POLITECNICO
DI TORINO

Corso di Laurea Magistrale in Ingegneria Biomedica

Automatic detection and study of Slow Biphasic Complexes (SBC) in electroencephalographic signal (EEG) of encephalitis patients

Relatore
Prof. Luca Mesin

Correlatore
Dott. Massimo Valerio

Laureanda
Agata Barbagini
matr. 262474

Dicembre 2020

Abstract

This dissertation focuses on an algorithm development, improvement, validation and test. The aim of the algorithm is automatic and early diagnosis of encephalitis.

Encephalitis is a parenchyma infection of the brain, which can have deleterious consequences if not promptly diagnosed. An innovative non-invasive diagnosis approach consists in EEG signal subjective evaluation by an expert physician with the purpose to detect definite slow waves, called Slow Biphasic Complexes (SBC). It has been demonstrated that the appearance of SBCs in EEG is related to the onset of pathology.

Our goal is the development of an algorithm able to recognize SBCs by itself. In this way, the diagnosis would be faster, objective and based on a completely non-invasive technique.

A pre-existing algorithm was the starting point of this study. Its functioning is based on thresholding of cross-correlation function computed between EEG signal and a prototype wave. Applied on a signal of high complexity, the same simple approach aims to emulate physicians' mental processes, which lead to the discrimination of waves only by observing them thanks to decades of experience.

The main body of research was developed following this thinking: we collaborated strictly with doctors, analysing data they provided, so to extract and quantify doctors' knowledge and to insert it in the research algorithm. After that, the algorithm has been improved and a post-processing has been implemented to distinguish between true and false positive waves.

In the following dissertation, three main chapters are presented: the first one is bibliographic and concerns electroencephalographic signal and slow biphasic complexes characteristics; second chapter describes methods used for the extraction of knowledge and approaches for algorithm improvement; the last chapter presents the results obtained concerning an accurate characterization of SBC, validation and testing of the algorithm.

SBCs' characterization allows to highlight interesting peculiarities of waves. Neither physicians were aware of some of these properties, they were only considered unconsciously; an example is the one related to power spectrum frequency of waves. The improved algorithm has been validated with the help of the doctor, demonstrating how it is able to discriminate adequately effective SBCs and waves that simply present a similar waveform. Finally, it has been tested for the recognition of healthy and pathological patients and for the identification of disease severity degree.

Contents

Introduction

This thesis focuses on the study of Slow Biphasic Complex (SBC), which are particular waves in electroencephalographic (EEG) signal, indicating encephalitis in the patient.

Encephalitis is a parenchyma infection of the brain, which can have deleterious consequences, if not promptly diagnosed. Encephalitis can occur in acute, subacute or chronic form. The aetiology is extremely varied: the disease can be caused by a viral infection or by an autoimmune disease [19]. The acute phase of the illness can provide headache, disorientation and state of unconsciousness; Patient presents these symptoms within a period of days or few weeks [1]. In the most serious cases, the situation can degenerate into necrosis of gray and white matter, respiratory arrest, coma and eventual death. In most cases, patients report neurological sequelae [2].
Encephalitis occurs at any age, but it is more common in children under three years of age and in elderly people, with an incidence in children about 10 per 100,000 [1, 2, 19]. This thesis will focus on childhood Encephalitis.

To avoid the worst consequences an early diagnosis is extremely important, firstly because most times this disease is initially confused with simple flu. The first step of the diagnosis is undoubtedly a subjective examination in order to assess the level of consciousness of the patient. Subsequently, if there is suspicion of encephalitis, the patient passes a blood count test, coagulation analysis, kidney and liver function tests, imaging using MRI (Magnetic Resonance Imaging) or CT (Computed Tomography) and a lumbar puncture which consents to detect the inflammation of cerebrospinal fluid [1, 19].These procedures are invasive and not comfortable for the patient.
Furthermore, there are several diseases that mimic encephalitis (such as meningitis or brain abscess) so the probability of errors in diagnosis is quite high [19].

A new, easier and non-invasive approach is required, and the best strategy could be the observation and evaluation of the EEG signal, whose acquisition is much more accessible than the currently used techniques and it provides detailed information on the state of the brain. In most encephalopathy cases (which include any disorder of the brain), EEG presents particular features and waves with specific shape, amplitude, and/or duration [17] .

Clinical evidence suggests that encephalitis can be identified by observing the presence and the occurrence of some specific waveforms in the EEG signal. These waves are called Slow Biphasic Complexes (SBC). Furthermore, it has been proved by pieces of evidence that there is a relation between the presence and the occurrence of SBCs and the clinical evolution of the pathology [8].
Today, clinicians detect waves through subjective EEG analysis, but this kind of procedure could be affected by a bias and it requires long time and resources. That is why automatic signal processing and slow biphasic complexes (SBC) identification would be a great improvement. It would allow the analysis of large amounts of data and to obtain information about the complexity or consistency between different channels, which could hardly be captured by visual analysis. Moreover, automatic methods make it possible to obtain quantitative and objective measurements [19, 13].
To this purpose, an algorithm has been developed, which intends to investigate the EEG signal, in order to detect Slow Biphasic Complexes. It uses a prototype waveform and compares the EEG signal using a cross-correlation threshold [13].
First part of this research concentrates on the study of some waves individuated by an eminent expert in the field, aiming to characterize SBCs in the most trustworthy way. The purpose is to construct the best wave prototype and to introduce some post-processing controls in the existing algorithm, to improve the automatic identification of SBCs.
Subsequently, thanks to the information extracted, the algorithm has been improved in order to obtain a better ability to discriminate waves.
The enhanced algorithm has finally been validated and tested, in order to prove its ability to distinguish between healthy controls' EEG signals and encephalitis patients' ones, characterized by different severity degree of the pathology.

Chapter 1

The EEG signal

> *Feeble currents of varying direction pass through the multiplier when the electrodes are placed on two points of the external surface, or one electrode on the grey matter, and one on the surface of the skull.*
> Richard Caton (1842 – 1926), Liverpool

Richard Caton was a British physiologist who firstly studied the electrical nature of the brain and performed some EEG measurements, presenting his founding on August 24, 1875, to the British Medical Association [16]. With this publication we fix the birth of the *electrophysiologram*, because that probably was the very first time that a needle moved thanks to the electrical energy produced by EEG phenomena [16].
Artifacts had presumably a greater amplitude than the actual signal; the used galvanometer was very rudimentary, and it had a very restricted frequency response, ranging about from 0 to 6 Hz; despite of all of this limitation, Canton is considered the discoverer of EEG fluctuating potential [16].
To observe a human EEG acquisition one has to wait until 1920, when Hans Berger (1873-1941), who was a German neuropsychiatrist, started his studies and was able to record the first one-channel human EEG, made on photographic paper and lasting one to three minutes. With this recording, alpha waves (8-13 Hz frequency band) and smaller amplitude beta waves (13-30 Hz frequency band) were described for the first time.[16].

Today we use electroencephalography to measures the spontaneous activity of neurons (mainly pyramidal cells) placed in the cerebral cortex, by placing electrodes on the scalp, with a non-invasive and low-cost method.
EEG measures total ions flux between intracellular and extracellular space, which constitutes excitatory and inhibitory synapses (respectively positive and negative flux) [12].
Essential potentials that can be demonstrated with intracellular recordings are characterized briefly. The membrane of a cell body, in its resting state,

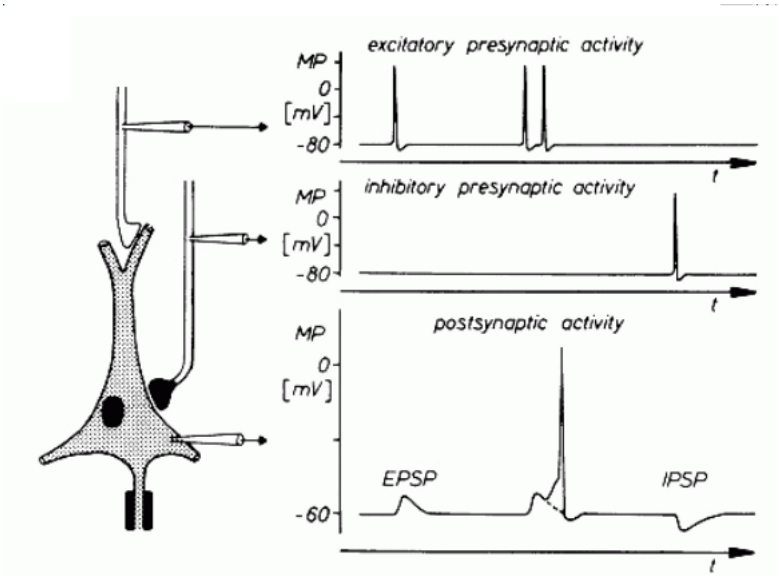

Figure 1.1: *Membrane potentials variation during synaptic activity. Membrane potentials are recorded using intracellular microelectrodes placed in presynaptic fibers and in postsynaptic neuron. A summation of two EPSPs is reported, with the resulting crossing of the threshold and generation of an action potential in the postsynaptic neuron.* [16]

presents a negative potential of about 60 to 70 mV in respect with extracellular space.
Synaptic activities affect the variation of this potential. The action potential travels along the presynaptic fiber, which can end either with an excitatory synapse or with an inhibitory synapse. In this way, an excitatory/inhibitory postsynaptic potential (EPSP/IPSP) occours in the following neuron. If two or more action potential travel along the same fiber during a short period, a summation of postsynaptic potentials will be observed. With an EPSP the membrane potential will become less negative, while with an IPSP a hyperpolarization will be observed (membrane potential becomes more negative). When the postsynaptic potential reaches the membrane threshold, a new action potential takes place, in the postsynaptic neuron, allowing action potential propagation [16].
This process is summarized in figure 1.1, which represents the potentials measured with invasive intracellular and extracellular electrodes [16].

Signal amplitude will be greater the more synapses activate synchronously [12].
Analysing EEG signal is not trivial. A lot of factors have to be taken in account [12], such as the wave occurrence and their frequency, the distribution of the waves in the hemispheres and their synchrony (EEG can be

symmetrical or asymmetrical), the signal amplitude and bandwidth, which depends on the synchronization degree of neurons. Furthermore, the expert who analyses the signal has to take in account the age of the subject and his state (awakeness, drowsiness, concentration), to better understand waves in the signal [16].

The importance of computational techniques in EEG waves analysis was clear very early in history, starting from 1932, when neurophysiologists and physicists firstly applied Fourier analysis to the signal. But computational techniques had a strong improvement with the computerization, starting from the 1970s [16]. Nowadays we are still working on it, with the awareness that EEG is far too complex for a complete automation, and a subjective point of view of an expert is always needed.

1.1 Properties of EEG

Brain waves are made up by undulation we can observe in EEG signal. In order to interpret, classify and finally understand them, an expert has to consider several aspects [12].

Amplitude. Waves amplitude is the most immediate aspect: it ranges from 10 to 500 μV, as regard non-invasive acquisition; even though waves recorded directly on the cortex surface (invasive method) can reach an amplitude of 10 mV [12].

Morphology. Wave morphology concerns their repeatability [12]:

- Monomorphus waves appear repetitively, presenting the same frequency and amplitude;
- Polymorphus waves are found in a signal that maintains the same frequency, but it is not periodic and changes its amplitude.

Topography. Waves topography relates their distribution in space (i.e. similarity in the two hemispheres). The EEG can be [12]:

- Symmetrical, if waves have the same characteristics on the two hemispheres;
- Asymmetrical, if a particular wave appear only on one side.

Frequency. All of previous aspects are detected in a time domain analysis. In such domain EEG signal is quite completely stochastic. While it is possible to extract very important knowledge about synchronization of

neurons activity by analysing the signal in frequency domain. Observing EEG spectrum, it is possible to identify different waves:

- Delta waves - frequency 1-4 Hz. Those are the slowest waves with biggest amplitude (75-200 μV). They are typical in infants and coma patients, while they indicate cerebral injury in awake adults [12].
- Theta waves - frequency 4-8 Hz. Theta waves are common in infants and coma patients, while unlikely in awake adults. Important theta rhythm is typically recorded in the hippocampus, during REM sleep, suggesting that these frequencies involve memory process [16].
- Alpha waves - frequency 8-13 Hz. Alpha waves were the first ones to be observed, as they are the most prevalent waves in closed-eyes and relaxed adults. These frequencies are observed during resting or in inactivated parts of brain during a task [12, 16].
- Beta waves - frequency 13-30 Hz. Beta waves are typically found in the frontal regions during tasks requiring concentration [12].
- Gamma waves - frequency >30Hz. Those waves are generally associated with an enhanced level of vigilance or concentration. Recent studies assumed that gamma oscillations could help a complex mechanism able to connect activities of spatially separate cortical areas [7]. Gamma rhythm is more easily recorded by the invasive procedure of electrocorticography (ECoG) rather than EEG [12, 16].

1.2 EEG recording

It is possible to record EEG either by placing electrodes on the patient's scalp (non-invasive scalp EEG) or inserting needle electrodes in cerebral cortex (invasive intracranial EEG). Those two measurements provide different informations and have their own advantages and disadvantages. Scalp EEG is surely more affected by artifacts, and it provides knowledge about the electrical activity generated by a wide population of neurons. Intracranial EEG has a greater amplitude, and it is less affected by physiological and non-physiological artifacts; it acquires the electrical signal generated by a single neuron or by a small population of those [12].
This thesis will focus on non-invasive EEG acquisition, which is the best way to achieve the purpose.

Electrodes and transfer chain for signal acquisition. Electrodes most widely used for scalp EEG acquisition are platinum cup electrodes, 5-10 mm diameter, filled with conductive gel, used as an ionic solution [12]. Electrodes constitute an interface between human tissue and wires, giving

the possibility to transduce a movement of ions in a current made by electrons flow [16].
Starting with electrodes, it is possible to observe the transfer chain, used to acquire EEG signal. In the following, the basic one is presented [16]:

- Electrodes;
- Amplifier, with the primary aim of magnifying the signal from microvolt to several volts (100,000 or more gain); other than the sufficiently high gain, the amplifier must present:
  - linearity across the entire frequency range of signals,
  - high common-mode signal rejection (CMRR), which summarizes its ability to pass differential-mode signals over common-mode signals,
  - high input impedance;
- Low-pass filter, which is useful to block those frequencies that are constituted mainly by noise, and to solve the eventual aliasing problem;
- High-pass filter, which is useful to clean some artifacts (such as movement artifact).

As regard digital acquisition of signal, an analog-to-digital converter (ADC) is needed [16].

The 10-20 International System. A new standard international electrodes placement was proposed in 1949, and it is still recommended by the International Federation of Clinical Neurophysiology. It is called 10-20 International System, and it is a protocol that aims to standardize electrodes positioning over different exams and laboratories [12]. It defines electrodes locations taking in account skull size and shape, providing a full head and physiologically important cortical areas covering [11].
Two basic anatomical measurements have to be done: the Nasion-Inion distance, and the ear-to-ear distance. These distances are divided into equal parts and the electrodes are placed every 10% or 20% of the lengths [12]. Each electrode position is coded with letters and numbers. The first letter shows the interest zone of the brain. The subsequent number indicates the hemisphere and the anterior or posterior position.
Letters:

- F is frontal lobe
- P is parietal lobe
- O is occipital lobe

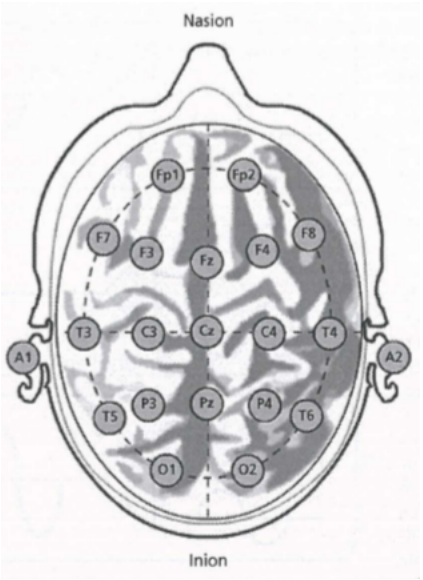

Figure 1.2: *10-20 International System electrodes placement* [11]

- T is temporal lobe
- C refers to central electrodes (placed on brain centerline)
- FP refers to electrodes place in frontal position, above eyes
- A is ear lobe

Numbers:
Even numbers refer to right hemisphere, while odd numbers refer to left one. Numbers increases from anterior to posterior of the head. In case of electrodes place in the middle, the number is substituted by a "Z" letter. At the end, 10-20 International System requires the placement of 21 electrodes [12].

Montages. Any electrophysiological record requires input from two different electrodes, to make up an EEG channel. Depending on how it is performed, it is possible to recognize different montages. Each one has its specific spatial filtering characteristics.

In the following, main montages are listed, ordering them from the lowest to the greatest in filtering of widespread fields [16]:

- The common reference montage (also called unipolar or monopolar setup): each electrode potential is referred to the same one. The reference electrode can be place on the earlobe (A1 or A2) or in some other 'neutral' position. In this way it is produced a simple view of

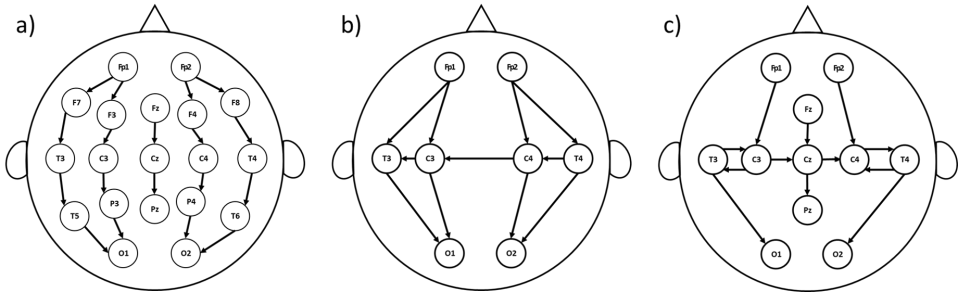

Figure 1.3: *Three principle montages used in this study. Each arrow indicates the couple of electrodes involved in the subtraction: to the electrode from which the arrow starts will be assigned a positive sign, to the one to which the arrow arrives will be assigned a negative sign.*
a) *Montage 3, longitudinal bipolar montage, most widely used for adults, it covers 19 electrodes;*
b), c) Montage 9 and 2, called reduced montages, suitable for children.

the scalp field, which allows a great overview of potentials, artifacts, symmetry over hemispheres and other general characteristics [11, 12, 16].

- The average reference montage: each electrode potential is referred to a linear combination of all other electrodes on the scalp, in particular mean value. It fits when a high amplitude and very focused activity has to be detected, but it loses information in case of widespread activity [11, 16].

- The bipolar montage: it is the most widely used, and it performs substruction between couple of subsequential electrode's potentials. It is possible to distinguish longitudinal and transverse bipolar montages, usually used together, as the voltage field can be asymmetric along anterior to posterior or transverse axis. It is useful to localize focal activities [11, 15, 16].
  Three widespread bipolar montages are graphically represented in figure 1.3.

- The Laplacian montage: it makes the linear combination with each electrode and the immediately surrounding ones. In this way each channel has a different reference. It is used to detect focal disorder on EEG [15, 16].

As we will see later, bipolar mounting will be the most widely used for the identification of specific waveforms, useful for this study. We focus on three particular bipolar montages: a longitudinal one that involves 19 electrodes (all 10-20 System electrodes except ear lobe ones) and other two, called

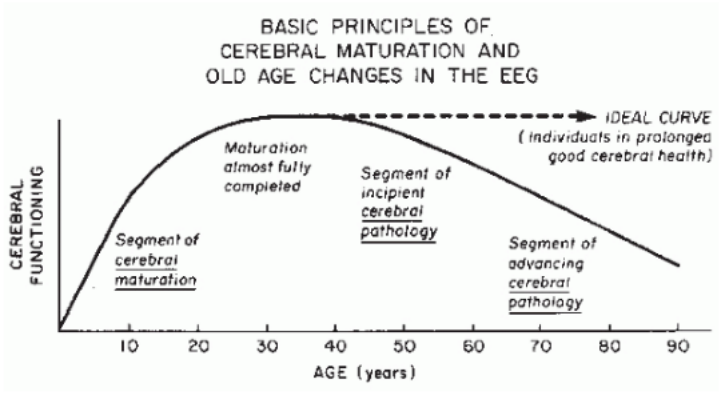

Figure 1.4: *Cerebral maturation on respect with age.* [16]

"reduced montages", that involve a lower number of electrodes, and are suitable for infant and children, whose cranium is smaller. With children, it is not possible to apply all the electrodes to the scalp. The electrodes that must in any case be positioned to ensure reliable acquisition are: Fp2, Fp1, C4, C3, T4, T3, O2, O1 and eventually Cz [11, 15].
Figure 1.3 shows the three montages.

1.3 Infant to adolescent EEG

As widely explained in manual *"Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields"* by Schomer and Lopes Da Silva [16] and reported in figure 1.4, aspect of EEG changes throughout whole life, as it undergoes a maturation process and then degradation caused by to cerebral pathology or age. Because of that it is very important to take in account the age of the patient, while analysing an electroencephalographic acquisition.
In this study, adolescent or younger patients have been considered.

EEG signal of a less-than-one-year infant is characterized by a posterior dominant slow rhythm, with a medium-high amplitude (50 to 100 uV) and a frequency ranging from 5 to 8 Hz (theta waves) [16].
With respect to pre-scholar age children, it is possible to observe an increasing in frequency of posterior dominant rhythm, settling in the lower alpha range during awakens. The older the child is, the most alpha rhythm is predominant over theta rhythm. In this case alpha waves can reach 100μV amplitude. Conversely, in drowsiness and sleep, desynchronized theta waves are still present and dominant. During this period of time, it is possible to recognize eye blinks as biphasic potential of medium to high voltage in

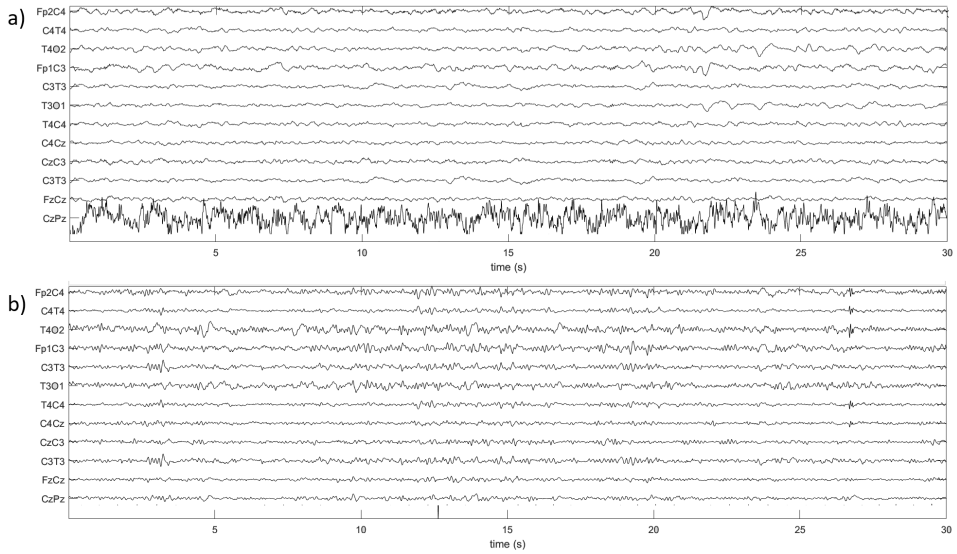

Figure 1.5: *Comparison between 30 seconds EEG signal of an awake 11 months old infant (a) versus 30 seconds EEG signal of an awake 15 years old adolescent (b).Both signals are acquired using montage 9.*

bilateral synchrony over occipital areas [16].
Considering older children and adolescent, dominant posterior alpha rhythm reaches a mean frequency of 10Hz, which is comparable with adult EEG mean frequency. Slow activity in posterior area still distinguishing [16].

1.4 EEG artefacts

EEG signal is easily corrupted by several artefacts. Since its amplitude is generally low, it is not unusual that an artefact completely overlaps the signal.

Artefact is anything recorded with scalp electrodes that is not generate in the brain. Artefacts can be physiological, if their origin is non cerebral but it is still inside the patient's body. Otherwise they are non-physiological, if they are generated from outside the body. [12]
More common artefacts are itemize below.

Physiological artefacts

- ***Ocular artefacts.*** Those artefacts generate because the eye constitutes a dipole, formed by the cornea, which is the positive pole, and the retina, which is the negative pole.
  Moving the eye or blinking cause a generation of an electric field which

is visible on EEG signal. The blink artefact is far predominant with respect to eye move one.
In particular, when eyes are opening the eyeball rotates upwards, the electropositive cornea nears the front electrodes, generating a positive peak on the signal (downwards). When eyes close vice versa happens, and a positive peak is visible on the signal. Peaks caused by blink are for sure more prominent in signal that involve frontal and temporal channels, which are the ones nearest to the dipole that origins the artefact.
Hence a blink will cause a bi-frontal potential with an amplitude ranging from 100 to 200 μV, which appears 20 times a minute if the patient is awake [11]. During drowsiness slow lateral movement of the eye are usual, producing an alternate oscillation and a reduction of alpha rhythm [12].

- ***Glossokinetic artefact.*** The tongue, as well, represents a dipole, positive on the tip and negative on the base. Moving it or chewing produces a small artefact that can be observed on frontal and occipital channels [12].

- ***Cardiac artefact.*** ECG signal is slightly recorded also by scalp electrode and, as it detected almost in the same way by all EEG electrodes, it reduces importantly if bipolar montages are used. In fact, single differential acquisition cuts off common mode potentials.
  Furthermore, this kind of artefact is easy to detect, as it appears synchronously with ECG signal potentials.
  Another artefact related to cardiac activity occurs if an electrode is placed over a vessel, that pulses according to the EEG rhythm. In this case artefact correction is due the technician, that should move slightly the electrode [12].

- ***Muscle artefact.*** Contracting muscles generate an electrical field that is sensed by scalp electrodes, as well as they sense ECG signal. EMG has a larger band frequency with respect to EEG, and a larger amplitude, it is superimposes on EEG recording and it is difficult to get rid of it.
  Moreover, contracting muscles can cause movement artefact.
  Because of this, it is very important that the patient stays still during acquisition [12].

Muscle and ocular artefacts are the most critical. If fact, many methods that try clean EEG signal from them have been developed. Mostly they are based on blink source separation and attempt to distinguish sources originating in the brain and artefactual ones [5, 9].

Non-physiological artefacts

- ***Electrode artefact.*** This kind of artefact occurs when an electrode pops or loses electrical contact with the scalp, making the impedance associated suddenly increasing. It is visible only on EEG channel that involve the electrode and it appears as few brief waveform [12].

- ***Electrical noise.*** This is widely the most common artefact, that affect all biomedical signals. It is caused by coupoling of electrodes with any other electrical object near the patient. This artefact present itself like a oscillation at 50 or 60 Hz.
  Generally this artefact doesn't constitute a big problem in EEG study, because normal condition EEG frequency band doesn't go above 40 Hz. Thus a low pass filter is sufficient to get rid of the artefact without loosing any information.

1.5 Slow Biphasic Complexes

As literature widely demonstrated [3, 4, 10, 8, 13, 19], Slow Biphasic Complex (SBC) in EEG signal is a bioelectric marker of acute processes of cerebral parenchyma, including encephalitis. This particular waveform is constant in subjects with encephalitis, and also appears to be related to the risk of developing neurological sequelae, such as epilepsy [19].
Here is a description of Slow Biphasic Complexes (SBCs), with a list of their principal features:

- SBCs are made up of two identical waves with inverted polarity [10];

- They typically present firstly the negative phase and secondly the positive one [8].

- Their amplitude ranges from 100 to 250 μV [8, 10];

- SBCs last from 300 to 600 ms. Mean duration is about 500 ms [8, 10, 13];

- If they have a low amplitude, they generally show up with a higher frequency and duration below 500 ms, and vice versa; [13, 19];

- They are not periodic [10];

- Identifying SBCs is easier during patient awakeness than during his sleep [13, 19];

- Their presence is more common in frontal and temporal regions [4].

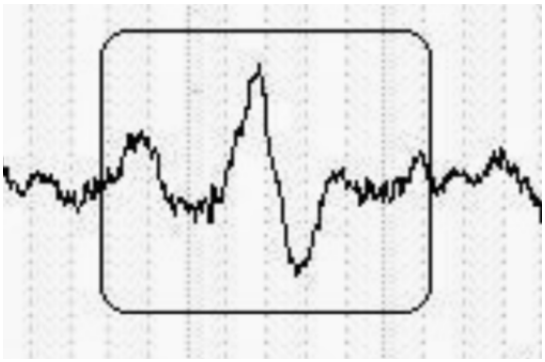

Figure 1.6: *Example of a Slow Biphasic Complex (SBC) in a EEG signal* [8]

Certainly a rapid, reliable, repeatable and automatic recognition of these anomalies could lead to a better understanding of their mechanisms, and of course it would be useful to speed up the diagnosis procedure, reducing subsequent complications.

Chapter 2

Materials and methods

First aim of this study was to improve the slow biphasic complexes detecting ability of a pre-existing algorithm, changing its signal pre- and post-processing while maintaining the fundamental idea. It is based on thresholding of cross-correlation between EEG signal and an SBC wave prototype [13]. This chapter presents data used to extract knowledge about SBCs and the initial algorithm employed to detect those particular waves. Moreover how extracted knowledge has been used to improved the method itself will be explained.
At the end, improved algorithm validation and test methods are presented.

2.1 Data and signals

The study was conduced on EEG data acquired from paediatric patients, whose clinical condition was consistent with the diagnoses of encephalitis.
One hundred and twenty-eight signals have been acquired from fourteen different paediatric patients, males and females.
Some of them have already been diagnosed with autoimmune encephalitis at the time of the acquisitions, while others have been tested for suspected encephalitis with EEG acquisition, due to suspicious symptoms such as headache, dizziness and vomiting. What caused encephalitis is generally unknown, as in 50-60% of this kind of patients [19].
In different patients it is possible to observe different severity of the pathology.

EEG duration goes from 8 to 32 minutes, with a mode of 17 minutes. The sampling frequency is 256 samples/s.
During acquisitions patients were awake, asleep or sedate; in some cases the waking state of the patient was not specified by the technician. In any case, as the expert has proved, SBCs are present in all EEG signals, in varying amounts.

Some of this signals have been used to extract new knowledge and to enhance the algorithm, while the totality of them has been utilized for algorithm test, as it is explained in Chapters 2.8 and 3.4.

2.2 Existing algorithm

Paper [13] presented the algorithm.
The paper explains that the algorithm *"is based on a set of match filters. Specifically, the raw EEG was cross-correlated with scaled versions of a prototype waveform, identified after decades of observation of SBC's in different pathologic traces"* [13].
Expert neuropsychiatrists designed the prototype, based on observations carried out throughout years of study. It has a total time duration of 1 second and a peak-to-peak amplitude of 90 μV. It is possible to identify the negative phase (upwards) and the positive one (downwards), with similar time durations and amplitudes. Prototype is reported di figure 2.1.
We are aware that Slow Biphasic Complexes can present themselves in the signal with different amplitude and duration: a larger duration is generally associated with a higher amplitude [19]. Because of that, the prototype is subjected to different time and amplitude scalings, before being compared with the signal through cross-correlation computation. The algorithm applies ten scales linearly distributed between 0,25 and 3 to the prototype, so that the smallest scaled waveform presents a time duration of 250 ms and an amplitude of 22 μV, while the largest one is 3 s long with an amplitude of 270 μV.

Cross-correlation is computed between each channel of EEG and scaled waveforms, using convolution, which allows to obtain a more efficient method:

$$C(t) = \frac{\int x(\tau)w(t+\tau)d\tau}{\|x\|^2\|w\|^2} \tag{2.1}$$

where t is time, $x(t)$ is the EEG signal and $w(t)$ is the proper scaled waveform prototype. Integral in equation 2.1 was implemented in the code using a finite sum, as the data are sampled.
Even if the likelihood is low, it is possible to find some SBC waveforms in the reverse form: first presenting the positive phase (downwards) and then the negative one (upwards) [19]. Because of this, it is necessary to investigate also the reverse form of prototype waveform in the signal.

Due to normalization, $C(t)$ goes from -1 to 1, where 1 indicates the perfect match between the signal at time point t and the prototype, and -1 indicates the perfect match between the signal and the reverse form of the prototype. As finding a perfect match is very unlikely, a threshold at 90% is imposed on the absolute value of $C(t)$. It allows to find in the signal waveforms which are similar to the prototype or its reverse.

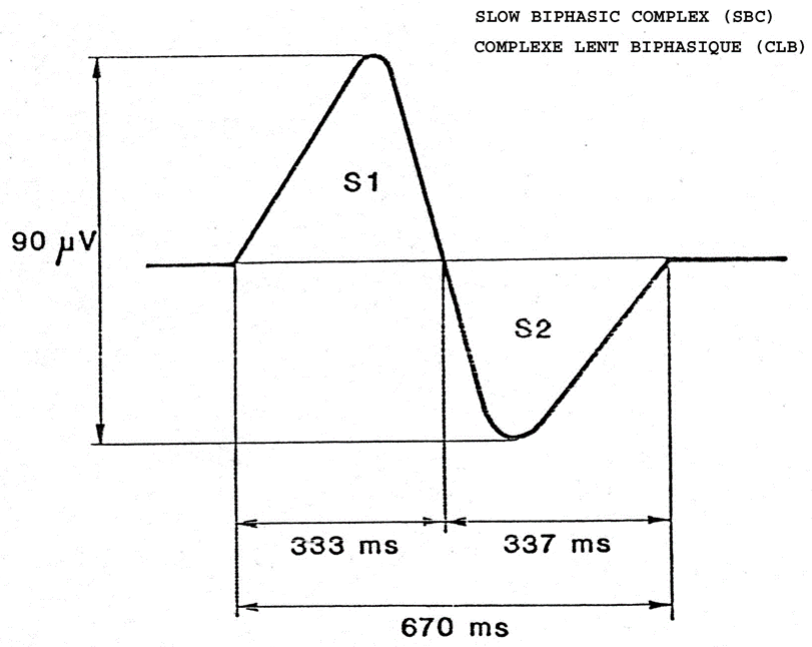

Figure 2.1: *Waveform prototype manually designed from M.D., used to search for slow biphasic complexes in EEG signal.* [13]

Therefore the method searches for peaks in absolute value of normalized $C(t)$ function and isolates the ones above 0,9. It considers the point where it detected the peak as the barycentre, and it selects a wave with a length equal to two times the distance between minimum and maximum point.

After that a post-processing step is provided, checking the amplitude of waves selected with cross-correlation. To do so, the envelope of the entire EEG channel is extracted by applying a low-pass filter with 1 Hz cut-off frequency; then the cumulative distribution of this envelope is estimated. The method considers a selected wave acceptable SBC if its amplitude is between 20% and 99% level of the cumulative distribution of total signal amplitude.

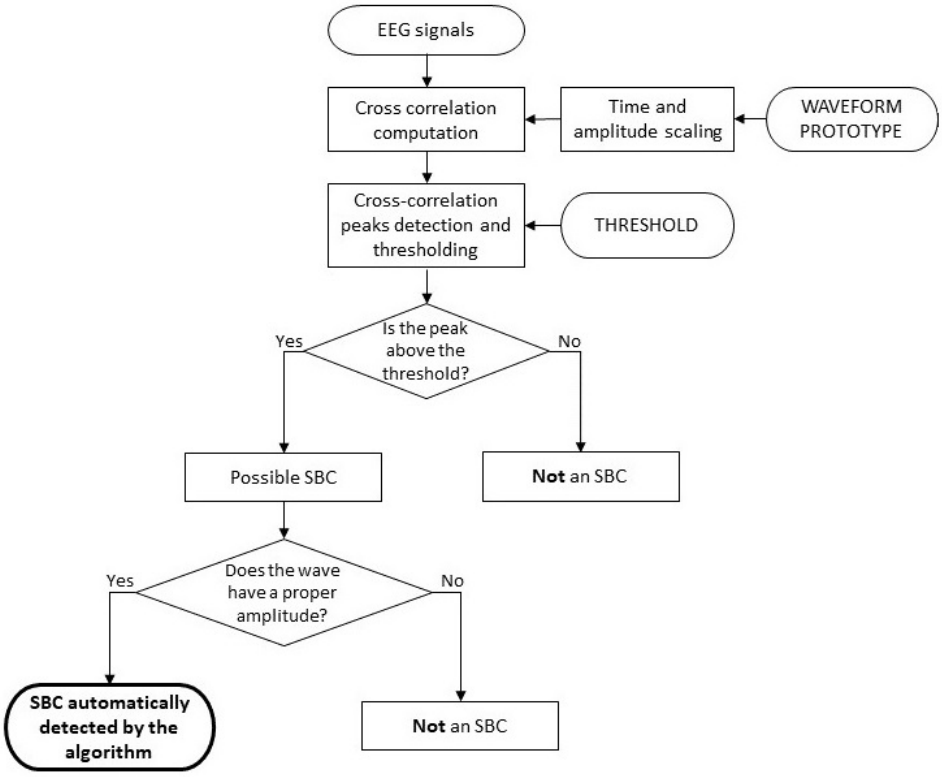

Figure 2.2: *Pre-existing algorithm flowchart* [13]

2.3 Data acquisition

In order to improve the pre-existing algorithm, it was essential to extract new knowledge about actual SBCs characteristics. To do so, an experienced neuropsychiatrist has been asked to manually select new waveforms in EEG signal, using a graphic user interface specifically built for this purpose.
The GUI, implemented in MATLAB *(Inc., Natick, Massachusetts, USA, ver. 2020a, interpreted single core implementation)*, allows to load EEG signals in *efd* format, either acquired in monopolar or bipolar mode. In case of monopolar acquisition, it is possible to visualize the signal in its original form, or using a desired bipolar montage (choosing between montage 3, montage 2 and montage 9), thanks to the application of required linear compositions of channels. The signal will be displayed in epochs. The user can select how many seconds should be displayed in each epoch. This allows to get a standard display, which will print 1.5 cm/s, regardless of the used screen size.
After choosing the most suitable visualization, it is possible to scroll EEG signal and select waves using mouse. If technician notes are available, it is possible to load a text file containing them and the GUI will display notes on the signal, permitting the expert to know about the subject status and significant events happened during acquisition. This is useful to the neuropsychiatrist, to reject artefacts and to better understand the signal.
Once selected, the user can navigate the waves and if necessary delete the wrong ones.
Results are automatically saved periodically in a file .mat, containing the signal, computed and visualized channels names, and references to the selected SBCs.

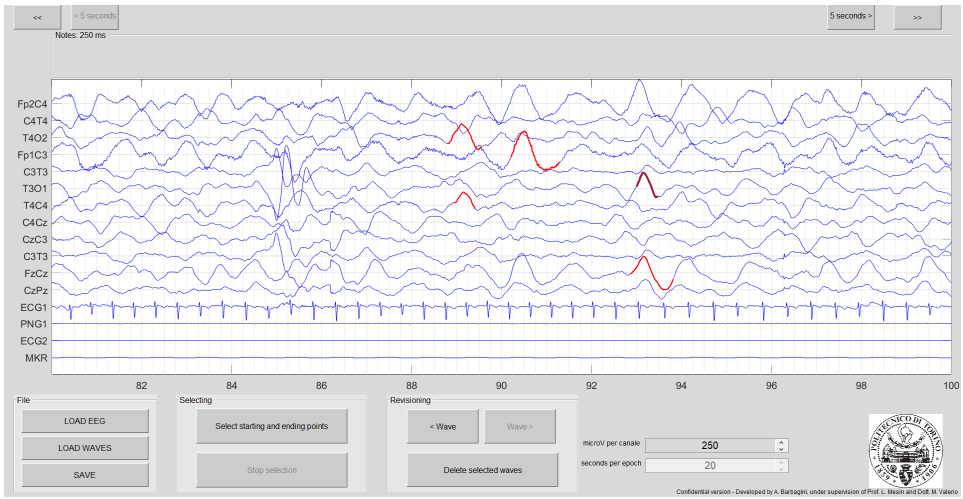

Figure 2.3: *Selection GUI: Usage example of data extraction with graphic user interface: selection of waves in a loaded EEG signal of a patient, visualizing it with a bipolar montage 3.*
'Seconds per epoch' parameter has been set as 20 seconds, because the screen was 30 cm long: 1,5 cm/s is displayed. Each second is divided in four parts, using a 250 ms grid, in order to better select the correct duration of waves.
'MicroV per channel' parameter represents the distance between two channels; it can be varied by the user to better visualize the signal.
Buttons in group 'Selecting' allow to use mouse to identify starting and ending points of a new waveforms, which will be red highlighted right after the selection. Buttons in group 'Revising' allow to navigate the already selected waves. The current one is highlighted in a darker colour and it can be deleted by pressing the dedicate button.
Buttons in group 'File' permit to save data and either to load a new EEG signal or to load previous selected waves on the same EEG, allowing to split the work in case of long signals.

2.4 Data processing

To improve the algorithm, it was necessary to process waves selected by the physician, in order to extract features that can characterize SBCs, hence that can help in the automatic identification of waves of interest.
In this thesis the carried out elaborations and the subsequently obtained results are presented straightforwardly. However, in reality, the research was carried out by a long collaboration with the physician. Results obtained up to a certain point have been presented to the doctor on a weekly basis. In this way he was able to take a critical look at what it had been obtained and what were the errors. In this way, combining technical and medical-physiological skills, it was possible to modify the algorithm little by little, always adding new controls.
Therefore data elaboration process and validation were not carried out once, but several time, in a loop. Information extracted from data processing were useful to achieve little enhancement of the method, checked from time to time by the expert.
This process of strict collaboration and communication could lead to the best possible result.

Of the one hundred and twenty-eight signals available, M.D. has processed forty-two through the Selection-GUI. He has selected 2049 waveforms, identifying them as representative slow biphasic complexes. 1846 waves detected in 39 signals were used for this elaboration. Therefore, three EEG signals from three different patients were excluded. The latter will be used as test.
Three acquisitions that constitute test set have been chosen in agreement with the doctor, because they have different characteristics, which present different challenges in the detection of waves by the algorithm.

1. One EEG signal is particularly slow, visibly clearly pathological. It is characterized by low frequency and high amplitude: delta band oscillation are predominant. Experts classified this EEG signal as correlated to a serious pathology.
   Within this signal SBCs don't particularly outstand from the background, but their waveform is clear and easy to recognize, because there are no high frequency oscillations which inject on them.
   Notes written by the treating physician about this EEG trace refer to: *"Persistent severe picture of widespread brain suffering"*. During the acquisition the patient was under the effect of an anti seizure drug.

2. The second signal has characteristics that lead to think that the patient's condition is not serious, in fact it is close to a normal path, classified as mild disease by experts. SBCs are few and present low

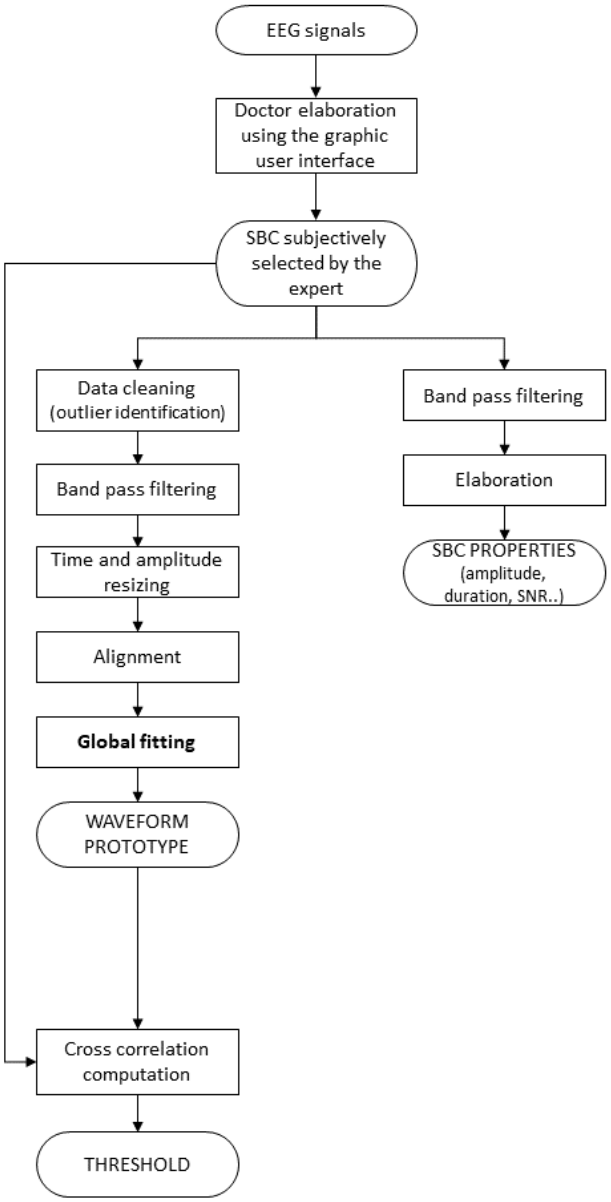

Figure 2.4: *Data processing flowchart*

amplitude, meaning that the pathology is in its early stages. During acquisition patient wasn't under any drugs.

3. The third EEG selected as test seems again a quite normal acquisition, which presents some SBCs, once again associated with a mild disease. In this case SBCs are a little more widespread then the second case, and their waveform is clearer. This considerations lead to think that the patient from whom this EEG was acquired is in intermediate pathological conditions. This signal is characterized by high frequency oscillation stretches and the presence of tri-phasic and poliphasic complexes. The latter should not be selected by the algorithm, even if a portion of them could seem a biphasic complex. There is no information available about the status of the patient during the acquisition.

Thirty seconds of each EEG in the test set are shown in figure 2.5.

While test acquisition were used to understand the algorithm performance, ones in the training set were used for all the elaborations presented in the following. Beginning from 1846 waves detected in 39 signals, those waves have been collected and tested, in order to reject outliers and to deal with fully SBCs, which include three zero crossings, thus a negative phase and a positive one. The check has been provide in part automatically and in part manually. Processing steps are shown below.

- Waveforms were firstly considered with their original length and time support.
- A low pass filter was applied, with cut-off frequency 7Hz, in order to extract waves envelope.
- The DC value has been removed, so that all waves average has been set to zero.
- Matlab code able to identify zero crossing points was developed.

```
function [zero_crossing] = find_zero_crossing(sig)
      zero_crossing=[];
      interval = 20;
      i=interval+1;
      while i<=length(sig)-interval
      % while loop used to scroll over signal
          if (sum(sign(sig(i-interval:i-1)))==-interval && sum(sign(sig(i+1:i+interval)))==interval) ...
                  || (sum(sign(sig(i-interval:i-1)))==interval && sum(sign(sig(i+1:i+interval)))==-interval
              zero_crossing = [zero_crossing, i]; %saving the point as a zero crossing
              i = i+interval; %skipping 20 samples, in order to count each zero crossing once
          else
              i=i+1;
          end
      end
end
```

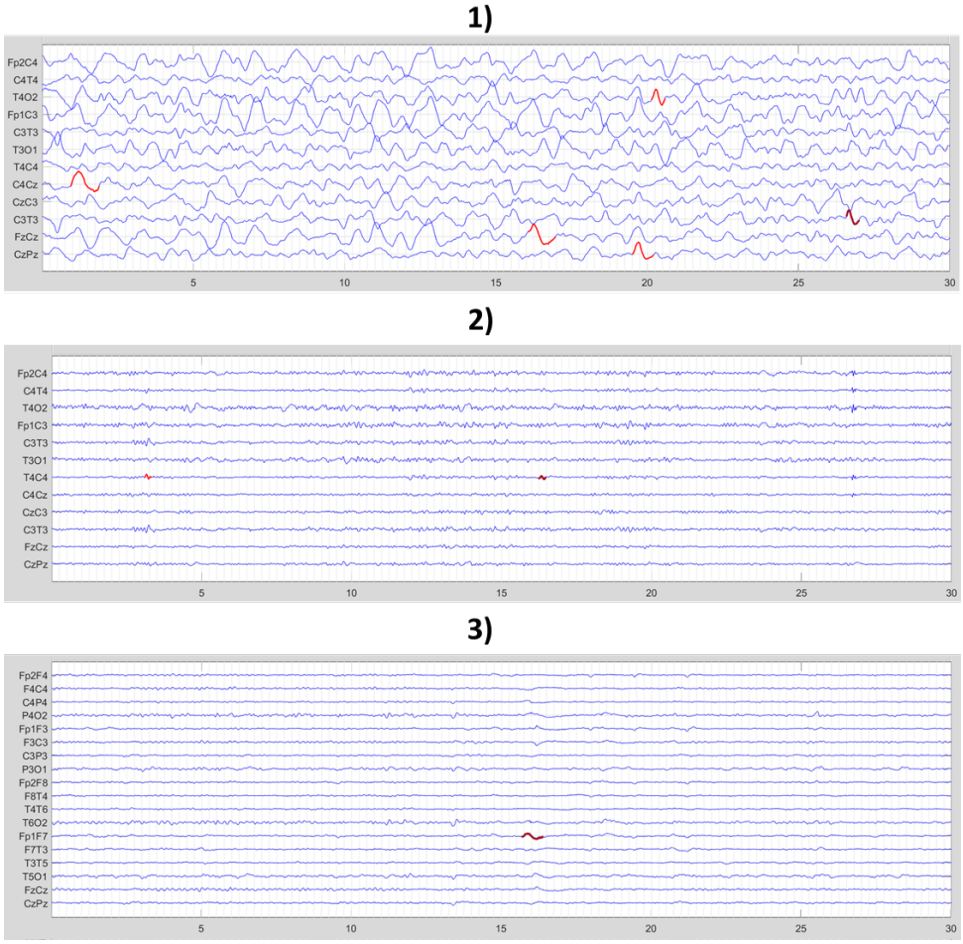

Figure 2.5: *Thirty seconds of each EEG selected to compose the test set. Waves manually selected by the physician are highlighted in red. Distance between channel is set at 300 μV, even if the default one is generallt 200 μV. This is for a better readability of the first signal.*
All EEG signals are acquired using a bipolar montage 3.
1) *EEG signal of the patient with the most serious pathology;*
2) *EEG signal of the patient with an early stage encephalitis;*
3) *EEG signal of a patient with an intermediate seriousness of the pathology. At this stage slow biphasic complexes still few, but they already present a clear and defined waveform.*

It looks for twenty positive samples followed by twenty negative ones (or vice versa), recognizing actual zero crossings and rejecting the ones caused by noise.

- If there are less than three zero crossing points, the time support length is iteratively increased until at least three zero crossing are found. It is initially set as two times the distance between the maximum and the minimum point in the waveform, and than increased at three times the distance, and so on, until three zero crossing points are found.

- If there are four zero crossing points, external signal stretches between two zeros RMS are computed. If one of them is significantly lower than RMS of the other two ones, the trait in object is not considered. Otherwise, the selection of the correct time support is carried out manually, with the possibility to decree an outlier;

This processing was inserted in a loop, which analyses all waveforms selected by the doctor, either adjusting them automatically or proposing them to the user, and finally saving the envelope of non-outlier ones in a Matlab structure.

2.4.1 New waveform prototype design

Starting from the waves selected by the expert and chosen as not-outlier in the previous processing, a new prototype has been developed, with the goal of maximizing its cross-correlation value with each SBCs. To do so, a global fit approach has been applied. The elaboration process is outlined as follows:

- Time support scaling so that all waves are 100 samples long, corresponding to a time duration of about 390 ms. Length of 100 samples has been chosen as it is near to the the mode of all waves length distribution, thus the most diffused one;

- Amplitude scaling so that all waves have a peak-to-peak amplitude of 200 μV;

- Whenever waves present firstly the positive phase and then the negative one, they are reversed. Code able to identify the correct point used to consider separately the phases and then decide to reverse the wave or not is reported below:

```
% GOAL: Identify the central zero crossing point, which divides the negative
% and the positive phase of the wave.
% Flip the wave in case it is reversed.

load('WAVEFORM')
WF_upright  = zeros(size(WF));

% FINDING THE CENTRAL ZERO CROSSING POINT
for i=1:size(WF,1)
    % Deleting the DC component
    s = WF(i,:)-mean(WF(i,:));
```

```
    % Finding zero crossing points
    zero_found{i} = find_zero_crossing(s);

    if length(zero_found{i})>3
    % In case of four or more zero crossing, I consider the one with the
    % steepest slope
        deriv = diff(s);
        slopes = deriv(zero_found{i});
        zerocross(i) = zero_found{i}(find(slopes==max(slopes)));
    elseif length(zero_found{i})==3
    % In case of three zero crossing point (ideal case), I consider the
    % central one
        zerocross(i) = zero_found{i}(2);
    elseif length(zero_found{i})==2
    % In case of two zero crosses point, I consider the one more in the
    % center (i.e. closer to the sample 50).
        [distanza_centro,ind] = min(abs(zero_found{i}-50));
        zerocross(i) = zero_found{i}(ind);
    elseif length(zero_found{i})==1
    % If there is just one zero crossing point, I consider that one
        zerocross(i) = zero_found{i};
    end

    % FLIPPING

    % Computing the sign of the first phase (wave segment before zero
    % crossing point detected)
    first_phase = sign(mean(s(1:zerocross(i))));

    if first_phase == 1
    % In this case the first phase is the negative one (upwards).
    % The wave remains as it is.
        WF_upright(i,:)=s;
    else
    % In this case the first phase is the positive one (downwards).
    % The wave has to be flipped.
        WF_upright(i,:)=(-1)*s;
    end
end
```

- The waves are aligned with respect to the central zero crossing point;
- Computation of a matrix C containing mutual cross-correlations be-

tween all waves:

$$C = \begin{bmatrix} 1 & C_{12} & \dots & C_{1N} \\ \vdots & \ddots & & \vdots \\ C_{N1} & C_{N2} & \dots & 1 \end{bmatrix} \tag{2.2}$$

where C_{ij} represents the maximum value of the cross-correlation function between waveforms i and j, calculated with formula 2.1. All elements on diagonal are ones, as they refers to the cross-correlation of wave i with itself;

- Considering the matrix C, it is possible to calculate the mean of the cross-correlation maximum values of each waveform with all others (excluding the diagonal). This is done substituting ones on diagonal with NaN's and calculating the mean of each row of the matrix C, excluding NaN's. If one presents this value lower than 75%, it is considered outlier and they are rejected. Value 0,75 has been chosen looking at the mean values distribution, to treat as outlier a low percentage of total waves.

- Starting from remaining waves, a new subset of waves is constructed iteratively, as it shown in flowchart reported in figure 2.6. A new wave is included if the maximum value of cross-correlation between it and each wave already present in the subset is lower than 88%. In this manner the built subset includes waveforms which are similar among each other, but still being able to represent the entire set with no redundancy.

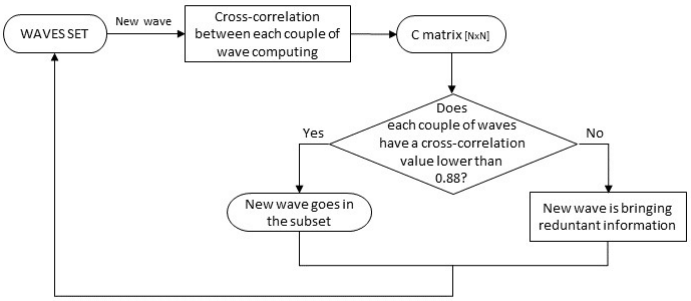

Figure 2.6: *Flowchart showing iteratively construction of the wave subsubset used to construct the new prototype.*

- A polynomial curve has been extracted from the subset of waves, globally fitting them all, using Matlab command *polyfit* and *polyval*. The

polynomial degree was tuned between 3 and 15, choosing the one that allows to maximize cross-correlation values between curve and all waves in the subset. In particular, a condition was imposed on the lower value over all, correlated to the wave that is less alike to the prototype.

Previously different approaches had been attempted to define the new prototype.
Firstly, optimization by changing one samples at a time had been attempted. Starting from a sine wave period, an iterative adjusting of single samples was performed, moving up or down each point by one small step, trying to raise the cross correlation value between the found prototype and whole set of waveforms. The stop condition was set on the minimum correlation between the curve and each SBC, accepting even a little worsening, if the mode of cross-correlation values distribution didn't lower.
This method has not led to adequate results, so it was discarded.

Next, clustering of SBCs has been tested, with the idea of different prototypes definition, that could fit different waves.
Selected SBCs have been divided in three clusters. The number of three was choosen in order to obtain homogeneous distribution of elements in clusters. Outliers were discarded, trying to lower the maximum intra-cluster variability (calculated for each wave as mean square error compared to the average of all curves). After that, three prototype were extracted as the average of all waves in a cluster.
These lead to the definition of prototypes highly correlated each other. So there was no point in applying over one of that.

2.4.2 Defining a new cross-correlation threshold

Having a new waveform prototype, calculation of a new and proper cross-correlation threshold is necessary.

The correlation between the entire signal and the new prototype wave subject to the different scaling has been evaluated, extracting the values at the indexes to which the physician selected the SBCs. In this way a matrix has been constructed, with as many rows as the number of scaling and as many columns as the SBCs identified by the doctor. Each column corresponds to an SBC, which presents a maximum value of cross-correlation with each scaling.
Finding the maximum value for each column, it is possible to identify the scaling that best matches a particular waveform. The threshold must be set so that the algorithm is able to detect a reasonably high number of SBCs.

2.4.3 Study of different scalings

The effect of number of scalings has been estimated. The simple idea was to examine the cross-correlation between adjacent scaling. If this correlation is largely above the threshold, it means that that scaling can be deleted. At the same time, cross-correlation between close scalings should not be strong below the threshold, so that no wave in EEG signal is excluded because it matches an intermediate scaling. The scalings does not necessarily need to be distributed linearly.
Several number and distributions of scalings of waveform prototype have been checked, eventually modifying them manually, evaluating the cross-correlation between each pair of adjacent ones.
To better approximate the SBC waves characteristics, time support of the prototype is actually scaled proportionally to each scaling, while the amplitude is compute starting from the duration and intercepting the line that interpolates the length-width trend of manually selected SBC.

Scalings distribution evaluation and threshold setting have been repeated iteratively, until finding a proper compromise.

2.4.4 Emergence of SBC from background

In neuropsychiatrists opinion, SBCs should have a considerably larger amplitude than the background. To pass this information to the method, it has been decided to impose a check on the ratio between RMS of the detected wave and one of the immediately preceding and following signal portions. To do so, RMS ratio has been studied in manually detected waves, considering different length of preceding and following signal portions. To find most suitable length for imposing a threshold, the quotient has been calculated from signal segments of length 50, 100, 150, 200 samples. The waves have been divided into slow waves (duration $\leq$ 1s) and fast waves (duration $<$ 1s), so that different background lengths can be considered depending on the temporal support size of the wave.
The formula is reported in 2.3.

$$\frac{RMS_{wave}}{\frac{1}{2}(RMS_{left} + RMS_{rigth})} \tag{2.3}$$

2.4.5 SBCs surrounded by high-frequency rhythm oscillation

According to the physician, it is possible to spot some SBCs within an oscillation characterized by a frequency higher that 5 Hz, even if it is odd. However, if it is the case, the SBC must be still typified by a low frequency and its amplitude must be significantly higher that the background.

To quantify this statements a study on manually selected waves has being developed.

All selected waves' adjacent trait (1 s) of signal were considered. Power spectral density of those signal stretches were computed. To do so, PSD was estimated by a simple periodogram, using a window of 1 s length and without any time-average. This was possible because each analysed signal has been considered stationary.
Percentage of spectral power which is over 5 Hz has been computed. If both signal stretches that surrounded a wave present more than 50% of the spectral power over 5 Hz, this particular wave is considered surrounded by high-frequency rhythm oscillation. Those waves' RMS ratio were evaluated once again, setting a new higher threshold, which will be impose on this specific case.

2.4.6 SBCs spectrum frequency

SBCs by their nature must be slow. Thus the low-frequency power percentage of the complexes selected by the physician was calculated.
To do so, PSD of each wave was computed, with a simple periodogram, using a window of the same length of the wave itself. Being shorter then 2 s, all waves have been considered stationary.
The power percentage has been extracted under 3 Hz and 5 Hz of all SBCs, in order to define a threshold imposable on waves selected by the algorithm, in order to discard wrong ones.

2.5 Detection of artefacts

Observing first algorithm results with the assistance of the physician, it has been noticed that some of the waves automatically selected did correspond to artefacts instead of slow biphasic complexes. The algorithm selected several kind of artefacts: ocular, muscular and non-physiological ones.
The latter did not represent a problem. In particular ones related to electrical noise are generally waves at a frequency of 50 Hz, which is far beyond the frequency band at which brain waves are shown. Applying a low-pass filter to signals was sufficient to get rid of this kind of artefacts.
As regard artefacts caused by poor contact between an electrodes and the scalp, those not have given any problems, as this kind of artefact don't generally present the waveform in which we are interested and, in any case their amplitude is generally low, meaning that they don't show any emergence from background.
Physiological artefacts have been more demanding, in particular the EOG artefact, which by its nature shares several characteristics with slow biphasic complexes. In fact, both SBCs and blink artefacts:

- Are characterized by a bi-frontal potential;
- Have a larger amplitude compared with normal EEG;
- Appear more often on frontal channels.

In order to get rid of this kind of artefacts, mistaken by the algorithm for SBCs, it has been tried to apply Gomez algorithm [9] to clean the signal form blink artefacts, before to elaborate it with the SBCs research algorithm.
Gomez method tries to remove ocular components, without relying on EOG signal that was not available in the acquisition used for this study. To do so, it implements blind source separation, and it considers as artefactual those components which have a low fractal dimension. In this way it reconstructs the signal not considering the selected components.
However, several slow biphasic complexes, which has properties in common with blink artefacts, were considered as artefacts by this method and deleted form the signal before the application of the searching algorithm.
A further attempt to clear the signal from blink artefacts was the application of the method presented in paper *"Detection of eye blink artefacts from single prefrontal channel electroencephalogram"* [5], which tries to detect slow ocular artefacts by computing summation of derivatives within window (SDW), using a dynamic size window and thresholding this function. Once again SBCs are erroneously considered as slow artefacts.
Consequently, it was decided to pass the signals to the algorithm without cleaning them from artefacts and only successively try to discriminate between real SBCs and waves produced by eye blinks.

To do so it has been established an amplitude check. Literature widely explains that duration and amplitude of slow biphasic complexes are strictly correlated: a short complex will show a lower amplitude compared to a longer one [13, 19], and vice versa. Because of that, starting from amplitude-duration data distribution extracted form waveforms selected by M.D., it was possible to define amplitude bounds which depend on duration of a particular wave.

2.6 Algorithm upgrading

Starting from the information extracted from the selected SBC, the algorithm has been enhanced, using a new waveform prototype, subjected to an appropriate set of scaling, and compared with the signal through cross correlation, on which a suitable threshold is imposed.
Before that a pre-processing filtering step has been added. Since the aim is detecting SBCs, which are composed of only slow activity, a filter has been inserted to eliminate everything that is not slow. Clinically, slow brain activity is defined as waves at a frequency lower than 7.5 Hz (Corresponding to the delta and theta bands). The signals are then filtered low pass with a cut-off frequency of 7.5 Hz.
After computing the cross-correlation a post-processing phase is been developed. Each wave automatically selected thanks its upper-threshold cross-correlation, is checked about its amplitude, its time duration, and the ratio between RMS of the wave itself and RMS of adjacent signal stretches. Furthermore,the relation between SBCs time duration and amplitude has been extracted: a shorter wave is associated to a smaller amplitude and vice versa. During SBC research, it was imposed that the selected waves respect this dependence.
In order to enhance the fact that SBCs have to distinguish themselves from the background amplitude and must not be too close each other, a series of zeros has been concatenated at the beginning and at the end of the waveform prototype, so that the cross-correlation between prototype and signal will be high only in case of those conditions respected. To do so the prototype was windowed, so that the first and the last samples were set to zero and there was not any step between wave and zeros.

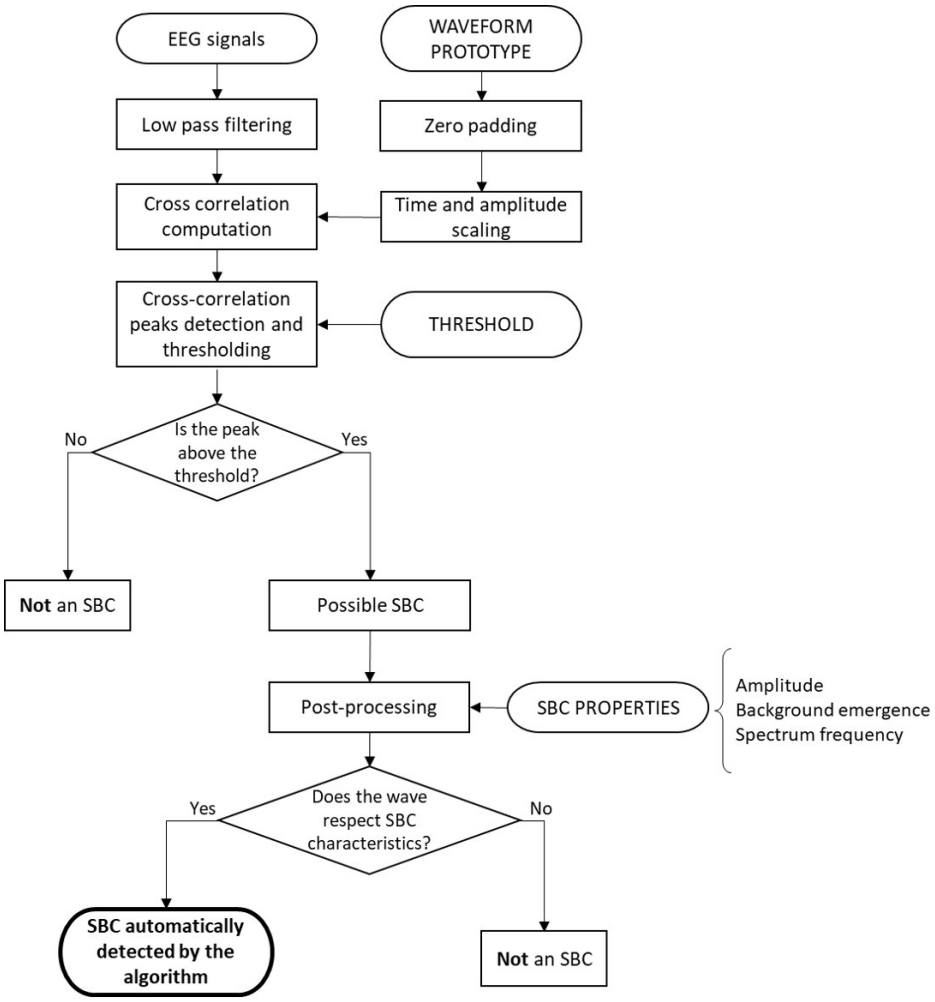

Figure 2.7: *Enhanced algorithm flowchart*

2.7 Validation of results

After elaborating all signals in the data set with the new enhanced algorithm, the results have been again submitted to the judgment of the doctor, who validate them, choosing for each selected wave if it was an actual slow biphasic complex or not. To this purpose a new graphic user interface has been built specifically, which allowed the M.D. to visualize signals, and scroll and evaluate all waves. An example of this second graphic user interface is reported in figure 2.8.

From the validation it was possible to extract for each EEG signal the number of true positive and false positive, meaning respectively the number of actual SBCs identified by the algorithm and the number of waves which were considered wrong.

Moreover, the ability of the algorithm to recognize SBC by distinguishing the severity of the pathology of different patients was evaluated.
Each EEG acquired by patients with encephalitis has been assigned by the doctors a number representing the severity of the disease:

Severity 0 Normal condition EEG signal. By default severity 0 has been assigned to all controls, other than some patients which not present any particular condition related to encephalitis at the moment of the acquisition.

Severity 1 Mild disease

Severity 2 Moderate disease

Severity 3 Severe condition

Severity 4 Serious pathology

Moreover, some EEG traces were considered as non-classifiable (NC). Analysing validation results divided in severity classes, it has been possible to understand if the performance of algorithm is better in case of advanced or mild pathology.

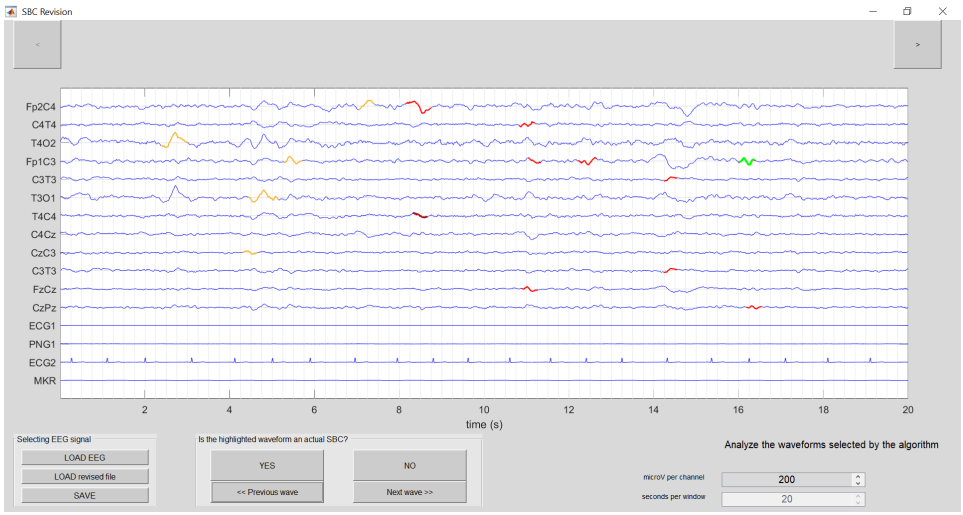

Figure 2.8: *Revision GUI: Usage example of SBC searching algorithm results reviewing with a graphic user interface. This was developed in order to allow M.D. to look at the automatic selection and decide ether if each wave is correct or not.*
Overall setup is similar to Selection GUI, as the physician used it to review same signal previously elaborated while selecting SBCs.
The GUI shows in red waves automatically selected and in green those selected manually during data acquisition. M.D. can scroll through red waves and give a 'Yes' or 'No' answer for the currently highlighted one. The waves for which a choice has been made change their color and become orange. Alternatively it is possible not to give an answer for a particular wave and use the 'Next Wave' button to move on. It is possible to save the given answers and load previously saved files in order to edit the responses or complete an already started work.

2.8 Algorithm test

After the validation of algorithm results, its diagnostic capability has been tested.
To do so, 128 spontaneous EEG signals were considered, as it was done in papers [14] and [13]. This signals were elaborated using the algorithm, in order to classify them either as healthy or encephalitis patients. Signals are acquired form 10 patients showing encephalitis symptoms, and 10 controls, which are healthy children or paediatric patients with diagnosis not consistent with encephalitis.
Characterization of *encephalitis patients* and *non encephalitis patients* was attempted considering different measures of identified SBCs. Same measure have been used to characterize different severity degree of the pathology.

Since signals present different number of channels and durations, the simple number of selected SBCs is not reliable. Because of that, to extract informations about each EEG signal, the following measures have been used:

- Averaged number of selected slow biphasi complexes;

$$M1 = \frac{SBC}{N_{Channels} \times N_{epochs}} \tag{2.4}$$

  Where SBC is the number of found slow biphasic complexes in a signal, $N_{Channels}$ is the number of acquired EEG channels and N_{epochs} is the number of 10 seconds epochs contained in the signal.

- Averaged number of selected slow biphasic complexes on frontal channels:

$$M2 = \frac{SBC_{frontals}}{N_{frontalChannels} \times N_{epochs}} \tag{2.5}$$

  Where $SBC_{frontals}$ is the number of found slow biphasic complexes in frontal channels, $N_{fornalChannels}$ is the number of frontal channels (typically two: Fp1 and Fp2), and N_{epochs} is the number of 10 seconds epochs contained in the signal.

- Mean RMS value of identified slow biphasic complexes, taking in account once again number of channels and time duration of singal;

$$M3 = \frac{\sum_{i=1}^{N_{SBC}} (RMS(SBC_i))}{N_{Channels} \times N_{epochs}} \tag{2.6}$$

  Where N_{SBC} is the total number of found waves, SBC_i is the waveform of the $i-th$ selected SBC, $N_{Channels}$ is the number of acquired EEG channels and N_{epochs} is the number of 10 seconds epochs contained in the signal.

RMS value of each SBC is computed as follow, using Matlab function *rms*:

$$RMS(SBC_i) = \sqrt{\frac{1}{T} \times \sum_{t-T}^{t} (SBC_i(t)^2)} \tag{2.7}$$

- Mean RMS value of identified slow biphasic complexes of frontal channels;

$$M4 = \frac{\sum_{i=1}^{N_{SBC_{Frontals}}} (RMS(SBC_{Frontals_i}))}{N_{FrontalChannels} \times N_{epochs}} \tag{2.8}$$

Where $N_{SBC_{Frontals}}$ is number of found waves on frontal channels, $SBC_{Frontals_i}$ is the actual waveform of the $i-th$ considered SBC, $N_{FrontalsChannels}$ is the number of frontal channels, and N_{epochs} is the number of 10 seconds epochs contained in the signal.
RMS value of each SBC is computed as it is shown in equation 2.7.

- Standard deviation of time interval between successive SBCs

$$M5 = \sqrt{\frac{\sum_{i=1}^{N_{SBC}} |interval_i - mean(interval)|}{N_{SBC}}} \tag{2.9}$$

Where N_{SBC} is the total number of found SBCs, $interval_i$ is the distance between the first sample of $i-th$ SBC and the first sample of the subsequent SBC on the same channel, and $mean(interval)$ is the averaged value of all intervals.

Using this measurements, alone or combined each other, it has been evaluated the possibility to classify patients.

Chapter 3

Results analysis

3.1 Data processing results

Extracted SBCs characteristics

3.1.1 Data cleaning

Initially, starting from data manually selected by physician using Selection-GUI, EEG signals were divided in training set and test set. The latter was leave as it was, in order to use it for the validation of final algorithm. Test set is composed by three signals acquired from three different patients. Training set contains 39 electroencephalographic acquisitions in which M.D. has been identified 1846 waves.
This waves have been taken in account: single waves were evaluated, in order to identify outliers and good ones, suitable for further elaboration and extraction of knowledge. After the evaluation, 1703 waves were considered suitable ones, while 142 were classified as outlier. That means 142 waves presents characteristics that are not representative of the entire set, so they were not suitable for the extraction of information useful to define a overall model of SBC. Many of them presented low frequency oscillation that engages on the main one, which was the actual slow biphasic complex. Other waves, considered good enough to be used in the following processing, are actually clear and evident SBCs, presenting two main phases and, in most cases, three zero crossing points. According to the literature, waves that present first the negative (upwards) and then the positive phase (downwards) should be more widespread, but collected data don't show this prevalence: straight and reverse waveforms appear to be almost equally distributed.
Figure 3.1 shows some examples of waves classified in the two groups.

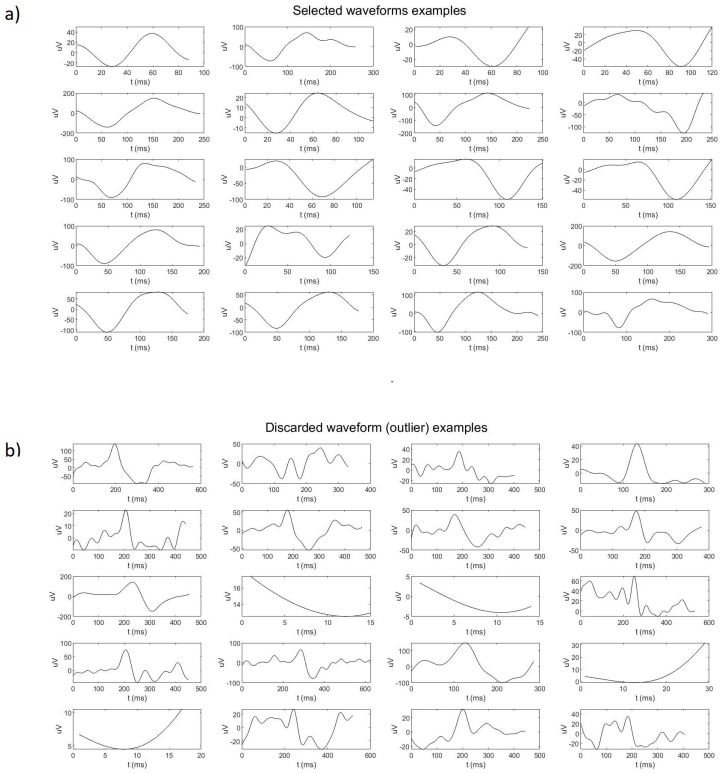

Figure 3.1: a) *Some of selected waves, subsequently used to characterize SBCs, are precedented in the upper figure. It is possible to observe that some waves first present the negative phase and then the positive one, while others are overturned. One type is not more clearly diffused than the other.*

b) *Some of discarded wave (considering them as outlier) are reported. As is evident, many of the outliers have a low frequency oscillation that engages on the SBC. Some other waves that were considered outliers, are probably errors committed during the selection.*

3.1.2 Amplitude and time duration

Time support length of each wave has been extracted. Their length ranges from 30 to 800 samples (from 120 ms to 3 s) with a mode of the distribution of about 130 samples (500 ms). They present a peak-to-peak amplitude going from 20 μV to 730 μV, with a 100 μV mode.
A representation of the distribution of length and amplitude of waves selected by the physician is reported in figure 3.2. It is obvious that a linear relation between time duration and magnitude does exist. This relation is further evidenced by the trend line superimposed over data distribution and

computed using Microsoft Excel:

$$A = 0.27 \times T + 11 \tag{3.1}$$

Where A is the amplitude of each SBC measured in μV, T is its time duration in ms. The coefficient of determination (R^2) associated with the trend line in equation 3.1 is 0.53.

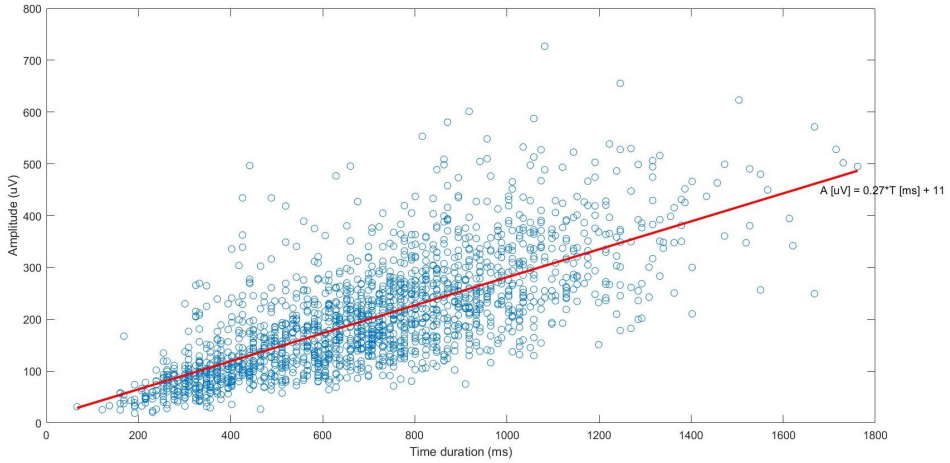

Figure 3.2: a) *Graphic representation of the linear relation between duration (in ms) and amplitude (in μV) regarding all slow biphasic complexes selected by the M.D.. Trend line equation was calculated using Excel.*

All 'good waves' have been scaled in time, so that they all provide a support 100 samples long (390 ms). This will be the prototype waveform length. Same logic has been applied to amplitude. All waves have been scaled in amplitude, in order to obtain a peak-to-peak amplitude of 200 μV, shared for all elements of the set. As all curves used to build it, the waveform prototype will present a time support 390 ms long and a peak-to-peak amplitude of 200 μV. Those measure have been chosen as intermediate, because the resulting prototype will be subjected to different scalings, above and below 1.

3.1.3 Definition of a new waveform prototype.

In order to the define the new prototype by global fitting, it was decided to use waves that were consistent with each other, but did not provide redundant information. To do so, matrix C has been computed, containing all mutual cross-correlation value between each wave and all others. Average of each column of C represents the mean cross-correlation between a wave and all others. So that a wave which has highly widespread characteristics

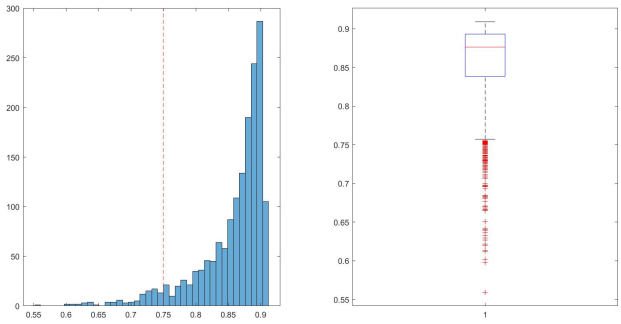

Figure 3.3: *Distribution of all waves averaged cross-correlation value is reported in the figure, visualizing it with an histogram (left) and with a boxplot (right). A threshold at 0.75 has been imposed, excluding 5.9% of waves, which have been considered too odd to be used in the new prototype construction.*

will present an high cross-correlation value with a lot of other waves, so an high averaged cross-correlation value. At the same time, a wave which is special in some way, will be associated with a low mean cross-correlation value. Distribution of mean cross-correlation values is reported in figure 3.3. Waves with mean cross-correlation value lower than 75% were again considered outliers. In this way 5.9% of the previously considered wave has been discarded.

From the remaining 94,01%, the aim is to extract a limited number of waves that is still representative of the entire set. To do so a subset has been built iteratively, adding a wave only if the maximum cross-correlation value between all waves in the subset was lower than 0,88 if the new one was considered, as it was shown by flowchart reported in figure 2.6 in the Chapter 2.4.1.
Thanks to this procedure, twenty-eight waves has been picked. Those are reported in figure 3.4, after being flipped in case they were reversed.

A global fitting procedure was applied at the extracted subset of wave, with the aim of design a polynomial curve which could fit at best all the waves. The polynomial degree that allowed to reach the higher cross-correlation value between the new curve and all the 'good waves' was 11. So that, the new waveform prototype has been designed with a polynomial curve of degree 11. The polynomial equation related to it is reported below:

$$\begin{aligned} p(x) =& 3.52 \cdot 10^{-16} \cdot x^{11} - 1.95 \cdot 10^{-13} \cdot x^{10} + 4.63 \cdot 10^{-11} \cdot x^{9} + \\ & - 6.17 \cdot 10^{-9} \cdot x^{8} + 5.04 \cdot 10^{-7} \cdot x^{7} - 2.6 \cdot 10^{-5} \cdot x^{6} + \\ & + 8.4 \cdot 10^{-4} \cdot x^{5} - 1.63 \cdot 10^{-2} \cdot x^{4} + 0.18 \cdot x^{3} - x^{2} + 10 \cdot x - 39 \end{aligned} \tag{3.2}$$

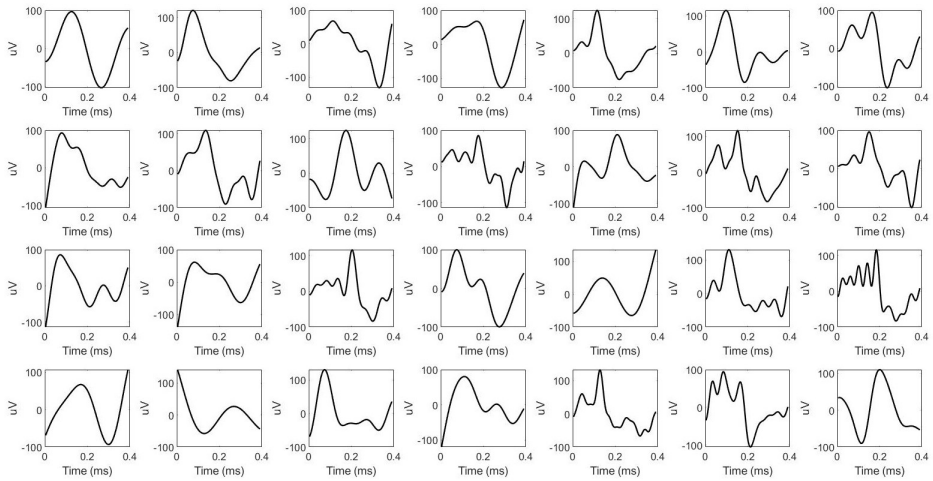

Figure 3.4: *Twenty-eight waves selected as representative of the entire set. This subset was used to create the new waveform prototype by global fitting.*

Subsequently the prototype has been scaled in order to obtain an amplitude of 100 μV, following the indication gave from the relation between time duration and amplitude reported in equation 3.1; to a time duration of 390 ms (100 sample) is associated an amplitude of 116 μV.
The prototype is reported in figure 3.5.
Subsequently a series of zeros has been concatenated at the beginning and at the end of the waveform prototype. To do so the prototype was windowed, so that the first and the last sample were set to zero and there was not any step between wave and queues of zeros. With this solution the algorithm will not select waves too close each other and, using this prototype to compute the cross-correlation, it is enhance the concept of emergence of a wave from the background in terms of amplitude.

3.1.4 Study of scalings distribution

The initial algorithm used ten scalings linearly distributed from 0.25 to 3. In the current application scalings have been redistributed, not necessarily linearly, from 0.3 to 4, so that the smallest wave prototype compared with the signal has a length of about 180 ms and an amplitude of 41 μV, while the largest is 2.37 s long and is 397 μV wide.
To obtain each scaled waveform, the original prototype time support was resize proportionally to the scaling value. To obtain the correct amplitude, the latter was computed starting from the duration in ms, using the formula in 3.1, trying to emulate as much as possible the trend of manually selected SBCs.

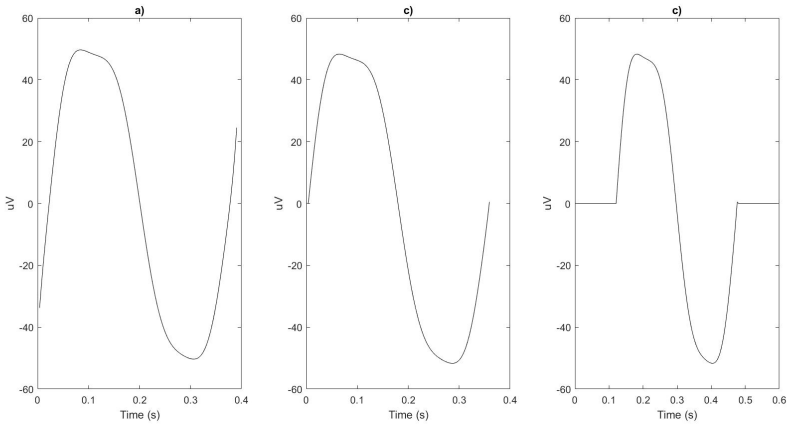

Figure 3.5: a)*New extracted and properly scaled in amplitude prototype;*
b) *Windowed prototype with first and last samples set at zero;*
c) *Prototype windowed and concatenated with two queues of zeros*

Basic idea was to examine cross-correlation of each pair of adjacent scalings. This cross-correlation value should be as equal as possible to the set threshold. If it is too above, it means that two considered scaled prototypes are almost the same and lot of waves will match with both of them; therefore it is possible to discard one of them, in order to reduce the computational complexity of the algorithm. Otherwise, if the cross-correlation value is below the threshold, it may happen that some waves are not recognized because they match with an intermediate scaling of the prototype between two adjacent ones considered. In this case it is necessary to add a scaling. Doing so, the idea is each wave should match with a scaling and be a little bit under the threshold if compared with the neighbours: this method uses the minimum computational complexity necessary to detect all the waves.
Obviously the ideal condition is difficult to achieve in practical terms; therefore, to define the scalings, it was imposed that each pair of neighbours had a cross-correlation value grater than or equal to a threshold set at 0.9, relying on the algorithm's internal routine that recognizes each wave that matches with more than one scalings, and considers it only once.
Therefore, scalings distribution was manipulated, in order to find the lowest number of scalings which allows to respect this condition.
Selected scaling values are the following:

0.3 0.38 0.48 0.61 0.77 0.98 1.2 1.5 1.9 2.4 3 3.5 4

Figure 3.6 presents graphically the chosen scaling values and shows how the constraint about cross-correlation has been respected.

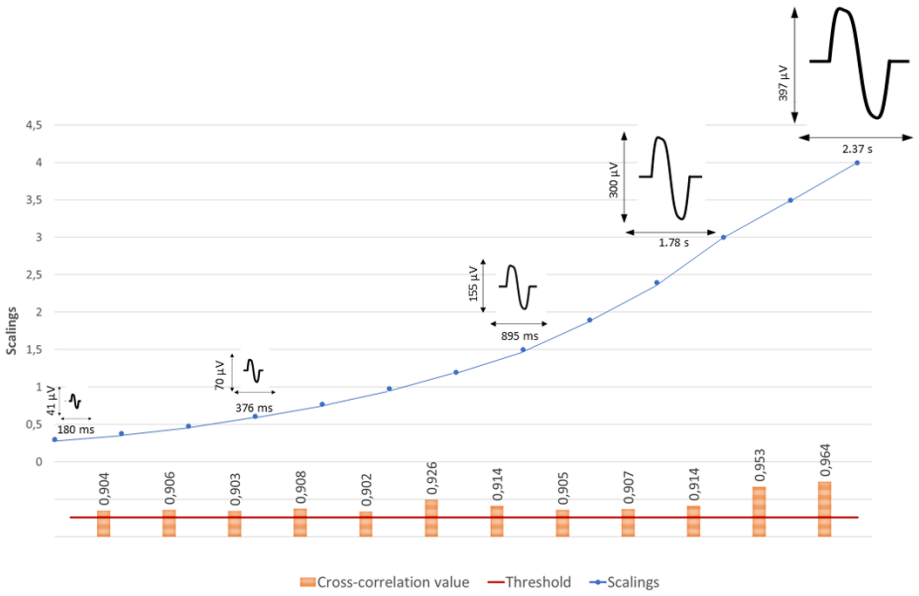

Figure 3.6: *Representation of scalings distribution and some examples of prototypes properly scaled. Bar diagram shows how all cross-correlation values between pairs of scalings are above the threshold set a 0.9. The smallest (180 ms, 41 μV) and the greatest (2.37 s, 397 μV) scaled prototypes are illustrated. Amplitude of each scaled prototype is obtain from its time duration, using the relation reported in equation 3.1.*

3.1.5 Definition of the cross-correlation threshold

The functioning of the algorithm is mainly based on the computation of cross-correlation function between EEG signal and each properly scaled prototype, in order to identify in the signal waves that present an high correlation with the prototype; thus it is possible to consider them SBCs.
In the ideal case, slow biphasic complexes in the signal, will reach a perfect match with almost one of the scaled prototypes, meaning a cross-correlation value equal to ± 1 in correspondence of the samples to which the complexes of interest are placed. Obviously, finding a perfect replica of scaled prototype within the electroencephalographic signal is very unlikely, because of the high complexity of this kind of signal. Because of that it is necessary to impose a threshold, in order to detect those waves which are similar to the SBC prototype.
Thus the choice of the most adequate threshold on the cross-correlation value is crucial. If it is too high a lot of possible SBCs will be discarded because of a low correlation. If the threshold is too low, the algorithm will identify as slow biphasic complexes many waves that are not of interest.
To define the new threshold the decision has been based on the observation of the distribution of cross-correlation values between all slow biphasic complexes selected by M.D. and the properly scaled prototype. Each value has been extracted form the computation of cross-correlation function between each scaled prototype and EEG signal, isolating the cross-correlation values to the samples to which SBCs identified by the physician are located. The maximum cross-correlation value among those calculated from the different scales has been considered.
An histogram representing the latter distribution is reported in figure 3.7.

With regard to the imposition of a threshold on the cross-correlation values, the percentage of waves selected by the doctor that respect the condition has been calculated for several different thresholds. Results are reported in the following table:

Threshold	**Percentage of included waves**
0.9	88.3 %
0.89	90.7 %
0.88	92.7 %
0.87	94.3 %
0.85	96.2 %

After this evaluation **the threshold on cross-correlation was set at 0.89**, as figure 3.7 shows, considering it a good compromise among a not too low threshold and an adequate number of included SBCs.

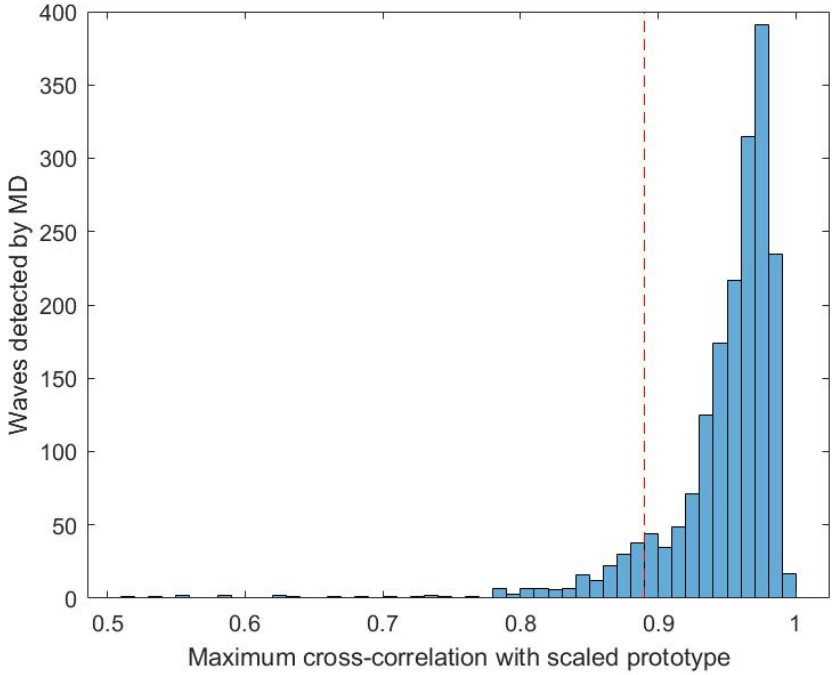

Figure 3.7: *Cross-correlation between scaled prototypes and SBCs manually selected by the physician. The threshold at 0.89 on the cross-correlation values is highlighted in red. The latter allows to include 90.7 % of the total waves.*

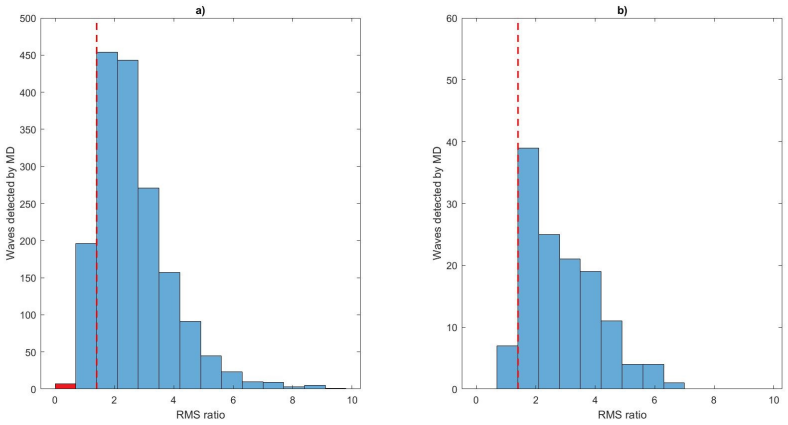

Figure 3.8: *Histograms of the distributions of RMS ratio, computed between each SBC selected by the physician's RMS and the one of the stretch of signal immediately adjacent.* a) *Fast waves (shorter then 1 s) RMS ratio. Waves selected by expert and associated with RMS ratio values lower than 1 (column highlight in red) are clearly outlier: their RMS is lower than the one of the background, meaning that they do not emerge from it, instead they are submerged in it.* b) *Slow waves (longer than 1 s) RMS ratio.*

New extracted prototype and its relate threshold were inserted in the initial and pre-existing algorithm, without changing its functioning, only modifying those parameters.
From now on, the elaboration results that will be presented have been used to extract new slow biphasic complexes' characteristics, useful to implement a post-processing of the algorithm, in order to decide if each wave which passes cross-correlation threshold is actually good or it has to be discarded.

3.1.6 RMS ratio for SBCs emergence from background

Even if literature doesn't mention it, neuropsychiatrists are aware that one of the most important slow biphasic complexes characteristic is that their amplitude must be significantly higher that the background one.

Since it was necessary to quantify this statement, the root mean square (RMS) values of SBCs selected by the M.D. were computed. They were compared with RMS of signal traits immediately adjacent to the wave of interest. To do so EEG signal were previously low-pass filtered with a cut off frequency of 7.5 Hz, as this will be the filter that will be applied in the algorithm method.
Several length of the considered adjacent traits were taken in account: form 50 samples (195 ms) to 256 samples (1 s). For each trait length the ratio

between RMS of each wave and it adjacent stretch of signal were computed. The distribution of all extracted values related to a specific trait length were analysed, by observing the histogram and evaluating the imposition of a threshold. Doing so, it was possible to select the most suitable length of the adjacent stretch of signal and the convenient threshold.
At the end the best solution was to divide manually selected waves in two groups: short SBCs, with a time duration lower than 1 s, and long SBCs, with a time duration equal o higher than 1 s (256 samples).

- Short (or fast) SBCs' RMS have been compared with RMS of an adjacent trait of length 100 samples (390 ms).
- Long (or slow) SBC's emergence have been checked referring to a signal trait of length 256 samples (1 s).

Figure 3.8 shows with histograms the distributions of RMS ratio computed for each selected SBC in the two groups.

It was necessary to impose a threshold, in order to either include or discard waves that gain a cross-correlation match in the algorithm. To do so, percentage of included SBCs selected by M.D. was taken in account.

Fast waves

Threshold	**Percentage of included waves**
1.3	91.6 %
1.4	88.3 %
1.5	84.7 %
1.8	74.2 %
2	65.8 %

Slow waves

Threshold	**Percentage of included waves**
1.3	96.9 %
1.4	94.6 %
1.5	91.6 %
1.8	76.3 %
2	66.4%

Threshold was set at 1.4 for both group, meaning that a wave, **to be selected by the algorithm must present a root mean square that has to be at leat 40% higher than the one of the background, considering as background 100 or 256 samples left and right adjacent to the wave itself**.

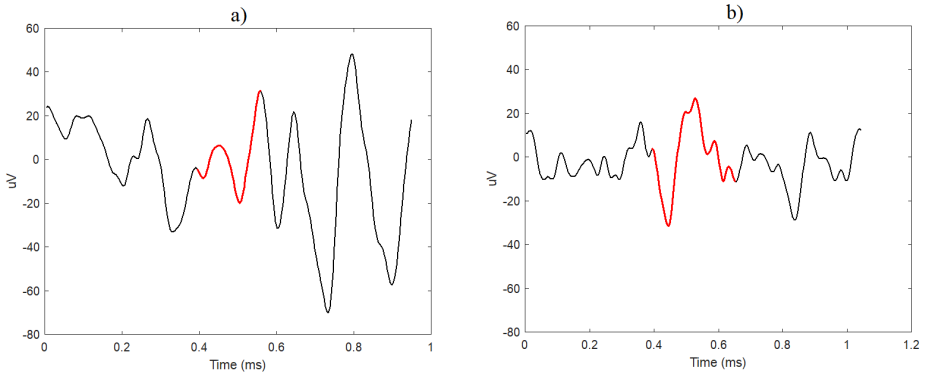

Figure 3.9: *Two fast waves selected by the algorithm, because of their above-threshold cross-correlation value, which are being checked about RMS ratio, representing their emergence from background*

a) *RMS ratio = 0.49. Wave on the left is discarded.*

b) *RMS ratio = 1.7. Wave on the right passes this check. It will be subjected to other controls.*

Figure 3.9 shows two examples of waves selected by algorithm, classified as fast. Both of them has a cross-correlation with the prototype which is above threshold, but just one of them passes the check regarding the emergence from background.

3.1.7 Waves surrounded by high frequency rhythm oscillation

Looking at the waves automatically identified, it was clear that the algorithm also selected some wave surrounded by high frequency rhythm. The latter has been quantified as 1 s oscillation, placed left and right the wave, with at least 40% of power spectrum frequency above 5 Hz.
In physician opinion, even if it is odd, finding some SBCs with this characteristic is not impossible. But they must present a robust emergence from background, stronger than common waves surrounded by slow oscillation.
However, this type of SBC is quite difficult to be identified through the observation of the signal; in fact, among those identified manually there are a very limited number of them. They are reported in figure 3.10.

As they are only twenty-five, extract a statistically significant measure of this kind of SBCs emergence from background is not possible. Therefore, according with the physician, looking at waves selected automatically, the minimum acceptable RMS ratio has been set at 1.8. Meaning that **a wave surrounded by oscillation of frequency higher than 5 Hz must**

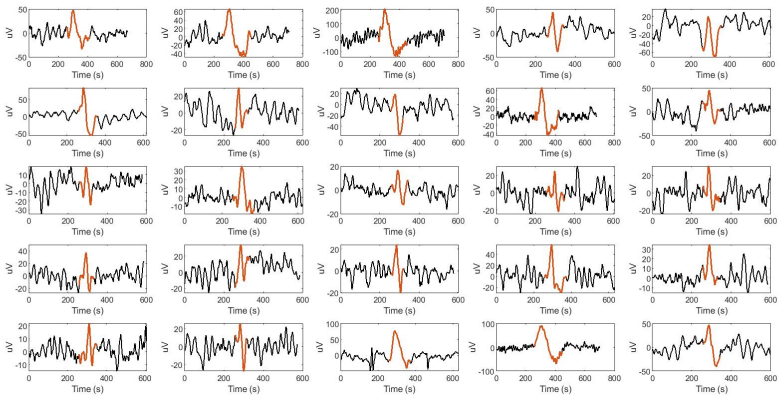

Figure 3.10: *Twenty-five waves selected by the doctor which correspond to the definition of waves surrounded by high frequency oscillation. Left and right adjacent signal straights of length 1 s (256 samples) present at least 40% of their power spectral frequency above 5 Hz.*

present a root mean square at least 80% higher than the one of the background, to be selected by the algorithm.

Figure 3.11 shows two examples of waves that respect this characteristic and that are being checked about their emergence from background.

3.1.8 SBCs spectrum frequency

SBCs must be slow waves. Since there was the will to control the frequency content of waves selected by the algorithm, in order to discard too high-frequency ones, PSD of waves selected by the doctor have been extracted, in order to quantify low frequency power percentage that characterize SBCs. Power percentage distributions extracted from all SBCs are reported in figure 3.12, both under 3 and 5 Hz.

Thresholding has been attempted on both conditions. Imposing that SBCs must present at least 75% of they power under 3 Hz, it is possible to include 77,7% of manually selected ones, while imposing the same threshold on power under 5 Hz, almost all (98.3%) selected SBCs are considered.
Second condition has been inserted in algorithm post-processing, choosing to be less restrictive in this control.
Figure 3.13 shows examples of two waves selected by the algorithm, which cross-correlation values were above-threshold and characteristics passes controls about amplitude and emergence from background, that are checked for the percentage of power in low frequency.

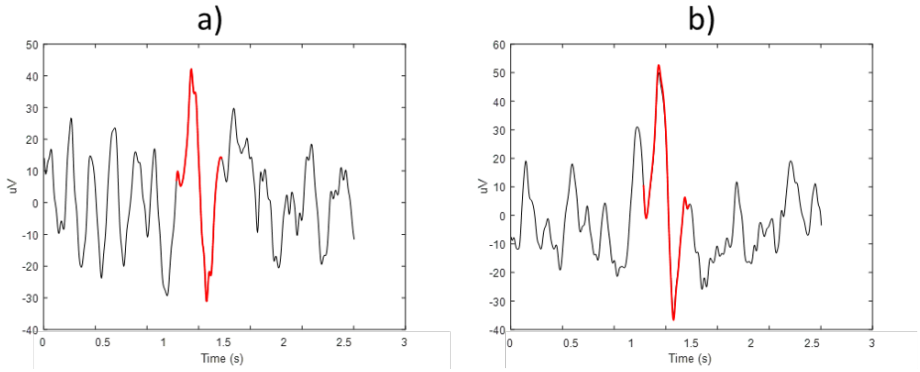

Figure 3.11: *Two waves selected by the algorithm thanks to their above-threshold cross-correlation values, that respect all previous requested characteristics. Both waves are surrounded by high frequency rhythm oscillation and because of that they are being checked about RMS rate, that must be greater with respect to waves surrounded by slow oscillations.*

a) *Wave represented on the left is considered surrounded by high frequency oscillation because its left adjacent trait presents 99% of power spectral density over 5 Hz, while the right one's power spectral density above 5 Hz is 44%. This wave's RMS is only 60% higher that background one (RMS rate = 1.6). So it is discarded.*

b) *Wave on the right is placed in fast oscillations, in fact the preceding and following adjacent segments present 41% and 66% spectral power above 5 Hz, respectively. Relative RMS ratio is 2.3: the wave passes this control.*

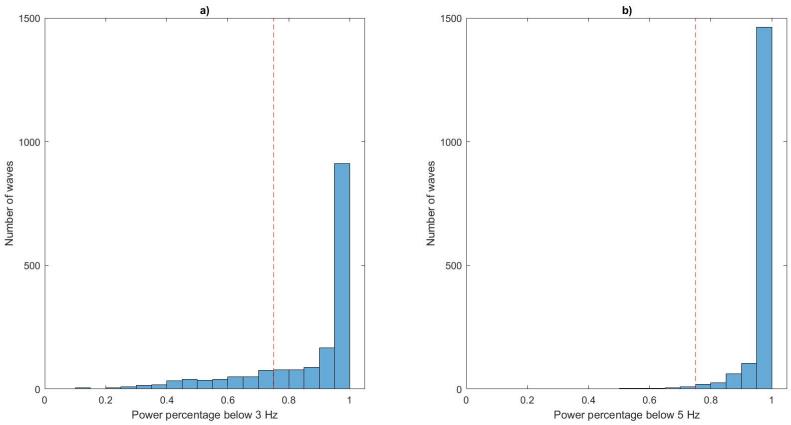

Figure 3.12: *Distribution of low-frequency power percentage extracted from all waves selected by the doctor.*

a) *Power spectral density percentage under 3 Hz. Setting a threshold at 0.75 (thus, imposing that SBCs must present at least 75% of they power under 3 Hz), 77.7% of manually selected waves are included.*

b) *Power spectral density percentage under 5 Hz. Setting a threshold at 0.75, 98.3% of waves selected by the physician are considered.*

3.1.9 Amplitude bounds and artefacts rejection

After first elaborations, it was clear that algorithm also selected a lot of artefacts.
As it was explained in Chapter 2, signals cleaning with different methods has been tried. This attempt failed, as all algorithms that should just delete ocular, muscular and other physiological artefacts, always effected also some SBCs, since they have many features in common with blinks, such as:

- They show an amplitude significantly higher that the background one;
- They appear often on frontal channel;
- They present a bi-frontal waveform.

Because of this, signals have been elaborated without cleaning them from artefacts. Those have to be discarded after the algorithm application, discriminating them from all selected waves.
To do so a threshold on the amplitude has been imposed. Being aware that amplitude of SBCs is strictly related to their time duration, amplitude bounds were imposed correlating them to length of the time support of each wave.

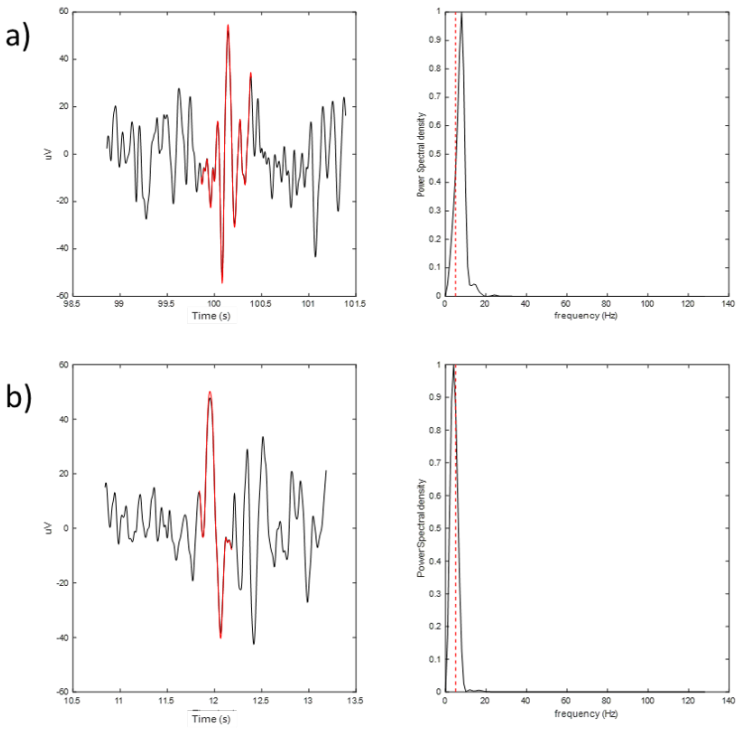

Figure 3.13: *Two waves selected by the algorithm, because of their above-threshold cross-correlation value and their emergence from background, which are being checked about their power spectral density.*

a) *On the left a wave selected by the algorithm, on the right the rappresentation of its power spectral density normalized with respect to its maximum value. 5 Hz line is highlighted. This wave is discarded during post-processing because only 25% of its spectral power is located under 5 Hz.*

b) *A wave selected by the algorithm and the relate power spectral density normalized with respect to its maximum value. In this case the wave will pass spectrum density check, as 75% of its spectral power is located under 5 Hz. It will be subjected to further controls.*

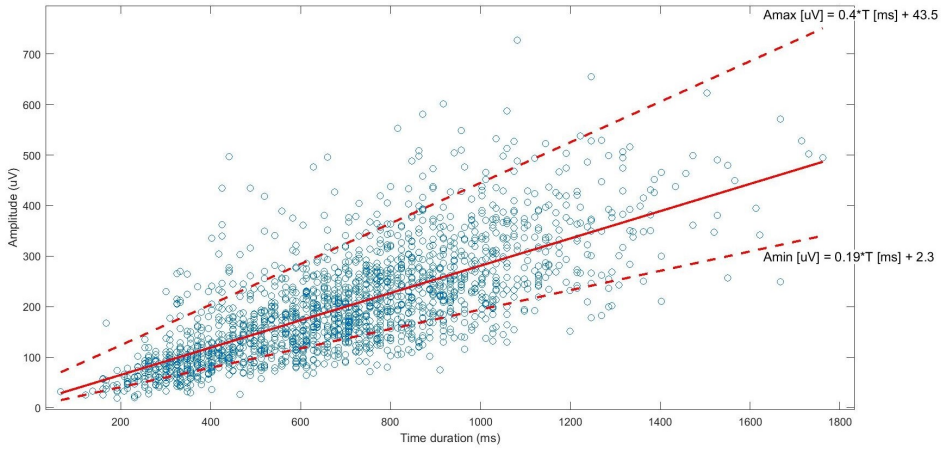

Figure 3.14: *Relation between duration (in ms) and amplitude (in μV) regarding all slow biphasic complexes selected by the M.D.. Amplitude bounds are represented. Applying this limits allows to include 80% of SBCs manually selected.*

Thus, if the algorithm identifies a short wave, this one will be discarded, considering it an artefact, if its amplitude is above a certain value. In another case, the very same amplitude value could be acceptable, if it is associated to a longer detected wave.

To define this bounds, distribution reported in figure 3.2 was the starting point. The definition of the upper amplitude limit was done by trial and error, observing the waves selected by the algorithm that represented artefacts rather than SBCs, and trying to interpolate all their amplitude values, in order to draw a line that was able to be a bound that excluded bad and kept good waves.
At the same time a lower bound has been defined as well, in order to exclude all those waves with a waveform similar to SBCs but whose amplitude weren't high enough.
Initially limits were established as parallel lines to the trend line reported in figure 3.2, keeping the same slope and making the intercept vary. Afterwards, it was decided not to keep the limits parallel to the trend line, in order to include more actual SBCs and to be more selective on the lower limit with regard to low amplitudes. Amplitude upper and lower limits are defined as:

$$A_{max} = 0.4 \times T + 43.5 \tag{3.3}$$

$$A_{min} = 0.19 \times T + 2.3 \tag{3.4}$$

Where A is wave's amplitude in μV and T is wave's duration in ms.
Figure 3.14 shows the defined lines.

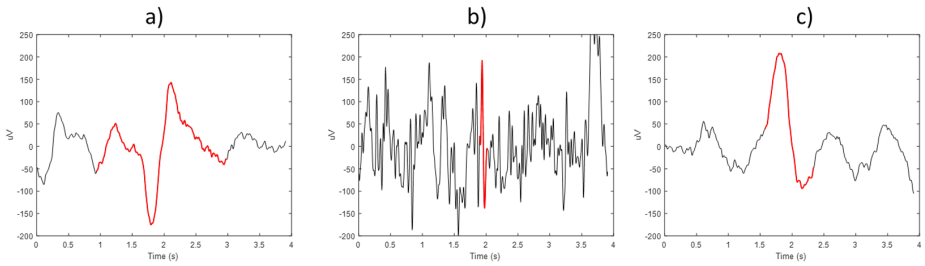

Figure 3.15: *Three waves selected by the algorithm. All of them present an amplitude between 300 and 350 μV.*

a) *Wave at the left presents an amplitude of 318 μV and a time duration of 2.04 s. It is discarded because its amplitude is below the lower limit, meaning that it has a width too small compared to its duration;*

b) *Center wave has a 330 μV amplitude and a 121 ms time duration. It is discarded as well, because its amplitude is placed above the upper limit, meaning that it is too wide with respect to its length;*

c) *Right wave is accepted by the algorithm post-processing, because it presents an amplitude of 302 μV and a time duration of 770 ms, which is included in both limits.*

Figure 3.15 reports some waves selected by the algorithm. They all have similar amplitudes, but just one of them passes amplitude check, because of the existence of a relation between amplitude limits and time duration of waves.

3.2 Algorithm upgrading

Collection of extracted characteristics about SBCs has been incorporated in the algorithm itself or in the post-processing. The aim was to select waves similar to SBCs and afterwards distinguish actual SBCs and what is not, despite the similar waveform.
In addition to the controls already explained, absolute limits have been introduced to discard waves that are too short, too long or insufficiently wide.

Here's an overview of requested characteristics:

Amplitude	Max	/
	Min	70 µV
Duration	Max	250 ms
	Min	1.6 s
Amplitude vs Duration	Max	Amax [µV] = 0.4 * T [ms] + 43.5
	Min	Amin [µV] = 0.19 * T [ms] + 2.3
Minimum RMS ratio wave/bacground	Slow oscillation (< 5 Hz)	1.4
	Fast oscillation (≥5 Hz)	1.8
Minimum power spectrum density below 5 Hz		75%

Moreover, a further control has been implemented, to recognize waves selected by algorithm that are poli-phasic, rather then biphasic.
To do so, waves have been low-pass filtered (5 Hz cut-off frequency) and, using a dedicated function, its local maximums and minimums have been calculated. If the number of local maximum and minimum sufficiently far each other was 3 or less, the wave has been considered biphasic. Two critical points are considered too near (and they are count together as once) if their distance on y-axes is lower then 10% of the total wave amplitude.
Figure 3.16 shows an example.

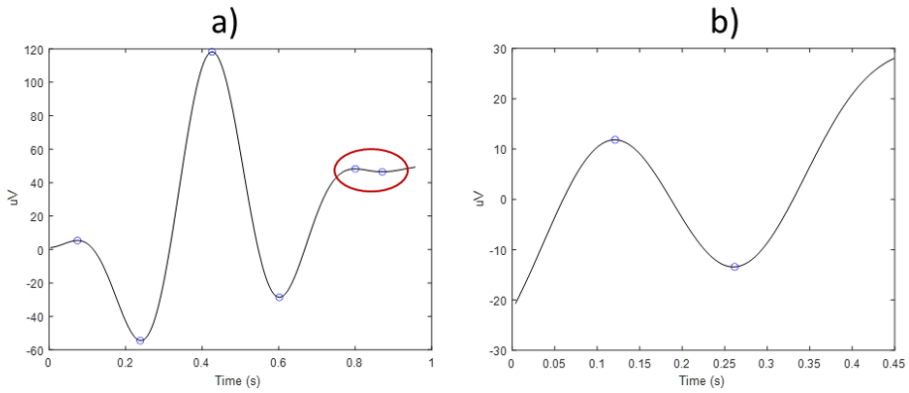

Figure 3.16: *Two waves selected by algorithm, that respect all previous requested characteristics. They have been low-pass filtered with a cut-off frequency of 5 Hz.*

a) *Left wave is considered a polyphasic one (triphasic in this case), so it is discarded. Five critical points are counted, because the two circles in red are worth as 1, in fact their distance in amplitude is less than 10% of the total.*

b) *Right wave is a biphasic (correct) one. Critical points are two.*

3.3 Validation of the improved algorithm

The physician has validated as many EEG traces as possible, consistent with his schedules, using the revision GUI presented in the chapter 2.7.

His general conclusion pointed out that the algorithm is quite well able to recognize SBCs in EEG traces of patients in non-serious conditions. On the contrary, in very slow signals, related to an advanced state of the pathology, method selects several false positive waves. In these EEGs many slow waves are easily mistaken with SBCs, both by the medical expert and by the automatic method.

Totally twenty-one EEG singals have been validated, from eleven different patients. Since validation process is very time-consuming, signals have not been validated in their entirety. Only the first 400 seconds of EEGs were analysed, preferring to examine a larger number of patients. Total validation results are reported in the table below:

	Number of waves	**Percentage**
Validated waves	5310	100%
True Positive	1394	26.6%
False Positive	3809	71.7%
Non-classified	107	2%

Moreover, validation results have been analysed taking into account the severity score assigned to each trace. As it was explained in chapter 2.8, each path has been assigned from the expert a severity score, ranging from 0 (normal condition) to 4 (serious pathology). Averaged value of true positive, false positive and non-classified percentages, divided for severity degree, are shown in the following table:

	Severity 1	Severity 2	Severity 3	Severity 4
TP	30.9%	16.2%	31.7%	21.16%
FP	69.1%	83.7%	65.8%	75.9%
NC	0%	0%	2.5%	2.9%

Looking at this results it is not possible to declare if the algorithm has a better performance with patient with mild disease rather then ones with severe pathology.

In the following some reflection about SBCs selection in EEG traces used as test and presented in Chapter 2.4.

1. First test EEG trace, classified with Severity score 4, characterized by low frequency and high amplitude, is the one that present the highest number of mistakes.
   The algorithm has found a total of 1048 waves in twelve EEG channels with a duration of 1147 seconds (about 19 minutes). The physician has analysed first 440 seconds, validating a total of 427 waves. Results are reported below:

	Number of waves	Waves per period of 10 s and channel	Percentage
Validated w.	427	0.82	100%
TP	92	0.15	21.5%
FN	325	0.62	76.1%
NC	10	0.019	2.3%

   As figure 3.17 shows, lots of slow waves are mistaken for SBCs. This is quite understandable, as all oscillation in this signal present common characteristics with SBCs, like slowness and larger than normal amplitude. Because of that, it is easy to erroneously consider each bi-frontal waves as a slow biphasic complex.
   While the classification of this trace as a serious pathological one is quite immediate, the punctual evaluation of single waves is hard even in the opinion of the physician.

2. The second path is characterized by severity score 1 and a few slow biphasic complex, not easy to identify, in the opinion of the expert, because their particularly low amplitude with respect to other patients' SBCs.
   Automatic selected waves are 145 in 814 seconds (13.5 minutes) of signal, composed of twelve channels. The physician has validated first 47 found waves in 440 seconds of signal. Algorithm results are the following:

	Number of waves	Waves per period of 10 s and channel	Percentage
Validated w.	47	0.089	100%
TP	13	0.025	27.6%
FN	34	0.064	72.3%
NC	0	0	0%

   Figure 3.17 shows 30 seconds of this signal, highlighting selected by the algorithm waves.

3. Third and last test EEG signal has been acquired from a patient in quite normal condition, infact it is normally fast and its amplitude is quite low. Despite the quite good condition, in the opinion of the expert, there are more SBCs with respect the second test EEG, and they presnt a clearer and more easy to recognize waveform.
   Results of the algorithm are in agreement with this consideration, in fact, although this signal has a slightly shorter duration than the second one, a greater number of waves have been identified: 245 waves have been selected in 777 seconds (13 minutes), in twelve channels.
   The doctor observed first 440 seconds, considering 94 waves.

	Number of waves	Waves per period of 10 s and channel	Percentage
Validated w.	94	0.017	100%
TP	29	0.082	30.8%
FN	65	0.036	69.2%
NC	0	0	0%

   Once again, figure 3.17, shows 30 seconds of this signal and the relative found waves.

Evaluation of this three representative signals support the qualitative initial statement of the physician: the algorithm makes more errors in the selection of waves in the EEG acquired from patients in more serious conditions.
What we can observe is that, even if number of false negative selected waves is still very high, overall number of found waves is correctly related to the patient condition and the proportions between the number of waves found in the three different EEGs correspond to what we expected.

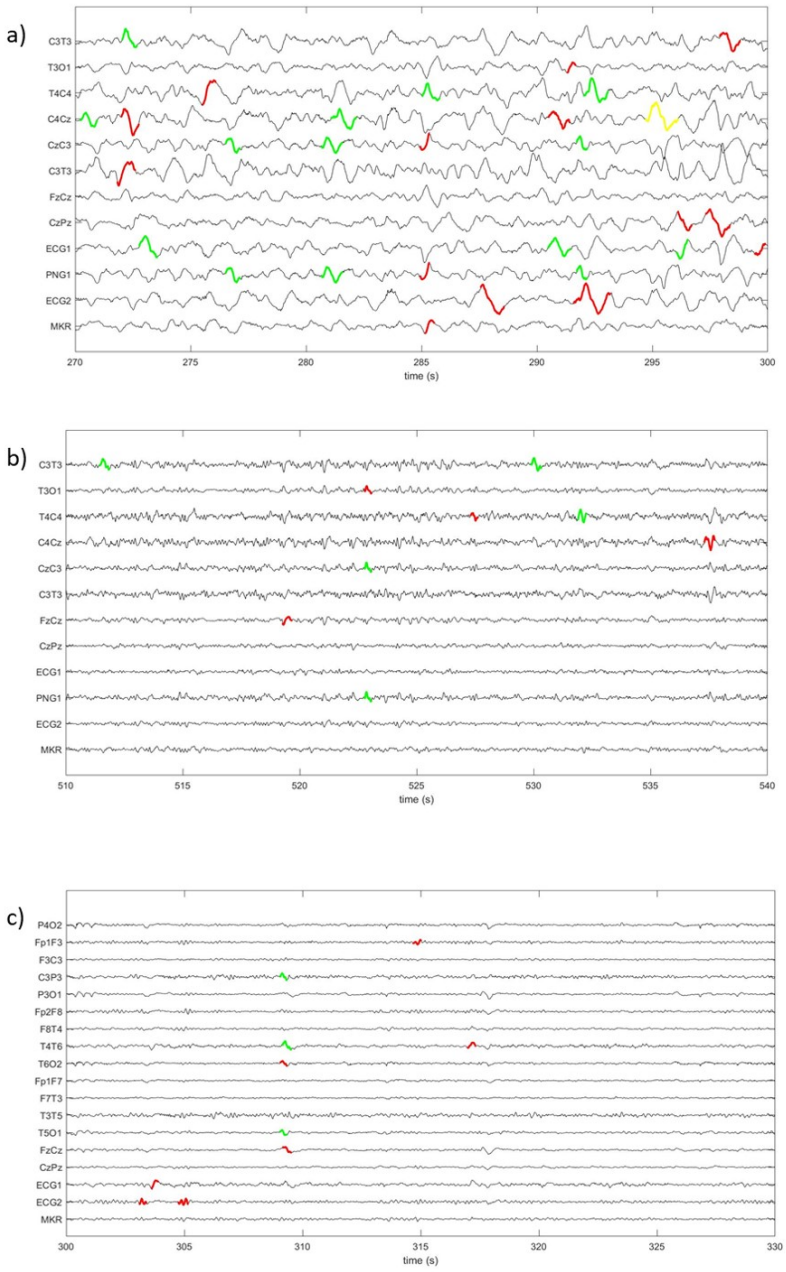

Figure 3.17: *Thirty seconds of each test EEG signal, with waves found by algorithm highlighted as true positive (green), false positive (red) or non-classified (yellow).*
Each signal present a montage 3, the first one is represented with a space between two channel of 300 μV, while other two show 200 μV per channel.
a) *First EEG trace, attributable to the patient in more severe conditions. It is possible to observe how the algorithm selects different false positives, which share different characteristics with actual SBCs.*
b) *Second EEG trace, similar to a normal signal, with few actual SBCs*
c) *Last test EEG, the patient's condition is not serious, despite the presence of some SBCs, which have a well-defined waveform.*

3.4 Algorithm test

In order to test the algorithm, the characterization of patients with different severity of the pathology has been attempted. To do so, five measures have been extracted from single EEG traces, starting from SBCs identified automatically by the algorithm. Computed measures are mean number of SBCs found in each channel every 10 seconds, number of waves found on frontal channels only every 10 seconds, mean RMS value of SBCs found on all channels or just in frontal ones (normalized per number of channels and 10 s periods) and the standard deviation of the distance (in seconds) between consecutive found SBCs. Formulas are presented in Chapter 2.8.
Ideally, this five features should characterize differently five category of EEG acquired from patients and a control group. A physician assigned each EEG trace an evaluation, depending on the severity of the disease. Classes goes from 0 to 4 and represent respectively: normal conditions, mild, moderate, severe and serious pathology.
Distribution of each measure values is analysed, in order to distinguish different classes of severity.
Figure 3.18 presents distributions of the five extracted features regarding severity classes into which patients and controls have been divided.

Kruskal-Wallis test has been applied on each pair of distributions related to near classes (Severity 0 vs. Severity 1, Severity 1 vs. Severity 2...), to identify couples of distributions that are significantly different each other.
Kruskal-Wallis method (also called "one way ANOVA on ranks" test) is a rank-based non-parametric test, that verifies the statistical difference of the medians of two distributions, proving that they belong to two different populations. It returns a p-value, which is able to accept or reject the null hypothesis, according to which the two samples were extracted from populations having the same median.
To apply Kruskal-Wallis, data must respect three assumption:

1. Dependent variables have to be measured at ordinal or continuous level.

2. Independent variables have to be categorical independent groups.

3. Independence between observation.

As it is possible to assume that our data respect this hypothesis, Kruskal-Wallis test is usable.
A resulting p-value lower than 0.05 represents a significantly difference between two distributions, while if it is lower than 0.01, distributions are considered highly significantly different [13].

With regard to distributions of Severity 0 and Severity 1 of each extracted

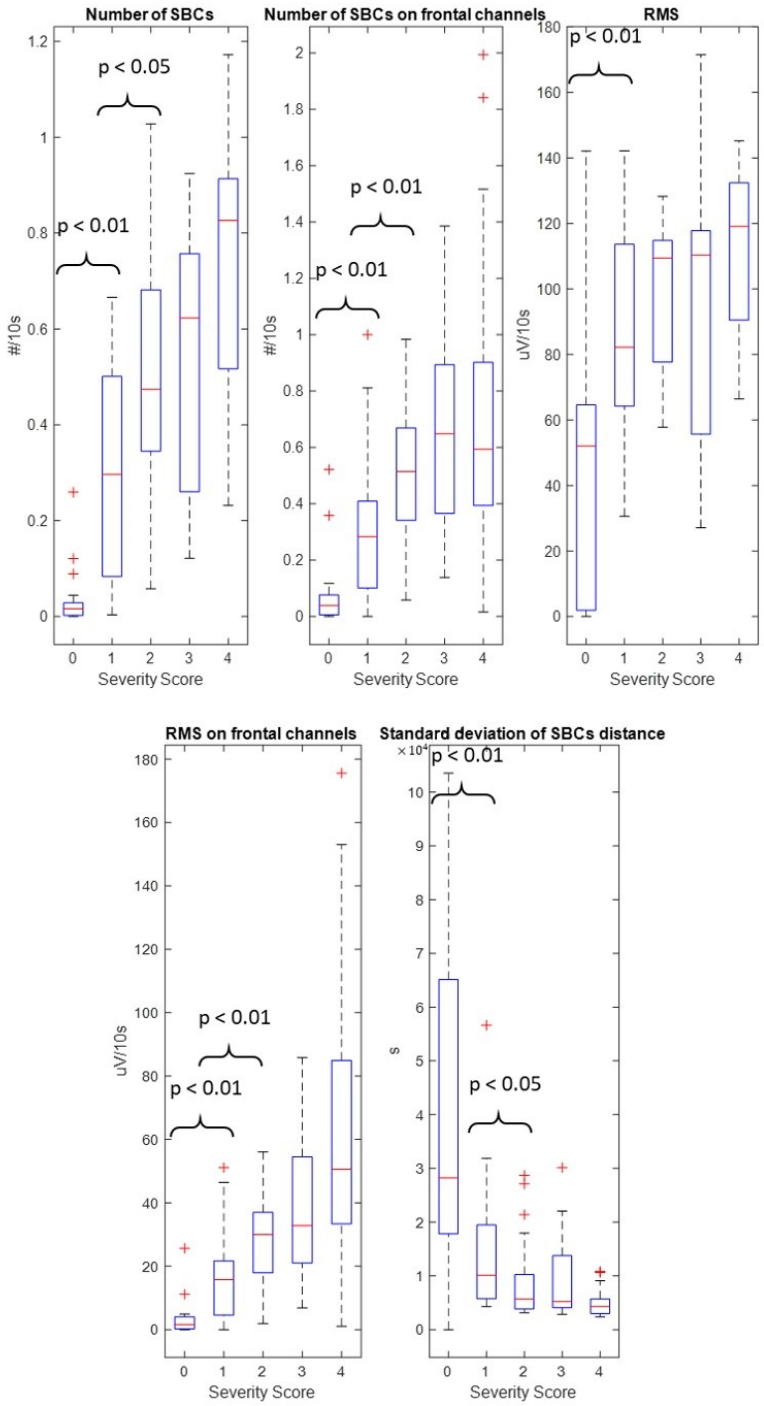

Figure 3.18: *Box-plots comparing distributions of values extracted from each EEG traces for different classes of severity: normal condition and mild, moderate, severe and serious illnesses respectively.*
Number and RMS value of SBCs (both total and in frontal channels only) are normalized for number of channel and time duration of the signal (periods of 10 seconds).
If two adjacent distribution are significantly different according to Kruskal-Wallis test, upper limit of p-value is reported, to indicate if the difference is either significant ($p<0.05$) or highly significant ($p<0.01$).

SBCs measures, p-value resulting from Anova test is always lower than 0.01, meaning that indexes of health subjects and lightly pathological ones are in any case statistically highly significantly different.
The latter and all other significant difference between couple of distributions are reported in figure 3.18, and summarize in the following.

		SBCs measures				
Comparison distributions		Number of SBCs	Number of SBCs in frontal channels	RMS of SBCs	RMS of SBCs in frontal channels	Standard dev. of SBCs distances
Severity 0 vs Severity 1	Normal vs Mild desease	✓✓	✓✓	✓✓	✓✓	✓✓
Severity 1 vs Severity 2	Mild vs Moderate desease	✓	✓✓		✓✓	✓

Where single V indicates a p-value lower than 0.05 (significantly difference), and a double V indicates a p-value lower than 0.01 (highly significantly different).

As in almost all cases two distributions belonging to non-adjacent classes (0 vs. 2, 1 vs. 3, 2 vs 4), are significantly different, a three severity classes division has been attempted.

Severity 0 Normal conditions

Severity 1 Not severe pathology: combines severity classes 1 and 2 from five-classes division.

Severity 2 Serious pathology: combines severity classes 3 and 4 from five-classes division.

Figure 3.19 uses box-plots to show distributions of measures extracted from each EEG signal, divided into three classes.

In this case *all* couple of adjacent classes are related to distributions that present statistically different medians, according to Kruskal-Wallis test. All differences between adjacent distributions result highly significant, except distributions relative to Severity classes 1 and 2 of SBCs RMS value, normalized per channel and period of 10 seconds.

As it is clear both observing the box-plots and the p-values, the most important differences are found between distributions relative to healthy subjects (severity class 0) and pathological ones (severity class 1, both in case of division into five and three classes). Figure 3.20 shows box-plots of distributions concerning the division in 2 classes: healthy (Class 0) and pathological (Class 1) subjects. Class 0 remains unchanged with respect to

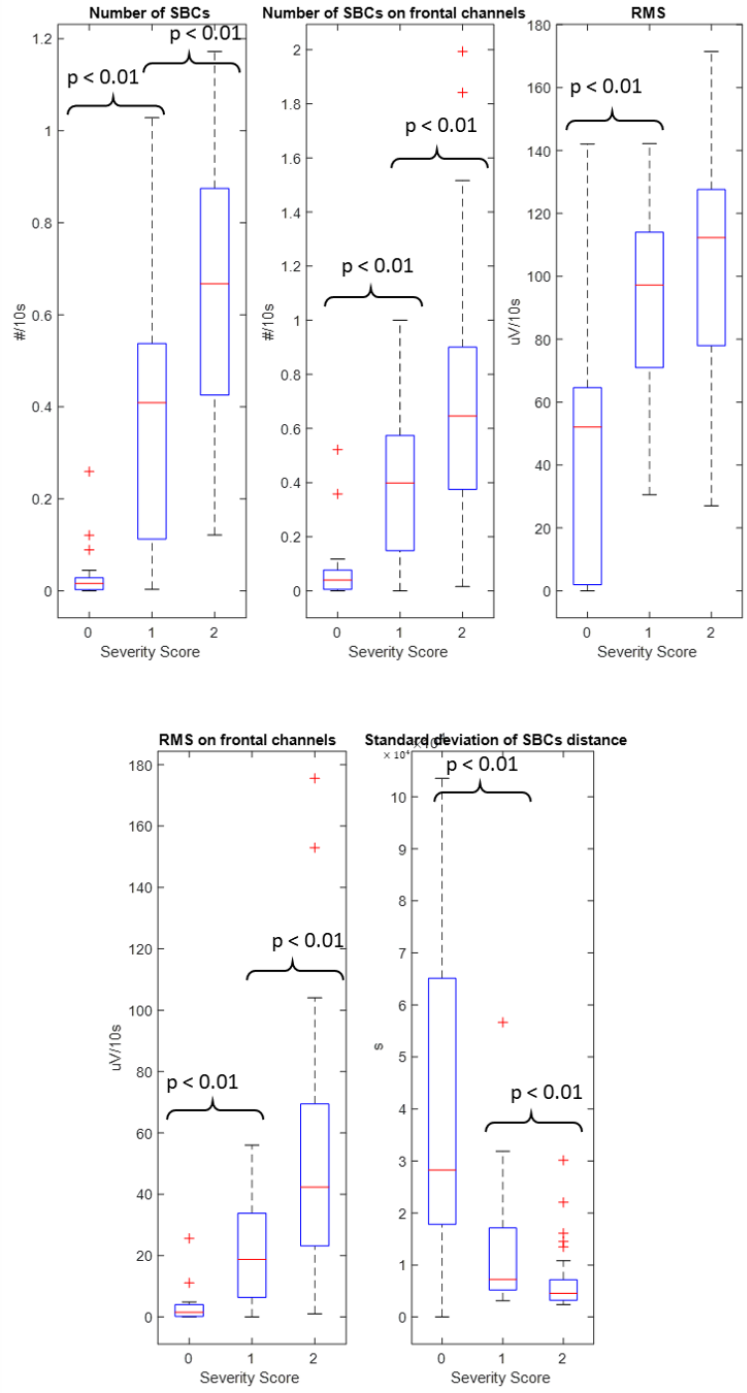

Figure 3.19: *Box-plots representing distribution of SBCs measures extracted form EEG signalas divided in three severity classes: normal conditions, non severe pathology, serious pathology respectively. Number and RMS value of SBCs (both total and in frontal channels only) are normalized for number of channels and time duration of the signal (periods of 10 seconds).*
Statistical differences between couples of distributions is highlight, reporting upper limit of p-values resulting from Kruskal-Wallis test.

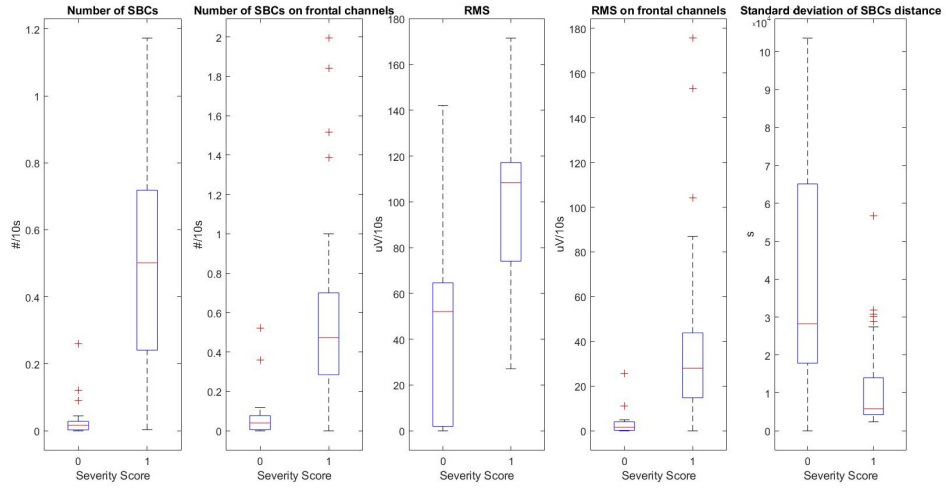

Figure 3.20: *Box-plots showing distributions of different SBCs measures, divided into two classes: healthy and pathological subjects. All distributions related to class 0 and class 1 are statistically highly different to each other, according to Kruskal-Wallis test.*

previous cases. The pathological class collects all the EEG traces that have been assigned a severity score greater than zero.

For all variable, distributions of Class 0 are statistically highly significantly different with respect to the distribution of Class 1, with a p-value always lower than 3×10^{-5}.

Because of this, the construction of a classifier has been tired, using the binary target (0 = healthy, 1 = pathological). To do so a construction of a binary decision tree has been attempted using Matlab. Decision tree structure obtained using the entire data set for the construction is reported in figure 3.21.

Classifying the entire data set (composed of 22 subjects in normal condition and 84 pathologic patients, thus excluding the non-classified ones) results reported in confusion matrix below have been obtained.

		Predicted		
		0	**1**	**True rate**
Target	**0**	20	2	**90,9 %**
	1	0	84	**100 %**
	Predictive value	**100 %**	**97,7 %**	

As it is possible to observe form the confusion matrix, 95.5% of EEG traces are correctly classified as normal or pathological. Moreover, it was possible to reach a 100% true rate with regard to pathological class, meaning that all signals classified in Class 1 were actually pathological. Predictive value of Class 0 reaches as well 100%, signifying that there isn't any pathological EEG trace that as been classified as normal.
It is possible to notice that reported true rate values represent respectively specificity and sensitivity, useful if this this classifier is considered as a screening test. Where specificity, equal to 90.9%, is the probability that a healthy subject is negative to the test (Class 0); sensitivity, equal to 100%, is the probability of a pathologic subject testing positive for the test (Class 1).

Subsequently, classification method with binary decision tree has been validated using a leave one out approach, obtaining the confusion matrix reported below:

		Predicted		
		0	1	True rate
Target	0	16	6	72,7 %
	1	3	81	96,4 %
	Predictive value	84,2 %	93,1 %	

As was to be expected, the results worsen but remain satisfactory. Average percentage of correct classified is 82.7%. As can be noticed decision tree manages to classify pathological subjects quite well, while it makes more mistakes in classifying subjects under normal conditions.

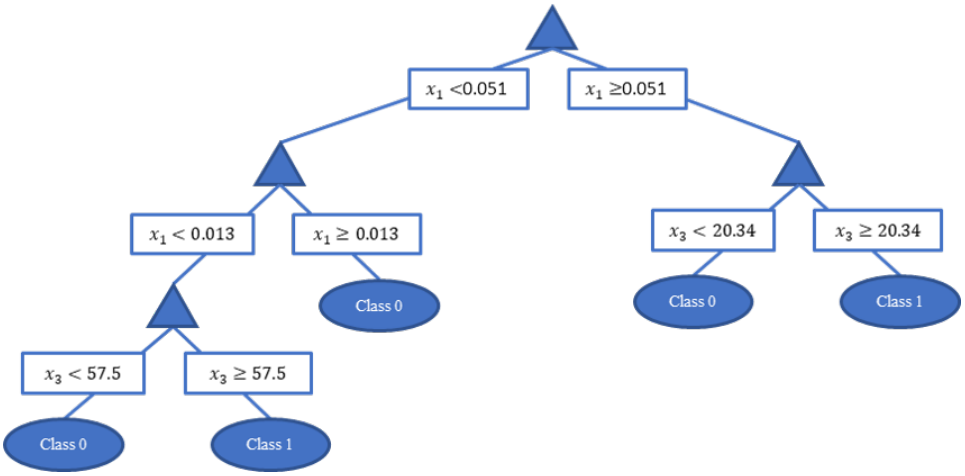

Figure 3.21: *Decision tree built with Matlab, in order to classify normal condition subjects (Class 0) and pathological ones (Class 1). The classifier uses variables x_1 as Mean number of found SBCs in each channel, every 10 seconds [#/10s], and x_3 as Mean RMS value of found SBCs, normalized per total number of channels and periods of 10 seconds [μV/10s].*

Discussion and conclusion

This is, for sure, only the first step of a study that could be carried out for a long time.
The algorithm still selects a significant number of false positive. However, in expert physician's opinion, to present day, the algorithm excludes a low number of SBCs actually present, selecting even those of low amplitude or ones which are difficult to detect by observation of signal alone. Artefacts rejection has certainly improved compared to the starting algorithm.
As it is obvious, clinicians do not rely on detection of SBCs only to evaluate an EEG signal. They also rely heavily on the evaluation of slow activity, and the distinction between physiological, inflammatory or pharmacological activity. Because of that, this instrument, once sharpened, would in any case not be able to classify a patient independently, but it could be used alongside a doctor to facilitate and streamline his work. For example, it could be used to train inexperienced doctors in SBCs identification.
In any case, developments of methods able to elaborate EEG data, are for sure interesting. In fact, EEG is the cheaper and less invasive method useful to evaluate a encephalitis patients; it allows a follow-up of the patient, monitoring the evolution of his/her condition. Moreover, EEG is the only possible examination that provides functional information, not anatomical ones. This aspect is extremely important when dealing with a disease that can cause loss of consciousness and neurological complications.

Moreover, this study has contributed to the detailed technical and mathematical characterisation of waves of interest, which until now had been described superficially and mainly from a clinical point of view only. This will be helpful in the standardization of subjective evaluation of EEG, and wave recognition by neurophysiologists. As we have been able to experience during the collaboration, often two experts disagree on the classification of some waves. This is confirmed by the considerable number of non-classified waves in the validation phase.
Having well-defined guidelines, with numerical indexes, could result very useful in the recognition of SBCs by signal observation.

If this study is meant to be carried out, it would be interesting to attempt

the characterization of slow biphasic complexes typical of different brain region. Brain could be divided either anatomically or functionally. In fact, it is clear that SBCs occurring on the channels related to the frontal regions present different characteristics with respect ones relative to parietal region. To do so, a large number of data would be necessary, in order to repeat all characterization steps for each brain area, taking in account a number of data sufficient to submit statistical considerations.
In addition, physiologists claim that within the same patient's EEG traces, waveforms always occur similar to each other. It would be worthwhile to try to personalize the research method for each patient. In this case the challenge would be to keep the method automatic, but capable of adapting to the individual patient.

Bibliography

[1] Encephalitis. *Paediatrics child health*, 3(1):47, 1998.

[2] A. K. Abbas, J. C. Aster, et al. *Robbins and Cotran pathologic basis of disease*. Elsevier/ Saunders,, 2015.

[3] A. Beaumanoir, D. Grioni, G. Kulimann, A. Tiberti, and D. Valseriati. Anomalies eeg dans la phase prémonitoire du syndrome de rasmussen. à propos de deux observations. *Neurophysiologie Clinique/Clinical Neurophysiology*, 27(1):25–32, 1997.

[4] S. Buzio, E. Bortone, G. Buccino, C. Giorgi, A. Maravita, U. Scoditti, and D. Mancia. Periodic eeg activity in the course of limbic encephalitis. a case report. *Electroencephalography and Clinical Neurophysiology*, 1(103):80, 1997.

[5] W.-D. Chang, H.-S. Cha, K. Kim, and C.-H. Im. Detection of eye blink artifacts from single prefrontal channel electroencephalogram. *Computer methods and programs in biomedicine*, 124:19–30, 2016.

[6] L. Chen, M. Zhu, H. Zhou, and J. Liang. Clinical study of west syndrome with ps and late-onset epileptic spasms. *Epilepsy research*, 89(1):82–88, 2010.

[7] P. Fries, D. Nikolić, and W. Singer. The gamma cycle. *Trends in neurosciences*, 30(7):309–316, 2007.

[8] A. Gatti, M. Guarneri, C. Motto, P. Vigliano, and A. Beaumanoir. Slow biphasic complex: electro-clinical considerations. 18, 01 1997.

[9] G. Gómez-Herrero. Automatic artifact removal (aar) toolbox v1. 3 (release 09.12. 2007) for matlab. *Tampere University of Technology*, 2007.

[10] D. Grioni, M. Contri, and G. Kullmann. Are slow biphasic waves (sbw) a possible bioelectrical marker of acute structural cerebral events? report of four cases with acute symptomatic seizures at the onset. *Bull Lega It Epil*, 138:23–24, 2009.

[11] O. Mecarelli. *Manuale teorico pratico di elettroencefalografia*. Wolters Kluwer Health, 2009.

[12] L. Mesin. *Neuromuscular System Engineering*. La community di ilmiolibro.it. ilmiolibro self publishing, 2019.

[13] L. Mesin, M. Valerio, A. Beaumanoir, and G. Capizzi. Automatic identification of slow biphasic complexes in eeg: an effective tool to detect encephalitis. *Biomedical Physics & Engineering Express*, 5(4):045006, 2019.

[14] L. Mesin, M. Valerio, and G. Capizzi. Detection and assessment of encephalitis from eeg. pages 1–4, 2019.

[15] M. Sazgar and M. G. Young. Overview of eeg, electrode placement, and montages. In *Absolute Epilepsy and EEG Rotation Review*, pages 117–125. Springer, 2019.

[16] D. L. Schomer and F. L. Da Silva. *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2012.

[17] R. Sutter, R. D. Stevens, and P. W. Kaplan. Clinical and imaging correlates of eeg patterns in hospitalized patients with encephalopathy. *Journal of neurology*, 260(4):1087–1098, 2013.

[18] K. Tanoue, H. Oguni, N. Nakayama, K. Sasaki, Y. Ito, K. Imai, and M. Osawa. Focal epileptic spasms, involving one leg, manifesting during the clinical course of west syndrome (ws). *Brain and Development*, 30(2):155–159, 2008.

[19] M. Valerio. Patologia infiammatoria del snc in età evolutiva e complessi lenti bifasici (clb): Quale rapporto? analisi e implementazione del riconoscimento delle anomalie eeg.

Acknowledgements

Per primi ringrazio senz'altro i miei genitori e mia nonna che, nonostante in gioventù non abbiano avuto la possibilità di studiare, sono riusciti a infondere in me un'immensa passione per lo studio e la scoperta, rimanendo sempre, instancabilmente e inesorabilmente i miei primi e più gradi sostenitori.

Vorrei, inoltre, spendere due parole per ringraziare ed esprimere la mia ammirazione nei confronti del mio relatore Luca Mesin e del mio correlatore Massimo Valerio, medico neuropsichiatra appassionato di ricerca e d'ingegneria. Entrambi, immensamente appassionati al proprio lavoro, non hanno mai mancato di farmi sentire al loro pari, dimostrando sempre che le mie idee e il mio lavoro erano valevoli quanto i loro.

Infine ringrazio chiunque sia entrato a far parte della mia vita in questi lunghi anni universitari, a partire da chi mi ha strappato un semplice sorriso, fino a chi è riuscito a lasciare un segno indelebile nel mio cuore.