# Analysis and Interpretation of Empirical Data Obtained by BCI Epoc 14+

Georgi P. Dimitrov[a], Galina Panayotova[a], Vasyl Martsenyuk[b], Inna Dimitrova[a], Eugenia Kovatcheva[a], Boyan Jekov[a] and Iva Kostadinova[a]

[a] *University of Library Studies and Information Technologies, 119, Tsarigradsko Shose, Sofia, Bulgaria*
[b] *University of Bielsko-Biala, 2 Willowa, Bielsko-Biala, 43-309, Poland*

**Abstract**

Brain signals based on effective computing are a new development in the research area, aimed at finding a correlation between human emotions and registered EEG signals. The Brain-Computer interface (BCI) would allow the users to control and manage external devices by brain signals emittance. These signals can be received and recorded by multiple special devices like EMotiv Epoc +14, Neuroscan, EasyCap and etc., but the reliable translation of the information obtained into computer commands is still a great challenge. This requires exceptional integration between the information emitted by the brain of the signal user, the BCI system, which transfers the information into digital signals and the respective algorithm translating the brain signals into commands. The analysis of incoming brain signals and the techniques for processing and classification of information are being actively explored in order to improve adaptability of BCI system to the end-user

In the present study, we propose an approach to the selection of characteristics based on descriptive statistics. Data streams were studied in order to take into account the time characteristic, the analysis and derivation of dependencies on time data, characterized by a relatively long duration of the experiment and short series of significant, useful data. This approach represents a good trade-off between prediction accuracy and numerical complexity.

**Keywords**

Mathematical models of objects and processes, Computer Science, Artificial Intelligence, Brain Wave, Machine Learning, Deep Learning, Robotic

## 1. Introduction

The use of data obtained from BCI is a complex process that requires multidisciplinary skills and knowledge in the field of computer science, signal processing, neurology, robotics, artificial intelligence and others [14]. The study is based on a fixed sequence, which usually consists of six steps, showing in fig.1: [6], [10] measuring brain activity, pre-processing data, extracting characteristics, classification, command translation and feedback:

- Receive Data: At this stage, different types of sensors are used to obtain signals that reflect the brain activity of the user [2]. In this study, we focus on BCI as the technology for obtaining data.
- Preprocessing: This step involves cleaning and removing noise from the input data to improve the quality of the received signals. [1], [3]
- Extraction of features: It aims to describe signals by several corresponding values, called “features” [4], [7].
- Classification: The classification stage determines the class based on the extracted characteristics of the signal [1]. The class corresponds to the type of pre-identified signal. This stage can also be referred to as “characteristic translation [11], [12] .Classification algorithms are known as “classifiers”.

*Information Technology and Implementation (IT&I-2021), December 01–03, 2021, Kyiv, Ukraine*
EMAIL: geo.p.dimitrov@gmail.com (A. 1), panayotovag@gmail.com (A. 2), vmartsenyuk@ath.bielsko.pl (A. 3), innavadi@yahoo.com (A.4), ekovatcheva@gmail.com(A. 5), b.jekov@unibit.bg (A. 6),i.kostadinova@unibit.bg (A. 7)
ORCID: 0000-0002-4785-0702 (A. 1);

- Command / application translation: Once the received command is identified, it is submitted for execution by the respective device [10].
- Feedback: Finally, this step provides the user with feedback on the identified command. This helps control the quality of the received signals processing [8], [9].

The electroencephalogram (EEG) is an excellent source for obtaining data related to human brain activity [13]. A typical EEG experiment can produce data described with a two-dimensional matrix based on brain activity every millisecond, projected onto the surface of the head at a spatial resolution of a few centimeters [15]. The placement of the electrodes is based on several circuits, the most commonly used of which is the Standard 10-20 EEG system [15]. As in other modern empirical sciences, EEG tools provide on the one hand an abundant flow of data and on the other - a corresponding need for new methods of data analysis.

An important stage of data Preprocessing is the selection and handling of the obtained data.

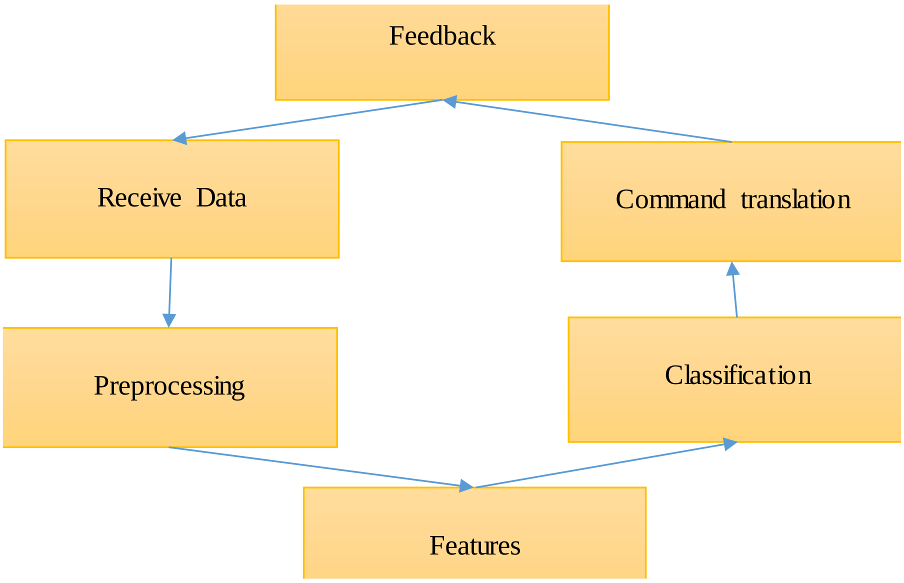


**Figure 1:** Steps for data analitics

## 2. Description of the research

### 2.1. Basic description

Our hypothesis is that data normalization simplifies the process of classification of brain signals considerably and leads to a significant simplification of computational procedures.

This study aims to simplify the incoming EEG signals preprocessing by normalization of obtained data, extracting certain characteristic values and subsequent signal classification . The level of signals related to specific events is registered by 14 channels of the EEG EMotiv Epoc 14+, while the subjects respond by giving mental commands to control the display of the corresponding command on the screen. 12 time characteristics (amplitudes and latencies) are calculated and used as descriptors of positive and negative emotional states in multiple subjects.

### 2.2. Collected data

The research includes analysis of raw data obtained from 21 physically and mentally healthy participants, without pre-existing neurological disorders and previous experience with using Brain-Computer Interface (BCI) devices [9]. The participants are in one age groups – 20 and 23 years. An Emotiv Epoc+ 14ch device is used for the purpose of the study. The device and the location of the electrodes is shown on fig. 2 The experiment was based on the display of static images (left , right arrows and Neutral state), where the participants in the experiment should mentally submit the appropriate command for the movement of a computer simulator - a motor boat. It is important to note that these are only mental commands, not movement of the arms or legs, which significantly complicates classification, since it is not related to limb activity. Additionally, the Neutral command is collection of all other commands, such as synchronization, relaxation etc.

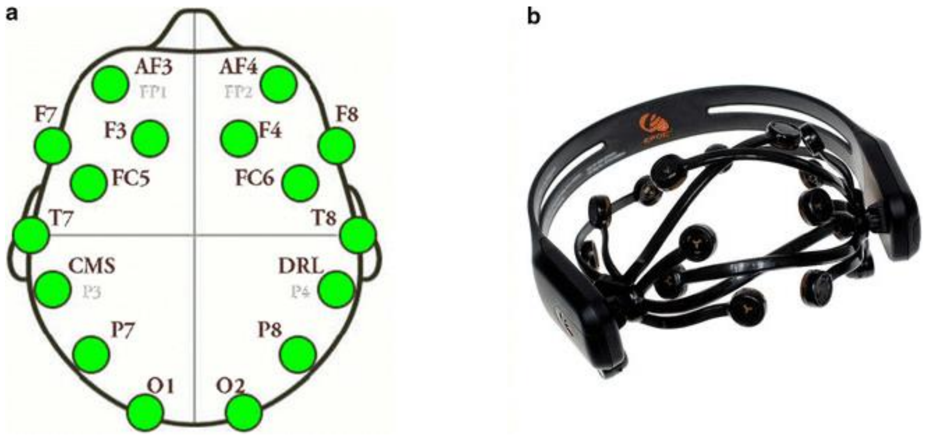


**Figure 2:** EMotiv Epoc 14+ and electrodes position schema

Each participant performed the experiment 3 times. Each experiment lasted 600 sec., or ~ 10 min. There were 30 min intervals between the different experiment in order to relax the participant. During the experiment, the respective images with written commands “left” and “right” were shown 20 times each. Each series consisted of a 3-second display of the respective image (epoch) and additional visual and audio signals. At the beginning of the series, a 1-second beep was sounded to alert the participant. Each test series lasted 15 seconds. This included 3 seconds to display the appropriate command and 12 seconds to perform synchronization actions, relaxing and etc. Because the experiment involved motor imagery, it was mainly focused on beta waves (12 - 30 Hz). That is, in each experiment we have 20 repetitions of Left and Right for 3 seconds (or a total of 60 seconds for each command separately. Commands received during the remaining time - 8 min (480 sec) are defined as Neutral command. Altogether, the duration of a given process (signal duration - epoch) is 3 sec. for Left and Right commands and 12 sec for Neutral. The average value for each condition is calculated and filtered. The maximum and minimum values of the ensemble of average signals are detected. The localization of the first minimum in the signals and the characteristics are determined by the latency and amplitude of successive minima (Amin1, ...) and successive maxima (Amax1, ...), and the associated latency (Lmin1, ..., Lmax1,. ..). Three circuits are implemented by selecting three different filters and detecting N maxima and N minima at the filter output. When this model is not implemented, the vector function is filled with zeros.

## 2.3. Processing and Norming Data

As a result, the initial data set was an X matrix with dimensions of 168 columns (14 channels x12 characteristics) and 52 rows (averaged positive and negative test classes of 26 subjects).

$$\bar{X} = \frac{X - mean(X)}{std(X)} \tag{1}$$

The vector space X is then normalized by subtracting the average value for each dimension and dividing the standard deviation of each column, see formula (1).

# 3. Result of the Experiment

## 3.1. Classification models

**Determining the set of characteristics by which the sample data will be evaluated.** The set of features is derived from the data stream registered for each EEG channel. The characteristics are determined on the basis of the first six local extremes - 3 minima and 3 maxima (Figure 3). The amplitudes of these initial extremes and the time of their occurrence (latency) are considered to be characteristics of the current data flow. Thus, each EEG channel is represented by 12 characteristics - the amplitudes and latency of the six extremes.

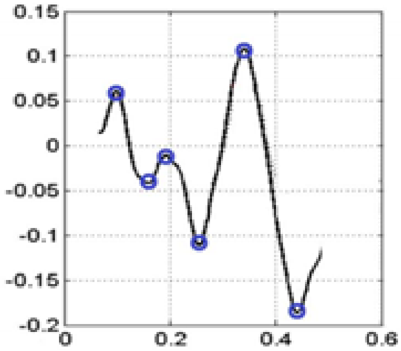


**Figure 3:** Amplitude Max and Min

By applying Butterworth fourth order filter with bandwidth [0.5 - 15] Hz, the number of preserved characteristics is 12, corresponding to latency (time of occurrence) and amplitude at N = 3 maxima and minima (see Figure 3); the characteristics correspond to the time and amplitude values of the first three minima that occurred after T = 0s., and the corresponding maxima between them.

When grouped by channels (Inter-subject), each object is represented by these 12 characteristics. [Amin1, Amax1, Amin2, Amax2, Amin3, Amax3, Lmin1, Lmax1, Lmin2, Lmax2, Lmin3, Lmax3]

## 3.2. Data analysis by characteristics, defined and extracted with descriptive statistics

We distinguish incoming commands basing on brain activity observed by electroencephalogram (EEG). The choice of features is important for signal classification. In the present study, we propose a selection technique based on descriptive statistics (mean and standard deviation) [22]. This approach represents a good compromise between the accuracy of prediction and numerical complexity. We propose to reduce obtained data volume by focusing on the central trend (arithmetic mean) and the variance (standard deviation) of the individual time characteristics and their distribution.

# 4. Real data application

## 4.1. Formation of databases by channels (Inter-subject)

For the purposes of this research, tree main commands were chosen, using antonymous words: LEFT, RIGHT and NEUTRAL. Each word is defined by a 14-dimensional vector of channels (x1, x2, ..., x14), where xj denotes the j- channel, of which we have made p observations. Thus, a matrix X of the type p × 14 is formed, the rows of which display the observations of the study. (Table 1)

**Table 1**
Row data

| AF3 | F7 | F3 | F5 | T7 | P7 | O1 | O2 | P8 | T8 | FC6 | F4 | F8 | AF4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 118,90 | 118,06 | 118,20 | 118,12 | 118,99 | 118,48 | 117,92 | 118,54 | 118,66 | 118,27 | 117,79 | 120,47 | 118,12 | 119,59 | H |
| … | | | | | | | | | | | | | | |

This database allows to reveal individual brain channel dependencies and conclude which of them are involved when a visual task of the described type is present.

## 4.2. Similarity measurement

Most statistical methods use correlation analysis to determine the similarity between different brain signals. The results are given in the form of correlation matrices. Table 2, Table 3 and Table 4 display the correlations between the individual channels of the selected words and their calculation results [5].

**Table 2**
Correlation Matrix of NEUTRAL

| AF3 | T7 | O1 | T8 | .... | AF4 |
|---|---|---|---|---|---|
| 1 | 0.1523 | 0.4581 | 0.5133 | | 0.6074 |
| | 1 | 0.4723 | 0.1661 | | 0.0908 |
| | | 1 | 0.3011 | | 0.3014 |
| | | | 1 | | 0.4674 |
| … | | | | | |
| | | | | | 1 |

Confidence level 95%. n = 8734.

**Table 3**
Correlation Matrix of Left

| AF3 | T7 | O1 | T8 | .... | AF4 |
|---|---|---|---|---|---|
| 1 | 0.1861 | 0.3060 | 0.6202 | | 0.6401 |
| | 1 | 0.2774 | 0.2355 | | 0.1884 |
| | | 1 | 0.2159 | | 0.2622 |
| | | | | | |
| | | | 1 | | 0.4902 |
| | | | | | 1 |

Confidence level 95%. n = 12480.

**Table 4**
Correlation Matrix of RIGHT

| AF3 | T7 | O1 | T8 | .... | AF4 |
|---|---|---|---|---|---|
| 1 | 0.1861 | 0.3060 | 0.6202 | | 0.6401 |
| | 1 | 0.2774 | 0.2355 | | 0.1884 |
| | | 1 | 0.2159 | | 0.2622 |
| | | | | | |
| | | | 1 | | 0.4902 |
| | | | | | 1 |

Confidence level 95%. n = 12870.

In this article we will use only channels with correlation > 0.5. Channels with correlation <0.5 are ignored. For all three commands we use channels AF3, T8 and AF4.

## 4.3. Data analysis by characteristics defined and derived from descriptive statistics

We calculate the mean and standard deviation of those channels that were selected according to clause 4.2. The results obtained for the three types of commands are given in Tables 5, 6 and 7.

**Table 5**
Statistical characteristics of Neutral

| chanel | mean | St.dev | max | min |
|---|---|---|---|---|
| AF3 | 8,61476139 | 593,0781423 | 884,5149578 | -1881,297142 |
| T8 | 8,561556105 | 588,5540203 | 879,4852859 | -1870,022561 |
| .. | | | | |
| AF4 | 8,844488861 | 596,8591494 | 890,2455 | -1893,2 |

**Table 6**
Statistical characteristics of Left

| chanel | mean | St.dev | max | min |
|---|---|---|---|---|
| AF3 | -0,6208525 | 10,55638 | 45,30853 | -92,1 |
| T8 | -0,63486515 | 11,15294 | 47,54477 | -89,764 |
| .. | | | | |
| AF4 | -0,6400505 | 13,02439 | 47,54477 | -92,3567 |

**Table 7**
Statistical characteristics of Right

| chanel | mean | St.dev | max | min |
|---|---|---|---|---|
| AF3 | 0,006265795 | 4,059517 | 18,61453 | -11,5468 |
| T8 | 0,033032954 | 7,721639 | 26,23524 | -23,1488 |
| .. | | | | |
| AF4 | 0,014057237 | 9,524141 | 39,64777 | -30,5498 |

### 4.4. Data normalization

The data is normalized (Table 8) in order to facilitate the calculation algorithms as much as possible. Processing of the data assumes that input does not depend on amplitudes but on the structure of the input value, which requires normalization.

**Table 8**
Normalized value

| chanel | AF3 | T8 | AF4 |
|---|---|---|---|
| Value | 118,903 | 118,272 | 119,5855 |
| normalized value | 0,192577 | 10,66148 | 29,45653 |

The most commonly used rationing is the statistical rationing, which is set by formula (1). Statistical normalization allows us to compute not the more extreme values but the statistically significant (typical) values.

## 5. Conclusion

The main contribution of this study is the method of identifying the most important characteristics that maximize the distinction between the individual commands issued after the corresponding brain stimulation. The proposed method is fast, simple and intuitive. It implements the individual distribution of features in multiple objects and offers an interpretation of the basic statistical information (mean and standard deviation). The method can be easily applied to other classification tasks, especially in the presence of high data variability, which usually occurs in a study that incorporates individual subjects.

The obtained results show suitable algorithms for the classification of EEG signals. This will help young researchers to achieve interesting results in this area faster.

## 6. Acknowledgements

This work is supported by the research program PPNIP-2021-09/12.14.2021 "Analysis and optimization of algorithms for classification of signals coming from Smart IoT devices" and National Science Program "Information and communication technologies for unified digital market in science".

## 7. References

[1] Akinyode, Babatunde & Khan, Tareef. (2018). Step by step approach for qualitative data analysis. International Journal of Built Environment and Sustainability. 5. 10.11113/ijbes.v5.n3.267.

[2] B Colombet, M Woodman, C G Bénar, J M Badier,"AnyWave: A cross-platform and modular software for visualizing and processing electrophysiological signals", HAL Id: hal-01323171, https://hal.archives-ouvertes.fr/hal-01323171, Submitted on 30 May 2016.

[3] Dr. Zhibin Tan, Dr. William H. Blanton, Miss Qianru Zhang, “Real-time EEG signal processing based on TI’s TMS320C6713 DSK”, 120th ASEE Annual Conference@Exposition, Frankly, 23-26 Jone, 2013.

[4] G. Schalk a,b,∗, P. Brunner a,c, L.A. Gerhardt b, H. Bischof c, J.R. Wolpawa,"Brain–computer interfaces (BCIs): Detection instead of classification Brain-Computer Interface Research and Development Program", Journal of Neuroscience Methods 167 (2008) 51–62.

[5] G. Panayotova, D.A. Dimitrov, "Modeling from time series of complex brain signals", International Journal of Signal Processing Systems Vol. 9, No. 1, March 2021 pp 1-6.

[6] Georgi P. Dimitrov,. Ilian Iliev, Front-end optimization methods and their effect, MIPRO 2014 - 37th International Convention, 26-30.06.2014

[7] Georgi P. Dimitrov,. Galina Panayotova, Stefkka Petrowa, Analysis of the Probabilities for Processing Incoming Requests in Public LibrariesThe 2 nd Global Virtual Conference 2014 (GV-CONF 2014) Goce Delchev University Macedonia & THOMSON Ltd. Slovakia, April 7 - 11, 2014,ISSN: 1339-2778

[8] Kryvonos, I.G., Krak, I.V, Barmak, O.V, Kulias, A.I.: Methods to create systems for the analysis and synthesis of communicative information. Cybern. Syst. Anal. 53(6), 847–856 (2017). https://doi.org/10.1007/s10559-017-9986-7

[9] I. Krak, O. Barmak, E. Manziuk. Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology, Computitional Intelligence (2020) 1-26. https://doi.org/10.1111/coin.12289

[10] Metcalf, Leigh & Casey, William. (2016). Introduction to data analysis. 10.1016/B978-0-12-804452-0.00004-X.

[11] O'Connor, H. & Gibson, Nancy. (2003). A Step-By-Step Guide To Qualitative Data Analysis. Pimatisiwin: A Journal of Aboriginal and Indigenous Community Health. 1. 63-90.

[12] Plesinger F1, Jurco J, Halamek J, Jurak P., "SignalPlant: an open signal processing software platform",Physiol Meas. 2016 Jul;37(7):N38-48. doi: 10.1088/0967-3334/37/7/N38. Epub 2016 May 31.

[13] Rieger, Josef & Kosar, Karel & Lhotska, Lenka & Krajca, Vladimir. (2004). EEG Data and Data Analysis Visualization. 3337. 39-48. 10.1007/978-3-540-30547-7_5.

[14] Silverman, B. W., Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986.

[15] Shu, Hong. (2016). Big data analytics: six techniques. Geo-spatial Information Science. 19. 1-10. 10.1080/10095020.2016.1182307.