MEG DATA ANALYSIS OF EVOKED POTENTIALS FROM VISUAL AND AUDITORY STIMULI

DIPLOMA THESIS by
MARIA K. PATSIOURA
CHANIA, 2015

Thesis Supervisor
Professor Michalis Zervakis

TECHNICAL UNIVERCITY OF CRETE
DEPARTEMENT OF
ELECTRONIC AND COMPUTER ENGENEERING

Thesis Committee

Professor Michael Zervakis, Supervisor
Professor Costas Balas
Associate Professor Katerina Mania

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Professor Michalis Zervakis, for his guidance, support, constructive remarks as well as for giving me the opportunity to expand my knowledge in the field of bioinformatics.

Moreover, I would like to thank M.Sc. Marios Antonakakis for sharing his knowledge and his devoted time. Furthermore I would like to thank the rest of my thesis committee: Professor Costas Balas and Associate Professor Katerina Mania for their contribution and insightful comments.

Last but not least, I would like to thank, my beloved friends for their love, support and encouragement through all these years.

ABSTRACT

Magnetoencephalogram (MEG) is a useful tool towards the direction of understanding of the mechanisms of human brain having been assessed for its superior accuracy over other modalities. The scope of the present study is the extraction of synchronization features using independent component analysis (ICA) on the MEG recording from normal people after being subjected to an experiment that involves exposure to visual and auditory stimulus. The purpose of this diploma thesis is to process MEG data in order to remove artifacts contaminating the brain activity recordings and then to classify the MEG channels in order to categorize the most important ones, that are most involved in the visual and the auditory brain response, into clusters. For the first task, proper filtering procedures, BSS and ICA methods were used. The application of ICA helps to emerge hidden cerebral and non-cerebral activity and the elimination of non-cerebral activity to the independent components. For the clustering purposes, the method that was used is k-means algorithm, with the help of measures like phase lag index (PLI), energy and kurtosis. Finally, a statistical analysis was performed for the extraction of statistical differences between auditory and visual results.

ΠΕΡΙΛΗΨΗ

Η μαγνητοεγκεφαλογραφία (ΜΕΓ) είναι ένα χρήσιμο εργαλείο προς την κατεύθυνση της κατανόησης των μηχανισμών του ανθρώπινου εγκεφάλου και έχει αξιολογηθεί για την μεγαλύτερη ακρίβειά της σε σχέση με άλλες μεθόδους. Το πεδίο εφαρμογής της παρούσας μελέτης είναι η εξαγωγή των χαρακτηριστικών συγχρονισμού χρησιμοποιώντας ανάλυση ανεξάρτητων συνιστωσών (ICA) σε ΜΕΓ καταγραφές από κανονικούς ανθρώπους, οι οποίοι έχουν υποβληθεί σε ένα πείραμα που περιλαμβάνει την έκθεση σε οπτικό και ακουστικό ερέθισμα. Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η επεξεργασία των δεδομένων ΜΕΓ, προκειμένου να αρθούν τα σήματα που μολύνουν τις καταγραφές της καθαρής εγκεφαλικής δραστηριότητας και, στη συνέχεια, η ταξινόμηση των καναλιών ΜΕΓ, προκειμένου να κατηγοριοποιηθούν οι πιο σημαντικές, αυτές που είναι περισσότερο εμπλεκόμενες στην οπτική και την ακουστική ανταπόκριση του εγκεφάλου, ομάδες καναλιών. Αρχικά, χρησιμοποιήθηκαν κατάλληλες διαδικασίες φιλτραρίσματος, καθώς και μέθοδοι όπως BSS και ICA. Η εφαρμογή της ICA βοήθησε στην εμφάνιση της κρυμμένης εγκεφαλικής και της μη-εγκεφαλικής δραστηριότητας καθώς και στην εξάλειψη των μη-εγκεφαλικών δραστηριοτήτων από τις ανεξάρτητες συνιστώσες. Για τους σκοπούς της ομαδοποίησης των καναλιών, η μέθοδος που χρησιμοποιήθηκε είναι ο αλγόριθμος k-means, με τη βοήθεια των μέτρων, όπως ο δείκτης υστέρησης φάσης (PLI), η ενέργεια και η κύρτωση. Τέλος, μια στατιστική ανάλυση έγινε για την εξαγωγή στατιστικών διαφορών μεταξύ των ακουστικών και των οπτικών αποτελεσμάτων.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

1.1 Human Brain

The human brain [1][2][3] is the most important part of our body and it constitutes a great part of the nervous system. It is positioned in the head, inside the skull which is protecting the brain. The human brain consists of two hemispheres, the left which is related with the language and the right which is said to be connected with the emotions. These two hemispheres are separated by the medial longitudinal fissure. Each hemisphere is divided into four lobes: the frontal lobe, parietal lobe, occipital lobe, and temporal lobe.

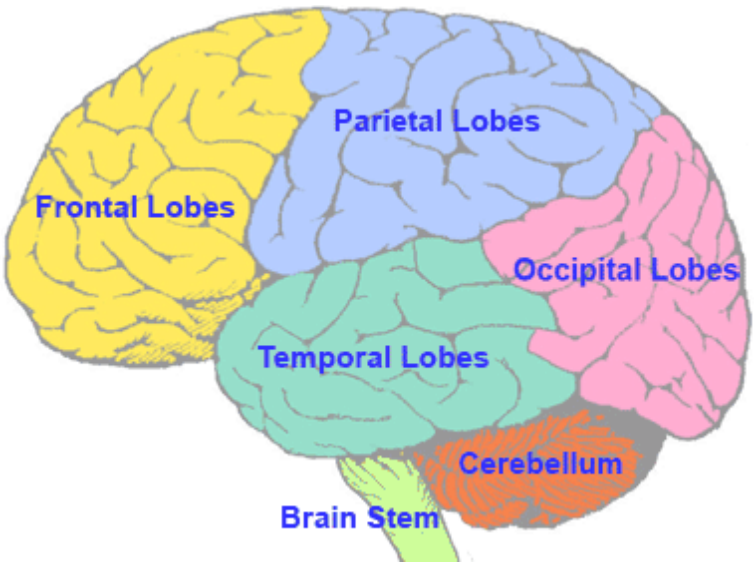

Figure 1.1: Lobes of Human Brain.

The frontal lobe is the biggest lobe and it is responsible for the speech, judgment and motor functions. The parietal lobe is responsible for the sense of touch, handwriting and body position. The occipital is responsible for the vision and the temporal lobe is responsible for the hearing and memory.

Furthermore, the human brain is divided into three parts called the forebrain, midbrain, and hindbrain. The forebrain (prosencephalon) contains the thalamus, hypothalamus, epithalamus and subthalamus. It plays a central role in the information processing that is usually related to complex cognitive activities, sensory and associative functions, and voluntary motor activities. The midbrain (mesencephalon), which is the smallest part of the brain, consists that part of the brain that is connected with the central nervous system that is responsible for vision, hearing, motor control, sleep/wake, arousal (alertness), and temperature regulation. Finally, the hindbrain (rhombencephalon) is a developmental categorization of portions of the central nervous system in vertebrates. It includes the medulla, pons, and cerebellum, which all together they support vital bodily processes.

Moreover, the brain's tissue is categorized into the grey and the white matter [4]. Grey matter is made of unmyelinated neurons, most of which are interneurons. The grey matter regions are the areas of nerve connections and processing. The white matter is made of myelinated neurons, which purpose is to connect the regions of grey matter to each other and to the rest of the body. The white matter also speeds the connections between the parts of the brain or the body which are distant.

The brain cells are divided into two categories. The first category consists of neurons, or nerve cells, which are the cells that carry out all the communication and the processing inside the brain. There is a specific kind of neurons, called sensory neurons that their purpose is to deliver information about the condition of the body and its surroundings through the peripheral nervous system. Another kind is the interneurons, which are located in the brain's grey matter, integrate and process the information delivered by sensory neurons. Lastly, interneurons are the ones that send signals to motor neurons, which are responsible for the muscles and glands.

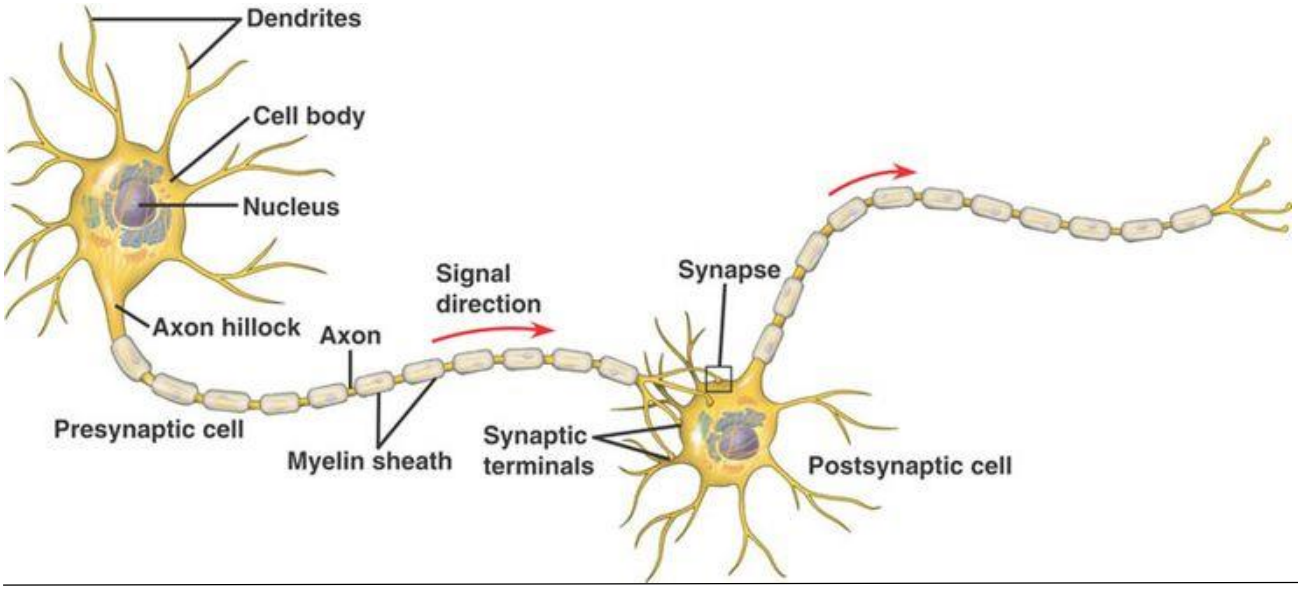

Figure 1.2: Structure of neurons.

Neuroglia, or glial cells, on the other hand, play a more auxiliary role. They are the ones that support and protect the neurons. We have four types of neuroglia in the human brain: astrocytes, oligodendrocytes, microglia, and ependymal cells. The protection is done by filtering all the nutrients out of the blood, or by destroying pathogens.

The human's brain physiology consists of many different operations. Motor control, Sensory, Reflexes, Memory, Processing and Sleep are just a few of them. In this diploma thesis, we are going to study one of that physiologies, the brain reflexes. A reflex is a fast and involuntary reaction to an internal or external stimulus. Many of the body of the face reflexes are integrated in the brain. All of these reflexes bypass the control center of the cerebral cortex and integrate in the lower regions of the brain, such as midbrain or limbic system.

1.2 Magnetoencephalography

There have been various approaches to understand the different functions of the human brain and its mechanisms, but without having the expected results. One reason might be that the researchers cannot apply all possible methods of studying the brain to human beings. Besides that, the neurology and physiology of the brain can be studied through various tools of biomedical technology such as the electroencephalogram (EEG) and the megnetoencephalogram (MEG). These are the two most significant non-invasive techniques that record the human brain activity. In this diploma thesis, we are going to work with the megnetoencephalogram and the processing of the electromagnetic signals which are produced by the human brain.

Magnetoencephalography [5][6] is a non-invasive neurophysiological technique that measures the magnetic fields generated by neuronal activity of the brain. MEG is a relatively new method of brain activity record, the first application was at 1968. The spatial distributions of the magnetic fields are analysed to localize the sources of the activity within the brain, and the locations of the sources are superimposed on anatomical images, such as MRI, to provide information about both the structure and function of the brain. MEG has a very high temporal resolution as well as an excellent spatial resolution, which lead to a very accurate picture of the brain. Moreover, MEG can record the brain oscillations simultaneously all over the scalp, because of the use of multiple channels. This kind of examination is also completely harmless, because it does not require any kind of injection or exposure to X-rays, in that way, repeated tests are possible even to children.

In the brain, there are individual neurons that have electrochemical properties that result in the flow of electrically charged ions through a cell. Consequently, this slow ionic current flow generates electromagnetic fields. These fields have insignificant magnitude at the level of an individual neuron, but when multiple neurons, in a specific area of the brain, are excited together then they produce a measurable magnetic field outside the head. These

signals are extremely small, in the range of 10fT to 10^3fT on the surface of the cortex. Thus MEG scanners require SQUID sensors – Superconducting Quantum Interference Device-, which are made of superconductive materials and they are bathed in a cooling unit at certain temperature, low enough that the SQUIDS can detect and amplify the desired magnetic fields.

Another essential parameter as to record MEG signals more efficiently is that the MEG scanner must be in a magnetically shielded room so that the environmental noise is diminished. These rooms have walls that are made of multiple layers of a high conductive material, so it can contribute to a noise reduction at scale of 100Hz to 0.1Hz.

The advantages of MEG over the other approaches of depicting the brain activity, such as EEG or fMRI, are multiple. First of all, only MEG provides both timing and spatial information about the brain signals. Moreover, MEG signals are obtained directly from the neuronal electrical activity, in contrast with fMRI signals which are obtained indirectly through the oxygenation of the blood flow near active neurons. Absolute neuronal activity can be shown through MEG signals, unlike the fMRI which shows only the relative neuronal activity. Finally, the subject is allowed to move during the MEG, while fMRI requires complete immobility.

1.3 Biosignal Processing

This diploma thesis is related to bioinformatics and furthermore to biomedical signal processing. The aim of this scientific area is to obtain biological signals from living organisms and then to apply any mathematical processing on them, so they can be easily analysed. By the term biosignals we refer to any signal that is produced in the living organism. These signals can be categorized according to the origin into bioelectrical signals (EEG, ECG etc.) and biomagnetic signals (MEG, MCG etc.), according to their application (neurology, cardiology) and their characteristics (continuous/discrete, stationary/non-stationary).

In general, the development of a biomedical method for biosignal analysis and exportation of results and conclusions includes the following steps:

- Record of biomedical signal with the help of the proper sensors.
- Filter and process of the signals for denoising and removing of the artifacts.
- Export of the characteristics that describe the conditions under which the biosignal is.
- Clustering of these characteristics for observation of the obtained information.
- Representation of the information in a more understandable way.

1.4 Thesis Outline and Innovation

The 2nd chapter covers the necessary theoretical background for the development of this thesis. In chapter 2 are outlined the different brain waves of the MEG dataset that was used for this study, as well as a description of the artifacts contaminating the brain activity recordings. Lastly, the event related potentials and the evoked potentials are reported.

The 3rd chapter presents all the proposed methodology that concerns this study. There is an introduction to each one step that we used for MEG data analysis, such as pre-processing, Blind Source Separation methods (PCA and ICA), a discussion over ICA algorithms and lastly an introduction to MEG channel clustering and the algorithms that were used.

In the 4th chapter are presented the results of each step that we used and was discussed in our methodology.

The innovative concept of this diploma thesis involves the process of MEG data in order to remove the artifacts from the pure signal that comes from brain activity. For that reason, ICA method is implemented for the removal of ocular and cardiac artifacts as well as the removal of power line noise. Moreover, in this thesis we present clustering of MEG channels with the use of a k-means algorithm. We also try to compare and discuss the clustering with the use of different criteria selection. Most of the published studies are involved with this kind of brain signal processing but under the scope of the electroencephalography (EEG).

2 THEORETICAL BACKGROUND

2.1 Brain Waves

As it was discussed above, when any brain activity occurs, the neurons get activated and then there is flow of current that produces magnetic field. If there are many neurons in a specific area of the brain that are simultaneously activated, the magnetic field has bigger amplitude and lower frequency. However, electromagnetic activity can be produced by the brain for several reasons spontaneously. Even when our body is idle there is electromagnetic activity at the cortex. This is due to multiple spontaneous energies our body performs, such as the eye movement or just blink and the heartbeat. These energies along with some other factors are named artifacts and will be analysed later.

In addition, the communication between the neurons produces the brainwaves. Brainwaves can be recorded by special sensors which are placed on the scalp. The main characteristic of the brainwaves is the frequency which is the criteria for separating them into 5 basic categories according to their bandwidth. We have alpha, beta, gamma, delta and theta waves that are measured in Hertz [7].

- Delta Waves: 0.5-4 Hz.
  Delta brainwaves are characterized by low frequency and they are generated in deep –dreamless- sleep or meditation and they are said to be connected with healing and regeneration.
- Theta Waves: 4-8 Hz.
  Theta brainwaves are also generated in idling in state of drowsiness. While in theta the brain is in a semi-hypnotic state and improves our intuition and creativity.

- Alpha Waves: 8-12 Hz
  Alpha brainwaves are in between of conscious think and subconscious mind that is why they are caught in relaxation and when closing our eyes.
- Beta Waves: 12-30 Hz
  Beta brainwaves are related mostly to active thinking, focus and alert. Beta are divided into three bands because of the big variance.
- Gamma Waves: 30-100 Hz
  Gamma brainwaves include frequencies bigger than 30Hz and they are related to simultaneous information processing from different brain areas, as well as to memory and learning. Gamma wave appears also in cases of psychosis and that is why it is used for clinical purposes in order to diagnose the disease.

2.2 Gradiometer

A gradiometer is a gradient, in other words a numerical rate of change, of a magnetic field. For MEG there are two types of gradiometer configurations, axial and planar. In first place, axial gradiometer MEG sensors are aligned orthogonally to the scalp and therefore record the same signal type at each channel (the gradient of the magnetic field along the radial direction). On the other hand, planar gradiometers MEG sensors record different signal types at pairs of channels (even and odd channels) because each channel correspond to one of the two orthogonal gradients on the plane tangential to the scalp.

2.3 Artifacts

The information that is extracted from the brain, through the sensors of MEG, may contain signals that come from extra-cerebral activity or from sensor imperfections. These contaminations interfere with the relevant physiological brain activity and need to be filtered out, so that the Signal-to-Noise Ratio (SNR) is increased and brain signal analysis may be facilitated.

The amplitude of the magnetic field generated by regular brain activity ranges between 10fT and 500fT. [8] The interference of these brain activity signals comes from two main sources. The first is the *Cardiac Artifact* (CA). It is caused by heartbeats and produces a magnetic field quite stronger than the one produced by the brain activity. Its amplitude can reach a few pT over the chest and might be of considerable magnitude around the head.

The CA is generated by the activity of the heart, which produces electrical signals. The channels placed on the left hemisphere of the brain are more affected by the cardiac signal due to the position of the heart with respect to the head. It is essential that we attach a great importance when we record MEG signals, because of these contaminations in the brain activity measurements. The magnitude of the CA is not of a constant value and therefore it varies from subject to subject.

The second main artifact is the *Ocular Artifact* (OA) that plays also a significant role as a contamination of the brain activity. OA is caused by the eye blinks and eye movements. Each eye can be considered as a small electrical dipole, with the positive pole oriented to the cornea and the negative to the retina. Consequently, the orientation of this dipole changes with every movement of the eyes and therefore the electromagnetic fields around the eyes are changed. On the other hand, the eye blinks cause the alteration in the intensity of the dipole created by the movement of the eyelid over the eyeball.

The amplitude of the electromagnetic fields caused by the ocular artifacts ranges between 3pT and 4pT over the lateral aspects of the orbits and the polarity of the magnetic fields is opposite to each hemisphere. Finally, the amplitude of the OAs can be affected by the brightness, increased with light and decreased with darkness.

Another notorious sources of artifact are the main power line frequency and the mechanical vibrations. In the case of this Diploma Thesis, the MEG measurements that we analysed are taken from American University and thus the main power frequency is equal to

60Hz, when in other parts of the world this frequency is at 50Hz apart from some places in Asia that follow the American system.

At last, artifacts in MEG channel signals can be produced by muscle activities. This activity, usually, is generated by muscle contractions in the face and neck areas. The muscle artifacts are weak in MEG measurements and they can be removed with the help of notch filters.

2.4 Event Related Potentials

Event Related Potential (ERP) is the potential difference that is measured usually on the scalp. An ERP is generated as the preparation or as a response to a specific event which is taking place either in the outer environment or as a psychological procedure. ERPs as categorized in Evoked Potentials (EPs), where the stimuli come from the environment and Emitted Potentials, where the stimuli come from psychological reasons.

It is possible that the researchers can study the brain activity that comes from an external stimulus with the use of prescribed experimental procedures. [9] The transient electric potential shifts, which are the ERP components, are time-locked to the stimulus onset. The amplitude of the event related potentials is of small value, i.e. 0.1 to 20 μV. ERPs are mostly characterized by simultaneous multi-dimensional online measures of polarity (negative or positive potentials), amplitude, latency, and scalp distribution. They can provide very high time resolution, in the range of one millisecond. ERPs are named with the letter P for positive or with N for negative, followed by a number showing the delay between stimulus and response measured in milliseconds (P300, N200, etc.).

2.5 Evoked Potentials

An Evoked Potential is an electrical potential recorded from the nervous system of a living being following presentation of a stimulus, as distinct from spontaneous potentials as detected by an electrophysiological recording method, such as EEG or EMG. The amplitudes of these potentials are usually low and they can be recorded from cerebral cortex, brain stem, spinal cord or peripheral nerves. With the term evoked we mean that the responses involve the stimulation of the central nervous system structures.
Depending on the type of stimulus, the Evoked Potentials can be divided into four categories:

- Visual Evoked Potentials, including projection of images, changing of colours, flashes of light etc. It is common the use of a black and white checkerboard or grating patterns with light and dark stripes and squares respectively.
- Auditory Evoked Potentials, including clicks, speech sounds, tones of different frequencies and in different volumes etc.
- Somatosensory Evoked Potentials, including electric current of short duration and intensity in order to irritate a specific nerve.
- Laser Evoked Potentials, including finely focused, rapidly rising heat to bare skin using a laser.

The first three categories constitute the Somatosensory evoked potentials (SEP). SEPs are recorded from the central nervous system and they cause sense organs stimulation. They are widely used as a diagnostic means of clinical medicine since 1970s and they can are elicited by electrical stimulation of peripheral nerves, as well as by mechanical stimulation.

In the context of this diploma thesis, we will try to study the behaviours of the human brain under the exposure of the subjects to visual and auditory evoked potentials and with the use of observations of the P300 waves or P3. The P300 wave only occurs if the subject is actively engaged in the task of detecting the target stimulus. Also, the P300 wave usually represents the transfer of information to consciousness, a process that involves many different regions of the brain.

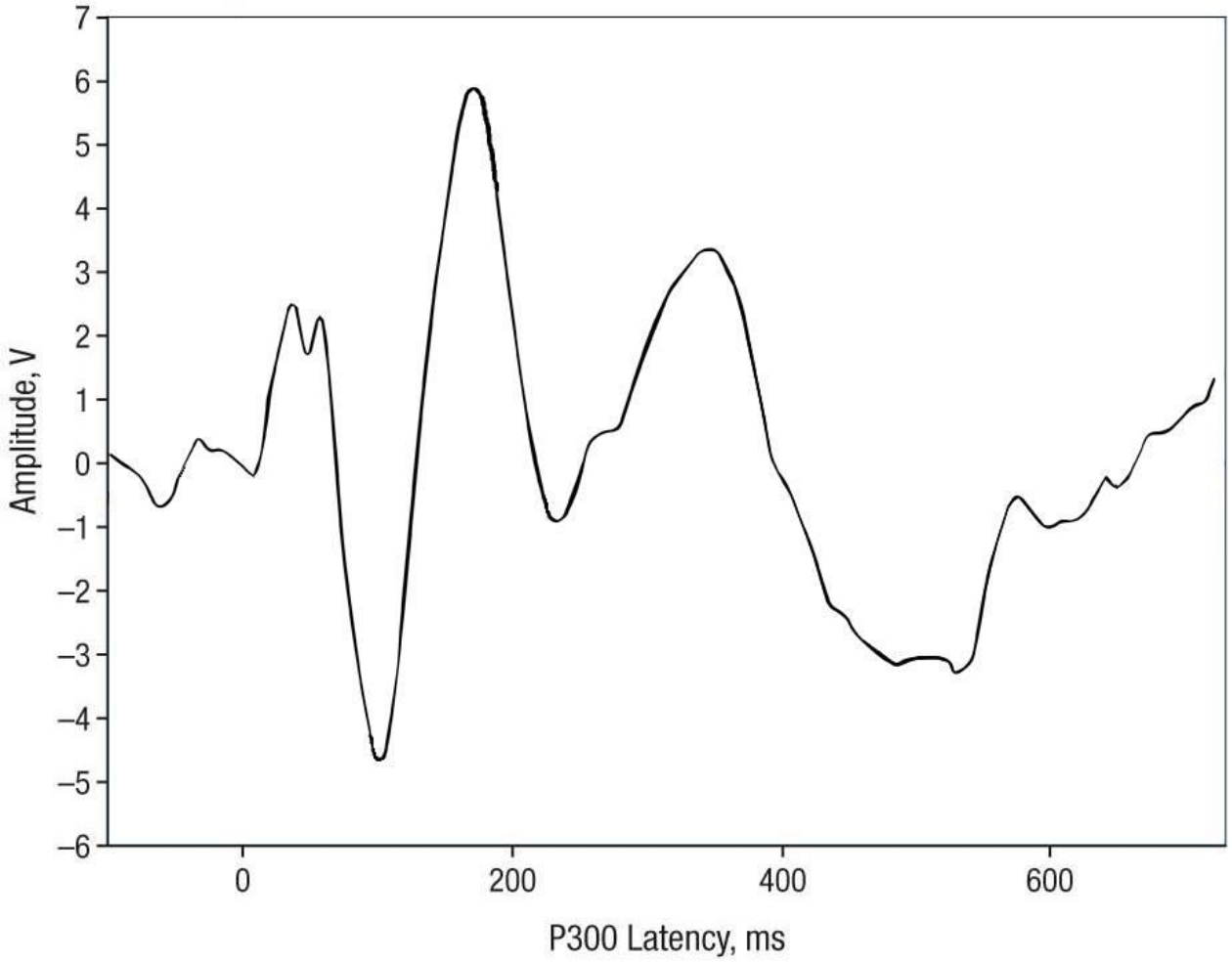

Figure 2.1: The P300 waveform.

3 METHODOLOGY

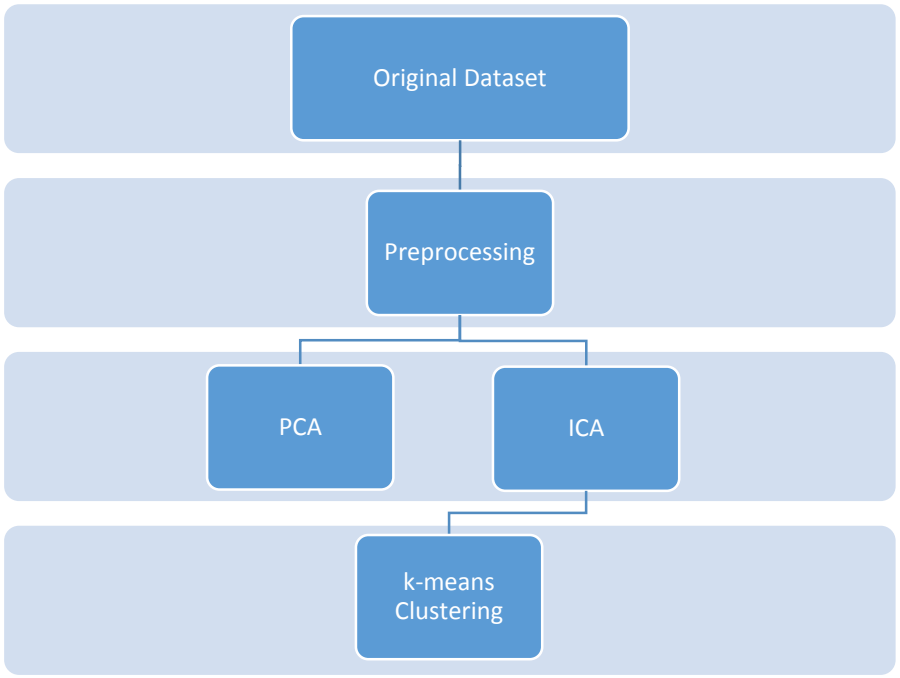

Figure 3.1: Methodology.

3.1 Pre- Processing

In this study, the MEG data are taken through 248 channels and the sampling frequency was set to 1017.25 Hz.

The software that was used for loading and pre-processing the files is MATLAB r2013b, with the help of Fieldtrip toolbox. Fieldtrip is a very useful tool specializing in the process and analysis of MEG as well as EEG signals. This toolbox contains many routines and functions which are essential to processing MEG data. The disadvantage of Fieldtrip is that there is no User Interface (UI) for the software that makes it a little harder to use. The MEG data which are loaded in the program have the form like the sample in the following figure.

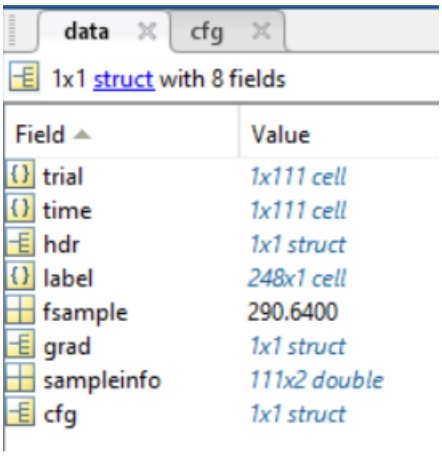

Figure 3.2: An example of structure of MEG data as constructed in Fieldtrip.

As it is shown in Figure 3.1, the data are saved in the workspace of Matlab in the form of a structure, along with a second structure (cfg) that contains the configurations needed for every function of Fieldtrip.

Firstly, it was essential that we should repair any bad or missing channel from the given MEG dataset. The procedure that was followed consisted of finding and replacing the bad or missing channels with the average of its neighbours. We consider as neighbours, the channels with the smallest distance from the channel to be repaired.

The second step in MEG data processing was spectral filtering. For this purpose a bandpass Butterworth filter was constructed in Matlab, which has Infinite Impulse Response (IIR). The bandpass filter was used in order to reduce the spectrum of MEG signal. After that, the next step was to cut off the main Power Line Noise. One of the methods used to eliminate this contamination in the MEG data is with a notch filter. As it was discussed later, in this case of MEG data we used a notch filter at 60Hz. It is obvious that the bandpass filter itself is not capable of cutting the PLN that is why it is preferred to be used in addition with the notch filter.

3.2 Blind Source Separation

In the last years, the use of the Blind Source Separation Methods (BSS) have become very extend in the field of bioinformatics and specifically for the analysis of biomedical signals. They are mostly used as statistical techniques for separating multichannel MEG or EEG records into independent components, without (or with little) aim of information about the source signals or the mixing process. This process is often essential because the signals from the brain activity contain extra cerebral information, like artifacts, as it was mentioned in chapter 2.

A classic example demonstrating the BSS is the Cocktail-Party Problem, where there is a lot of people talking simultaneously. In this case we have multiple speakers (sources) and our ears (sensors) and we have to isolate every speaker in the room by suppressing the other sounds. This problem can be extended to an M-sensors and N-sources problem which makes the BSS problem.

The modelling of the BSS problem is presented in the following mathematical equation:

$$x(t) = As(t) + n(t)$$

, where $n(t)$ is the noise which interfere with the signal, due to artifacts or the existence of insufficient information.

The BSS methods target to calculate the mixing matrix $A(MxN)$, which contains weights in form of vectors $a_k = (a_{1k}, a_{2k}, \dots, a_{mk})$ and also to estimate the components $s(t)$.

3.3 Independent Component Analysis

The Independent Component Analysis –also known as ICA method, is a technique for decomposing a multivariate signal into additive subcomponents. ICA is a special case of Blind Source Separation (BSS) and was the first method to be able to solve the Cocktail-Party Problem with the separation of the speech signals, as recorded by microphones, into their components. The ICA problem generally involves estimating a set of source signals when only their mixtures are observed. According to the BSS problem, M time signals $x_1(t), x_2(t), \dots, x_m(t)$ as vector $x = (x_1, x_2, \dots, x_m)^T$ called channels or mixtures result from the linear combination of N time signals $s_1(t), s_2(t), \dots, s_n(t)$ as vector $s = (s_1, s_2, \dots, s_n)^T$ called sources or components. The signals $x(t)$ are the ones to be measured from MEG, but the components $s(t)$ and the factors a are unknown values.

$$x_i(t) = a_{i1}\, s_1 + a_{i2}\, s_2 + \cdots + a_{in}\, s_n \text{ , } i \in [1, \text{m}] \text{ and } a_i \in \mathbb{R}$$

OR

$$x(t) = As(t)$$

If the reversibility of the matrix A is taken into account, we can write the above equation as:

$$s(t) = Wx(t) => \sum_{i=1}^{n} a_i\, s_i$$

, where $W(MxN)$ is the matrix of the inverse linear transform of A.

The aim of the ICA method is to solve the BSS problem, as well as to reconstruct the components with the help of the signals from the channels. There are some certain conditions under which the ICA can be applied. In other words, ICA's task is to estimate the mixing matrix A and the sources s given the model and the samples x. The first is that the components s_i must be statistically independent and that they are non-Gaussian signals. Secondly, the matrix A must be a square matrix and of full rank, that means that the channels must be more or equal than the sources (m ≥ n). Lastly, the channels must have mean m_μ=0 and variance v=1. If the given MEG data do not have these statistical properties then they have to be a proper transformation before the application of the ICA method.

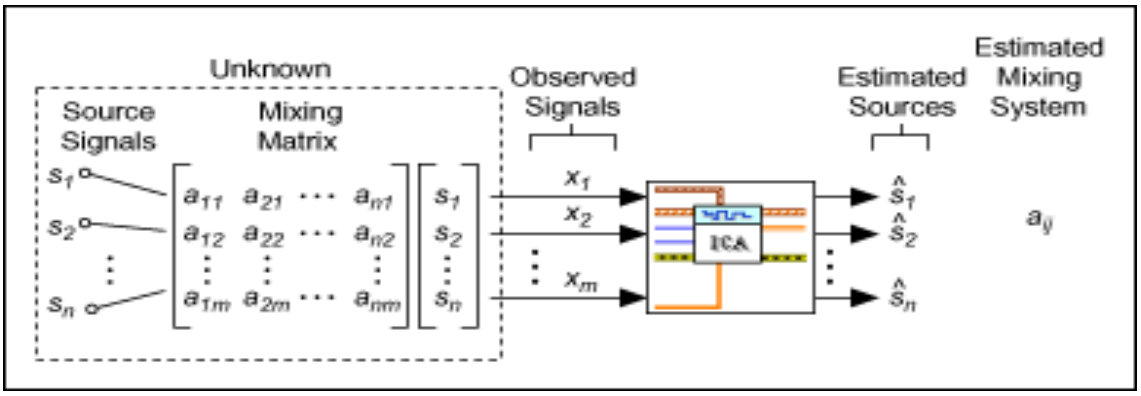

Figure 3.3: ICA computing source signals flow chart.

A more accurate and realistic procedure of ICA would be described by the following mathematical equation:

$$x(t) = f\{s(t)\} + n(t)$$

, where f is an unknown transformation function and $n(t)$ is the noise in the channels.

Then, the aim of ICA would be the construction of a function approximate to the f^{-1} function, where $f^{-1}()$ is the inverse $f()$, with the help of which the unknown components $s_1(t), s_2(t), \dots, s_n(t)$, will be defined. Because the known information that we have from the channels and the noise are insufficient, we have to make several assumptions about the signals and the methods. These assumptions, has been proved that they do not change the sufficiency or the accuracy of the ICA method.

The assumption of statistical independence is the basis of ICA method. This means that the sources contain no information about the relations between them. According to the next equation, we may assume that s components are mutually independent if and only if the following is valid:

$$p(s_1, s_2, \dots, s_n) = \prod_{i=1}^{n} p_i(s_i)$$

, where $p(s_1, s_2, \dots, s_n)$ is the joint probability density function of the N components and $p_i(s_i)$ is the probability density function for every s_i.

It may be essential to follow a specific process with the data before the application of ICA method, called Whitening, under which the data are transformed into having zero mean value and unit variance. This procedure will limit down the free variables and will achieve a more accurate estimation of the problem. Moreover, it must be taken into account, the assumption that the values of the transformation matrix A are invariant over time. The independency between the variables is enshrined in the zero value of the mutual information.

3.3.1 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This technique is mostly used in multivariate analysis and data-mining, but in this case PCA is used for the Whitening method that was mentioned before, which is significant to MEG data pre-processing before the ICA application. The aim of PCA is to analyse the data into a set of successive orthogonal components, in order to identify and find patterns to reduce the dimensions of the dataset with minimal loss of information.

If $x(t) = [x_1, x_2, \dots, x_m]^T$ a multidimensional signal with the assumption of zero mean. The eigenvectors and eigenvalues of the estimated covariance matrix are computed by:

$$C_{xx}^0 = E\{x(t)x^T(t)\} = \mathrm{V}\Lambda\mathrm{V}^T \in \mathbb{R}^{mxm}$$

, where $\mathrm{V} = [v_1, v_2, \dots, v_m]$ the corresponding unitary matrix containing the principal eigenvectors and $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]$ a diagonal matrix with the eigenvalues sorted in descending order. Again, the case holds that all variables have zero mean.

The above equation succeeds in the transformation of signal x into uncorrelated variables y, with the condition that Y is a diagonal covariance matrix such that

$$Y = PX$$

As well as,

$$C_{xx}^{0} = \mathrm{V}\Lambda^{1/2}\Lambda^{1/2}\mathrm{V}^{T}$$

$$E\{x(t)x^{T}(t)\} = E\{P^{-1}YY^{T}P^{-T}\} = P^{-1}E\{YY^{T}\}P^{-T}$$

So we conclude that:

$$P = \Lambda^{-1/2}\mathrm{V}^{T}$$

And

$$Y = \Lambda^{-1/2}\mathrm{V}^{T}\ X$$

Y is the whitening matrix, where it is noticed that the directions of the principal components claim the basis of non-association (i.e. the covariance is identical) and the variances show the gradation of the transformation matrix. The final step for the whitening method is that the mixing matrix A, which was mentioned in the beginning of chapter 3.3, can be transformed as:

$$\hat{A} = \Lambda^{-1/2}\mathrm{V}^{T}\ A$$

At last, if we follow this described analysis to find principal components after the process of whitening, the number of free parameters will be reduced and the evaluation of the independent components (ICs) will become much easier.

3.3.2 Kurtosis and Skewness

Kurtosis is a statistical measure of the probability distribution of a real valued random variable. Kurtosis is just fourth-order statistics

and its value for a Gaussian variable limits to zero. Its normalized equation is given below:

$$kurtosis(s) = \frac{E[y^4]}{(E[y^2])^2} - 3$$

OR

$$kurtosis(s) = \frac{m_4}{{m_2}^2} - 3$$

, where y is a random variable with zero mean.
The measure of kurtosis shows how sharp are the peeks of the distribution of a random variable. The value of kurtosis is relevant to 3. If $kurtosis(s) > 3$ then s follows the super-Gaussian distribution (leptokurtic), if $kurtosis(s) < 3$ then s follows the sub-Gaussian distribution (platykurtic) and if $kurtosis(s) = 3$ then s follows the normal (Gaussian) distribution (mesokurtic).

Skewness is a statistical measure of the asymmetry of the probability distribution of a real valued random variable about its mean. Symmetric is a distribution or a dataset if and only if there exists a value x_o such that $f(x_0 - \delta) = f(x_0 + \delta)$, $\delta \in \mathbb{R}$ and f the probability density function (pdf).

$$skewness = \frac{m_4}{{m_2}^{3/2}}$$

If $skewness < 0$ then the distribution of s is concentrated on the right of the figure, if $skewness > 0$ then the distribution of s is concentrated on the left of the figure and if $skewness = 0$ then the distribution of s is normal (Gaussian).

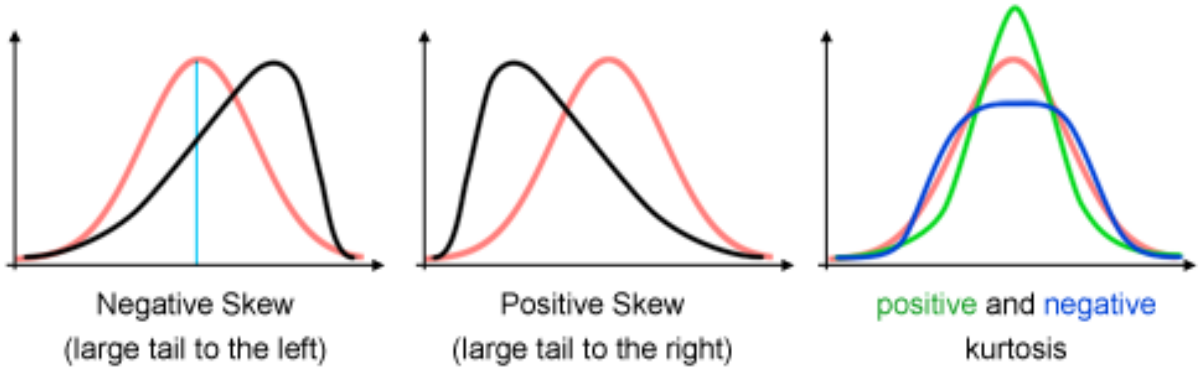

Figure 3.4: Graphs of Skewness and Kurtosis.

3.4 ICA Algorithms

There are many popular algorithms for the implementation of ICA method. The first algorithms are the ones that are based on High Order Statistics (HOS), which aim to find the statistically independent components. HOS algorithms are separated into three categories depending on the criteria used to determine the statistical independence and the method used to maximize it.

1. **FastICA**: an efficient and popular algorithm by Hyvärinen-Oja in 1997. It seeks an orthogonal rotation of pre-whitened data, through a fixed-point iteration scheme, that maximizes a measure of non-Gaussianity of the rotated components. It finds one component at a time and uses the statistical measure of kurtosis. Purpose of the algorithm is the maximisation of kurtosis in order to make the component signals independent.

2. **InfoMax** (Information Maximisation): introduced by Bell and Sejnowski in 1995. It finds the independent components by maximising the differential entropy. The bigger the entropy the smallest amount of information of the variable in the same variance has. As differential entropy of a random variable s is named the difference of entropy $H(s) = E[\ln \left|\frac{\partial s}{\partial x}\right|] - E[\ln f_y(s)]$, where f_s the pdf of s. If $H(s) = 0$ that means that the variable s follows the normal distribution. Purpose of the algorithm is the maximisation of entropy kurtosis in order to make the component signals independent.

3. **JADE** (Joint Approximate Diagonalization of Eigenmatrices): introduced by Cardoso in 1999. It aims to extract independent non-Gaussian sources from signal mixtures with Gaussian noise. It is based on the construction of a fourth-order cumulants array from the data. This algorithm uses the covariance matrix, which is the cumulant tensor of second range and the cumulant tensor of fourth range.

3.5 Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group that is called a cluster, are more similar to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including image analysis, information retrieval and bioinformatics.

In this thesis, the purpose of using cluster analysis is to group MEG independent components (ICs) or MEG channels in general. This grouping of the channels is depending on various signal characteristics, such as energy, kurtosis or other statistical measures, but also is depending on the signal activity that comes from the scalp.

There are various cluster models depending on the different properties of cluster algorithms. These cluster models are the connectivity models, centroid models, distribution models, density models, subspace models, group models and graph-based models. For each of these models, there are corresponding algorithms. Every method of these has a common aim, which is to analyse the components into clusters in order to remove the artifacts.

3.5.1 Hierarchical Clustering

The hierarchical clustering is a connectivity based clustering which lies on the idea that the relation of two objects is based on the distance between them. Objects that are closer to each other form a cluster, which is basically described by the maximum distance needed to connect parts of the cluster. The name "hierarchical" is justified in the fact that the clusters' form can be presented into a dendrogram.

The hierarchical algorithm follows the next steps:

1. Each observation is consider as a unique group.
2. Calculation of a matrix containing every distance between all observations.

3. Grouping of the observations that have the minimum distance between them.
4. Continue with step 1 until every observation is grouped into an existing cluster.
5. If two groups are characterized by the same minimum distance, then these groups are merged.

Hierarchical methods will not produce a unique partitioning of the inserted dataset and they are quite slow for large datasets, considering their time complexity $O(2^{n-1})$ that is why they are often considered obsolete.

There are six different ways to compute the distance between two observations. Assuming that we have a matrix $x(t) = [x_1, x_2, \dots, x_m]$ of observations, the distance between x_t and x_s can be described with one of the following similarity metrics:

- Euclidean distance: $d^2 = (x_s - x_t)(x_s - x_t)'$
- Cityblock distance: $d = \sum_{j=1}^{n} |x_{sj} - x_{tj}|$
- Chevychev distance: $d = max_j\{|x_{sj} - x_{tj}|\}$
- Minkowski distance: $d = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - x_{tj}|^p}$
- Cosine distance: $d = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}}$
- Correlation distance: $d = 1 - \frac{(x_s - \overline{x_s})(x_t - \overline{x_t})}{\sqrt{(x_s - \overline{x_s})(x_s - \overline{x_s})'}\sqrt{(x_t - \overline{x_t})(x_t - \overline{x_t})'}}$

The metrics that are named above can be used in many methods which are applied in order to separate the given signals into clusters:

- Nearest neighbour method (single linkage)
- Furthest neighbour method (complete linkage)
- Average group method
- Ward's method …etc.

It is obvious that the hierarchical clustering has advantages as well as disadvantages as a method. Firstly, the advantages are that the algorithm does not require any apriori information about the number of clusters and also, it is easy to implement and may give best result in some cases. On the other hand, one disadvantage is the time complexity of the algorithm which is $O(n^2 \log n)$ in best case. Moreover, after applying the algorithm, changes cannot be undone. Also, this kind of method are susceptible to noise and outliers, they

have in breaking large clusters and in handling different sized clusters and convex shapes.

3.5.2 k-means Clustering

K-means clustering is a centroid based clustering. In general, centroid based clustering methods, clusters are represented as a central vector, which in some cases may not be a part of the dataset. K-means clustering is a popular data mining method of vector quantization from signal processing. K-means method is a popular unsupervised learning method, meaning that the algorithm divides a set of features into k clusters automatically and it was introduced by McQueen in 1967.

As a clustering method, it aims to partition N observations into k clusters. Each cluster contains the observations that have the nearest mean, which is the prototype of the cluster. For this diploma thesis we used an approach of k-means algorithm which includes computing a brain mapping called topology. For that reason there are constructed the mixing matrices A in such way that they are similar for every subject.

The k-means clustering method uses as parameter a matrix which consists of the fingerprints of every subject. A fingerprint is a group of variables that are the metrics that can describe the brain activity of the subject.

The k-means algorithm follows the next steps:

1. Set k to the expected number of clusters.
2. Calculation of the centroids, i.e. the centre, of every cluster.
3. Assignment of the data points to the closest cluster centre.
4. Recalculate the new cluster centroids.
5. Continue with step 3, until convergence is reached.

The k-means clustering has both advantages and disadvantages as a method. Firstly, the advantages are that the algorithm is fast and robust and also it is relatively efficient with time complexity $O(kndi)$ (where k: the number of clusters, n: the number of objects, d: the dimension of each object and i: the iterations). Also, it gives best results when the dataset is distinct. On the contrary, a disadvantage can be the fact that it requires apriori information

about the number of clusters. Moreover, the algorithm fails for the non-linear datasets and are unable to handle data that contain noise.

For the purposes of this diploma thesis, the cluster analysis algorithm that was used is the k-means clustering method. The reasons are firstly, that k-means produces a single partitioning unlike the hierarchical algorithm that can give different partitioning depending on the level of resolution. Secondly, k-means method is usually faster than hierarchical one.

3.6 Statistical Analysis

A statistical test was applied to each one of the 248 channels, for every brain wave –δ, θ, α, β, γ- and for every measure i.e. PLI, energy and kurtosis. Firstly, for this statistical test proper filter methods must be used, which are divided into multivariate and univariate methods. Multivariate methods are able to find relationships among the features, while univariate methods consider each feature separately. Univariate filter techniques can be divided into two categories: parametric and model-free methods. In parametric methods the data is drawn from a given probability distribution while in model-free methods, or non-parametric, the data may not follow a normal distribution.

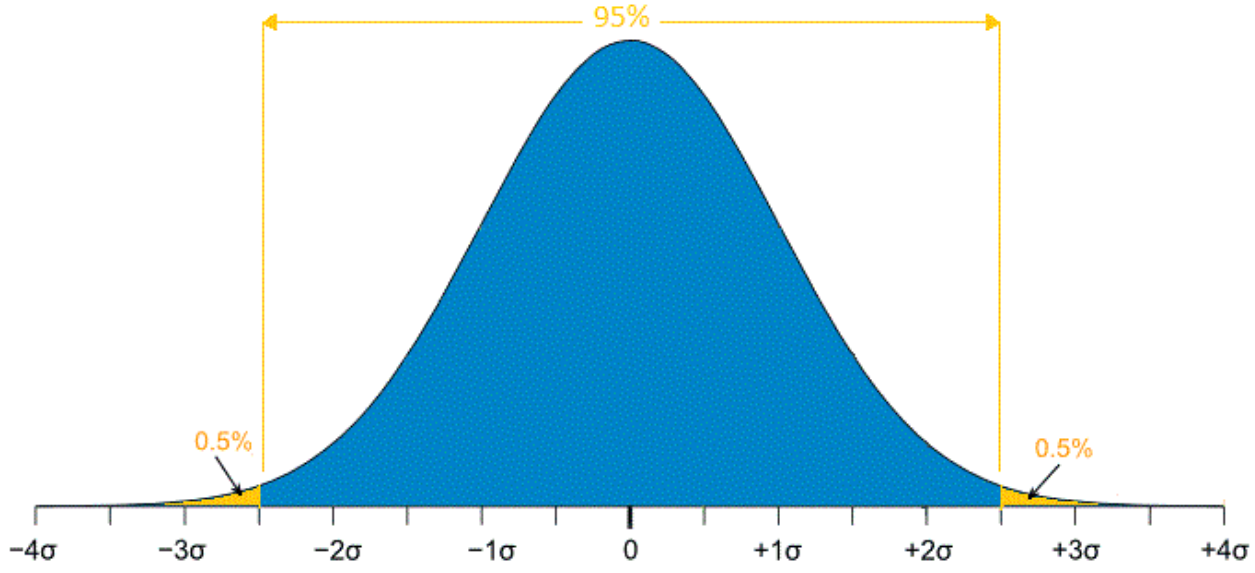

Figure 3.4: Normal distribution.

It is essential to take into consideration the significance level (α, or alpha), which may be defined as the probability of making a decision to reject the null hypothesis when the null hypothesis is actually true (a decision known as a Type I error, or "false positive determination"). Popular levels of significance are 5%, 1% and 0.1%, empirically corresponding to a "confidence level" of 95%, 99% and 99.9%. The term "null hypothesis" refers to a general statement or default position that there is no relationship between two measured phenomena, or no association among groups.

If the point estimate of a parameter is P, with confidence interval [x, y] at confidence level C, then any value outside the interval [x, y] will be significantly different from P at significance level $\alpha = 1 - C$, under the same distributional assumptions that were made to generate the confidence interval. That is to say, if in an estimation of a second parameter, we observed a value less than x or greater than y, we would reject the null hypothesis. In this case, the null hypothesis is: "the true value of this parameter equals P", at the α level of significance; and conversely, if the estimate of the second parameter lay within the interval [x, y], we would be unable to reject the null hypothesis that the parameter equaled P.

For the purposes of this study the confidence level is 95%. If the p-value is smaller than the critical alpha-level (0.05) then we conclude that the data in the two experimental conditions are significantly different, so we reject the null hypothesis. It is common practice to say that a channel is significant if its p-value is less than the critical alpha-level. This practice suggests that it is possible to do spatio-spectral-temporal localization by means of the cluster-based permutation test.

4 RESULTS

4.1 Materials and Methods

- **Participants**

  Ten normal subjects in the age of 22-43 years participated in the study. With the term normal subjects we mean that they have no history of neurological disorders and have normal or corrected-to-normal vision and normal hearing.

- **Procedure**

  These records are taken under certain circumstances, which include evoked potentials from visual and auditory stimuli. Through the procedure, a standard oddball paradigm was used involving 500 stimuli. The probability of occurrence of rare visual or auditory stimuli was 20% (~110-120 trials). For the visual task, the stimuli were a black-white checkerboard (frequent) or a yellow-black (rare) on a projection screen placed 60cm from the subject. For the auditory task, the stimuli were a 1 kHz tone (frequent) or a 2 kHz (rare) delivered binaurally. Each stimuli had a duration of 100ms and was presented in a random order with a variable inter-stimulus interval of 1 - 2s. Each participant was instructed to raise the index finger of their right hand on the occurrence of a rare stimulus.

- **Data Acquisition**

  MEG recordings were conducted using a whole-head neuromagnetometer containing an array of 248 sensors housed in a sound-damped and magnetically shielded room. The magnetic flux measurements were digitized at 290 Hz, bandpass filtered from 0.1-20 Hz and subjected to a noise reduction algorithm that is part of the 4D-Neuroimaging WHS3600 software. The data are P300 wave, which means that

the signal is an event related potential (ERP) component. The ERP P300 has a positive polarity and the latency is approximately 300ms after the stimulus. P300 components may exhibit a peak anywhere between 250ms and 700ms.

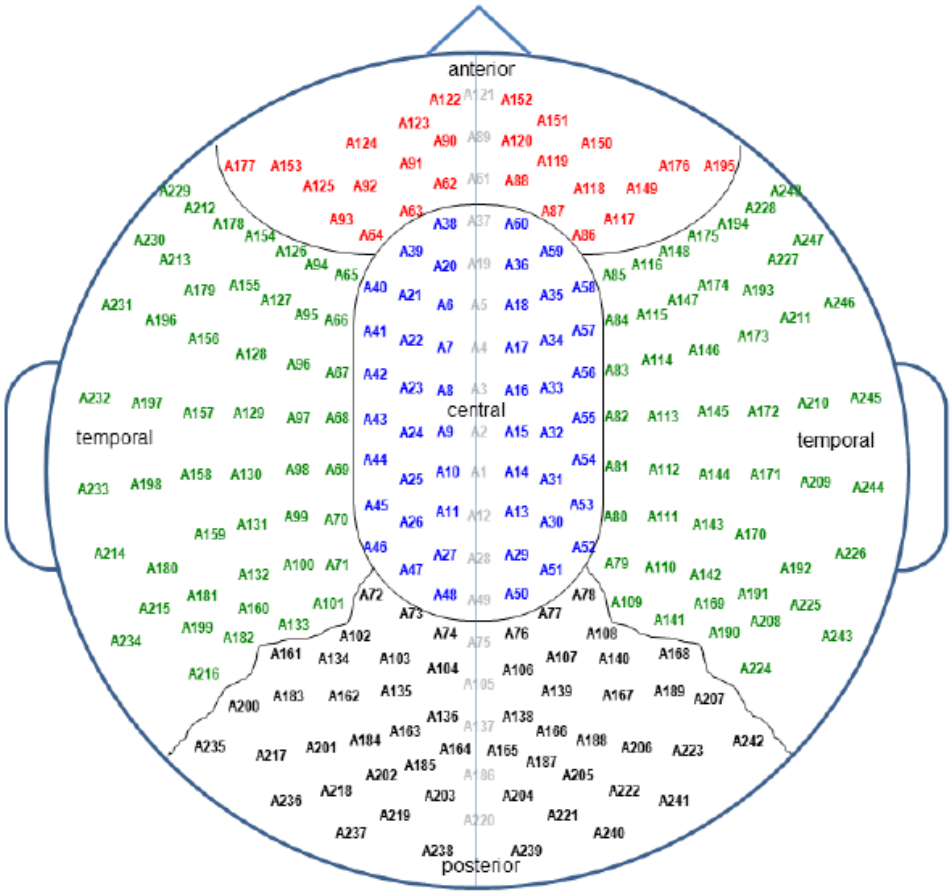

Figure 4.1: Allocation of channels, as in 4-D Neuroimaging, Magnes WH3600 Whole Head MEG Scanner.

The original dataset consists of 40 MAT-files which contain the recordings from the MEG scanner that were taken from ten people who were subjected to the experiment, as it was described above. For each subject, there are two files for the auditory stimulus, one for the frequent brain activity signals and one for the rare brain signals. Respectively, there are two files for the frequent and rare brain activity signals from the visual stimulus. As it was described, the frequent signals come from the normal brain activity when the subject sees a static picture (visual) or when the subject hears a monotonic sound (auditory) and the rare signals are derived from the brain reaction to the change of the projected picture (visual

stimuli) and the change of the frequency of the heard sound (auditory stimuli).

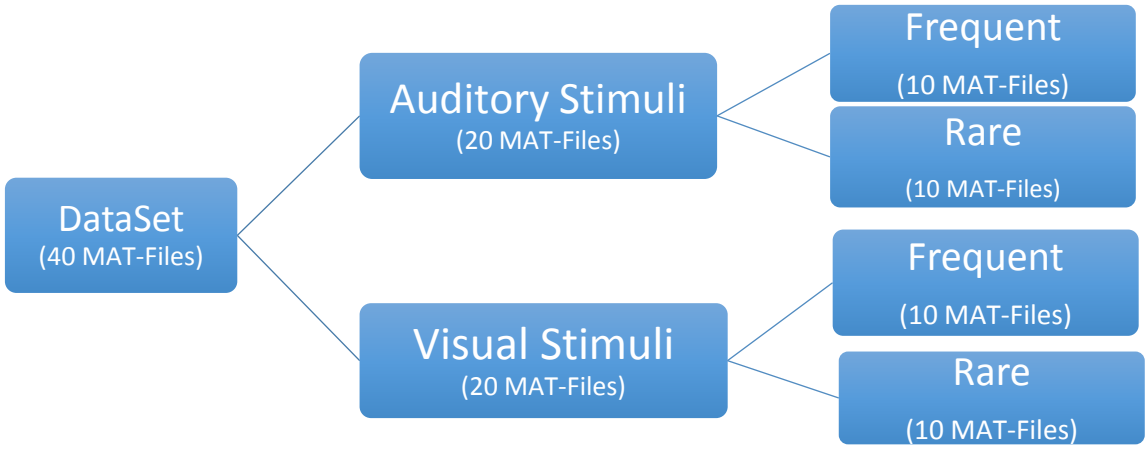

Figure 4.2: Structure of Dataset.

4.2 Pre- Processing

The form of the original MEG data required a pre-processing procedure. This procedure will transform the data in a way that the data will be more easily manipulated in order to extract more precise results. For the purposes of this study certain steps were followed in order for the data to be appropriately processed before the classification into clusters. The followed steps are shown in figure 4.3.

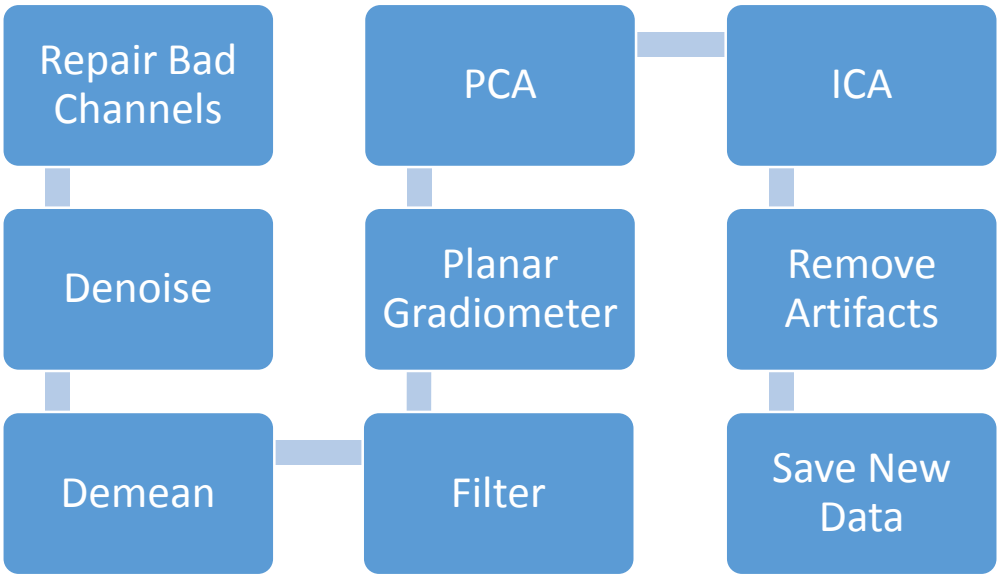

Figure 4.3: Pre-Processing Flow Chart.

Repair Bad Channels

The first step of the pre-processing was to repair the bad channels and construct the missing ones. For that reason, firstly, there were found all the neighbours of the channels based on the triangulation method, which calculates a triangulation based on a two-dimensional projection of the sensor position, in order to find the nearest neighbours. The reconstruction of the bad or missing channels is done by calculating and replace them with the average of its neighbours weighted by distance.

Denoise

Like it was mentioned in chapter 3, the brain activity signals that were recorded from the MEG scanner, contain several artifacts. One of them is the main power line noise, which frequency is at 60Hz due to the fact that our data come from the USA. A notch filter was designed for that purpose and it was applied for the frequencies [59 61] Hz.

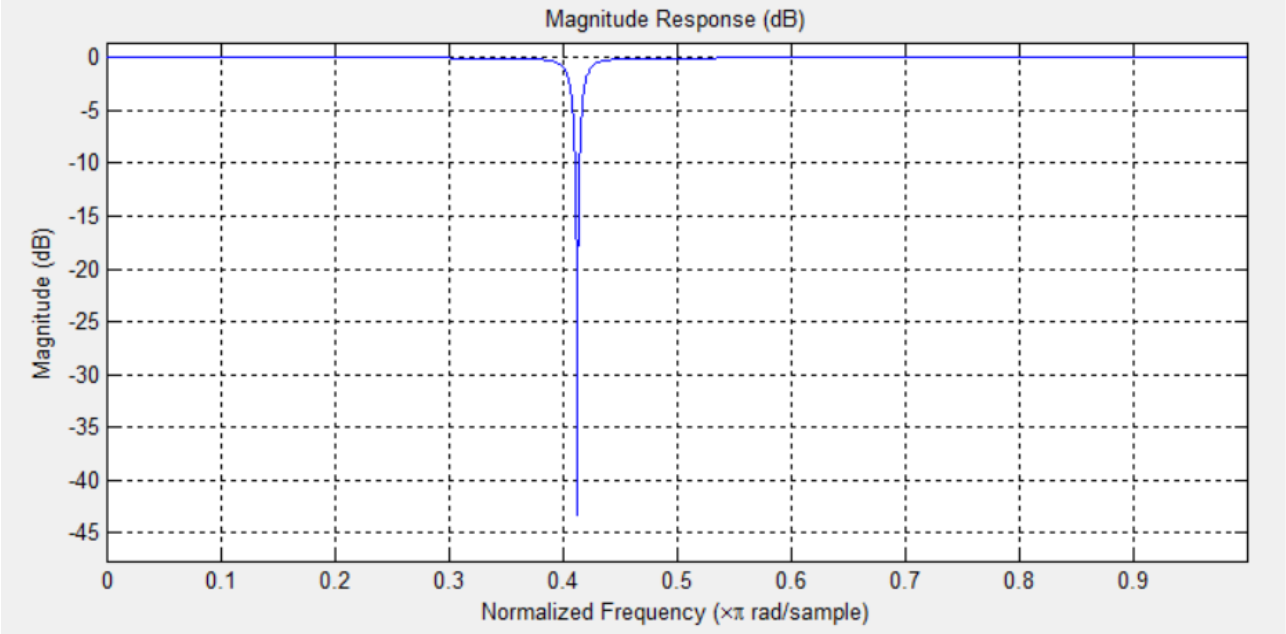

Figure 4.4: Second-order IIR notch filter at 60 Hz or 0.41π radians per sample.

Demean

Demeaning is the removal of the mean value of each trial. In other words it is the removal of the offset voltage from MEG signals

and it was accomplished by the application of baseline window to all trials.

Filtering

The next step that was followed was spectral filtering. The human brain activity frequencies range between 0.5Hz and 100Hz, though for the uses of this study the significant frequencies of the brain activity signal are found in between 0.5Hz and 80Hz. A lowpass butterworth 4^{th} order filter was used in order to reduce the spectral range into the frequency band of brainwaves (delta to gamma) and cut off the unneeded signal frequencies.

Convert to Planar Gratiometer

With the use of Fieldtrip toolbox we managed to convert MEG axial gradiometer to planar.

PCA

The first step for performing PC Analysis was PC squash. Its purpose is to compress MEG data into a principal component subspace. Firstly, the covariance matrix of the channels is computed, in which the 95% of the total variance of our data is used and then the eigenvalues and eigenvectors of the data are exported.

ICA

ICA method was performed for decomposition, with the help of Fieldtrip library and for the purposes of the study we used the Infomax algorithm (runica) as described in chapter 3.

Artifact Removal

Lastly, we had to remove all the trials that contained artifacts in our signal. That was accomplished with the use of the linear measures of skewness and kurtosis, through which the relevant epochs were suggested to be rejected from the analysis.

Below there are some example figures that show the Energy Spectral Density (ESD) of the signals before and after the preprocessing, where we can see how the frequencies are decreased after the application of filtering and denoising methods.

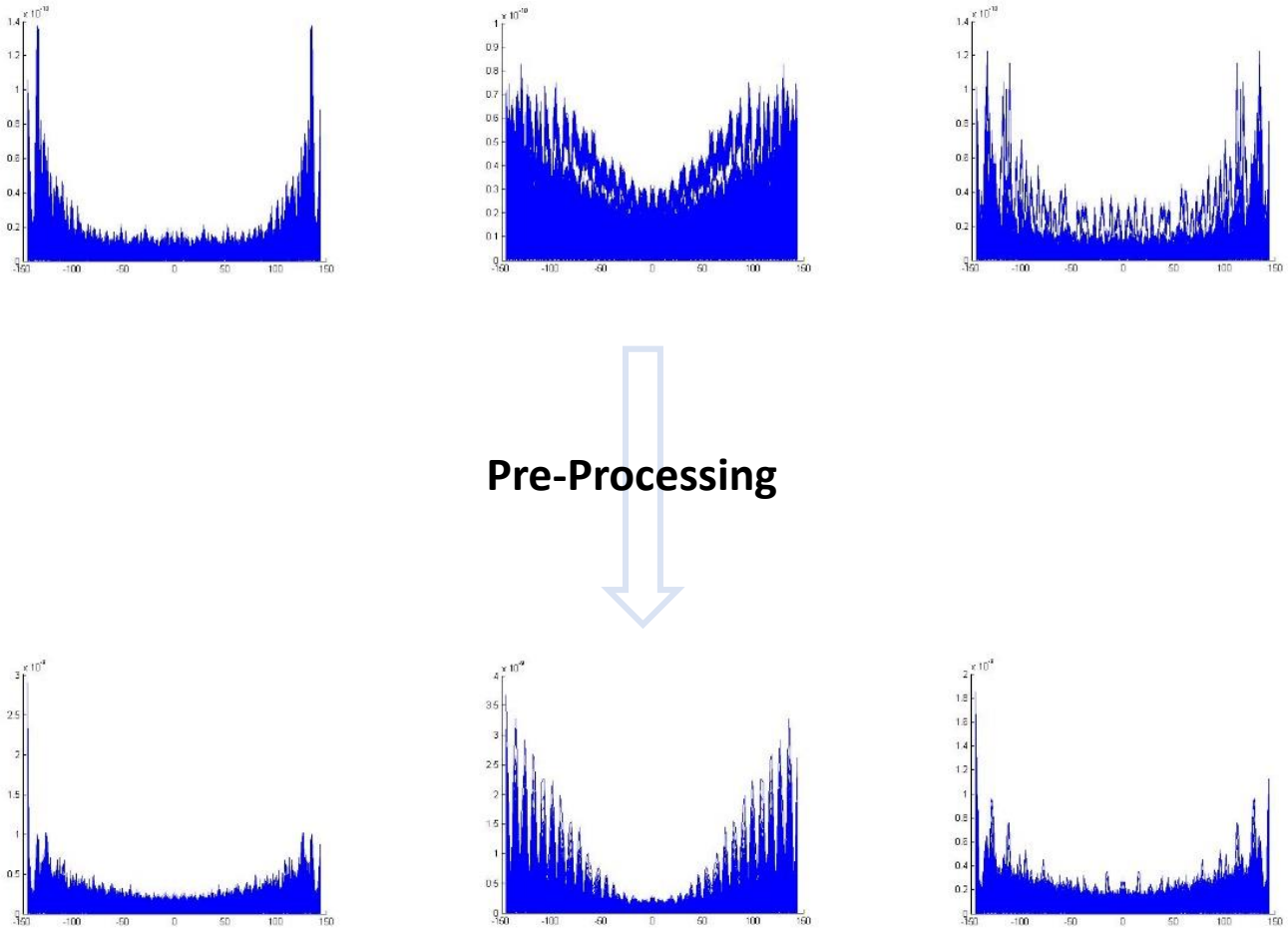

Figure 4.5: ESD before and after data preprocessing.

4.3 Clustering

After following all the necessary steps for the MEG data pre-processing, it was managed to reach a point where all the data that come from the 248 different channels is normalised, denoised and filtered properly. As it was discussed previously, all this procedure was carried out so we may conclude to a clean signal which will contain only pure brain activity, without artifacts.

The next step as to produce a clean brain activity signal without artifacts is the application of clustering. For the means of this study, the algorithm that was used is k-means clustering and the criteria for the artifact removal that were taken into account and used as input to k-means clustering method were kurtosis, energy and PLI (Phase Lag Index). The above measures were extracted from our signal and fed into the k-means algorithm for every brain wave separately and in the whole range of frequencies in total.
The PLI is a measure that is relatively insensitive to the effects of volume conduction and ranges between 0 (no phase locking) and 1 (total synchronization).

The calculation of the k number of cluster was made with the help of silhouette method. The silhouette value is computed from the type $s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)}$ and ranges from -1 to +1. A high silhouette value indicates that i is well-matched to its own cluster, and poorly-matched to neighboring clusters. In order to give the best possible number of clusters, we had to run the algorithm in a loop for a range of k=1:20 for each brain wave individually. Then we evaluated the clusters by the mean value of the silhouette and the chosen number of clusters was that with the biggest mean silhouette value.

The results that we have extracted from the k-means clustering algorithm can be categorized into three groups according to its input, i.e. we took the results with an input of each criterion (PLI, energy and kurtosis) separately and for every brain wave individually.

Below we present the figures with the results of the k-means clustering for every case of metric and brain wave, for visual and auditory stimulus individually. These figures show the different clusters with different colours as presented at the colorbar at the bottom of each head. The clusters being presented on the heads are the one that are marked as the most valuable channels.

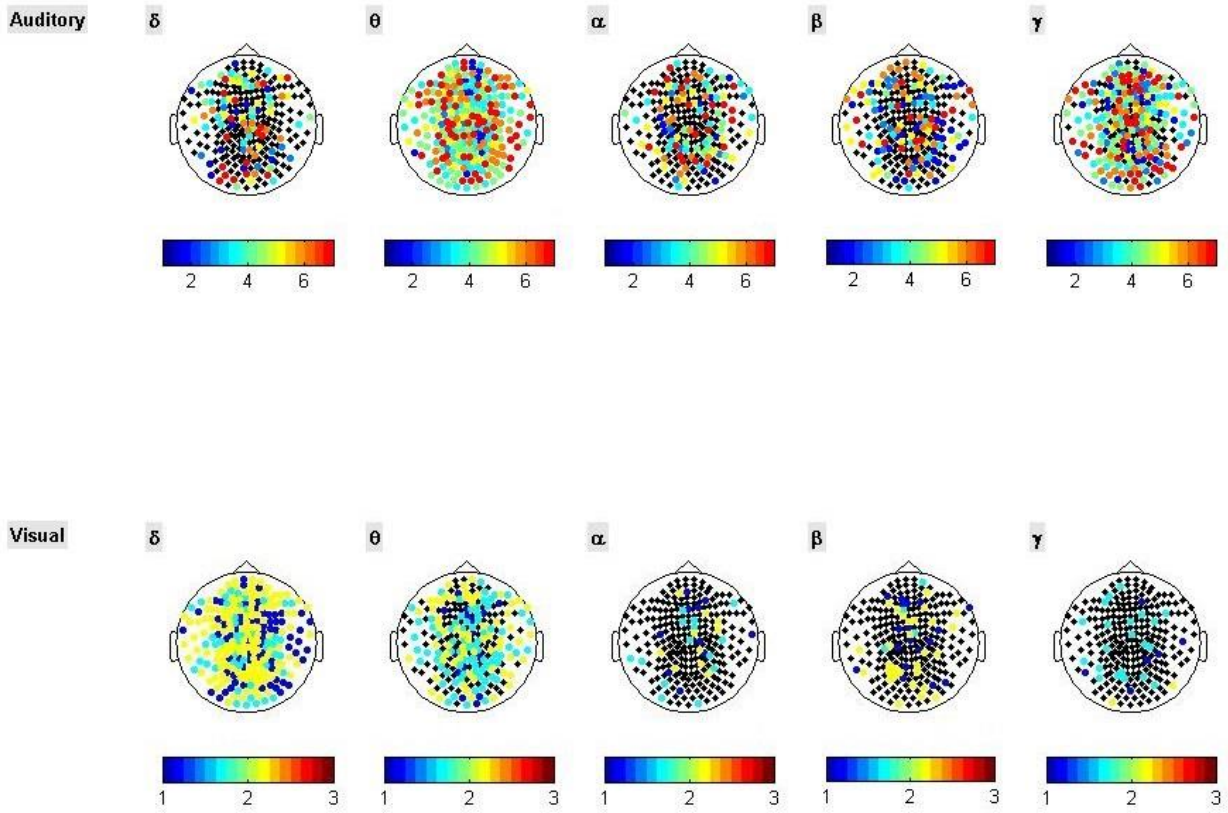

Figure 4.6: Clustering with PLI input.

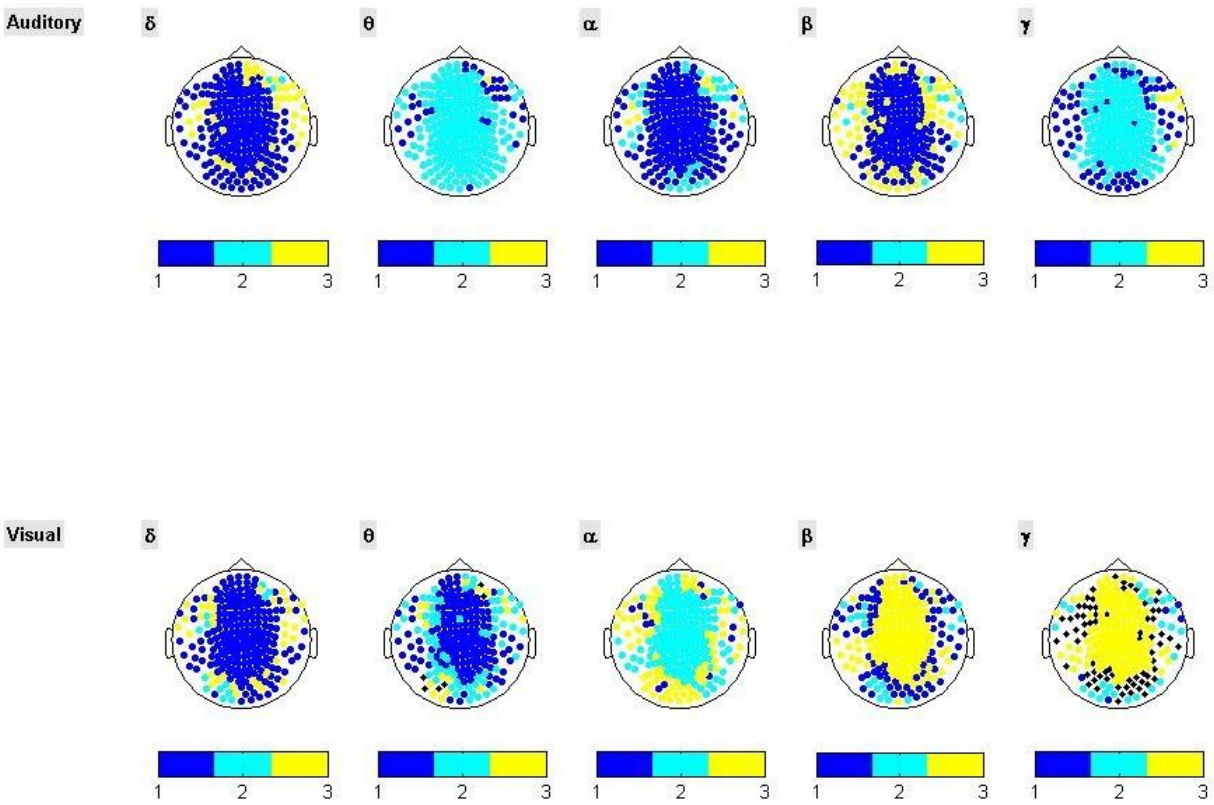

Figure 4.7: Clustering with energy input.

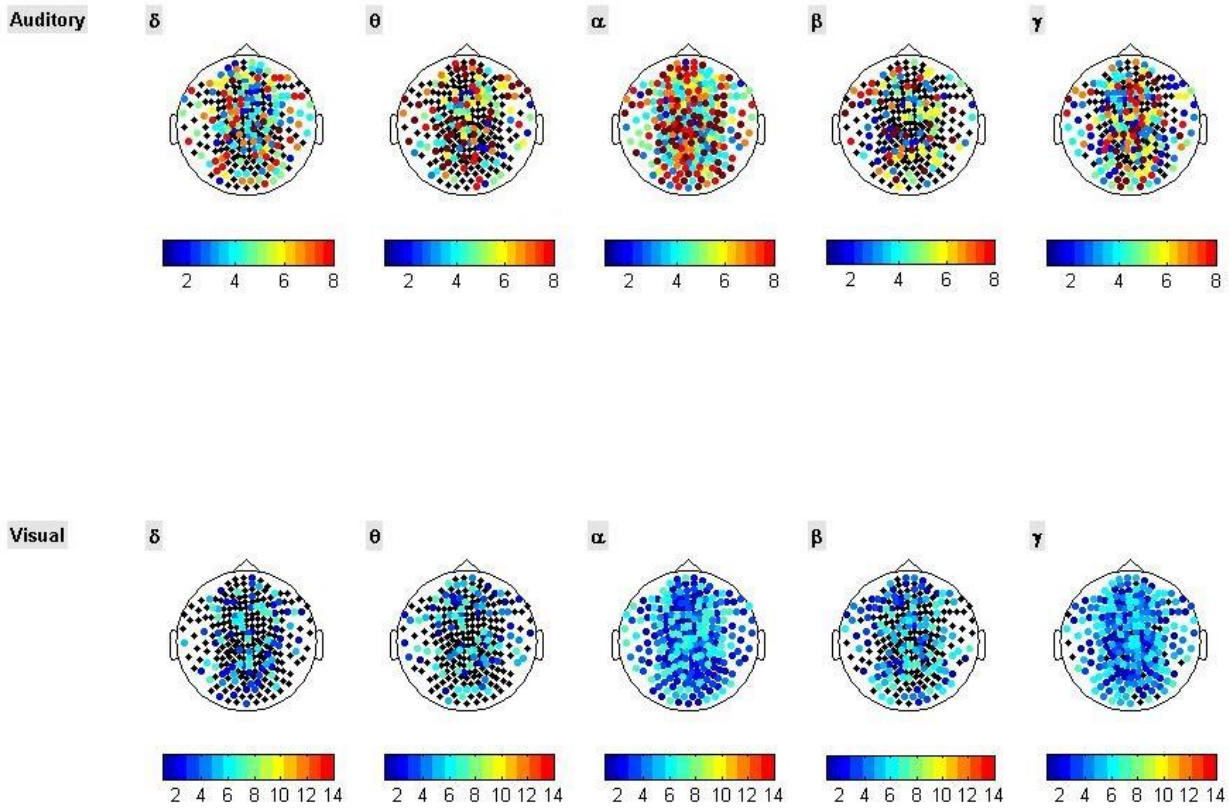

Figure 4.8: Clustering with kurtosis input.

4.4 Statistical Analysis

The next step is to make a statistical analysis of these results. We accomplished this task with the use of lilliefors test, to check if our data follow the normal distribution. If so we continue with a parametric test, else we do a non-parametric test to the data. For that reason we have set the alpha value equal to 5% and compare that to the p-value.

Below we present the figures with the results of the statistical test for every case of metric and brain wave, for visual and auditory stimulus individually.

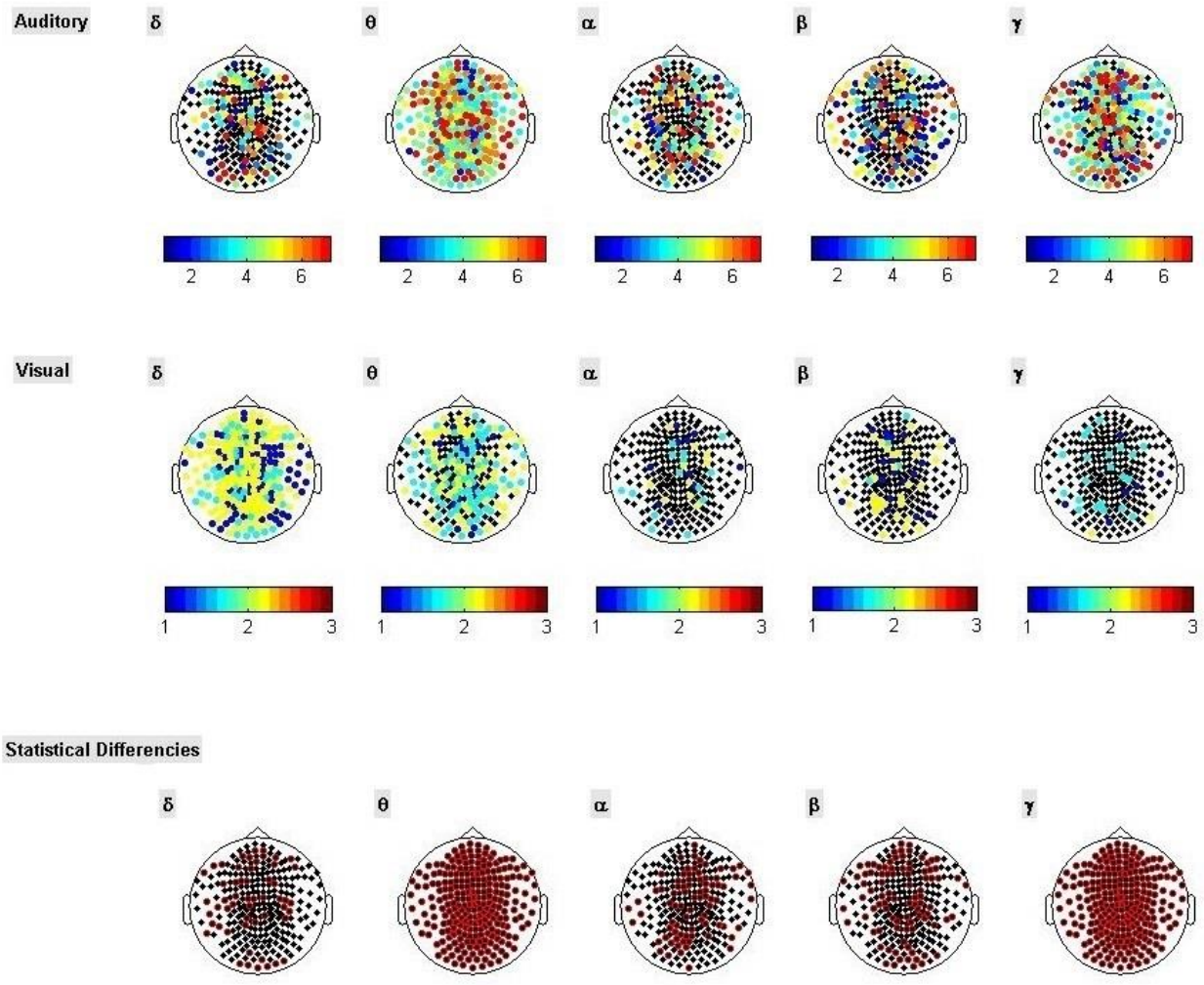

Figure 4.9: Statistical differences for PLI.

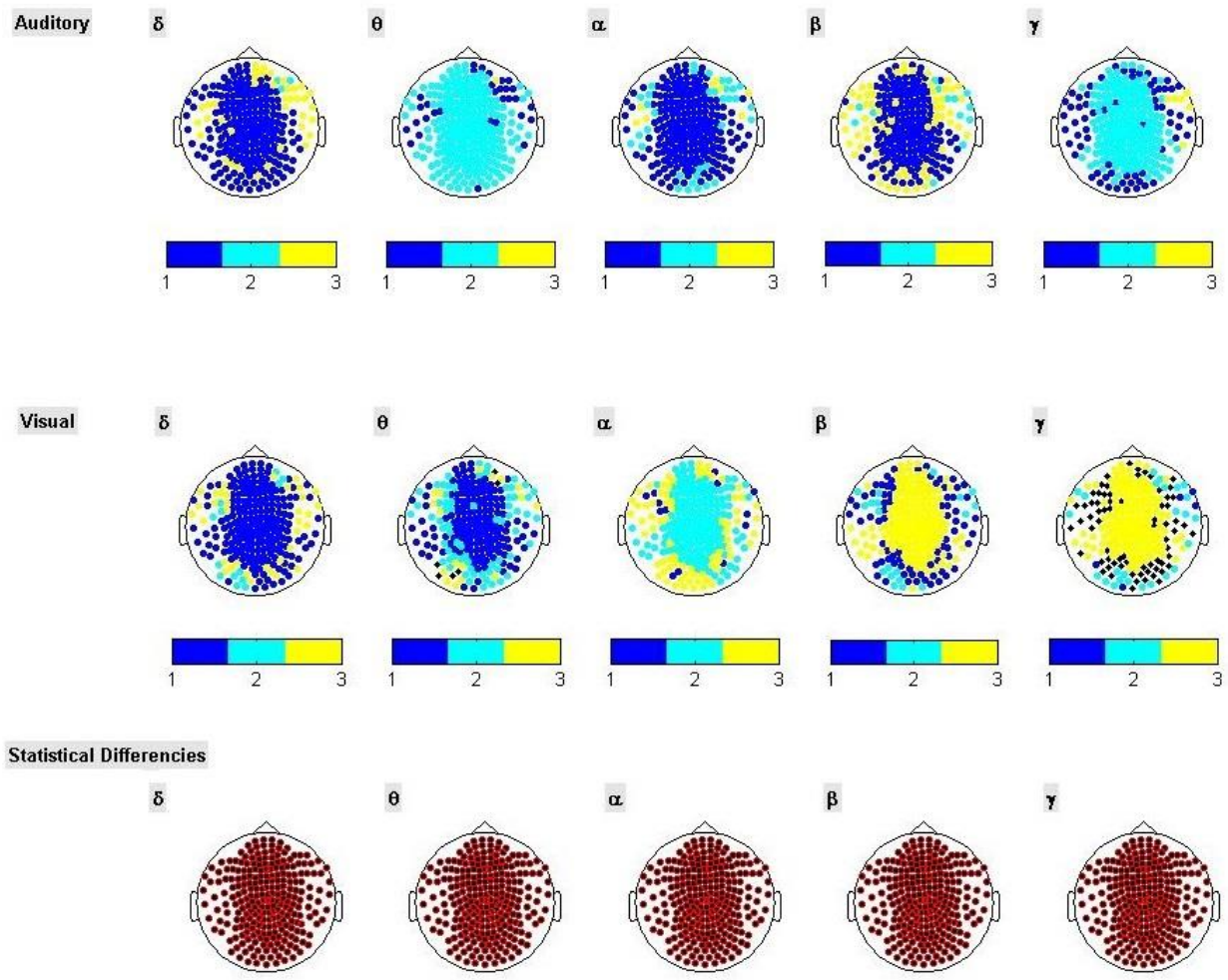

Figure 4.10: Statistical differences for energy.

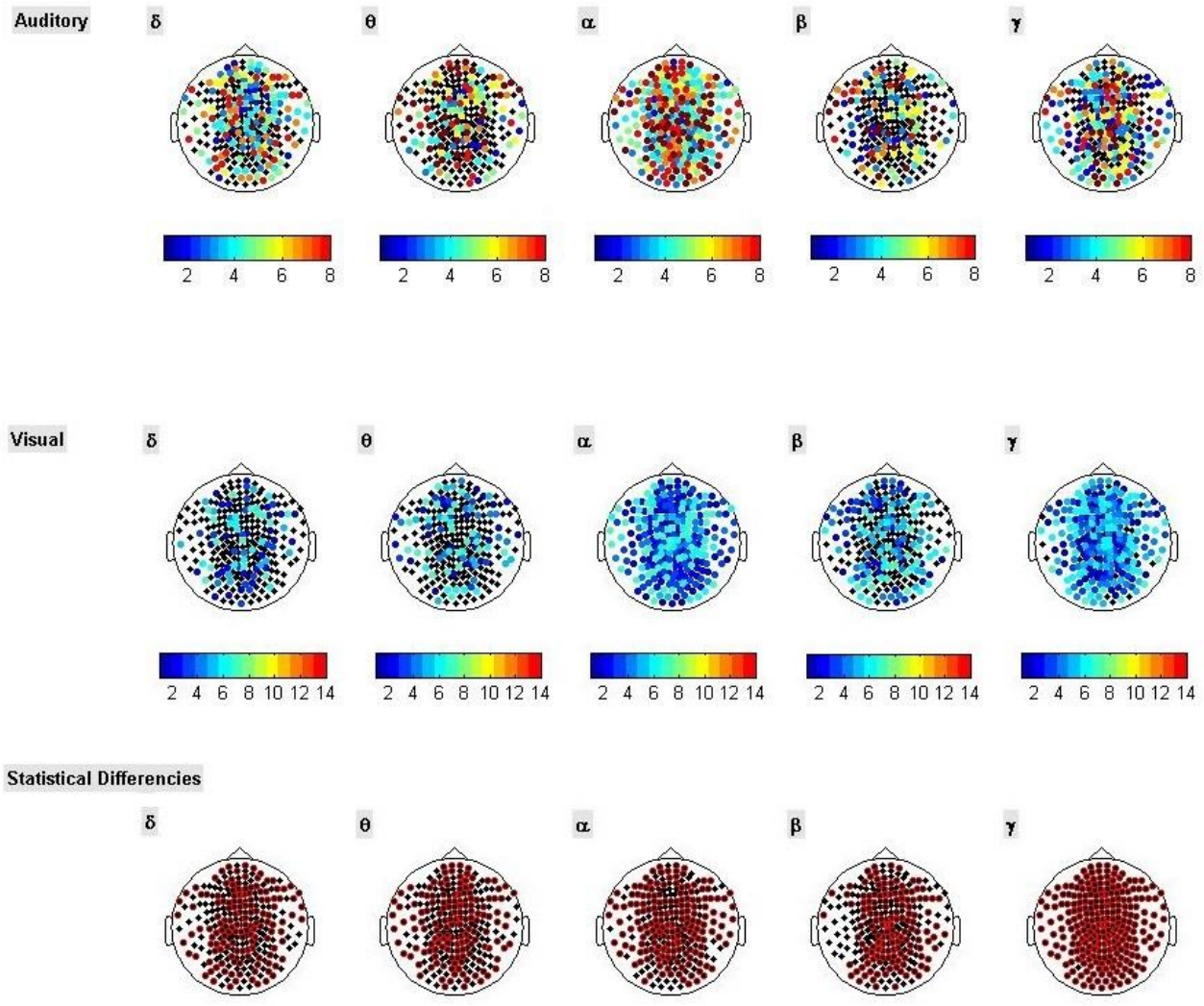

Figure 4.11: Statistical differences for kurtosis.

5 Conclusion and Future Work

5.1 Conclusions

With the process and analysis of MEG data in this diploma thesis, we resulted in some interesting conclusions. First of all, ICA method is a very strong computational method. Its role in this study is to separate our MEG multivariate signal into its additive subcomponents, which contain either noiseless brain activity, or non-cerebral activity, or combination of these signals.

K-means clustering showed the significant clusters in which MEG channels were classified in. From the bibliography we know that the most essential channels for the auditory stimulus are located to both sides of the brain and also the most essential channels for the auditory stimulus are located to the rear side of the brain.

Last but not least, the statistical significant test that was taken concluded to the assumption that our data reject the null hypothesis, so there is ground to believe that there is a significant relationship between the two groups of data –signals taken after visual stimulus and signals taken after auditory stimulus. Signals from auditory stimulus are known to be in front of the signals from visual stimulus. So it is reasonable that we have so many statistical differences in our dataset and with the specific threshold that was used in this study.

5.2 Future Work

In the effort to analyse and process MEG data for an outcome of conclusions which come to further understanding of the brain activity, there are many methods and different parameters to examine.

In this study we tried to cluster the channels of the brain, with a purpose to extract information about the significant channels that may contain most valued information about brain activity after visual or auditory stimuli. Also, another purpose of the study was to detect and exclude all the outliers or artifacts of the signals that are extracted out of a MEG. For this aim, we took advantage of methods as ICA and k-means algorithm for clustering, with the help of the metrics of PLI, energy and kurtosis.

For the artifact removal, alternative methods could be used, such as Kalman filter, or the use of topologies and neural networks.

Another procedure that can be introduced as a way of channel clustering may be the Hierarchical as well as the Two-Phase algorithms. The advantage of these methods over the k-means algorithm is that they come to more stable results, as the k-means method concludes to different results for every run of the algorithm.

REFERENCES

[1] "Brain facts and figures".https://faculty.washington.edu/chudler/facts.html
[2] Human Brain: Facts, Anatomy & Mapping Project by Tanya Lewis, Staff Writer
[3] National Institute of Neurological Disorders and Stroke, NIH Publication No.11-440a
[4] Taylor, Tim. "Respiratory System." *InnerBody.com*. HowToMedia, Inc., Oct 2012. Web. 14 Nov 2013. < http://www.innerbody.com/anatomy/respiratory>.
[5] J. Bronzino., The biomedical engineering handbook. Springer, Boca Raton, FL, U.S.A., 3rd, edition, 2006
[6] Carlson, Neil R. (2013). Physiology of Behavior. Upper Saddle River, NJ: Pearson Education Inc
[7] Antonakakis Marios, MEG Data Analysis with use of ICA, 2013
[8] I. Tal, M. Abeles, Cleaning MEG artifacts using external cues, doi:10.1016/j.jneumeth.2013.04.002
[9] Michael Ullman, ELECTROENCEPHALOGRAPHY/EVENT RELATED POTENTIALS (EEG/ERP) LABORATORY