a band-pass filter, selecting frequencies between , covering the typical brain activity frequency range. After this procedure, two criteria were employed to remove channels: (1) the channel was flat for more than 5 seconds; (2) the high-frequency noise standard deviation (noise to signal ratio) of the channel was greater than the standard value of 4. Then, the average re-referencing was computed, which consists in averaging all the channels signals and subtract it to each channel.

After filtering the signal, Independent Component Analysis was used to identify and reject artifacts.

In [signal processing](https://en.wikipedia.org/wiki/Signal_processing), ICA is a computational method for separating a [multivariate](https://en.wikipedia.org/wiki/Multivariate_statistics) signal into additive subcomponents. This is done by assuming that at most one subcomponent is Gaussian and that the subcomponents are [statistically independent](https://en.wikipedia.org/wiki/Statistical_independence) from each other. Therefore, the components flagged by EEGLAB as heart, muscle and eye blinking artifacts were removed from the signal if their probability to occur was greater or equal to 70%.

After the pre-processing, the main goal was to distinguish different scenarios seen by each subject using Power Spectrum Analysis (PSA) and Connectome Analysis.

The PSA of a [time series](https://en.wikipedia.org/wiki/Time_series) describes the distribution of [power](https://en.wikipedia.org/wiki/Power_(physics)) into frequency components composing that signal. According to [Fourier analysis](https://en.wikipedia.org/wiki/Fourier_analysis), any physical signal can be decomposed into a number of discrete frequencies, or a spectrum of frequencies over a continuous range. The statistical average of a certain signal or sort of signal (including [noise](https://en.wikipedia.org/wiki/Noise_(electronics))) as analyzed in terms of its frequency content, is called its [spectrum](https://en.wikipedia.org/wiki/Spectrum).

PSA is based on the Power Spectrum Densities (PSD) calculation which consists of the integration over time of the time-series autocorrelation function.

The PSA procedure was performed on clean EEG signal for each subject and scenario, thus obtaining the PSDs (see appendix A).

Thereafter, to have a quantitative analysis of the data a connectome analysis was performed.

A connectome is an adjacency matrix built computing Pearson’s correlation between each pair of signals. The absolute value of each matrix element was then calculated because the correlation sign is not significant for the purpose of this analysis. Moreover, the diagonal elements were put to zero to avoid self-correlation.

To prevent noise from affecting data, elements showing a correlation coefficient under the threshold of 0.20 were put to zero.

A weighted graph was built for each adjacency matrix, where each node represents an EEG channel and each link the correlation between every pair of nodes.

For each graph the different calculated metrics were: (1) *Betweenness* of a node v is the sum of the fraction of all-pairs shortest paths that pass through; (2) *Closeness* of a node v is the reciprocal of the average shortest path distance to v over all n-1 reachable nodes; (3) *Clustering* *Coefficient* of a node v in a graph quantifies how close its neighbors are to being a clique; (4) *Degree* of a node v is the number of nodes it is connected to; (5) *Edge Betweenness* is a measure describing the frequency at which an edge lies on the shortest path between pairs of nodes in a network; (6) *Eigenvector* centrality computes the centrality for a node v based on the centrality of its neighbors; (7) *Information* is a variant of closeness centrality based on effective resistance between nodes in a network.

For each resulting metric distribution and for each subject, a boxplot was produced comparing the various scenarios (see Appendix B).

In order to verify if there was some difference among the stimuli, the Kolmogorov-Smirnov statistical test was employed because the data samples had small dimension and their statistical variables were continuous.

The tests had a confidence level of 0.05, so the null hypothesis that the distributions were not statistically different was rejected if the p-value was under the confidence level. The statistical test was repeated between each possible pair of scenarios, fixing one subject and one graph metric at the time. Results are shown in tables reported in Appendix C, where 1 stands for statistical difference while 0 for no difference.

**RESULTS**

Due to a problem with the Cython board during the acquisition, subject 7 data were found to be in complete discordance with the behavior of the other ones; thus, they have been rejected.

The figures reported in Appendix A show also the brain activity heatmaps for the specific frequencies of 5.9, 9.8 and 22.5 Hz.

The rest state figures in Appendix A point out that six subjects show a similar behavior which results in a peak in the region of the alpha waves, in agreement with the expectations for this kind of stimulus.

In figures reporting the first part of the snow stimulus, it is possible to observe a flattening of the response in the alpha frequency range, whereas no relevant peaks activity has been observed in the beta frequency range for the majority of the subjects.

Figures concerning the second part of the snow stimulus display a similar trend compared to the previous one.

Also for the urban stimulus, no relevant peaks seem to be present in the beta waves region.

For the connectome analysis, from the boxplots reported in Appendix B, it was not possible to find a common pattern for the all subjects with different stimuli.

In tables reported in Appendix C, statistical tests for the various metrics are displayed for each subject. It can be inferred that the edge betweenness and the information metrics distinguish more effectively the various scenarios, but due to the small subject sample no significant conclusions can be drawn.