POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'informazione

Corso di Laurea Magistrale in Ingegneria Biomedica

Analyzing Seizure-related Brain Synchronicity levels from Stereo-EEG signals: a PCA-based K-means Clustering Approach

Relatori: **Prof.ssa Anna Maria BIANCHI**
Dott. Lino NOBILI

Tesi di Laurea di:

Juan Camilo AVENDANO DIAZ

Matricola. **798195**

Anno Accademico 2014 - 2015

Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Epilepsy is a complex neurological disease in which seizures constitute the main clinical manifestation. Nowadays it is one of the most frequent neurological disorders and affects around 50 million people worldwide. For about two thirds of the patients, seizures can be controlled through the use of antiepileptic drugs. However, for the remaining one third, the surgical resection of the region responsible of the production and propagation of seizures (Epileptogenic Zone, EZ) may be the only way to suppress or reduce seizures.

Bancaud & Talairach, as part of the Saint-Anne group, developed a methodology called Stereoelectroencephalography (SEEG), which aims to establish the extent of the EZ, in order to plan its surgical resection. SEEG is ideal for studying the relationships between the structures concerned in seizure production and propagation, and can potentially contribute to the understanding of the origin and spread of seizures.

The present study, aims to obtain a quantitative description of the evolution of the synchronicity levels among brain areas before and during seizures by analyzing SEEG data.

CLINICAL DATA

The data used in the present study comes from the pre-surgical evaluation performed in 14 neurosurgical patients affected by Focal Cortical Dysplasia Type II. All the patients had a history of drug-resistant epilepsy and were candidates for the surgical removal of the epileptic focus. The recordings were performed during sleep by stereotactically implanted depth multi-lead electrodes (Stereo-EEG, SEEG). The SEEG signals were recorded using a 192-channel recording system (NIHON-KOHDEN NEUROFAX-110) with a sampling rate of 1000Hz. For each patient and recording, the information regarding the temporal moments in which the seizures started (ictal events) were provided by an expert neurologist. This info was obtained from the videos of the patients corresponding to each recording (Video-SEEG). Finally, bipolar montages formed by adjacent SEEG leads (of the same depth-electrode) in grey matter were suggested by the neurologist in order to minimize the

common electrical noise and maximize the spatial resolution. All the recordings and data treatment procedures were approved by the corresponding Ethical Committee and all the patients provided a written informed consent.

METHODS

After converting the SEEG data from the Nihon-Kohden format into Matlab arrays and calculating the bipolar montages suggested by the expert neurologist, standard pre-processing techniques were applied over the data under the Matlab environment using the EEGLAB toolbox. The data was down-sampled to 250Hz in order to reduce the computational cost of the processing. The resampling was performed after low-pass filtering (120Hz, order: 110) to avoid aliasing. Then two notch filters (at 50Hz and 100Hz, order: 414 each one) were used in order to suppress the noise of the electrical power line and its harmonics. In addition, a high-pass filter (0.01Hz, order: 82500) was used in order to attenuate the DC component without removing the information related to the infraslow oscillations. Finally, the data were normalized by subtracting the mean value and dividing by the standard deviation. All the implemented filters were Zero-Phase Finite Impulse Response (FIR) ones.

Moreover, a Principal Component Analysis-based K-means clustering method is proposed, to be applied into the SEEG data. Three time intervals of interest are defined; they are consecutive time spans, 100 seconds each, starting 100 seconds before the beginning of the seizure. On each of the intervals previously specified, the data is divided into short epochs (4 seconds each). Next, for each epoch, the decomposition into Principal Components is performed. Then, the SEEG-channels are grouped using the contribution (i.e. the weights of the mixing matrix) of the Principal Components which explain the 90% of the variance in the original data as features for a K-means clustering. Since the K-means algorithm requires the prior selection of K (the number of clusters), an exploration and evaluation of different K (from 2 to 15) is performed and the “ideal” one is selected based on the mean value of their silhouette coefficients. Finally, the representative SEEG channels per each partition are chosen as those closer to the centroid of each cluster.

Along the epochs, the different partitions obtained by applying the previously explained method could represent an indicator of the evolution of the levels of synchrony between diverse brain areas (SEEG-channels) before and during seizures. To my knowledge, in literature this kind of clustering hasn't been applied yet with this purpose.

Finally, since the PCA decomposition is performed separately for each epoch, a further indication of the evolution of this synchronicity levels might be given by the changes in the number of principal components needed to explain the 90% of the total variance over the epochs.

RESULTS

The evolution of the K-values along each of the time epochs was analyzed for the three pre-defined intervals. This evolution shows that, once the seizure starts, the Stereo-EEG contacts tend to be grouped into a lower number of clusters per partition. This also means that the average number of Stereo-EEG bipolar leads grouped per cluster along the epochs increases in relation to the ictal event. Furthermore, the evolvement of the number of principal components needed to explain the 90% of the variance on the original data for each of the time epochs was studied. The results exhibit a decline towards a lower number of components needed after the ictal event. All these findings suggest a higher seizure-related synchronicity among the electrical activity of the different brain areas covered by the Stereo-EEG electrodes. The statistical significance of these findings was positively evaluated according to the Kruskal-Wallis Test followed by a Bonferroni corrected Multiple Comparisons procedure.

Moreover, the way in which the Stereo-EEG bipolar leads are grouped along the partitions was described (i.e. the names of the SEEG channels forming each group were presented) and the representative channels per each partition were calculated. However, some additional medical information should be known in order to stablish further conclusions in this regard.

Key words: *Epilepsy, Stereo-EEG, PCA, K-means Clustering, Brain Synchronicity, Sleep.*

SOMMARIO

L'epilessia è una malattia neurologica complessa di cui le convulsioni costituiscono la principale manifestazione clinica. Oggi è uno dei disturbi neurologici più frequenti e colpisce circa 50 milioni di persone in tutto il mondo. Per circa due terzi dei pazienti, le convulsioni possono essere controllate attraverso l'uso di farmaci antiepilettici. Tuttavia, per il rimanente terzo, la resezione chirurgica della regione responsabile della produzione e propagazione delle convulsioni (Zona Epilettogena, EZ) può essere l'unico modo per eliminare o ridurre le convulsioni.

Bancaud & Talairach, come parte del gruppo Saint-Anne, hanno sviluppato una metodologia denominata Stereoelettroencefalografia (SEEG), mirata a stabilire l'entità della Zona Epilettogena, al fine di programmare la sua resezione chirurgica. La SEEG è ideale per lo studio dei rapporti tra le strutture interessate nella produzione e propagazione delle crisi epilettiche, e ha il potenziale per contribuire alla comprensione dell'origine e la diffusione di queste crisi.
Nel presente studio, si propone di ottenere una descrizione quantitativa dell'evoluzione dei livelli di sincronicità tra aree cerebrali, prima e durante la crisi, analizzando dei dati SEEG.

DATI CLINICI

I dati utilizzati in questo studio derivano dalla valutazione pre-chirurgica eseguita in quattordici pazienti affetti da Displasia Corticale Focale (FCD) di tipo II. Tutti i pazienti avevano una storia di epilessia farmaco-resistente e sono stati candidati per la rimozione chirurgica del focus epilettico. Le registrazioni sono state effettuate durante il sonno tramitte l'impianto stereotassico di elettrodi di profondità (Stereo-EEG, SEEG). I segnali SEEG sono stati registrati usando un sistema a 192 canali (Nihon-KOHDEN Neurofax-110) con una frequenza di campionamento di 1000Hz. Per ciascun paziente e registrazione, le informazioni riguardanti i momenti in cui le convulsioni sono state iniziate (ictal events) sono stati forniti da un esperto neurologo. Questa informazione è stata ottenuta dai video dei pazienti corrispondenti ad ogni registrazione (Video-SEEG). Infine, una serie di montaggi bipolari formati da contatti SEEG adiacenti (dello stesso elettrodo) allocati nella materia grigia sono stati

suggeriti dal neurologo con lo scopo di minimizzare il rumore di modo comune e massimizzare la risoluzione spaziale. Tutte le registrazioni e le procedure di trattamento dei dati sono state approvate dal Comitato Etico corrispondente e tutti i pazienti hanno fornito un consenso scritto.

METODI

Dopo convertere i dati SEEG dal formato Nihon-Kohden in array Matlab e calcolare dei montaggi bipolari suggeriti dal neurologo, delle tecniche di pre-elaborazione standard sono state applicate sui dati tramitte Matlab e il Toolbox EEGLAB. I dati sono stati sottocampionati a 250Hz per ridurre il costo computazionale delle elaborazioni successive. Il ricampionamento è stato eseguito dopo un filtraggio passa-basso (120Hz, ordine: 110) per evitare l'aliasing. Poi due filtri notch (a 50Hz e 100Hz, ordine: 414 ciascuno) sono stati utilizzati per sopprimere il rumore della linea di alimentazione elettrica e le sue armoniche. Inoltre, un filtro passa-alto (0,01Hz, ordine: 82500) è stato utilizzato per attenuare la componente continua senza rimuovere le informazioni relative alle infraslow oscillations. Infine, i dati sono stati normalizzati sottraendo il valore medio e dividendo per la deviazione standard. Tutti i filtri utilizzati sono dei Finite Impulse Response (FIR) a fase nulla.

Inoltre, un metodo di clustering (k-means) basato sull'Analisi delle Componente Principali (PCA) da applicare nei dati SEEG e' stato proposto. Tre intervalli di tempo da analizzare sono definiti; sono intervalli di tempo consecutivi, di 100 secondi ciascuno, partendo da 100 secondi prima dell'inizio delle crisi. Su ciascuno degli intervalli specificati in precedenza, i dati vengono suddivisi in brevi epoche (4 secondi ciascuna). Quindi, per ogni epoca, viene eseguita la decomposizione in Componenti Principali. Poi, i canali SEEG vengono raggruppati utilizzando il contributo delle componenti principali che spiegano il 90% della varianza nei dati originali (cioè i pesi della 'mixing matrix') come 'features' per una clusterizzazione alla K-means . Poiché il algoritmo K-means richiede la preventiva selezione di K (il numero di cluster), un'esplorazione e valutazione di diversi K (da 2 a 15) viene eseguita e quello "ideale" è selezionato in base al valore medio dei suoi coefficienti di Silhouette loro. Infine, i canali SEEG rappresentativi per ogni partizione sono scelti come quelli più vicini al centroide di ogni cluster.

Lungo le epoche, le diverse partizioni ottenute applicando il metodo precedente potrebbero rappresentare un indicatore dell'evoluzione dei livelli di sincronia tra diverse aree cerebrali (SEEG canali) prima e durante la crisi. A la mia conoscenza, in letteratura, un metodo di questo tipo non è stato ancora applicato con lo stesso proposito.

Infine, poiché la decomposizione PCA viene eseguita separatamente per ogni epoca, una ulteriore indicazione della evoluzione di questi livelli sincronicità potrebbe essere fornita dalle variazioni nel numero di componenti principali necessari per spiegare il 90% della varianza totale sui epoche.

RISULTATI

L'evoluzione dei valori di K lungo ciascuna delle epoche è stata analizzata per i tre intervalli predefiniti. Questa evoluzione dimostra che, una volta iniziata la crisi, i contatti Stereo-EEG tendono ad essere raggruppati in un minor numero di cluster per partizione. Questo significa anche che il numero medio contatti SEEG raggruppati per cluster lungo le epoche aumenta in relazione alle crisi. Inoltre, l'evoluzione del numero di componenti principali necessari per spiegare il 90% della varianza dei dati originali per ciascuna delle epoche di tempo è stato studiato. I risultati mostrano un declino nel numero di componenti necessari dopo l'inizio delle crisi. Tutti questi risultati suggeriscono una maggiore sincronicità, tra l'attività elettrica delle diverse aree cerebrali coperte dagli elettrodi stereo-EEG, in relazione alle crisi. La significatività statistica di questi risultati è stata positivamente valutata secondo il test di Kruskal-Wallis seguito da una procedura di comparazioni multipli con correzione alla Bonferroni.

Inoltre, il modo in cui i contatti Stereo-EEG sono raggruppati lungo le epoche è stato descritto (cioè i nomi dei canali SEEG formanti ciascun gruppo sono stati presentati) e sono stati calcolati i canali rappresentativi per ogni partizione. Tuttavia, ulteriori informazioni mediche al riguardo dovrebbero essere note per stabilire ulteriori conclusioni in questo senso.

Parole chiavi: *Epilessia, Stereo-EEG, PCA, K-means, Brain Synchronicity, Sonno.*

INTRODUCTION

The brain is the most complex part of the human body, and it is an exceptional one. This three-pound organ allows us to perform every particular task of our daily lives. It is the source of "intelligence, interpreter of the senses, initiator of body movement, and controller of behaviour" [1]. It is the origin of all the qualities that define our humanity [2].

When the brain is healthy, it works perfectly. However, when problems occur, they could lead to different kind of neurological disorders and the results can be devastating. Some of the most common types of disorders include: "neurogenetic diseases (such as Huntington's disease and muscular dystrophy), developmental disorders (such as cerebral palsy), degenerative diseases of adult life (such as Parkinson's disease and Alzheimer's disease), metabolic diseases (such as Gaucher's disease), cerebrovascular diseases (such as stroke and vascular dementia), trauma (such as spinal cord and head injury), convulsive disorders (such as epilepsy), infectious diseases (such as AIDS dementia), and brain tumors" [1].

Epilepsy in particular, has probably been present in humans since their early evolution 5 million years ago [3] and for centuries the understandings of its origins was confounded by the dramatic and bizarre manifestations of seizures [4]. Epilepsy was associated with possession by evil spirits and patients with epilepsy were discriminated and prosecuted as being affected by magic or supernatural entities [3]. In more recent times the understanding of the neurological nature of epilepsy has evolved and it is now perceived as a focal or brain network disease [5].

Epilepsy is a generic name given to a batch of intricate brain disorders, comprising a broad range of causes and manifestations. The occurrence of epileptic seizures is the characterizing factor in this large number of disorders. [3]. Nowadays it affects around 50 million people worldwide [6] and it is one of the most frequent neurological disorders, with an incidence of 50/100000/year and a prevalence of 0.5-1% [5].

In many cases epileptic seizures can be controlled by antiepileptic drugs, however, they are ineffective in about one third of the patients [6] [5]. In these patients the surgical resection of the epileptogenic zone (EZ), "the cortical region responsible for the onset, early organization, and propagation of seizures" [7], may be the only way to suppress or reduce seizures. Stereoelectroencephalography (SEEG) aims to establish, based on ictal anatomo-electro-clinical correlations, the extent of the EZ [8] [9], in order to plan the surgical resection of the cortical area [9]. It constitutes a mean for studying the relationships between the structures concerned in seizure production and propagation [5] and can potentially contribute to the understanding of the origin and spread of seizures [10].

In the present study, the SEEG data obtained from the pre-surgical evaluation performed in 14 neurosurgical patients affected by Focal Cortical Dysplasia Type II are used in order to obtain a quantitative description of the evolution of the levels of synchrony among brain areas before and during seizures. This is achieved by applying a Principal Component Analysis-based K-means clustering algorithm over different epochs along the seizure events.

The present thesis work is structured as follows:

The Chapter 1 is an introductory one. Firstly, it gives an introduction to epilepsy, its classification and its relation with sleep. Then, Stereo-EEG, the concept of Epileptogenic Zone (EZ) and their importance in epilepsy surgery are presented. Finally, a general overview about the main brain rhythms found in literature is performed.

Chapter 2 describes the Principal Component Analysis-based K-means clustering method applied/proposed in the present thesis work. In order to do it, first, it illustrates each method (PCA and K-means clustering) separately. Then, it elucidates how to combine them in order to achieve the goals of the present study.

Chapter 3 presents the results of the work. First of all, it describes the data used and the pre-processing methods performed over it. Then it shows the outcome of each step of the implemented algorithm and discusses the final results obtained.

Finally, the conclusions of the work are presented and some possible future developments of the work are suggested.

CHAPTER 1

An Introduction to Epilepsy, Stereo-EEG and Brain Rhythms

1.1 Epilepsy

Epilepsy is the name given to a set of complex brain disorders, comprising a broad range of manifestations provoked by a wide variety of causes [3]. The common factor in this large number of disorders is the occurrence of epileptic seizures [3]. “Epilepsy is a chronic condition of repeated seizures” [4]; seizures constitute the defining symptom [3], the principal clinical manifestations [8].

Seizures are transitory brain disruptions resulting in abnormal neuronal activity [4]. They can be characterized either by neuronal underactivity, generating negative signs and symptoms such as paralysis, impairment of consciousness and blindness, or over-activity, producing positive symptoms such as perception of flashing lights or the jerking of an arm .[3]. These signs, however, depend on the location and extent of the affected brain regions and may be consequence of the activity of normal tissues (with normal cellular and network properties), considering that in some cases (as in focal or partial seizures), the abnormal activity can spread beyond its original boundaries [4].

Seizures characterized by an uncontrolled excessive activity of either part or all of the central nervous system constitute the most common form of epileptic seizures [11] [3], and “is one of the most dramatic examples of the collective electrical behaviour of the mammalian brain”. “They can quite literally hijack the normal functions of the brain” [4].

Finally, it’s important to mention that, although Epilepsy is defined by the occurrence of seizures (which can generate further symptoms), these (seizures) could be just “the tip of the iceberg” of the manifestations of epilepsy. Epilepsy implies a disturbance of

the brain function, which could be expressed in various different ways, from the most observable ones (seizures), to less obvious symptoms like sensory/psychic experiences, or apparently unrelated ones like depression, cognitive, behavioural disorders, and even sudden death. [3].

1.1.1 Epilepsy Classification

Epilepsy has probably been present in humans since their early evolution 5 million years ago. The first recorded evidence of epilepsy is found in the ancient Indian medicine (around 4500–1500 BC) [3]. However, the first trial to describe and classify the disease is dated from 2000 BC. The Sakikku (‘all diseases’) (1067 BC), a Babylonian cuneiform text shown in Figure 1, is considered as the oldest available written documentation regarding seizures. It deals mainly with the clinical manifestations of epilepsy in a very descriptive and well-structured way. Even a distinction between seizures with and without loss of consciousness is stablished [8]. In this text, seizures are attributed to the activities of demons [3]. The description of the disease includes, for example, a warning sensation by the patient that is possessed by a demon, recognizes the onset and cries ‘it is he again’ [8]

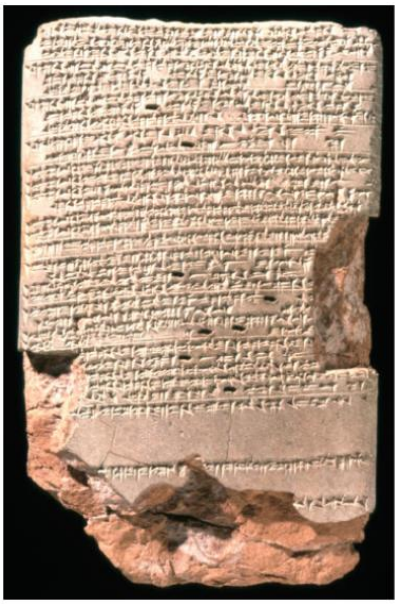


Figure 1. One face of the Babylonian cuneiform text of the Sakkiku (c.1050 BC) [8].

The first differentiation of seizure events from a semiological point of view is found in Hippocrates’ famous treatise “On the sacred disease” [8], which is also the first to clarify that epilepsy is just a natural brain disease and not a ‘sacred’ one [3]. After that, many attempts were carried out in order to classify the disease. Particularly relevant is the one performed in 1824 by Louis Florentin Calmeil, who individuated 3

types of epilepsy based on the severity of their clinical manifestations (semiology): *'Grand mal'*, *'Petit mal'* and *'Abscences'*. [3]

- *'Grand mal'* epilepsy, also known as 'generalized tonic-clonic' [3], is characterized by severe neuronal discharges in all areas of the brain. They may include the cortex, deeper parts and even in the brain stem [11]. Sometimes the discharges transmitted through the spinal cord cause generalized tonic seizures (the patient falls to the ground and rigidly extends all extremities) followed by clonic spasmodic muscle contractions (jerking in all extremities) [4]. During the attack the patient may 'swallow' his tongue, have difficulty for breathing, and the signals transmitted from brain to the viscera might cause urination or defecation. It usually lasts from few seconds to some minutes (3 or 4) and is also characterized by a post-seizure depression of the entire nervous system [11].
- *'Petit mal'* involves the thalamo-cortical brain activating system [3] and is defined by the presence of dizziness, head turning, arm extension, or unresponsiveness. The patient usually experience from 3 to 30 seconds of reduced consciousness with some twitch-like contractions of muscles usually in the head region (for example, blinking of the eyes). It could or not precede grand mal attacks [11].
- *'Absences'* are characterized by short interruptions of consciousness. The patient stop his current activity, and although his senses are aroused, he is temporary impeded from impressions [3].

The current classification of epileptic seizures was proposed in 1981 by the Commission on Classification and Terminology of the International League Against Epilepsy. It was predetermined to be phenomenological, taking into account that at that time, the knowledge about the underlying neural mechanisms and anatomical aspects of seizures was limited. This classification is widely accepted nowadays and is presented in Table 1 [4].

In this way, although the details of the classification are under continuous discussion, seizures can be allocated into two main categories: focal (or partial) and generalized [4].

- *Focal seizures* derive from a small group of neurons (the seizure focus) and their symptoms depend on the location of the focus in the brain. They can be classified as simple partial when there is no alteration of consciousness, or complex partial when there is an alteration [4]. However, there is still some deliberation in literature regarding this classification, taking into account that consciousness impairments are difficult to define [3].

I. Partial (focal, local) seizures
(A) Simple partial seizures
1. With motor signs
2. With somatosensory or special sensory symptoms
3. With autonomic symptoms or signs
4. With psychic symptoms
(B) Complex partial seizures
1. Simple partial onset followed by impairment of consciousness
2. With impairment of consciousness at onset
(C) Partial seizures evolving to secondarily generalized seizures
1. Simple partial seizures evolving to generalized seizures
2. Complex partial seizures evolving to generalized seizures
3. Simple partial seizures evolving to complex partial seizures evolving to generalized seizures
II. Generalized seizures (convulsive or nonconvulsive)
(A) Absence seizures
1. Typical absences
2. Atypical absences
(B) Myoclonic seizures
(C) Clonic seizures
(D) Tonic seizures
(E) Tonic–clonic seizures
(F) Atonic seizures (astatic seizures)
III. Unclassified epileptic seizures

Table 1. International classification of epileptic seizures proposed by the Commission on Classification and Terminology of the International League Against Epilepsy (1981) [4].

Very often focal epilepsy is caused by localized lesions or abnormalities in the brain such as a tumor, a destroyed area of brain tissue, a scar tissue affecting the adjacent areas or a congenitally unbalanced circuitry [11]

The onset of a focal seizure is generally anticipated by symptoms called auras. They represent the earliest manifestation of a focal seizure and are caused by the electrical

activity originated in the focus. Examples of auras include peculiar sensations such as a sense of fear or the perception of a specific odor [4].

In some cases, the strong electric signals originated in the focus could spread further, excite the mesencephalic portion of the brain and generate a grand mal attack [11]. In this case it is said that the focal seizure has ‘secondarily generalized’ [4].

- *Generalized seizures* comprise both brain hemispheres since the beginning and start without an aura. They are also called ‘primary generalized’ in order to differentiate them from seizures that originate in a focus and then generalize secondarily. They can be also divided into convulsive or non-convulsive depending on whether the clinical manifestations include tonic or clonic movements. [4]

Finally, in the Appendix A, a 5-tier classification of epilepsy is reported. It is presented here since it considers epilepsy from different levels, providing a holistic view of the disease. In fact, two of these five tiers (semiological characteristics of the seizures and seizure frequency) define the epileptic seizures, while the other three (etiology, related medical conditions and location of the epilepsy) define what is producing the epilepsy and the location of the brain abnormality.

1.1.2 Focal Cortical Dysplasia (FCD)

Focal cortical dysplasia is a malformation of cortical development [12]. It was first identified by Taylor and colleagues in 1971. At that time they reported the morphological and cellular abnormalities found in brain tissue from therapeutic resections. Taylor considered FCD as a developmental malformation, a view that is still shared by researchers nowadays. [13]

The cause of FCD is not well established neither an explanation for its ability to cause seizures. Apparently both neurodevelopmental abnormalities and possible premature neurodegeneration are related [13]. Moreover, FCD may involve any part of the brain, may vary in size and location and may be multifocal. It is also responsible for

almost half of the intractable epilepsy cases, but at the same time it is characterized by good treatment outcomes [12]. In fact, drug treatment is usually ineffective, while surgical treatment has proved to be curative in many cases [13]. According to the literature, 60–80% of patients remain seizure-free after surgery [12].

Various classifications of the complex structural abnormalities of focal cortical dysplasia have been proposed. Currently, three types of cortical dysplasia are recognized according to the histopathological findings [12]. The most recent classification (2011) is presented in Table 2.

Type	Characteristic features
I	a – focal cortical dysplasia with abnormal radial cortical lamination b – focal cortical dysplasia with abnormal tangential 6-layer cortical lamination c – focal cortical dysplasia with abnormal radial and tangential cortical lamination
II	a – focal cortical dysplasia with dysmorphic neurons b – focal cortical dysplasia with dysmorphic neurons and balloon cells
III	a – architectural distortion of cortical layer in temporal lobe with hippocampal atrophy b – architectural distortion of cortical layer adjacent to glial or glioneuronal tumor c – architectural distortion of cortical layer adjacent to vascular malformation d – architectural distortion of cortical layer adjacent to other lesions acquired in early childhood such as trauma, ischemic event, encephalitis

Table 2. Classification system of focal cortical dysplasia by Blumcke et al. 2011 [12].

In particular, FCD type II, also referred as FCD type Taylor [12], "is one of the most common neuropathological findings in tissue resected therapeutically from patients with drug-resistant epilepsy". Its semiology however, is diverse, and not specific to the particular pathology [13].

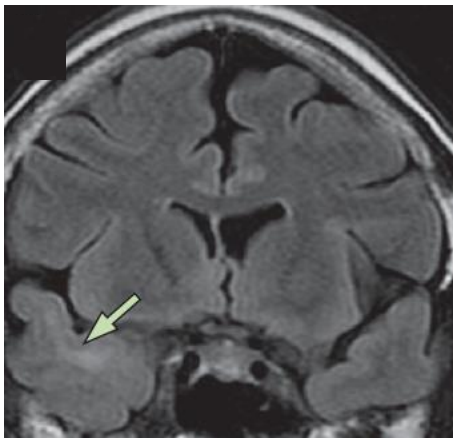


Figure 2. T1 weighted volumetric coronal image. Typical imaging appearance of FCD type II. An area of increased signal in the subcortical region is shown [12].

As stated before, epileptic seizures in FCD are difficult to control with pharmacological treatment and are usually intractable. Then, the surgical treatment emerges as the next therapeutic option; Procedures as the resection of lesion, lobectomies and even hemispherectomies are usually performed [12]. In order to do them, pre-surgical evaluations oriented to determine the epileptogenic focus must be done. Positron Emission Tomography (PET), Diffusion tensor imaging

(DTI), Magnetoencephalography (MEG) and Magnetic resonance imaging (MRI) are often the first methods to be used [13], while Intracranial EEG can be performed when the other methods result insufficient [12]. Figure 2 shows the typical Magnetic Resonance Imaging appearance of FCD type II.

1.1.3 Epilepsy and Sleep

Sleep is a set of active brain states in which many important biological mechanisms take place (e.g. memory consolidation or plasticity). Based on EEG, sleep can be classified into two main categories: rapid eye movement (REM) sleep and non-REM (NREM) sleep. In addition, NREM sleep can be subdivided into four stages. Stages I/II are considered light sleep, while stages III/IV are part of the deep sleep phase. This set of stages is cyclic throughout the night, with a period of around 90 minutes. Deep sleep appears mainly in the first part of the night, while REM sleep tends to occur during the morning. Moreover, "the transition between wakefulness and sleep and between different sleep stages is often gradual and the mechanisms controlling these transitions are poorly understood" [14]. Figure 3 shows a typical hypnogram describing the different stages of sleep.

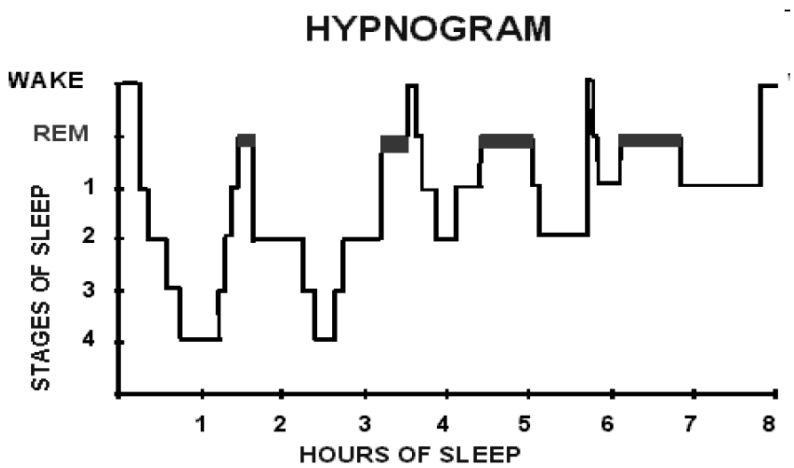


Figure 3. Normal hypnogram. Notice how deep sleep occurs in the first parts of the night and REM sleep towards morning [14].

The potential relationship between sleep and epilepsy has been recognized since ancient times [8]. For more than 2000 years, scientists have perceived the bidirectional connection among them [15]. However, till some decades ago, the

knowledge in this field was based only on clinical observations [8] regarding mainly the relationship between seizures timing and the sleep-wake cycle [15]. While in the last years, the use of video-EEG and polysomnography has importantly enlarged the previous observations [8].

In 1880, Féré followed the seizure occurrence for a three-month period in hospitalized patients, identifying that nearly two-thirds of them occurred between 8 p.m. and 8 a.m. and that insufficient sleep seemed to activate seizures. He further observed that the sleep onset and awakening were periods of particular vulnerability for nocturnal seizures, that nocturnal seizures most often occurred near the end of the sleep period (around 5 a.m. to 6 a.m.) and less often 1–2 hours after sleep onset and that diurnal seizures were clustered in the early morning and late afternoon [8].

More recently, video-EEG monitoring techniques suggest that "sleep appears to activate frontal seizures more often than temporal seizures. Secondary generalization of partial seizures tends to occur more often during sleep (28%) compared to wakefulness (18%), and frontal lobe seizures tend not to secondarily generalize during sleep. In addition, 57% of frontal lobe seizures arose from sleep compared to only 44% of neocortical temporal, 40% of mesial temporal, and 13% of parieto-occipital lobe seizures" [8].

In literature it is commonly accepted that epileptic seizures are more likely to happen in NREM than in REM sleep [16]. NREM sleep tends to facilitate partial seizures, especially at stages 1 and 2 (lighter stages, non-SWS), while REM sleep seems to inhibit them [17]. In fact, seizures rarely occur during REM sleep [14]. Moreover, the neural networks generating wakefulness, NREM and REM sleep engender diverse physiological characteristics which influence the likelihood of having a seizure. Two of the main state-specific factors that determine seizure propagation are the degree of synchronization of cellular discharges and the presence or absence of antigravity muscle tone. On the one hand, NREM sleep is characterized by a state of EEG synchronization and a relative preservation of antigravity muscle tone. Thalamocortical rhythms and synchronous oscillations of cortical neurons that

generate sleep spindles, K complexes and tonic background slow waves are activated during NREM, promoting seizure propagation during this particular stage [8] [14]. In this phase, "the recruitment of a critical mass of neurons needed to initiate and sustain a seizure occurs" [16]. On the other hand, REM sleep is characterized by desynchronization of the EEG and loss of skeletal muscle tone. This desynchronization of the EEG blocks seizure propagation during REM phases and wakefulness. The absence of antigravity muscle tone during REM sleep impedes the expression of seizures, while its preservation during NREM allows the appearance of seizure-related movements [8].

Finally, it is particularly important for the present study to mention that in some types of epilepsy, such as FCD Type II (Taylor), has been demonstrated an increase in the risk of sleep-related seizures [18].

1.2 Stereo-EEG (SEEG)

During the last decades significant improvements have been done in the field of epilepsy surgery, leading to the development of various techniques, concepts and methodologies aimed to increase the safety, accuracy and efficacy of both presurgical evaluations and surgical interventions [9].

Back in the first half of the last century Penfield & Jasper were the firsts to point out the usefulness of preoperative electrocorticography in epileptic patients [9] [8]. Since them, many studies have indicated the potential advantages of recording electrical activities of brain structures using implanted cortical electrodes. However, the first attempts were based on 'free hand' techniques, leading to an imprecise targeting of the intracerebral structures [9]. Essential enhancements were achieved thanks to the employment of stereotactic methods for targeting intracranial structures, which allow a precise identification of non-visualized anatomic structures by using a three-coordinate system and by the introduction of the concept of Epileptogenic Zone (EZ) [8].

Human-oriented stereotactic devices were designed by Spiegel and Wycis in 1947 and the first reports regarding their employment in recordings from deep brain structures are dated to 1950 [8]. Without a doubt, the arrival of stereotactically guided studies decidedly improved the safety and accuracy of these procedures [9]. In addition, the use of stereotactical methods applied to epileptic patients in the Neurosurgical Unit of the Saint-Anne Hospital in Paris inspired the view of epilepsy as a "dynamic process, with a spatialtemporal, often multidirectional, organization, which is best defined referring to a 3-dimensional arrangement" and lead to the origin of the concept of Epileptogenic Zone (EZ) [8].

Hence, Bancaud & Talairach, as part of the Saint-Anne group, developed a complete new methodology called Stereoelectroencephalography (SEEG), which aims to stablish, based on ictal anatomo-electro-clinical correlations, the extent of the cortical areas involved in the ictal discharge (the EZ) [8] [9], in order to plan the surgical resection of the cortical area [9]. SEEG makes possible to define in 3D the spatio-temporal organization of the epileptic activity by using arrangements of intracerebral electrodes placed in precise determined locations [8] [19]. Moreover, it makes possible to access deep cortical structures, which is not feasible by other methods, like subdural grids recordings [8]. And it's well tolerated by most of the patients, with overall complication rates around the 5% (while 13% using subdural grids) [19]

SEEG is ideal for studying the relationships between the structures concerned in seizure production and propagation [19] and has notably contributed to the understanding of the origin and spread of seizures [10], leading to the creation of some initial models of seizure organization [19]. However, as a presurgical tool, this kind of invasive recordings may be used only when the non-invasive techniques fail to correctly localize the EZ, i.e. when there are incoherencies among the anatomical, electrical and clinical findings. Currently around 35% of the patients require SEEG evaluation [8].

1.2.1 Electrodes Placement

Once a patient is classified as candidate for SEEG evaluation, the electrodes implantation strategy is established depending on the clinical, neurophysiological and anatomical information characterizing each particular patient [9]. Then, prior to the surgery for implanting the electrodes, fiducial markers are placed in the patient's head and MRI analyses are executed. After that, with the patient under general anaesthesia, a stereotactic frame is placed on the head of the patient, a computer tomography (CT) and a cerebral angiogram are performed and the electrodes are placed according to the information obtained from both CT and angiography and the previous plan established by the doctors. Once the electrodes have been placed, the head frame is removed and postoperative CT scans and scull X-rays are executed in order to confirm the surgery results. After the surgery (usually the day after), the patient is transferred to an Epilepsy Monitoring Unit, were the recordings are performed. The monitoring phase is generally no longer than a month and after it finishes, the electrodes are removed under local anaesthesia and the patient goes in a 'holiday' period prior to the surgery oriented to the resection of the epileptogenic zone [20]. Figure 21 illustrates the electrode implantation procedure.

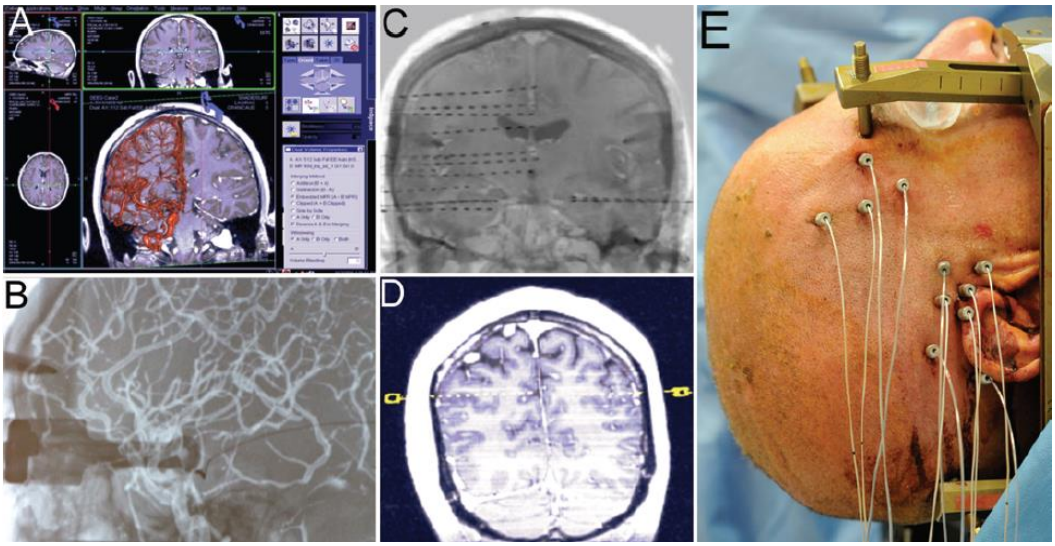


Figure 4. Example of SEEG electrode implantation. A) Preoperative MR image and intraoperative digital angiogram fused. B) Fluoroscopic image and digital angiogram fused during electrode implantation. C) Live fluoroscopic image and preoperative MRI fused during implantation. D) Intraoperative post-implantation image MRI – CT fused. E) Final implantation [20].

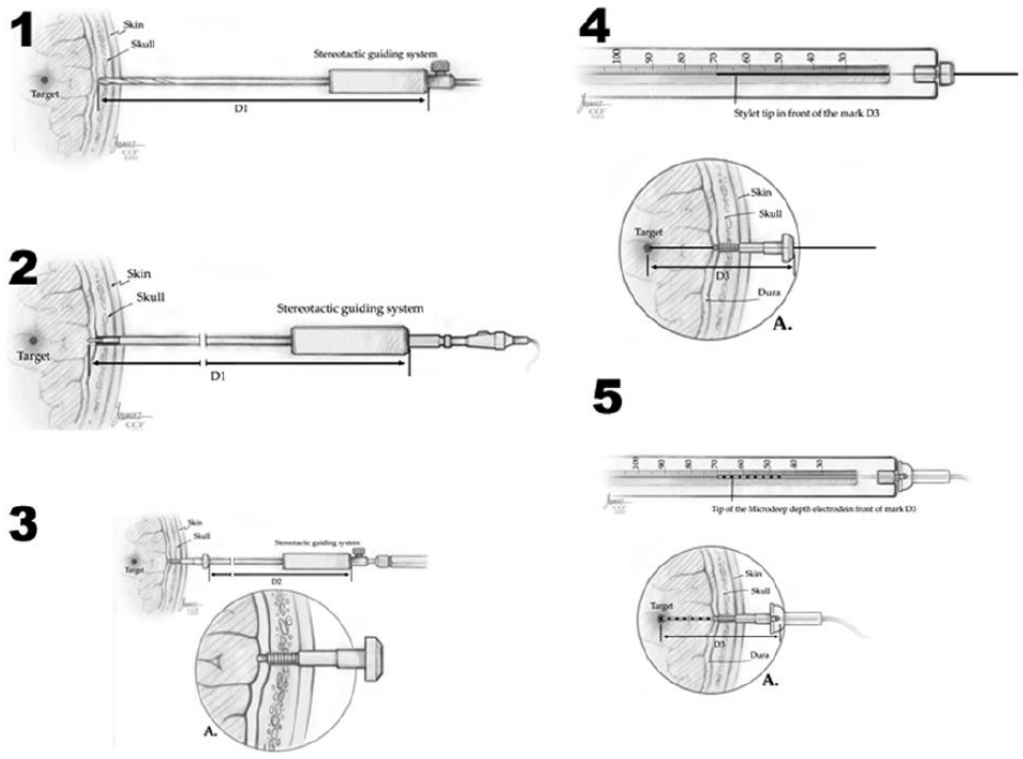


Figure 5 . Electrode implantation. 1) The skull is drilled, guided by the stereotactic system. 2) A monopolar coagulator probe is inserted and the dura is opened. 3) The implantation bolt is screwed into the skull, guided by the stereotactic system. 4) The final depth distance for the electrode (D3) is measured/calculated [(Target-DuraDistance +D1)-D2=D3]. 5) Final position and fixation of the electrode, preventing displacements and CFS leaks. [23]

1.2.2 Electrophysiological Seizure Stages

From an electrophysiological point of view, the temporal evolution of seizures can be described by four phases [21, 22]:

- *Preictal phase*, is the time span before the actual seizure. Its duration is variable; it can last from minutes to days. Moreover, patient specific mood changes are related with it; however, they are not experienced by all the subjects. From an electrical point of view, the presence of high power spikes is a characterizing sign. By the end of this phase, high frequency ripples appear, indicating the beginning of the seizure onset.
- *Ictal phase*, is the interval in which the seizure is really developed. In this stage the brain literally experiences an ‘electrical storm’, which leads to important

cardiovascular, metabolic and electrophysiological changes that could go together with symptoms like depression, anxiety or fear (during the 'auras' mainly).

- *Postictal phase,* is a neuronal recovery/refractory stage which occurs immediately after the seizure. It is characterized by a slight electrical activity.
- *Interictal phase,* is the stage between seizures (between the postictal and the preictal phases). Particular symptoms could be present also in this phase in a subject specific way.

1.2.3 The Epileptogenic Zone (EZ)

The epileptogenic zone (EZ) is the "cortical area(s) capable of generating epileptic seizures whose surgical removal or disconnection will result in seizure freedom" [8] even after taking out all the antiepileptic medication. The scope of epilepsy surgery is to completely remove the epileptogenic zone while saving the eloquent cortex (i.e. "areas of cortex that, if removed, will result in loss of sensory processing or linguistic ability, minor paralysis, or paralysis" [24]). Therefore, the presurgical evaluation in patients with intractable epilepsy is oriented to identify the EZ. However, diagnostic methods nowadays are not able to determine directly the exact location and extent of the EZ. Indeed, in patients who become seizure free after surgery, the only conclusion that can be done is that the EZ was included in the resected area, while is not possible to figure out if a smaller resection would have had the same result. So that, diverse diagnostic methods should be applied in order to estimate in the best possible way the location and extent of the EZ. Actually, combining the information about the localization of some other measurable cortical zones (ictal onset zone, irritative zone, epileptogenic lesion, ictal symptomatogenic zone, and functional deficit zone), the hypothesis about the localization of the EZ can be improved. If there is certain concordance regarding the localization of these zones, increases the probability of including the epileptogenic zone in the resected area [8].
Below, a short description regarding these zones is performed, according to [8]:

- *The symptomatogenic zone,* is 'the cortical area generating the initial ictal symptomatology when activated by epileptic seizures'. Usually, the ictal symptoms

start after the propagation of the seizure activity. So that, it is rare to find an overlap between the symptomatogenic zone and the epileptogenic one. However, the EZ is generally located close to the symptomatogenic zone. For example, a localized somatosensory aura at the beginning of a seizure suggests that the EZ lies located in the vicinity of the corresponding primary sensory area.

- *The irritative zone,* is the cortical region which generates interictal epileptiform discharges. It is not necessary part of the epileptogenic zone and can be affected by different factors as the type of epilepsy, the state of consciousness, changes in temperature, age, presence of anesthesia, and use of anticonvulsants etc.

- *The ictal (or seizure) onset zone,* is the cortical area that 'initiates clinical seizures'.

- *The epileptogenic lesion,* is the name given to a cerebral abnormality detected by the surgeon (as a reminder, morphological brain abnormalities are the prevalent cause of focal epilepsies). The lesion is potentially detected using neuroimaging techniques as CT or MRI. However, epileptogenic lesions are usually small and difficult to detect. In addition, their extent may be underestimated using MRI, especially in patients with MCD. In MCD in particular, the cortex beyond the lesion may appear normal on imaging but abnormal at a microscopic level and may be part of the epileptogenic zone. Generally, the epileptogenic lesion is part of the epileptogenic zone or at least lies close to it (it is an immediate neighbor). However, not all the lesions are epileptogenic and it is impossible to predict their epileptogenic nature using neuroimaging techniques. So that, video-EEG monitoring, and sometimes invasive recordings, are needed in order to find out the relationship between the irritative and seizure onset zones and the structural lesion obtained through imaging. These zones (irritative and seizure onset zone) must be overlapping the lesion or be immediate neighbors in order to consider the lesion as an epileptogenic one.

- *The functional deficit zone*, is the brain area with abnormal functioning during the interictal period. Finding this zone has a limited value in the localization of the epileptogenic zone. According to its correlation to the other zones it can simply support the previous hypothesis regarding the location of the EZ, or lead to additional tests (like iEEG).

The main techniques used in order to measure the different cortical zones are: MRI, SPECT, PET, EEG, Video-EEG and iEEG. Invasive monitoring is used when unexplainable discrepancies in the other tests are found. However, since iEEG cover a limited cortical area, a prior clear hypothesis regarding the location of the seizure onset zone should exist in order decide where to place the intracranial electrodes.

Finally, it is important to mention that the epileptogenic zone is composed by an 'actual' and a 'potential' seizure onset zones. The actual seizure onset zone is the area of cortex in which the seizures currently recorded emerge. Sometimes, seizures could persist even after the complete resection of the actual seizure zone, because other region in the cortex that previously did not generate seizures starts generating seizures. It is said that the potential seizure onset zone reaches the threshold for seizure generation. Nowadays it is impossible to measure these potential seizure onset zones prior to surgery [8].

1.3 Brain Rhythms

The brain cortical electrical activity presents different kind of oscillations characterized by certain ranges of amplitudes and frequencies. These oscillations are referred as brain rhythms or brain waves. The intensity of these waves is determined primarily by the number of neurons and fibres that fire in synchrony with others, not by the total level of electrical activity in the brain. In fact, strong asynchronous signals could nullify each other since they have opposing polarities. In normal healthy people, the main EEG rhythms are called *alpha*, *beta*, *theta*, and *delta* [11].

Alpha and beta waves were introduced by Berger in 1929. The delta rhythm was suggested by Walter (1936) to nominate the frequencies below the alpha range. He also considered the theta waves as those having frequencies within the range of 4–7.5 Hz. The term ‘gamma’ (to refer to the waves of above 30 Hz) was proposed by Jasper and Andrews in 1938. And the notion of a theta wave was introduced in 1944 by Wolter and Dovey [25]. They are all presented in Figure 6.

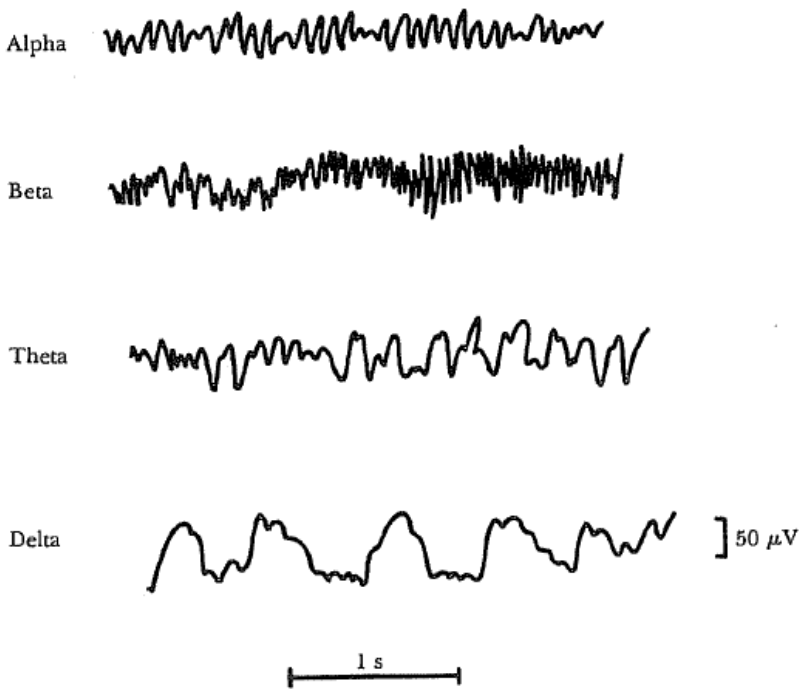


Figure 6. Different types of EEG waves/rhythms [26]

- **Alpha waves (α)** are the most prominent rhythm in the brain [25]. They occur at frequencies between 8 and13 Hz and are typically sinusoidal (there is often an asymmetry in the rhythm with the right side being lightly higher in voltage [27]). Alpha rhythms appear in the posterior half of the head and are usually found over the occipital region of the brain [25], reason why they are also known as the posterior dominant rhythm [27] (anyhow, they can also be recorded from the parietal and frontal regions [11]).

They are typical in normal adult people in awakeness, and in a quiet, resting state. During deep sleep, the alpha waves disappear [11]. Usually, they are also produced when the eyes are closed, and are reduced or eliminated by opening the eyes, hearing unfamiliar sounds, by anxiety, mental concentration or attention [25]. When the attention is directed to some specific mental activity, the alpha waves are replaced by asynchronous, higher-frequency, lower-voltage *beta waves*. Figure 7 shows the effect of the eyes opening-closing on the alpha waves [11].

The origin and physiological significance of alpha waves is still unknown [25]. It is recognized that Alpha waves don't occur in the cerebral cortex without cortical connections with the thalamus. So that, it is believed that they result from spontaneous feedback oscillation in the thalamocortical system [11].

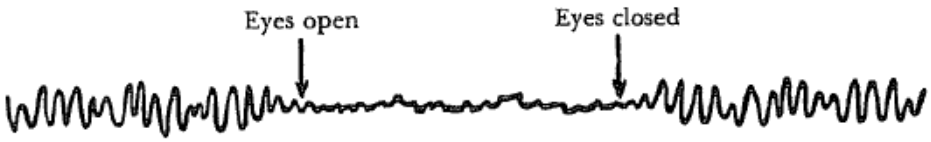


Figure 7. Effect of eyes opening-closing on alpha waves. The alpha rhythm is replaced by an asynchronous, higher frequency, lower voltage activity (beta waves) [11].

- **Beta waves (β)** occur at frequencies above 13Hz [27] (in the range 14-26Hz in some literature [25]). They are recorded mainly in the frontal and central-parietal regions [25] [11] during the specific activation of these parts of the brain [11]. They are associated with active thinking, active attention, focus on the outside world, or solving concrete problems [25]. Their activity is often increased during drowsiness and in patients receiving sedating medication, particularly barbiturates or benzodiazepines [27].

- **Theta Waves (θ)** have frequencies from 4 Hz to less than 8Hz [27]. They commonly occur in the parietal and temporal regions [11] and appear while consciousness shifts towards drowsiness. They are associated with the access to unconscious material, creative inspiration and deep meditation [25]. They also occur during emotional stress, particularly during disappointment and frustration, and are

characteristic of many brain disorders [11]. Apparently they have a thalamic origin. [25]

- **Delta waves (δ)** refer to frequencies below 4Hz. They are primarily associated with deep sleep [27] but may be also present in waking state [25]. They usually have greater voltages (two to four times) than the other types of brain waves and can occur in the cortex independent of the activities in lower regions of the brain [11]. In fact, the transection of the fiber tracts from the thalamus to the cerebral cortex, which blocks thalamic activation of the cortex and eliminates the alpha waves, does not block delta waves in the cortex. This indicates that there is some mechanism in the cortical neuronal system, which can cause delta waves, independent of lower structures in the brain [11].

- Frequencies above 30 Hz correspond to the gamma range (sometimes called 'fast beta'). The amplitudes of these rhythms are very low and their occurrence is rare, however, their detection can be used as confirmation in some brain diseases. [25]

- Frequencies under 1Hz (infraslow oscillations), have been recently raising the interest of the scientific community. Infraslow oscillations are prominent during sleep, but their functional meaning is still unknown [28]. In addition, they could explain the "yet enigmatic aggravation of epileptic activity during sleep" [29]. More about infraslow oscillations can be found in [28-31].

- Rhythms at frequencies higher than the normal activity range of EEG, mainly in the range of 200–300 Hz, have been found in animals, but they don't play any role in clinical neurophysiology [25].

CHAPTER 2

Methods

This chapter aims to introduce the mathematical and statistical methods chosen for the processing of the Stereo-EEG data. Their actual application to the SEEG data and the results obtained will be described in the next chapter.

2.1 Principal Component Analysis (PCA)

"Principal component analysis (PCA) is the most widely known technique of attribute reduction by means of projection" [32]. It is an orthogonal linear transformation that transforms data to a new coordinate system in a way that the greatest variance comes to lie on the first axis, the second greatest variance on the second axis, and so on [33]. The purpose of this method is to obtain a transformation that replaces a set of attributes with a lower number of new attributes obtained as their linear combination, without causing a loss of information [32].

Let's denote $X' = [X_1, X_2, \dots, X_p]$ as a random vector with covariance matrix Σ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let's consider then the linear combination:

$$
\begin{aligned}
Y_1 &= a_1'X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
Y_2 &= a_2'X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \\
Y_p &= a_p'X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
\end{aligned}
$$

Under this consideration, the first Principal Component corresponds to the linear transformation $a_1'X$ which maximizes $Var(a_1'X)$, subject to the constraint $a_1'a_1 = 1$ (unit norm constraint):

$$\max_{a_1} Var(a_1'X) = a_1'\Sigma a_1$$
$$s.t.\ a_1'a_1 = 1$$

The first principal component represents a vector in the direction of maximum variance in the space of the original attributes [32]. The second Principal Component corresponds to the linear transformation $a_2'X$ which maximizes $Var(a_2'X)$, under the constraints $Cov(a_1'X, a_2'X) = a_1'\Sigma a_2 = 0$ and $a_2'a_2 = 1$. In this way, the i-th Principal Component is the $a_i'X$ which maximizes $Var(a_i'X)$, under the constraints $a_i'a_i = 1$ and $Cov(a_i'X, a_k'X) = a_i'\Sigma a_k = 0$, for $k < i$. Where $a_i'X$, with $i = 1, \dots, p$ are the Principal Components (or scores) and the coefficients a_i are the corresponding weights [34].

From another perspective, the principal components could be considered as virtual sources which mixed in a weighted manner ($X' = MY'$) would generate the original signal X. Under this consideration, the matrix M takes the name of "mixing matrix" [35].

The p principal components are orthogonal to each other. So that, they are also uncorrelated and can be ordered according to a "relevance indicator". In fact, the first principal component explains the greatest proportion of variance in the data, the second explains the second greatest proportion of variance, and so on [32].

The coefficients a_i can be interpreted as the weights of the attribute X_j in determining the component Y_i. At the same time, $Var(Y_i) = \lambda_i$ represents a measure of the proportion of total variance explained by the principal component Y_i. For this reason, the index

$$I_k = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$$

expresses the percentage of total variance explained by the first k principal components and provides an indication of the amount of information preserved by the first k components.

In order to determine the number of principal components to be used, it is possible to consider the components which exceed a threshold I_{min} considered reasonable. Usually, the number of components to be used is chosen considering the components which explain around $80\% - 90\%$ of the variance, in order to avoid excessive loss of information.

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} * 100 = 80\% - 90\%$$

2.2 Clustering: K-means Algorithm

The aim of clustering analysis is to find out the "natural grouping(s)" of a set of objects. It is a "statistical classification technique for discovering whether the individuals of a population fall into different groups by making quantitative comparisons of multiple characteristics" [36]. Cluster analysis does not use any category labels to tag objects. It is an unsupervised learning technique with an exploratory nature; it aims to find a structure in the data without any prior labeling or information about the groups.

Clustering algorithms can be divided into two categories: hierarchical and partitional. Hierarchical algorithms find the clusters either in an agglomerative or divisive way; in the former one, each data point corresponds to one cluster and they merge successively the most similar pairs of clusters, forming a cluster hierarchy. In the latter, all the data points belong initially to one cluster and they divide recursively into small clusters. On the other hand, partitional algorithms find all the clusters concurrently as a partition of the data, without imposing a hierarchical structure [36]

The "most popular and the simplest partitional algorithm is the K-means", first published in 1955 [36]. The main reasons for its popularity are its simplicity, ease of implementation, efficiency and empirical success [36] [37].

The scope of the K-means algorithm is to divide M points in N dimensions into K non-empty, non-overlapping, non-subordinated clusters, minimizing the within-cluster sum of squares; each observation belongs to the cluster with the nearest mean [38]. In order to do it, the algorithm follows three main steps [32]. The clustering procedure is summarized in Figure 8.

1. K elements of the dataset are randomly chosen as the cluster centroids/means (the sample space is initially partitioned into K clusters).

2. Iteratively each element/point is assigned to the cluster with the most similar centroid (i.e. the one which minimizes the distance to the observation/point). The distance from the observation to the centroid is calculated. If the observation is closest to its own cluster, then it stays there, otherwise, it goes to another cluster.

3. For each new cluster, the new centroid is calculated as the mean of the elements in the cluster. Then, it goes back to the second step. If at any iteration no element changes cluster in respect to the previous iteration, the algorithm stops.

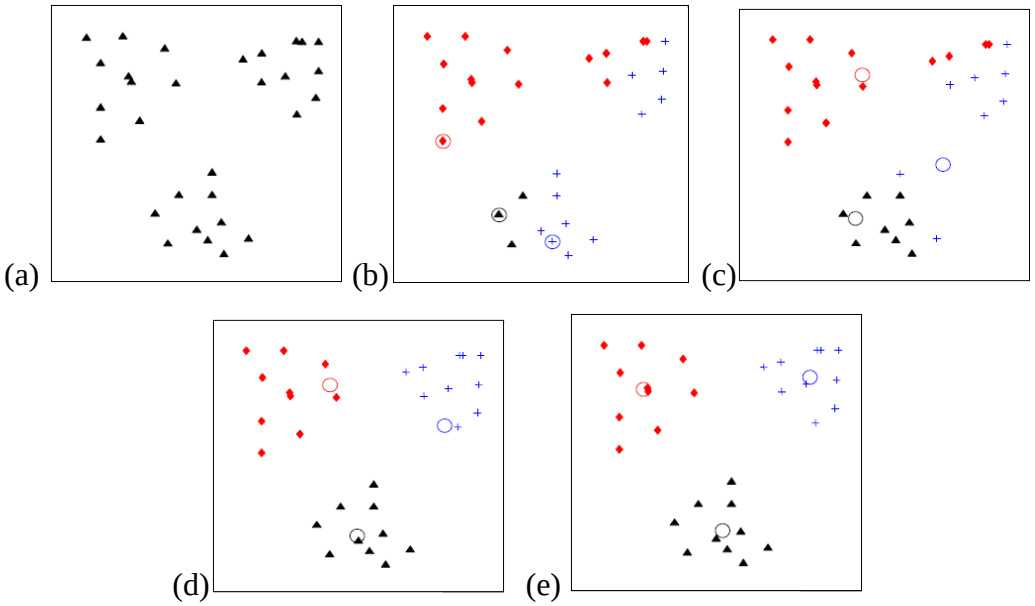


Figure 8. K-means application example. (a) Input data set (two dimensions, three clusters). (b) three seed points are selected as center of the clusters and there is an initial assignment of the data points to the clusters. (c) (d) intermediate iterations. (d) the algorithm converges, it's the final iteration [36].

The algorithm has two main drawbacks. First, the K number of clusters must be specified in advance. In addition, as the iteration proceeds, the solutions could be trapped into local minima due to the non-exhaustive search nature of the algorithm [37] [39] [40]. In order to avoid being trapped in local minima, the algorithm could repeatedly start from different randomly selected sets of initial centroids and at the end, return the partition with the lowest sum of distances over all replicates [41].

2.2.1 Measures of Distance

The most common distance measures implemented in cluster analysis algorithms are the Euclidean distance, the Euclidean squared distance and the Manhattan or City distance.

The Euclidean measure corresponds to the shortest geometric distance between two points. For an n-dimensional dataset, the distance between two observations $x_i = (x_1, x_2, \dots, x_n)$ and $y_i = (y_1, y_2, \dots, y_n)$ is defined as:

$$d = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}$$

A faster way of determining this distance is by using the squared Euclidean distance which calculates the above distance squared:

$$d_{sq} = \sum_{i=1}^{N} (x_i - y_i)^2$$

Finally, the Manhattan measure calculates a distance between points based on a grid; for going from one point to the other it considers the path including two sides of a rectangle having as vertex the points of interest as shown in Figure 9.

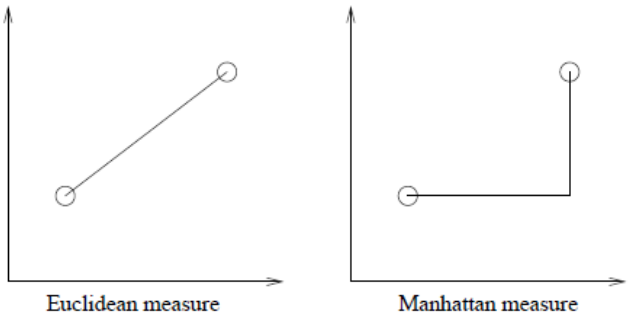


Figure 9. Euclidean distance measure vs. Manhattan distance measure [42].

2.2.2 Silhouette Coefficients and Silhouette Diagram

There are some performance indicators which could be calculated about the final set of K clusters generated; cohesion, overall cohesion, separation, overall separation and the silhouette coefficient are part of them [32]. Considering $C = \{C_1, C_2, \dots, C_k\}$ as the set of K clusters,

The cohesion is defined as

$$coh(C_h) = \sum_{\substack{x_i \in C_h \\ x_k \in C_h}} dist(x_i, x_k)$$

While the overall cohesion is given by

$$coh(C) = \sum_{C_h \in C} coh(C_h)$$

The first one is an indicator of homogeneity of the observations within each cluster C_h, while the second one is referred to the overall partition. One cluster is preferable over others in terms of homogeneity if it has a smaller cohesion.

The separation on the other hand, is an indicator of inhomogeneity between pairs of clusters. It is defined as

$$sep(C_h, C_f) = \sum_{\substack{x_i \in C_h \\ x_k \in C_f}} dist(x_i, x_k)$$

While the overall separation of the partition is given by

$$sep(C) = \sum_{\substack{C_h \in C \\ C_f \in C}} sep(C_h, C_f)$$

In terms of inhomogeneity, one partition is preferred over others if it has a higher overall separation.

One more indicator of the quality of the clustering is represented on the silhouette coefficient, which is a combination of cohesion and separation measures. Three steps are necessary in order to calculate the silhouette coefficient for an observation x_i [32]:

1. Calculate the mean distance u_i of x_i from all the remaining observations belonging to the same cluster.

2. For each cluster C_f different from the cluster to which x_i belongs, calculate the mean distance w_{if}between x_i and all the observations in C_f. Determine the minimum among the distances w_{if} by varying the cluster C_f (let's call v_i that minimum distance).

3. The silhouette coefficient of x_i is defined as

$$silh(x_i) = \frac{v_i - u_i}{\max(u_i, v_i)}$$

It varies between −1 and 1. Negative values indicate that "the mean distance u_i of the observation x_i from the points of its cluster is greater than the minimum value v_i of the mean distances from the observations of the other clusters", so that, the

membership of x_i in its cluster is not desirable and it's not well characterized. Under ideal conditions, u_i should be as close as possible to 0 and the silhouette coefficient should be positive.

The coefficients can be plotted in silhouette diagrams, in which the observations are in the vertical axis, divided by clusters, and the values of the coefficients are presented in the horizontal axis. The silhouette diagram shows how well the data are separated into the K clusters [42]. In addition, the overall silhouette coefficient of a partition can be calculated as the mean of the silhouette coefficients for all the observations in the dataset. In fact, a quantitative way of comparing different solutions (partitions) is to look at the average silhouette values for the different cases; higher values are preferable [41]. Figure 10 shows some examples of Silhouette diagrams.

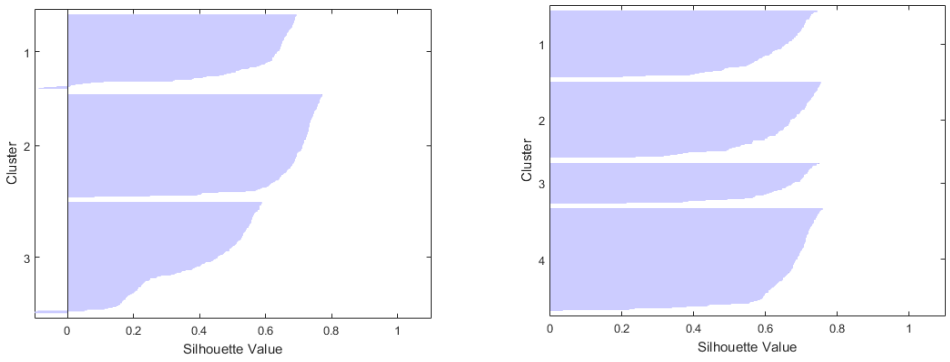


Figure 10. Examples of silhouette diagrams. In the diagram on the left, the third cluster shows some points with low silhouette values and some with negative ones, indicating that the cluster is not well separated. In the diagram on the right, the clusters show high silhouette values, indicating that the clusters are well separated [41].

2.3 PCA-based K-means Clustering

A Principal Component Analysis-based Clustering method is proposed, to be applied into the SEEG data. In this section, a general overview of the process is presented. The details of its implementation are described on Chapter 3.

The idea is to obtain a representation of the evolution of the synchronicity levels among brain areas before and during seizures. In order to do it, three time intervals of

interest were defined; they are consecutive time spans, 100 seconds each, starting 100 seconds before the beginning of the seizure. On each of the intervals previously specified, the data is divided into short epochs (4 seconds each). Next, for each epoch, the decomposition into Principal Components would be performed. Then, the SEEG-channels would be grouped using the contribution (i.e. the weights of the mixing matrix) of their Principal Components (from those which explain the 90% of the variance in the original data) as features for a K-means clustering. Since the K-means algorithm requires the prior selection of K (the number of clusters), an exploration and evaluation of different K (from 2 to 15) would be performed and the "ideal" K would be selected based on the mean value of their silhouette coefficients. Finally, the representative SEEG channels per each partition would be chosen as those closer to the centroid of each cluster. The entire procedure is summarized in Figure 11.

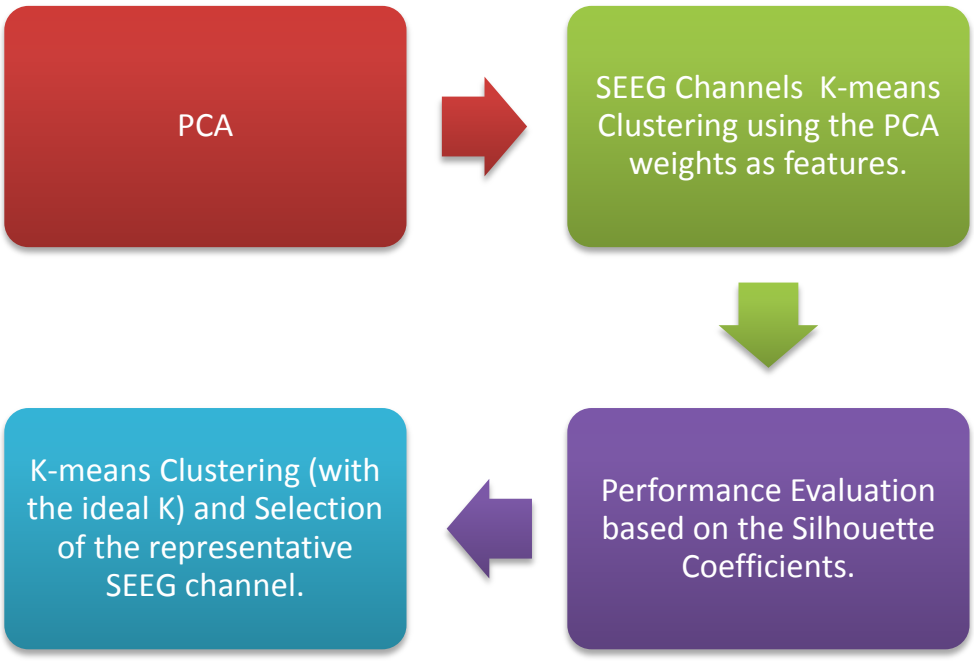


Figure 11. PCA-based K-means clustering to be applied over each epoch of the SEEG signal time intervals previously defined.

Along the epochs, the different partitions obtained by applying the previously explained method could represent an indicator of the evolution of the synchronicity levels between diverse brain areas (SEEG-channels) before and during seizures.

Finally, since the PCA decomposition is performed separately for each epoch, a further indication of the evolution of this synchronicity levels might be given by the changes in the number of principal components needed to explain the 90% of the total variance over the epochs.

CHAPTER 3

Data Analysis and Results

3.1 Patients and Data Acquisition

The data used in the present study comes from the pre-surgical evaluation performed in 14 neurosurgical patients affected by Focal Cortical Dysplasia Type II. All the patients had a history of drug-resistant epilepsy and were candidates for the surgical removal of the epileptic focus.

The recordings were performed during sleep by stereotactically implanted depth multi-lead electrodes (Stereo-EEG, SEEG) in order to precisely localize the epileptogenic zone and related/connected areas. The position of the electrodes was determined based on non-invasive clinical assessment.

SEEG activity was recorded from platinum-iridium semiflexible multi-contact intracerebral electrodes, with a diameter of 0.8mm, a contact length of 1.5 mm, an inter-contact distance of 2mm and a maximum of 18 contacts per electrode (Dixi Medical, Besancon, France). The placement of the electrodes was verified by post-implantation CT scans.

In addition, EEG (from the scalp), Electro-ocular activity and subelemental electromyographic activity were recorded. Both EEG and SEEG signals were recorded using a 192-channel recording system (NIHON-KOHDEN NEUROFAX-110) with a sampling rate of 1000Hz. The data was recorded and exported in the EEG Nihon-Kohden format.

All the recordings and data treatment procedures were approved by the corresponding Ethical Committee (Niguarda Hospital, Milan, Italy) and all the patients provided a written informed consent.

Moreover, for each patient and recording, the information regarding the temporal moments in which the seizures started (ictal events) were provided by an expert neurologist. This info was obtained from the videos of the patients corresponding to each recording (Video-SEEG). Finally, bipolar montages formed by adjacent SEEG leads (of the same depth-electrode) in grey matter were suggested by the neurologist in order to minimize the common electrical noise and maximize the spatial resolution [43].

3.2 Data Preprocessing

Firstly, the acquired data was converted from the Nihon-Kohden format into Matlab arrays in order to do all the processing under the Matlab environment. This was done through customized Matlab scripts. Then, all the non-SEEG channels were discarded, the bipolar montages suggested by the expert neurologist were calculated and standard pre-processing techniques were applied over the SEEG data. All the pre-processing was executed using the EEGLAB toolbox.

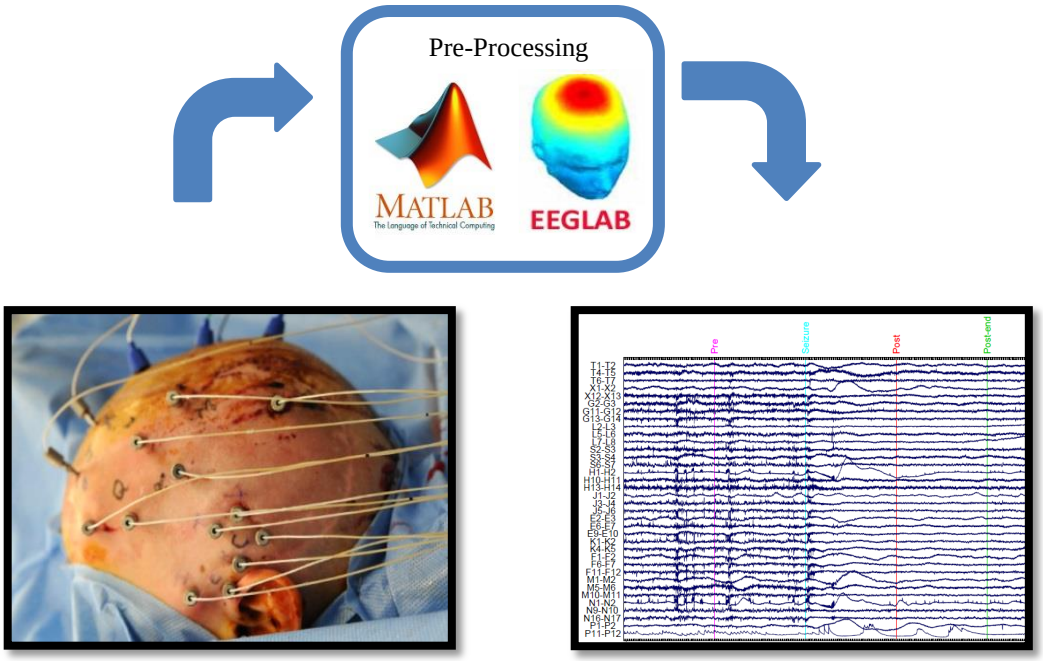


Figure 12. Data pre-processing. It was performed under the MATLAB environment and the EEGLAB Toolbox. On the left: electrode arrangement in the patient. On the right: SEEG signal for all the channels.

The data was down-sampled to 250Hz in order to reduce the computational cost of the processing. The resampling was performed after low-pass filtering (120Hz, order: 110) to avoid aliasing. Then two notch filters (at 50Hz and 100Hz, order: 414 each one) were used in order to suppress the noise of the electrical power line and its harmonics. In addition, a high-pass filter (0.01Hz, order: 82500) was used in order to attenuate the DC component without removing the information related to the infraslow oscillations. Finally, the data were normalized by subtracting the mean value and dividing by the standard deviation. All the implemented filters were Zero-Phase Finite Impulse Response (FIR) ones. The complete pre-processing procedure is synthetized in Figure 13. Finally, all the events of interest were manually added to the main dataset into the EEGLAB toolbox.

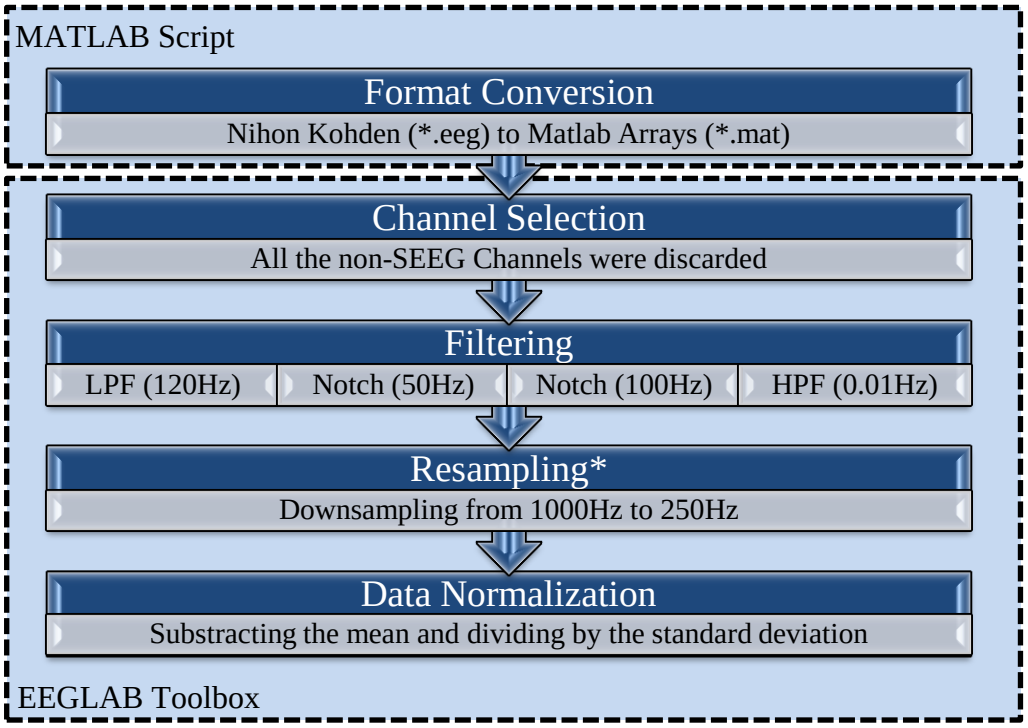


Figure 13. Data pre-processing sequence performed under the MATLAB environment and the EEGLAB Toolbox. *In the actual procedure, the resampling was performed after the LPF to avoid aliasing and before the remaining filters in order to reduce their computational cost.

3.3 Application of the PCA-based K-means Clustering method

The implemented algorithm uses the main Matlab structure defined by the EEGLAB toolbox. It reads the SEEG signal, detects a set of events (“pre”, “seizure”, “post”, ”post_end”) previously inserted into EEGLAB, reads the corresponding event latencies and according to them, defines the time intervals to be studied (Pre, Seizure, Post) (See Figure 14). Each one of these time intervals is 100 seconds long. The “seizure” interval starts in the moment in which the seizure begins, according to the information provided by the neurologist. The “pre” interval commences 100 seconds before the seizure begins, and the “post” interval initiates immediately after the end of the “seizure” one.

Then, over the time intervals of interest, the data is divided into short epochs (4 seconds each one). Then, for each epoch, a decomposition into Principal Components is performed. Next, considering only the first Principal Components which explain the 90% of the variance, the SEEG-channels are grouped using the contribution (i.e the weights of the mixing matrix) of these Principal Components as features for a K-means clustering. The distance measure implemented in the cluster analysis algorithm is the Euclidean squared distance.

The scope of the K-means algorithm is to divide M points (SEEG Channels) in N dimensions (Number of Principal Components) into K clusters. The clustering has an exploratory nature; it aims to find a structure in the data without any prior labeling or information about the groups. The idea is to find out the “natural grouping” of the set of SEEG Channels.

The k-means algorithm has two main drawbacks. First, the K number of clusters must be specified in advance. In addition, as the iteration proceeds, the solutions could be trapped into local minima due to the non-exhaustive search nature of the algorithm [55, 57-58].

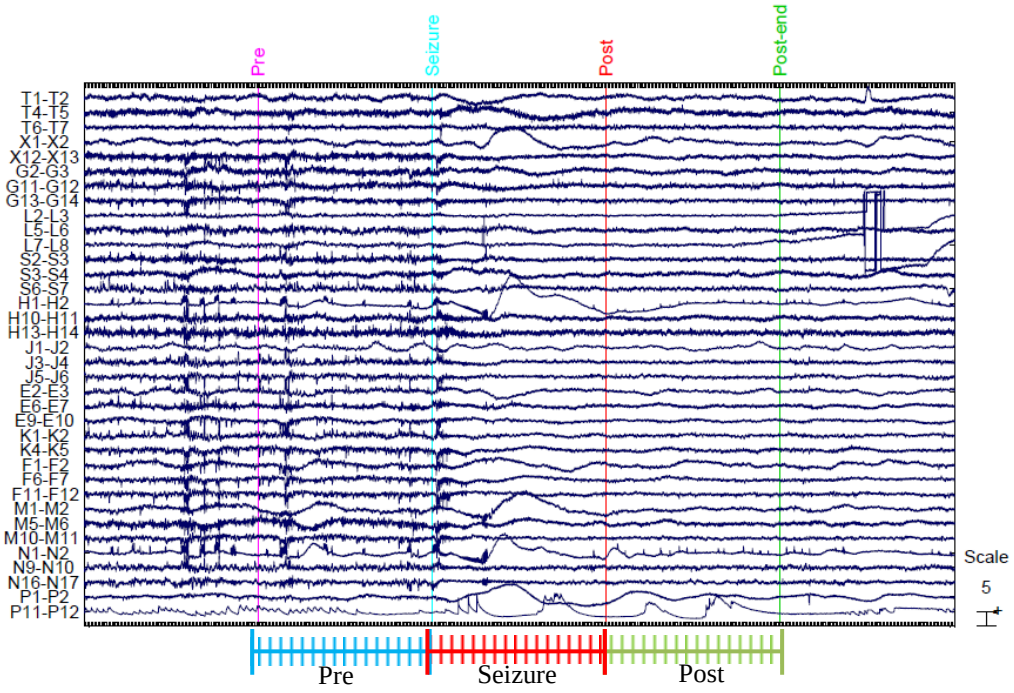


Figure 14. Example of Stereo-EEG signal corresponding to one of the seizures under analysis (Patient 2, seizure #1). A 500 time window is presented. The events and time intervals of interest (Pre, Seizure, Post) are specified. Each one of the three time intervals is 100 seconds long. Over each interval the data was divided into short epochs (4 seconds each one).

In order to avoid being trapped in local minima, the algorithm repeatedly starts from different randomly selected sets of initial centroids and at the end, returns the partition with the lowest within cluster distance (i.e. the sum of the distance between the centroid and all the samples of the cluster) over all replicates. In the present implementation the algorithm performs 500 replicates and gives as output the 'best partition' according to the previously expressed criteria.

On the other hand, in favor of dealing with the need of specifying 'a priori' the K number of clusters, a performance measure based on the silhouette diagram was implemented and applied over a set of different K partitions. In particular, the k-means clustering was applied with different K, ranging from 2 to 15 and the 'best' one according to the performance indicator remained as the ideal K for the correspondent epoch.

As stated before, the performance measure implemented in order to choose the ideal K is based on the silhouette diagram. Explicitly, after performing all the possible partitions (with K from 2 to 15), their corresponding silhouette coefficients and

silhouette diagrams are obtained. If there is no partition containing miss-classified channels (i.e. there are no negative silhouette coefficients), the partition having the higher mean of the silhouette coefficients is considered the ideal one. If all the partitions have miss-classified channels, the one with the highest mean of the negative region of the silhouette diagram (i.e. the one with the highest mean among the negative silhouette coefficients, which would be also the one with the lowest miss-classification) is considered the 'best' one. Finally, if some partitions have miss-classified channels and some do not, the partitions lacking of miss-classification are considered the best and among them, the one with the highest mean of the silhouette coefficients is chosen as the ideal one.

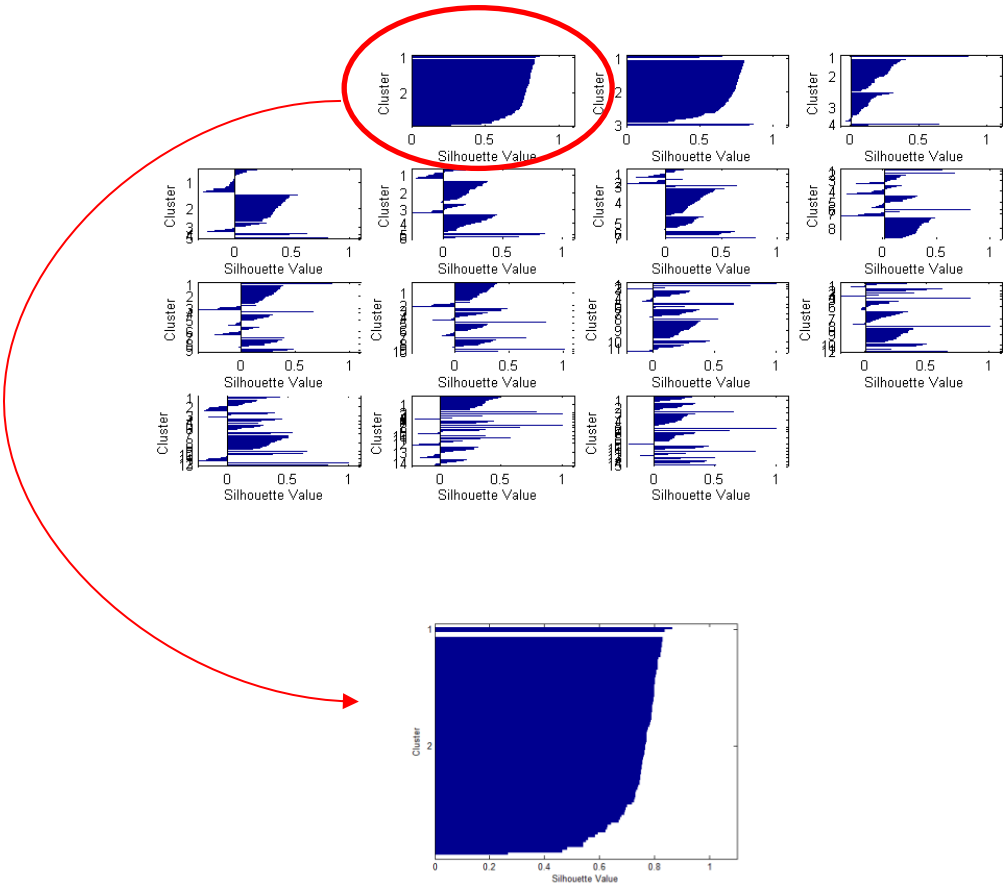


Figure 15. Example of the application of the performance measure implemented in order to choose the ideal K based on the silhouette diagram. Top: the k-means clustering is applied with different K, ranging from 2 to 15. Bottom: the 'best' one according to the performance indicator remains as the ideal K for the correspondent epoch (K=2 in this case).

Then, once the ideal K is selected, for each cluster inside the partition, the representative channel is chosen as the closest one to the centroid of the corresponding cluster.

An overview of the different steps applied over each signal epoch is presented on Figure 16.

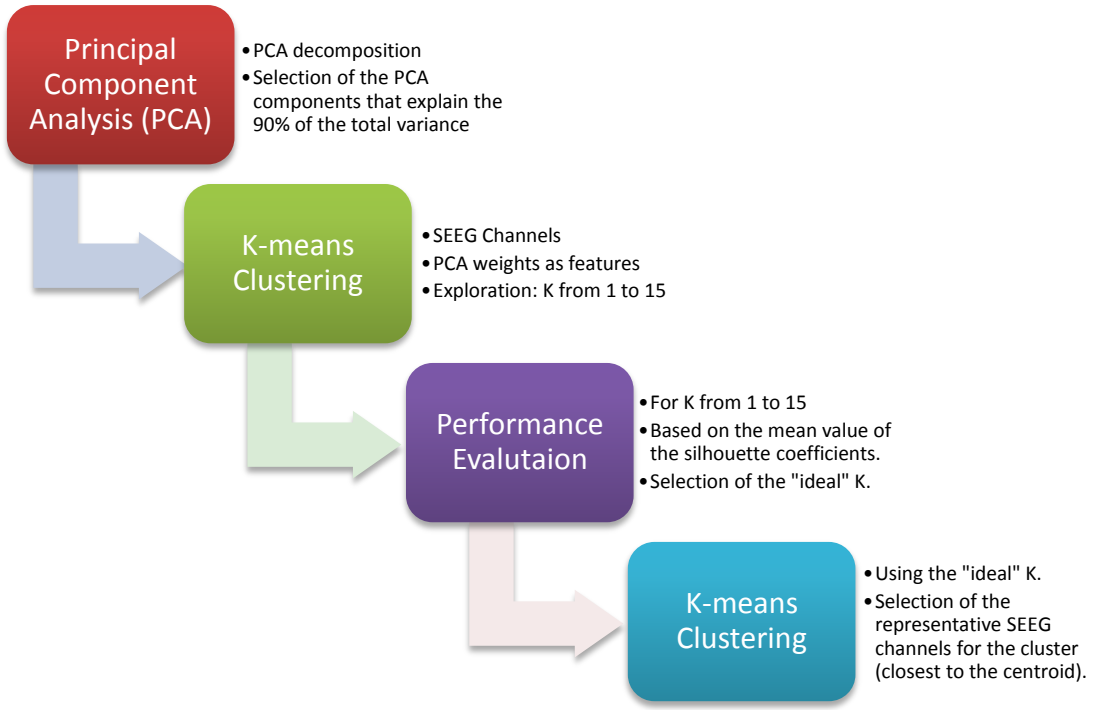


Figure 16. Pipeline of the algorithm applied over each signal epoch.

Moreover, since this procedure is applied to all the epochs along the time intervals corresponding to the period before and during the epileptic seizures, the changes of K over the epochs could describe the evolution of the synchronicity levels between diverse brain areas (SEEG bipolar leads) along the epochs (and along the seizure). An additional indicator related to this evolution may be found on the variation of the number of PCA components required in order to explain the 90% of the variance on the original data over the different epochs, taking into account that the PCA decomposition is performed separately for each epoch.

3.4. Results and analysis for one seizure example

In this section the final results regarding the application of the PCA-based K-means Clustering algorithm to one of the seizure cases analyzed will be discussed in detail. In particular, the Stereo-EEG signal corresponding to the seizure example presented previously in Figure 14 will be studied. For that particular patient, 36 bipolar leads were analyzed according to the suggestion made by the neurologist. On Figure 14, a 500 time window is presented. The events and time intervals of interest (Pre, Seizure, Post) are specified. Each one of the three time spans is 100 seconds long and over each interval the data was divided into short epochs (4 seconds each one).

3.4.1 Evolution of the SEEG bipolar leads grouping along the time epochs.

In Figure 17-19 is presented the sequence of silhouette diagrams for each epoch along the three time intervals under study. Figure 17(a) shows the silhouette diagrams progression for the "pre" time interval, while (b) and (c) exhibit the sequence found for the "seizure" and "post" intervals respectively. In the figure, the epochs evolve from left to right and from top to bottom.

The "Pre" interval shows different partitions with K ranging from 2 to 15, some of them including miss-classified observations (SEEG bipolar leads) according to the silhouette criteria (i.e. they have negative silhouette coefficients). In this interval, during the first 23 epochs it is possible to observe high values of K (14 or 15), while in the last two epochs the K drops till the lowest value allowed (K=2). The "Seizure" interval exhibits a K value equal to 2 for most of the 25 epochs forming the 100 seconds time interval. Next to it, the "Post" time span shows K values from 2 to 15, but opposite to the "seizure" interval behavior, in the "post" one most of the K have the maximum value allowed (K=15).

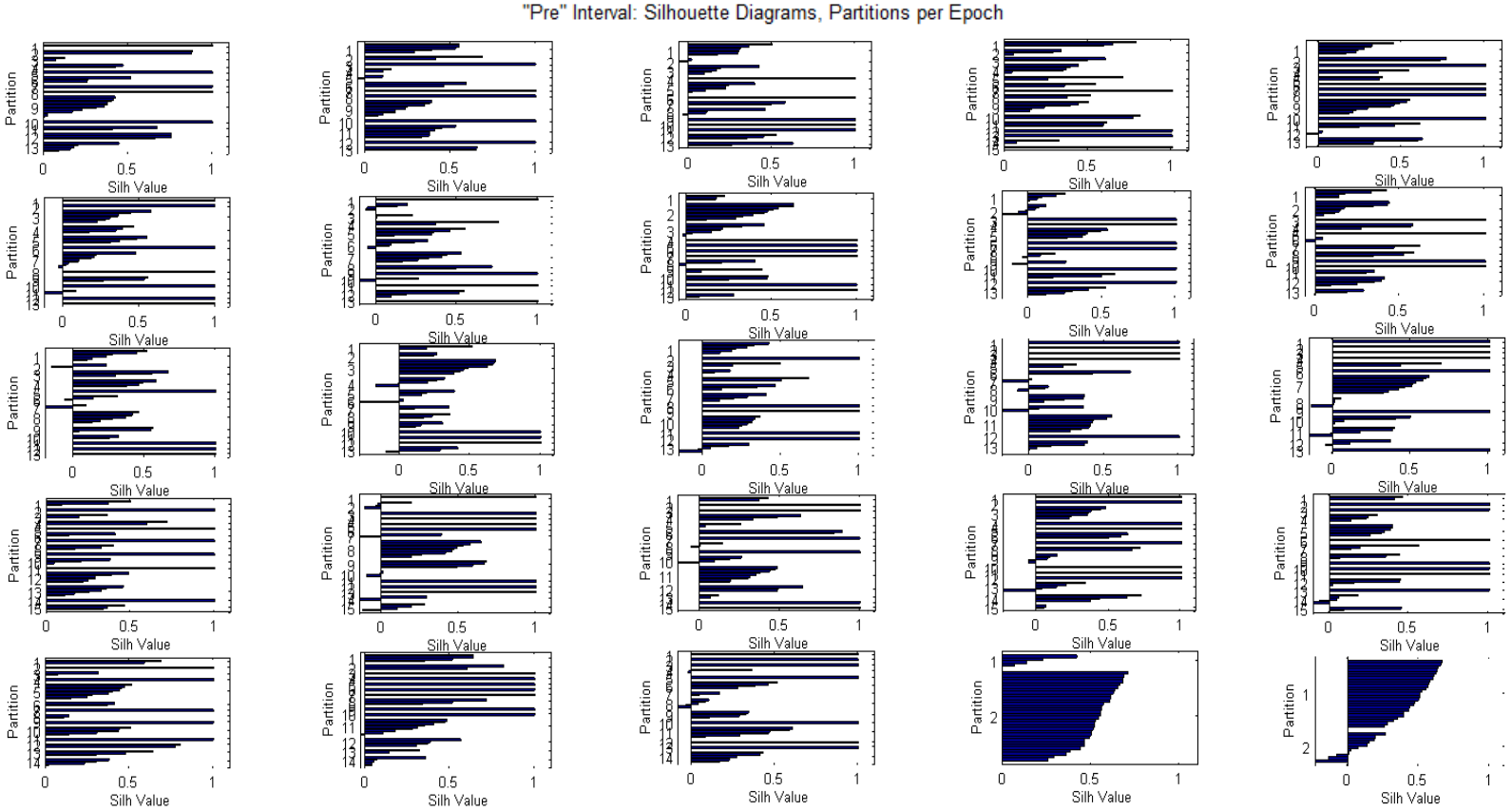


Figure 17. Sequence of silhouette diagrams for each epoch along the 'Pre' interval. The epochs evolve from left to right and from top to bottom.

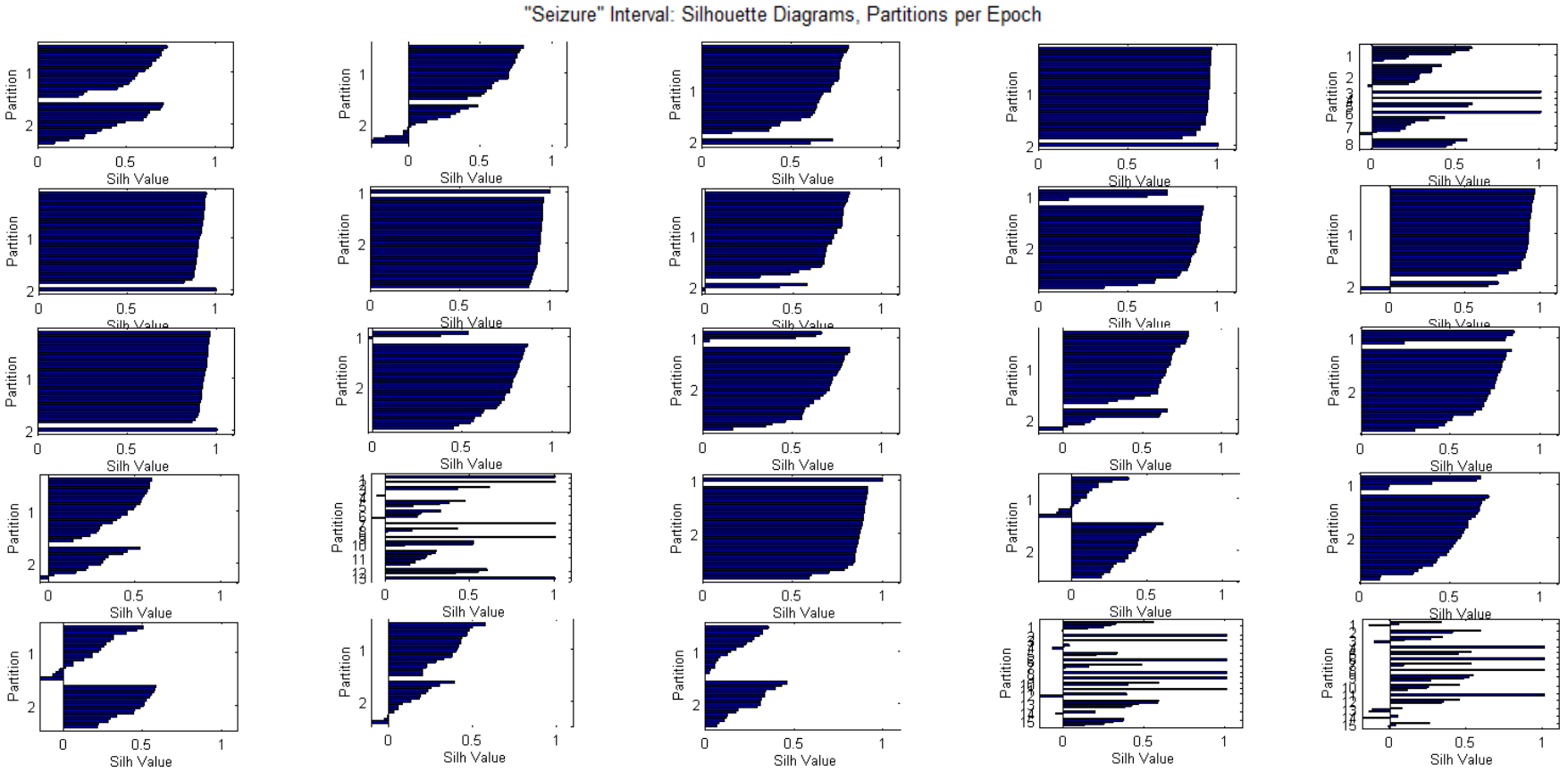


Figure 18. Sequence of silhouette diagrams for each epoch along the 'Seizure' interval. The epochs evolve from left to right and from top to bottom.

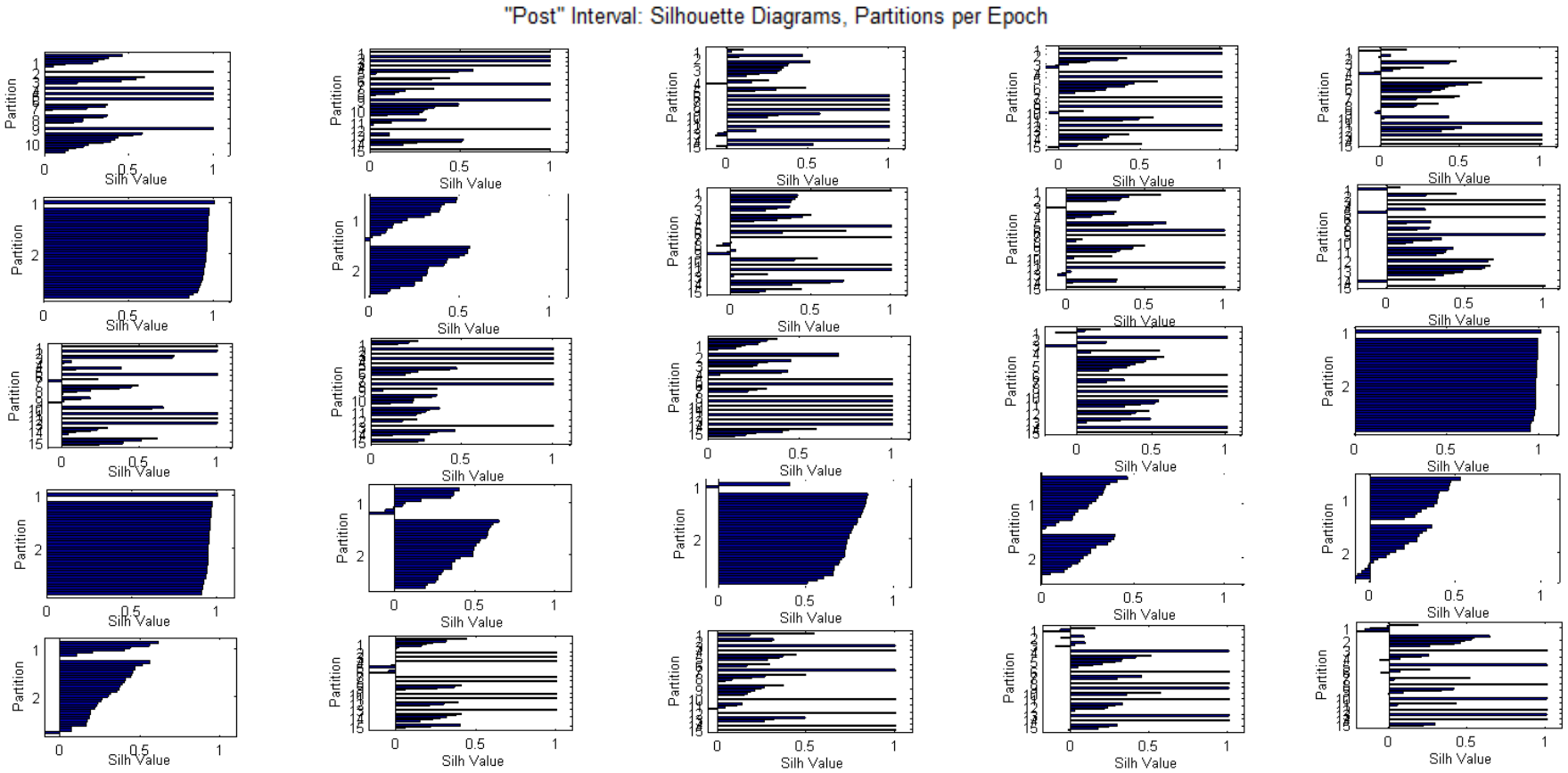


Figure 19. Sequence of silhouette diagrams for each epoch along the 'Post' interval. The epochs evolve from left to right and from top to bottom.

The K values evolution previously described is presented in a clearer way on Figure 20. It shows the changes of the number of groups (K value) found per partition along the epochs. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist.

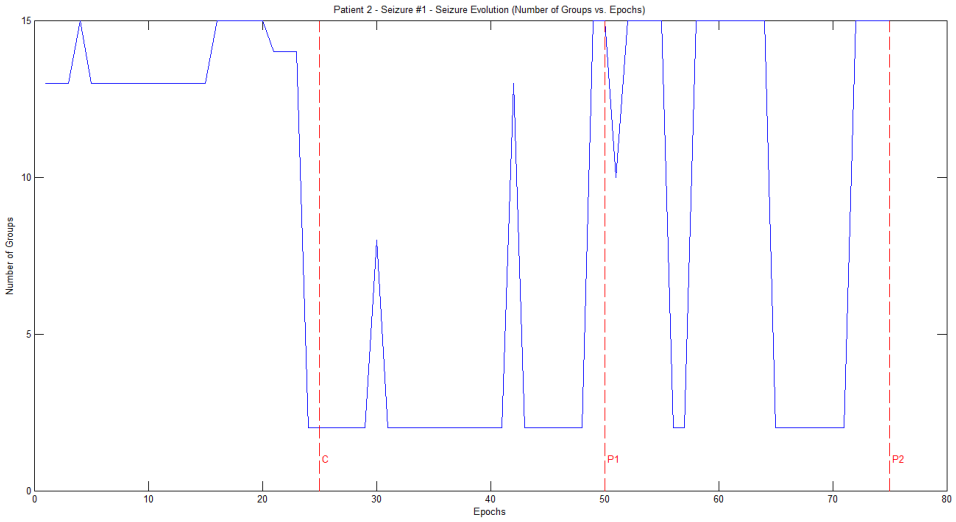


Figure 20. K values evolution along the time epochs. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist (or the end of the "pre" time interval). While P1 and P2 constitute the end of the "seizure" and "post" intervals respectively.

Considering that, once the seizure starts, the Stereo-EEG bipolar leads are grouped into a lower number of clusters per partition, the results obtained may suggest an increase in the synchronicity levels among brain areas related with the seizure beginning for this particular case (i.e. for the patient and seizure under analysis).

On Figure 21 the number of observations (SEEG contacts) per cluster along the epochs is presented. Each circle represents one cluster; its position indicates the number of observations forming that particular group, while the number inside each circle represents the number of clusters with the same number of elements for a certain epoch. In this way, for example, for the first epoch of the 'pre' interval, the algorithm found a partition composed by 12 clusters; five of them formed by one observation (SEEG leads), four groups with two elements, two clusters with three, one with five and one with twelve.

In addition, in the same figure, explicit differences on the behavior among the three time intervals can be distinguished. The “pre” interval is characterized by having a high number of clusters with a small number of elements per group. This behavior changes radically with the beginning of the seizure. So that, during most of the ‘seizure’ interval the partitions found are composed by a few groups containing more observations if compared to the previous interval. Finally, the ‘post’ time span exhibits similar characteristics to the ‘pre’ interval (i.e. high number of groups with few elements forming each of them) for most of the epochs. Moreover, the average number of channels per cluster along the epochs is shown on Figure 22. Again, it becomes clear how, once the seizure starts, the average number of SEEG bipolar leads grouped per cluster increases. Those findings suggest once more a higher ictal-related synchronicity among the contacts under study.

The Appendix B shows, for each one of the 14 patients analyzed, the average results obtained regarding the K values evolution and the average number of observations (SEEG channels) per cluster along the epochs. Similar behaviors to the one previously described were found. The statistical significance of these results will be analyzed later in this Chapter.

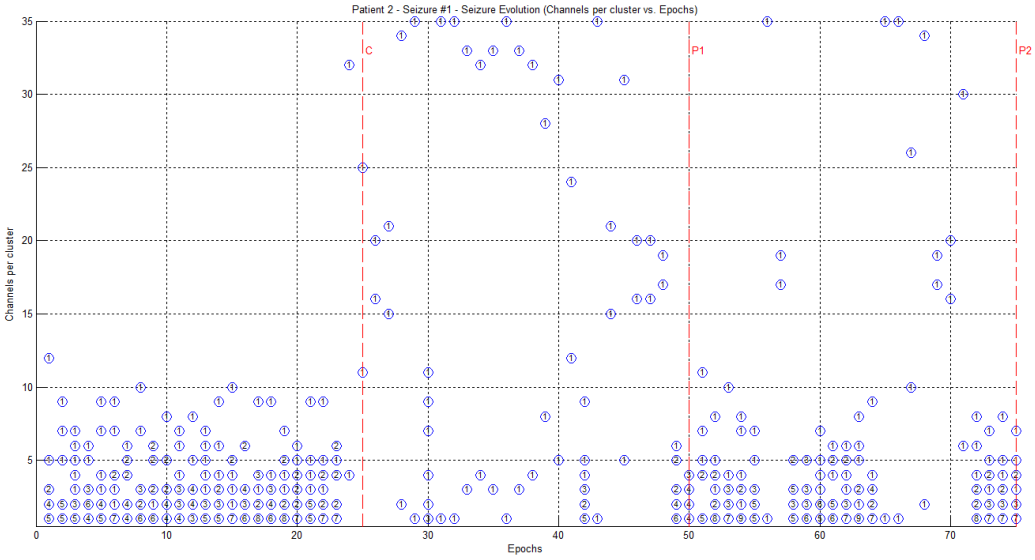


Figure 21. Number of observations (SEEG channels) per cluster along the epochs. Each circle represents one cluster; its position indicates the number of observations forming that particular group, while the number inside each circle represents the number of clusters with the same number of elements for a certain epoch. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist (or the end of the “pre” time interval). While P1 and P2 constitute the end of the “seizure” and “post” intervals respectively.

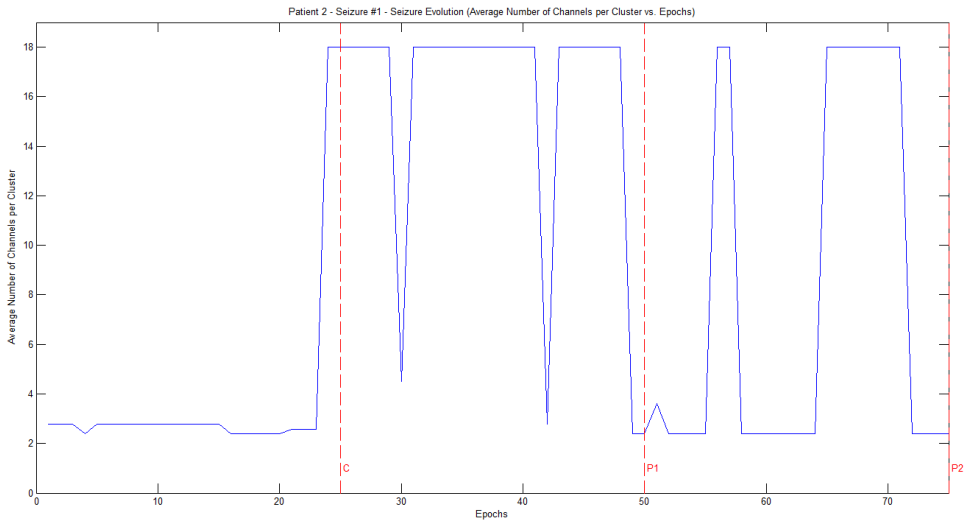


Figure 22. Average number of observations (SEEG channels) per cluster along the epochs. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist (or the end of the "pre" time interval). While P1 and P2 constitute the end of the "seizure" and "post" intervals respectively.

It is important to mention that although the temporal moments corresponding to the beginning of the seizure were indicated by an expert neurologist. The postictal starting points (the end of the seizure) were not provided. This causes some uncertainty while trying to associate the findings of this study with the end of the ictal phase and the beginning of the postictal one.

Besides, the way in which the SEEG bipolar leads are grouped on the partitions found by the algorithm along the 75 time epochs is presented on Figure 23. In this case, for each epoch, the channels belonging to the same cluster have the same color. It is clear again how a decrease in the number of groups appears in correspondence to the seizure starting point. An example of this grouping is shown on TABLE XX. On the left, are presented the names of the SEEG bipolar leads forming each cluster inside the partition found for the first epoch of the 'pre' interval and the number of channels per each group. On the right, the same information is provided regarding the first epoch of the 'seizure' interval. The bipolar leads written in bold characters constitute the representative channels for each cluster (i.e. those which are closer to the centroid of each cluster).

Furthermore, Figure 24 exhibit the representative channels found per each partition along the different time epochs. The blue dots correspond to channels which are equally distant to the centroid of their cluster and are considered both as representative ones for that cluster. The figure provides another perspective regarding the grouping evolution along the three intervals; since there are more groups during the 'pre' and 'post' intervals when compared to the 'seizure' one, the same happens with the number of centroids found.

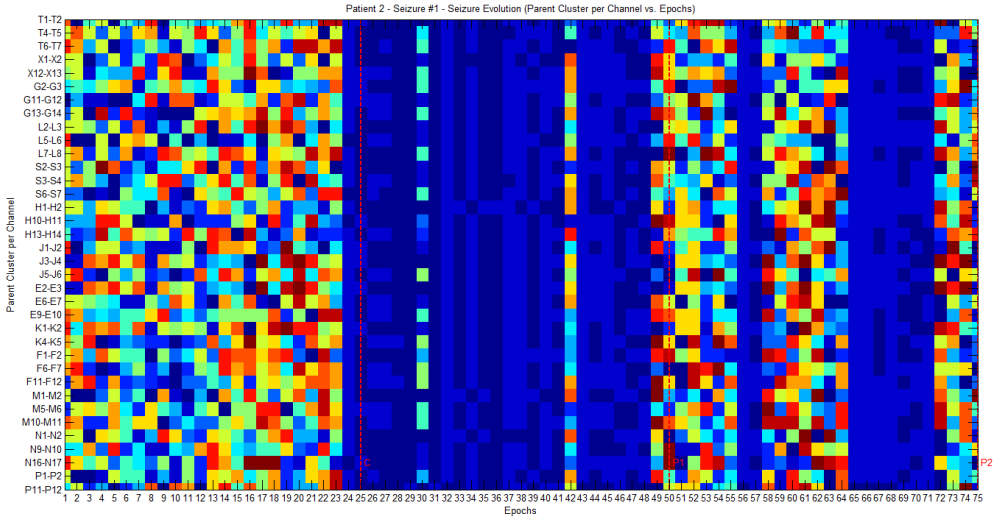


Figure 23. Stereo-EEG leads grouping along the partitions found. For each epoch, the channels belonging to the same cluster have the same color. The dotted red lines constitute the end of each of the three time intervals under study.

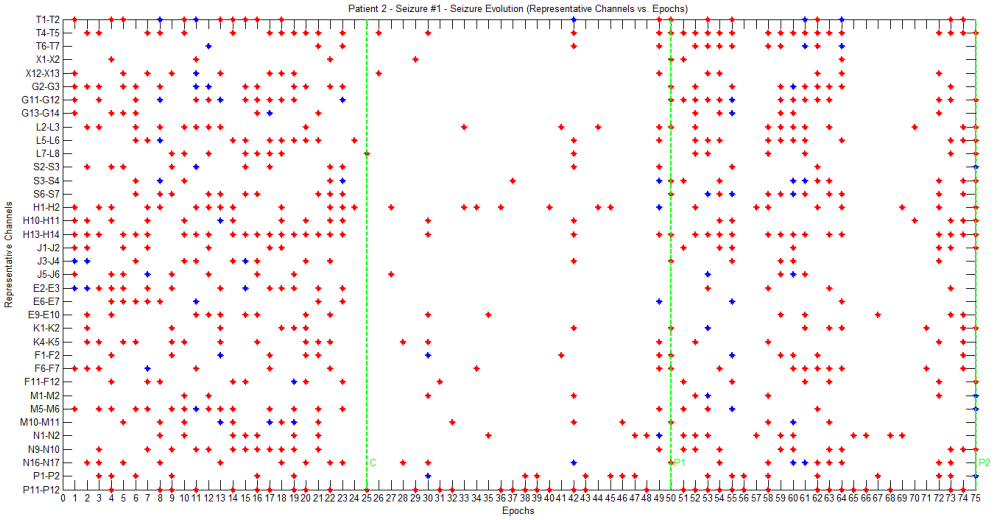


Figure 24. Representative channels found per each partition along the different time epochs. The blue dots represent channels which are equally distant to the centroid of the cluster and are considered both as representative ones for that cluster. The dotted green lines constitute the end of each of the three time intervals under study.

'Pre' Interval – 1ST Epoch	
SEEG Bipolar Leads per Cluster	Number of SEEG Leads per Cluster
G11-G12	1
J3-J4 **E3-E3**	2
F11-F12 **M5-M6**	2
G13-G14 S6-S7	2
H10-H11	1
G2-G3 N9-N10 P11-P12	3
H13-H14	1
X12-X13	1
T4-T5 X1-X2 L2-L3 L7-L8 S2-S3 S3-S4 **H1-H2** E6-E7 F1-F2 M1-M2 N1-N2 P1-P2	12
J5-J6	1
K4-K5 **F6-F7**	2
T1-T2 E9-E10 M10-M11	3
T6-T7 L5-L6 **J1-J2** K1-K2 N16-N17	5

'Seizure' Interval – 1ST Epoch	
SEEG Bipolar Leads per Cluster	Number of SEEG Leads per Cluster
T1-T2 **T4-T5** T6-T7 X1-X2 G2-G3 L2-L3 L7-L8 S2-S3 H1-H2 H10-H11 J1-J2 J3-J4 E2-E3 K1-K2 F1-F2 M1-M2 N1-N2 N9-N10 N16-N17 P11-P12	20
X12-13 G11-G12 G13-G14 L5-L6 S3-S4 S6-S7 H13-14 J5-J6 E6-E7 E9-E10 K4-K5 F6-F7 F11-F12 M5-M6 M10-11 P1-P2	16

Table 3. SEEG bipolar leads grouping example. Left: names of the SEEG bipolar leads forming each cluster inside the partition found for the first epoch of the 'pre' interval and the number of channels per each group. Right: the same information is provided regarding the first epoch of the 'seizure' interval. The bipolar leads written in bold characters constitute the representative channels for each cluster.

Some additional medical information should be known in order to stablish further conclusions about the way the SEEG channels are grouped and about the representative channels found after the analysis carried out. For example, information with reference to the SEEG channels corresponding to the epileptogenic areas according to the pre-surgical evaluation performed by the neurologist or the position of the SEEG electrodes inside the brain would surely contribute to a better understanding of the present results and on the setting down future development tracks for the present project. This will be further discussed on Chapter 4.

3.4.2 Principal Components evolution along the time epochs.

As stated before, the weights of the mixing matrix corresponding to the components which explain the 90% of the total variance in original data were used as input for the K-means clustering in order to obtain a quantitative description of the brain synchronicity evolution over the established time epochs.

Since the PCA decomposition is performed separately for each epoch, a further indication of the evolution of this synchronicity levels might be given by the changes in the number of principal components needed to explain the 90% of the total variance over the epochs. Principal Component Analysis is usually used to re-express large datasets in a way that only the first few components account for as much of the variability possible. So that, if the number of components needed to explain the 90% of the total variance is compared from the decomposition results for two epochs, the one with the lower number of components may be characterizing a higher synchronicity among the electrical signals originating on the different Stereo-EEG leads considered in the decomposition.

Figure 25 shows, for the seizure example under study in the present chapter, the average evolution of the number of principal components needed to explain the 90% of the variance on the original data for each one of the different time epochs considered for the analysis. This evolution shows that the number of principal

components needed to explain the 90% of the variance over the epochs reaches the minimum value closely after the beginning of the seizure.

In the Appendix B, the mean principal components evolution obtained for each one of the 14 patients is presented. These results replicate the behaviour previously found. This decline towards a lower number of components closely after the ictal event may represent a higher seizure-related synchronicity among the electrical signals (brain areas) originating on the different Stereo-EEG bipolar leads considered in the decomposition.

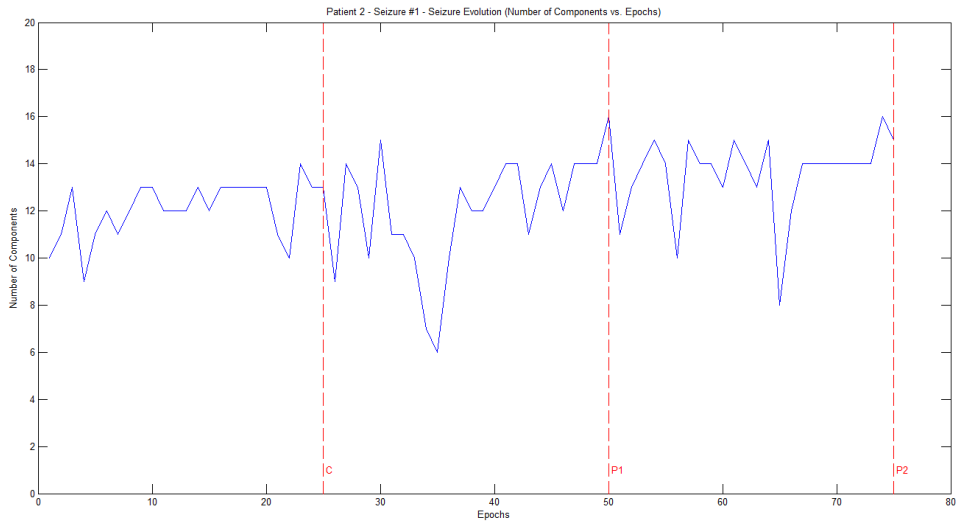


Figure 25. Evolution of the number of principal components needed to explain the 90% of the variance for each one of the different time epochs considered in the present study. In this case, the evolution for the seizure presented in Figure 14 is described.

3.5 Statistical Significance of the results for the population analyzed

In the same Appendix (B) are shown the figures corresponding to the average results obtained for each one of the 14 patients analyzed. In particular, the appendix shows the evolution of the K values and the changes in the average number of observations (SEEG channels) per cluster and in the number of principal components explaining the 90% of the variance in the original data along the epochs. In this section, the statistical significance of these results will be discussed.

From the average results obtained for each patient, the mean number of groups found during the ‘pre’ and ‘post’ intervals was determined. The same calculation was performed for the evolution of the average number of channels per cluster and the number of components. In addition, from the results regarding the evolution of K and the principal components, the minimum value over the ‘seizure’ interval was computed. In the case of the average numbers of channels per cluster, the maximum value corresponding to the ‘seizure’ interval was calculated. The results are presented on Table 4 and were used in order to statistically test the differences found between the ‘seizure’ interval and the other ones. On Figure 26 the corresponding boxplots are presented.

	Number of Groups vs. Epochs			Average Number of Channels per cluster vs. Epochs			Number of Principal Components (90%Var) vs. Epochs		
Patient	Mean Pre	Min Seizure	Mean Post	Mean Pre	Max Seizure	Mean Post	Mean Pre	Min Seizure	Mean Post
1	6.14	2.00	7.20	8.86	18.50	9.10	14.34	10.00	15.56
2	7.66	2.00	6.85	10.62	18.00	10.79	11.88	6.67	13.29
3	13.36	2.00	4.84	3.07	17.00	9.44	12.08	12.00	18.92
4	2.00	2.00	5.52	20.50	20.50	11.49	16.52	13.00	16.60
5	7.02	2.33	5.38	9.99	14.67	11.34	12.12	8.00	11.71
6	11.72	2.00	3.32	5.02	16.00	10.77	12.36	4.00	12.72
7	5.50	2.00	3.56	10.13	16.00	10.87	13.96	4.50	11.52
8	13.96	2.00	6.32	3.45	17.00	11.83	11.72	1.00	11.12
9	2.00	2.00	-----	21.50	21.50	-----	12.76	5.00	-----
10	4.12	2.00	12.88	10.56	16.00	3.26	14.60	3.00	11.80
11	8.30	2.00	7.80	7.77	16.00	8.58	12.84	6.00	12.07
12	6.24	2.00	2.28	12.12	17.50	15.87	15.42	9.50	14.74
13	8.52	6.88	8.88	8.48	8.49	8.03	12.87	10.67	15.04
14	10.56	2.00	7.12	5.80	17.50	10.49	9.76	3.00	9.84
Mean	7.65	2.37	6.30	9.85	16.76	10.14	13.09	6.88	13.46
SD	3.76	1.30	2.74	5.47	3.00	2.83	1.73	3.71	2.56

Table 4. Parameters considered in order to perform the statistical testing procedure.

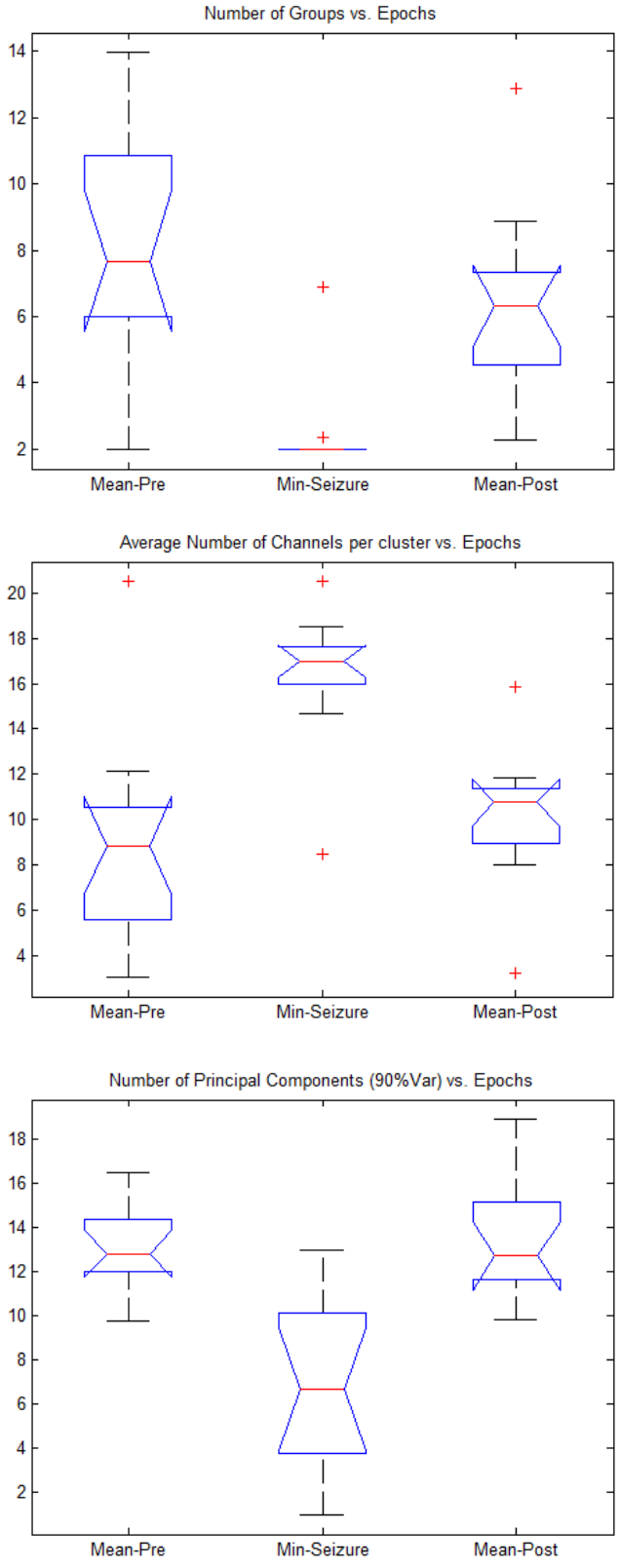


Figure 26. Boxplots corresponding to each of the parameters' evolution under analysis.

Considering the low number of samples (14 patients) analyzed on the present study, it is not proper to assume that the data obtained is normally distributed. So that, it is necessary to perform a statistical test which doesn't have any particular constraint regarding the data distribution. The Kruskal-Wallis Test fills this requirement. It is the non-parametric alternative to the one-way Analysis of Variance (ANOVA) and an extension of the Wilcoxon rank sum test for more than two groups. It compares the medians of the ranks associated with the group samples in order to test for the null hypothesis that all samples are drawn from the same population (or from populations with the same distribution) [44]. This hypothesis would be rejected if $p < 0.05$. The present testing procedure aims to evaluate the statistical significance of the difference found between the ictal-related behavior and the behavior along the 'pre' and 'post' intervals.

On Table 5 the p-values obtained for each of the three cases under consideration (i.e. number of groups, average number of channels per cluster and number of components along the epochs) are presented. The p-value is considered significant if lower than 0.05.

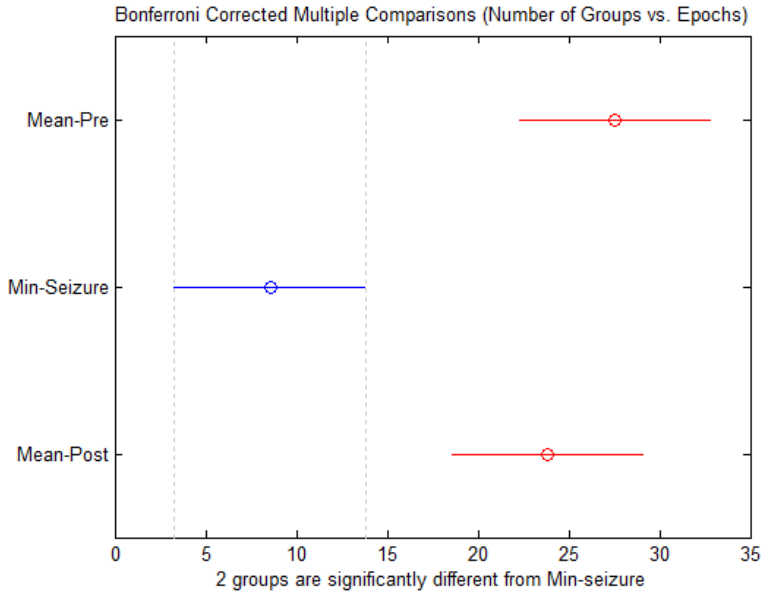
Number of Groups vs. Epochs		Average Number of Channels per cluster vs. Epochs		Number of Principal Components (90%Var) vs. Epochs	
p-value	0.00002933*	p-value	0.0001*	p-value	0.0001*

Table 5. P-values corresponding to the Kruskal-Wallis Test performed for the 3 cases under consideration. *The information regarding the patient 9 was not included in these calculations since no information regarding the 'post' interval is available.

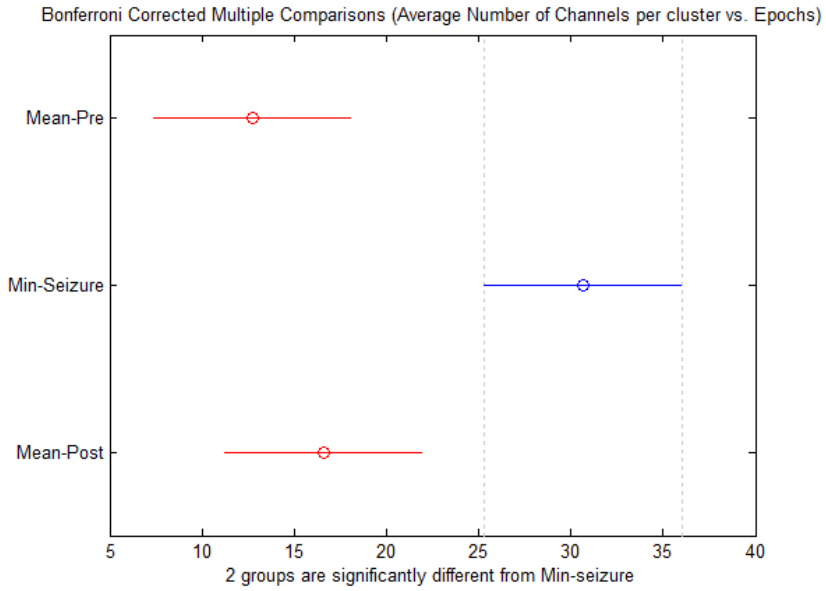
According to the p-values obtained, the null hypothesis can be rejected in all the cases. Then, for the three parameters analyzed, the 'pre', 'seizure' and 'post' intervals are statistically different.

The analysis performed compared the several groups together, but the conclusion obtained in this way might be too general. For the present analysis in particular, a specific comparison between the 'seizure' interval and each of the other ones is

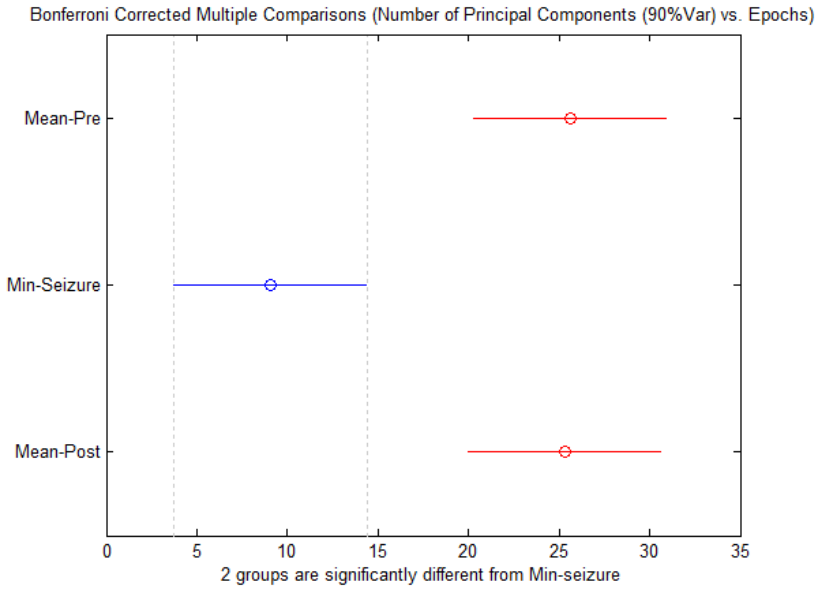
needed. So that, a Bonferroni corrected Multiple Comparisons approach was executed. This allows for the pairwise comparison required. In Figure 27 the outcome provided by the multiple comparison test performed in Matlab is provided. In the figure are shown, for each group, the estimated medians with the comparison intervals around them. Disjoint comparison intervals (i.e. no overlapping, lacking of intersection) indicate a significant statistical difference among groups, while those overlapping imply that there is no evidence for rejecting the null hypothesis when comparing that particular couple of groups.



A)



B)



C)

Figure 27. Bonferroni corrected Multiple Comparisons outcome. a) Number of Groups vs. Epochs. b) Average Number of Channels per Cluster case vs. Epochs. c) Number of Principal Components (90%var) vs. Epochs. The estimated medians with comparison intervals around them are shown. Disjoint comparison intervals indicate a significant statistical difference among groups, while those overlapping imply that there is no evidence for rejecting the null hypothesis when comparing that couple of groups.

The results indicate, for each of the three cases analysed, that there is a significant difference between the 'seizure' and 'pre' groups. So the test rejects the null hypothesis that the data in these two groups comes from the same distribution. The same is true when comparing the 'seizure' group and the 'post' one. This suggests that the previously discussed seizure-related increase on the level of synchronization among brain areas is statistically significant. Nonetheless, since no information was provided regarding the end of the seizures, the minimum value of the 'seizure' interval was used for the testing procedure. A more proper evaluation could be performed by using the mean values for the whole duration of the seizure, if it was known.

Finally, from the multiple comparisons procedure it is also possible to conclude that there is not a significant difference between the 'pre' and the 'post' groups, so the test does not reject the null hypothesis that these two groups come from the same distribution. There is no convincing evidence to affirm that they differ. This may suggest that the alterations on the synchronicity levels related with the ictal event are not preserved during the 'post' interval.

CHAPTER 4

Conclusions and Future Developments

4.1 Discussions and Conclusions

Stereoelectroencephalography (SEEG) is a methodology used in the pre-surgical evaluation of epileptic patients in order to establish the extent of the Epileptogenic Zone (EZ) and plan its surgical resection. SEEG is ideal for studying the relationships between the structures concerned in seizure production and propagation, and can potentially contribute to the understanding of the origin and spread of seizures.

In the present study, the SEEG data obtained from the pre-surgical evaluation performed in 14 neurosurgical patients affected by Focal Cortical Dysplasia Type II are used in order to obtain a quantitative description of the synchronicity levels evolution among brain areas before and during seizures. This is achieved by applying a Principal Component Analysis-based K-means clustering algorithm over different epochs (4 seconds each) along three consecutive predefined time intervals (‘pre’, ’seizure’, ‘post’) around the seizure events. Each interval is 100 seconds long. The “seizure” interval starts in the moment in which the seizure begins, according to the information provided by the neurologist. The “pre” interval commences 100 seconds before the seizure begins, and the “post” interval initiates immediately after the end of the “seizure” one. So that, a 300 seconds time span is covered, starting 100 seconds before the beginning of the seizure.

Since the population under analysis is affected by focal seizures, which derive from a small group of neurons (the seizure focus) and then spread further, the initial hypothesis was to expect a progressive increase in the levels of synchronization among brain areas when moving from the ‘pre’ interval towards the ‘seizure’ one, reaching a maximum during the ictal event.

The evolution of the K-values along each of the time epochs was analyzed for the three pre-defined intervals. This evolution shows that, once the seizure starts, the Stereo-EEG contacts tend to be grouped into a lower number of clusters per partition. This also means that the average number of Stereo-EEG bipolar leads grouped per cluster along the epochs increases in relation to the ictal event. Furthermore, the evolvement of the number of principal components needed to explain the 90% of the variance on the original data for each of the time epochs was studied. The results exhibit a decline towards a lower number of components needed after the ictal event. All these findings suggest a higher seizure-related synchronicity among the electrical activity of the different brain areas covered by the Stereo-EEG electrodes. The statistical significance of these findings was positively evaluated according to the Kruskal-Wallis Test followed by a Bonferroni corrected Multiple Comparisons procedure.

Moreover, the way in which the Stereo-EEG bipolar leads are grouped along the partitions was described (i.e. the names of the SEEG channels forming each group were presented) and the representative channels per each partition were calculated as those closer to the centroid of each cluster. However, some additional clinical information should be known in order to stablish further conclusions in this regard.

In fact, although the temporal moments corresponding to the beginning of the seizure were indicated by an expert neurologist, further information would be needed in order to improve the understanding of the results here obtained:

- The postictal starting points (the end of the seizure) were not provided. This causes some uncertainty while trying to associate the findings of this study with the end of the ictal phase and the beginning of the postictal one.
- Information regarding the SEEG channels corresponding to the epileptogenic areas according to the pre-surgical evaluation performed by the neurologist would be useful in order to make a comparison between the representative channels found in the present study and the epileptogenic ones identified by the clinicians, allowing finding out if there is any potential relationship between them.

- Knowing the position of the SEEG electrodes inside the brain would make possible to relate the present findings to anatomical brain areas, allowing to have a clearer idea about how to proceed with future steps/developments of the present study (e.g. to define how to perform a connectivity analysis over the brain regions involved in the seizure process).
- Since the SEEG signals from the population analyzed were recorded during sleeping periods, obtaining the hypnogram describing the sleep stages' evolution for each patient would be fruitful in order to evaluate the influence of the sleep stage to the synchronicity levels found in the present thesis work.

4.2 Future Developments

In this section some possible future developments for the present thesis work are suggested. Some of them were already briefly discussed in the previous section in regard to the additional medical information needed in order to improve the understanding of the results here obtained.

- *Modifying the epochs' length*

Here, fixed 4 seconds long time epochs were analysed. However, the Stereo-EEG signals exhibit different kinds of electrical activity. By defining shorter time epochs a more detailed description of the seizure evolution could be performed. Another possible track would be to develop an adaptive algorithm able to stablish the epochs' length according to the SEEG signal evolution.

- *Relationship with Sleep*

As presented before, the potential relationship between sleep and epilepsy has been recognized since ancient times [8] [15]. In literature it is commonly accepted that epileptic seizures are more likely to happen in NREM than in REM sleep [8, 14-16]. In fact, NREM sleep tends to facilitate partial seizures, especially at stages 1 and 2 (lighter stages, non-SWS), while REM sleep seems to inhibit them [17]. Moreover, in

some types of epilepsy, such as Focal Cortical Dysplasia Type II, has been demonstrated an increase in the risk of sleep-related seizures [18].

If the related sleep stages corresponding to the acquired data are not provided by the neurologist, they could be identified from the recordings by using one of the spectral procedures found in literature [45-48]. Analysing that information together with the results obtained in the present study would be helpful in order to evaluate the contribution made by sleep to the synchronicity levels found in the present thesis work and would either confirm what has been previously found in literature or suggest new findings regarding the relationship between the different stages of sleep and seizures.

- *Connectivity Analysis*

"Neurons and neural populations do not function as islands onto themselves. Rather, they interact with other such elements through their afferent and efferent connections in an orchestrated manner" [49]. Nowadays, how these cortical regions communicate one to each other, i.e. the concept of brain connectivity, beyond the simple mapping of their activity, is becoming a central topic in neuroscience [50] [51].

The concept of brain connectivity can be subdivided into three main categories: Structural, Functional and Effective Connectivity. Effective connectivity is particularly interesting for our purposes, since it describes the interactions between brain regions, taking into account the direction of the information flow among them [7]. So that, it aims at describing causal influences [5]. Effective connectivity can be estimated using model-based techniques, or data-driven ones. Model-based techniques, such as Structural Equation Modeling (SEM) and Dynamic Causal Modeling (DCM) are based on an a priori model specifying the causal links (i.e. the structural graph must be previously known), allowing the assessment of some a priori proposed causal interactions [5, 51-53]. They are considered confirmatory methods for some hypothesis, rather than exploratory ones [52]. On the other hand, data-driven techniques do not specify any a priori model about neuronal interactions, allowing the estimation of effective connectivity directly from the signals when no prior knowledge is available [51].

A large group of effective connectivity measures are based on the concept of Granger causality (GC) [7]. Some of the techniques developed under the GC method are: the Directed Transfer Function (DTF), the Partial Directed Coherence (PDC) and the direct DFT (dDFT). All of them are frequency domain approaches, and all are based on a Multivariate Autoregressive (MVAR) modelling [54]. More detailed information about connectivity analysis can be found in [55-66].

As a further development for the present study, a Granger Causality connectivity analysis could be performed over the previously identified brain areas involved in the seizure evolution (among or within them). This approach would probably contribute to the understanding of the mechanisms underlying the origin and spread of epileptic seizures, focusing on the relationship between the cortical areas involved before, during and after them.

- *Phase Coupling Within and Between Brain Regions*

In order to continue contributing to the understanding of the mechanisms underlying the origin and spread of epileptic seizures, a further study regarding the phase coupling between brain oscillations (within and between the brain regions involved in the seizure) during the whole process could be performed.

On brain dynamics, diverse transient local oscillations are generated. These are able to "mediate coordinated interactions within and between different neuronal subsystems" [67]. In order to study their relationship, some already available phase coupling methods might be useful. Three main approaches are distinguished in literature: They are based either on the Hilbert Transform [68, 69], Wavelets [67, 70] or High Order Spectra Analysis [71-75]. The previously cited methods could be also used in order to improve the understanding about the influence of the Infraslow Oscillations in the production of seizures.

BIBLIOGRAPHY

[1] Brain Basics: Know Your Brain. National Institute of Neurological Disorders and Stroke (NINDS). Last updated April 17, 2015. Available at: http://www.ninds.nih.gov/disorders/brain_basics/know_your_brain.htm.

[2] Shannon JB. Brain Disorders Sourcebook. Omnigraphics Incorporated; 2010.

[3] Panayiotopoulos, C. P., Benbadis, S. R., Beran, R. G., Berg, A. T., Engel Jr, J., & Galanopoulou, A. S. (Eds.). (2010). *Atlas of epilepsies* (Vol. 3). Springer Science & Business Media.

[4] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science, 5th edition*, McGraw-Hill, 2012.

[5] F. Pittau, P. MÃ©gevand, L. Sheybani, E. Abela, F. Grouiller, L. Spinelli, C. M. Michel, M. Seeck, and S. Vulliemoz, "Mapping Epileptic Activity: Sources or Networks for the Clinicians?," *Front. Neurol.*, vol. 5, no. November, pp. 1–21, Nov. 2014.

[6] M. Papadopoulou, K. Vonck, P. Boon, and D. Marinazzo, "Mapping the epileptic brain with EEG dynamical connectivity: Established methods and novel approaches," *Eur. Phys. J. Plus*, vol. 127, no. 11, p. 144, Nov. 2012.

[7] F. Panzica, G. Varotto, F. Rotondi, R. Spreafico, and S. Franceschetti, "Identification of the Epileptogenic Zone from Stereo-EEG Signals: A Connectivity-Graph Theory Approach.," *Front. Neurol.*, vol. 4, no. November, p. 175, Jan. 2013.

[8] Bingaman, W. *Textbook of epilepsy surgery*. Ed. Hans Lüders. London: Informa Healthcare, 2008.

[9] Munari, C.; Hoffmann, D.; Francione, S.; Kahane, P.; Tassi, L.; Lo Russo, G. & Benabid, A. Stereo-electroencephalography methodology: advantages and limits. *Acta neurologica Scandinavica. Supplementum*, 1994. PMID:8209659

[10] Wieser, H. & Elger, C. *(Eds.)*Stereo-Electroencephalography *Presurgical Evaluation of Epileptics, Springer Berlin Heidelberg,* 1987, 192-

[11] Guyton, Arthur, and John Hall. "Textbook of medical physiology, 11th." (2006).

[12] J. Kabat and P. Król, "Focal cortical dysplasia - review.," *Pol. J. Radiol.*, vol. 77, no. 2, pp. 35–43, 2012.

[13] S. M. Sisodiya, S. Fauser, J. H. Cross, and M. Thom, “Focal cortical dysplasia type II: biological features and clinical perspectives.,” *Lancet Neurol.*, vol. 8, no. 9, pp. 830–843, 2009.

[14] M. C. Walker and S. H. Eriksson, “Epilepsy and sleep disorders,” *Eur. Neurol. Rev.*, vol. 6, no. 1, pp. 60–63, 2011.

[15] S. S. Tallavajhula and J. D. Slater, “Sleep and Epilepsy,” *Sleep Med. Clin.*, vol. 7, no. 4, pp. 619–630, Dec. 2012.

[16] B. A. Malow, “The interaction between sleep and epilepsy,” in *Epilepsia*, 2007, vol. 48, no. SUPPL. 9, pp. 36–38.

[17] S. T. Herman, T. S. Walczak, and C. W. Bazil, “Distribution of partial seizures during the sleep-wake cycle,” *Neurology*, vol. 56, pp. 1453–1459, 2001.

[18] Nobili, L., Cardinale, F., Magliola, U., Cicolin, A., Didato, G., Bramerio, M., ... & Cossu, M. (2009). Taylor’s focal cortical dysplasia increases the risk of sleep-related epilepsy. *Epilepsia*, *50*(12), 2599-2604.

[19] C. D. Binnie, R. D. Elwes, C. E. Polkey, and A. Volans, “Utility of stereoelectroencephalography in preoperative assessment of temporal lobe epilepsy.,” *J. Neurol. Neurosurg. Psychiatry*, vol. 57, no. 1, pp. 58–65, 1994.

[20] Cleveland Clinic. A Guide to Stereoelectroencephalography. http://my.clevelandclinic.org/ccf/media/files/Epilepsy_Center/SEEG%20Fact_Sheet_022210.pdf

[21] Fontana, Luca. Automatic classifier of stereoelectroencephalographic signals for epileptogenic zone localization in drug resistant epilepsy. 2013.

[22] J. Engel, T. A Pedley, and J. Aicardi. Epilepsy: a comprehensive textbook, volume 1. Lippincott Williams & Wilkins, 2008.

[23] J. Gonzalez-Martinez, J. Mullin, S. Vadera, J. Bulacio, G. Hughes, S. Jones, R. Enatsu, and I. Najm, “Stereotactic placement of depth electrodes in medically intractable epilepsy.,” *J. Neurosurg.*, vol. 120, no. 3, pp. 639–44, 2014.

[24] The Eloquent Cortex. Available at: http://en.wikipedia.org/wiki/Eloquent_cortex

[25] Saeid, Sanei, and J. A. Chambers. "EEG signal processing" *Chichester: John Willey & Sons. Ltd* (2007).

[26] Webster, John G., ed. *Medical Devices and Instrumentation*. Wiley-Interscience, 1988.

[27] Hirsch, Lawrence, and Richard Brenner, eds. *Atlas of EEG in critical care*. John Wiley & Sons, 2011

[28] Picchioni, Dante, et al. "Infraslow EEG oscillations organize large-scale cortical–subcortical interactions during sleep: a combined EEG/fMRI study."*Brain research* 1374 (2011): 63-72.

[29] Vanhatalo, Sampsa, et al. "Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep." *Proceedings of theNational Academy of Sciences of the United States of America* 101.14 (2004): 5053-5057.

[30] Rodin, E., and P. Modur. "Ictal intracranial infraslow EEG activity." *Clinical Neurophysiology* 119.10 (2008): 2188-2200.

[31] Keković, Goran, et al. "The slow and infraslow oscillations of cortical neural network." *Neurology, Psychiatry and Brain Research* 18.4 (2012): 175-180.

[32] Vercellis, Carlo. *Business Intelligence: modelli matematici e sistemi per le decisioni*. McGraw-Hill, 2006.

[33] Mafessanti F, Slides Laboratorio di elaborazione di bioimmagini, Politecnico di Milano, 2014.

[34] Antonini, Marco. "Analisi statistica dei segnali elettrici cerebrali in pazienti in stato vegetativo e di minima coscienza." (2012).

[35] Richardson, Mark. Principal component analysis, “Special topic”. University of Oxford, 2009.

[36] Jain, Anil K. "Data clustering: 50 years beyond K-means." *Pattern recognition letters* 31.8 (2010): 651-666.

[37] Ding, Chris, and Xiaofeng He. "K-means clustering via principal component analysis." *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004.

[38] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Applied statistics* (1979): 100-108.

[39] Zha, Hongyuan, et al. "Spectral relaxation for k-means clustering." *Advances in neural information processing systems*. 2001.

[40] Caiani E, Slides Laboratorio di elaborazione di bioimmagini, Politecnico di Milano, 2014.

[41] MathWorks, Inc. K-means Clustering. Matlab. Available at: http://it.mathworks.com/help/stats/kmeans.html

[42] The K-means Clustering Algorithm. Department of Electronic Systems. Aarbor Universitet. kom.aau.dk/group/04gr742/pdf/kmeans_worksheet.pdf

[43] Pigorini, A., Sarasso, S., Proserpio, P., Szymanski, C., Arnulfo, G., Casarotto, S., ... & Massimini, M. (2015). Bistability breaks-off deterministic responses to intracortical stimulation during non-REM sleep. *NeuroImage*, *112*, 105-113.

[44] MathWorks, Inc. Statistics Toolbox for use with Matlab. Available at: http://www.pi.ingv.it/~longo/CorsoMatlab/OriginalManuals/stats.pdf

[45] Swarnkar, Vinayak, and Udantha R. Abeyratne. "Bispectral analysis of single channel EEG to estimate macro-sleep-architecture." *International Journal of Medical Engineering and Informatics* 6.1 (2014): 43-64.

[46] R Roessler, F Collins, R Ostman, A period analysis classification of sleep stages, Electroencephalography and Clinical Neurophysiology, Volume 29, Issue 4, October 1970, Pages 358-362, ISSN 0013-4694.

[47] L.E Larsen, D.O Walter, On automatic methods of sleep staging by EEG spectra, Electroencephalography and Clinical Neurophysiology, Volume 28, Issue 5, May 1970, Pages 459-467, ISSN 0013-4694.

[48] Thien Thanh Dang-Vu, Martin Desseilles, Steven Laureys, Christian Degueldre, Fabien Perrin, Christophe Phillips, Pierre Maquet, Philippe Peigneux, Cerebral correlates of delta waves during non-REM sleep revisited, NeuroImage, Volume 28, Issue 1, 15 October 2005, Pages 14-21, ISSN 1053-811.

[49] B. Horwitz, "The elusive concept of brain connectivity," *NeuroImage*, vol. 19, no. 2. pp. 466–470, 2003.

[50] L. Astolfi, F. Cincotti, D. Mattia, M. G. Marciani, L. A. Baccalà, F. D. V. Fallani, S. Salinari, M. Ursino, M. Zavaglia, and F. Babiloni, "Assessing Cortical Functional Connectivity by Partial Directed Coherence : Simulations and Application to Real Data," vol. 53, no. 9, pp. 1802–1812, 2006.

[51] L. Astolfi, F. Cincotti, D. Mattia, F. D. V. Fallani, S. Salinari, M. Ursino, M. Zavaglia, M. G. Marciani, and F. Babiloni, "Estimation of the cortical connectivity patterns during the intention of limb movements," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 4, pp. 32–38, 2006.

[52] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG.," *Comput. Biol. Med.*, vol. 41, no. 12, pp. 1110–7, Dec. 2011.

[53] Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (Eds.). (2011). *Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images*. Academic press.

[54] L. Astolfi, F. Cincotti, D. Mattia, F. de Vico Fallani, M. Lai, L. Baccala, S. Salinari, M. Ursino, M. Zavaglia, and F. Babiloni, "Comparison of different multivariate methods for the estimation of cortical connectivity: simulations and

applications to EEG data.," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 5, pp. 4484–7, Jan. 2005.

[55] J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. a Morettin, and L. a Baccalá, "Frequency domain connectivity identification: an application of partial directed coherence in fMRI.," *Hum. Brain Mapp.*, vol. 30, no. 2, pp. 452–61, Feb. 2009.

[56] A. Korzeniewska, M. Mańczak, M. Kamiński, K. J. Blinowska, and S. Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," *J. Neurosci. Methods*, vol. 125, no. 1–2, pp. 195–207, May 2003.

[57] K. J. Blinowska, "Review of the methods of determination of directed connectivity from multichannel data.," *Med. Biol. Eng. Comput.*, vol. 49, no. 5, pp. 521–9, May 2011.

[58] M. G. Tana, R. Sclocco, and A. M. Bianchi, "GMAC: a Matlab toolbox for spectral Granger causality analysis of fMRI data.," *Comput. Biol. Med.*, vol. 42, no. 10, pp. 943–56, Oct. 2012.

[59] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econom. J. Econom. Soc.*, vol. 37, no. 3, pp. 424–438, 1969.

[60] M. Ding, Y. Chen, and S. L. Bressler, "Granger Causality: Basic Theory and Application to Neuroscience," no. February, Aug. 2006.

[61] L. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biol. Cybern.*, vol. 474, pp. 463–474, 2001.

[62] K. Sameshima and L. A. Baccala, "Using partial directed coherence to describe neuronal ensemble interactions," vol. 94, pp. 93–103, 1999.

[63] P. Franaszczuk and G. Bergey, "Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures," *Brain Topogr.*, vol. 21201, no. October 1997, pp. 1–17, 1998.

[64] P. J. Franaszczuk, G. K. Bergey, and M. J. Kamiński, "Analysis of mesial temporal seizure onset and propagation using the directed transfer function method.," *Electroencephalogr. Clin. Neurophysiol.*, vol. 91, no. 6, pp. 413–427, 1994.

[65] M. Kamiński, M. Ding, W. Truccolo, and S. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 157, pp. 145–157, 2001.

[66] R. Kuś, M. Kamiński, and K. J. Blinowska, “Determination of EEG activity propagation: Pair-wise versus multichannel estimate,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 9, pp. 1501–1510, 2004.

[67] Le Van Quyen, Michel, and Anatol Bragin. "Analysis of dynamic brain oscillations: methodological advances." *Trends in neurosciences* 30.7 (2007): 365-373.

[68] Miyakoshi, Makoto, et al. "Automated detection of cross-frequency coupling in the electrocorticogram for clinical inspection." *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013.

[69] Voytek, Bradley, et al. "Shifts in gamma phase–amplitude coupling frequency from theta to alpha over posterior cortex during visual tasks." *Frontiers in human neuroscience* 4 (2010).

[70] Li, Xiaoli, et al. "Interaction dynamics of neuronal oscillations analysed using wavelet transforms." *Journal of neuroscience methods* 160.1 (2007): 178-185.

[71] Sigl, Jeffrey C., and Nassib G. Chamoun. "An introduction to bispectral analysis for the electroencephalogram." *Journal of clinical monitoring* 10.6 (1994): 392-404.

[72] Bianchi, A. M., et al. "Spectral and bispectral analysis of the EEG rhythms in basal conditions and during photic stimulation." *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*. Vol. 1. IEEE, 2004.

[73] MathWorks, Inc, et al. *Higher-order Spectral Analysis Toolbox: for Use with MATLAB: User's Guide*. Mathworks, Incorporated, 1998.

[74] Marceglia, S., et al. "Dopamine-dependent non-linear correlation between subthalamic rhythms in Parkinson's disease." *The Journal of physiology* 571.3 (2006): 579-591.

[75] Marceglia, Sara, et al. "Interaction between rhythms in the human basal ganglia: application of bispectral analysis to local field potentials." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 15.4 (2007): 483-492.

APPENDIX A

Five-Tier Epilepsy Classification

In [8] a 5-tier classification of epilepsy is proposed. It is presented here since it considers epilepsy from different levels, providing a holistic view of the disease. In fact, two of these five tiers (semiological characteristics of the seizures and seizure frequency) define the epileptic seizures, while the other three (etiology, related medical conditions and location of the epilepsy) define what is producing the epilepsy and the location of the brain abnormality. The 5 proposed tiers are presented below:

Tier 1: Epilepsy

The first Tier defines the location of the epileptogenic zone. The precision of the location varies depending on the number and type of diagnostic procedures performed (EEG, MRI, PET, SPECT, iEEG, etc.). The following subdivisions can be defined:

1. Focal: epileptogenic zone located within one cortical lobe
    a. Frontal
    b. Perirolandic
    c. Temporal
2. Neocortical temporal
    a. Mesial temporal
    b. Parietal
    c. Occipital
    d. Other
3. Multilobar: epileptogenic zone affects more than one brain lobe
    a. Bilobar homotopic: epileptogenic zone affects two homotopic brain lobes (bitemporal, bifrontal, etc.)
    b. Other
4. Generalized: epileptogenic zone is bilateral, diffusely distributed affecting most of or the entire brain cortex.

Tier 2: Semiological seizure classification

It refers to the clinical expression of a patient with epilepsy. Here the classification is made according exclusively to the clinical semiology, independent of the results of other tests (EEG, MRI, PET, SPECT, etc). It relies only on the *clinical observation made* by the physician (video recordings), by the patient itself or other direct observers.

1. Auras
    a. Somatosensory auras
    b. Visual auras
    c. Auditory auras
    d. Gustatory auras
    e. Olfactory auras
    f. Autonomic auras
    g. Abdominal auras
    h. Psychic auras
2. Autonomic Seizures
3. Dialeptic Seizures
4. Motor Seizures
    a. Simple Motor Seizures
        i. Myoclonic seizures
        ii. Clonic Seizures
        iii. Tonic Seizures
        iv. Versive Seizures
        v. Tonic-Clonic Seizures
        vi. Epiletic Spasms
    b. Complex Motor Seizures
        i. Automotor seizures
        ii. Hypermotor seizures
        iii. Gelastic Seizures
5. Special Seizures
    a. Atonic seizures

b. Akinetic seizures
c. Astatic seizures
d. Negative myoclonic seizures
e. Hypomotor seizures
f. Aphasic seizures

According to this model, the evolution of a seizure could be expressed by linking the seizure components/stages by arrows. For example:
Left visual aura⇨left hand clonic seizure (*AOC**) *⇨ generalized tonic-clonic seizure.*
*AOC = Alteration of consciousness

Tier 3: Etiology

Epilepsies are produced by multiple etiologies. Sometimes one of the causes could play a dominant role (for example, seizures caused by a tumor) while in other cases it's almost impossible to isolate a single etiological factor (for example, 'genetic' epilepsy, where multiple genes contribute to the generation of seizures). However, there are always multiple causes, even if there is a single dominant one.

1. Hippocampal sclerosis
2. Tumor
    a. Glioma
    b. Dysembrioplastic neuroepithelial tumor
    c. Ganglioglioma
    d. Other
3. Malformations of cortical development (MCD)
    a. Focal MCD
    b. Hemimegalencephaly
    c. MCD with epidermal nevi
    d. Heterotopic grey matter
    e. Hypothalamic hamartoma
    f. Hypomelanosis of Ito
    g. Other

4. Malformations of vascular development
    a. Cavernous angioma
    b. Arteriovenous malformation
    c. Sturge–Weber syndrome
    d. Other
5. Central nervous system infections
    a. Meningitis
    b. Encephalitis
    c. Abscess
    d. Other
6. Central nervous system inflammation
    a. Rasmussen encephalitis
    b. Vasculitis
    c. Other
7. Hypoxic-isquemic brain injury
    a. Focal ischemic infarction
    b. Diffuse hypoxic-ischemic injury
    c. Periventricular leukomalacia
    d. Hemorrhagic infarction
    e. Venous sinus thrombosis
    f. Other
8. Head trauma
    a. Head trauma with intracranial hemorrhage
    b. Penetrating head injury
    c. Closed head injury
    d. Other
9. Inheritable conditions
    a. Presumed genetic cause
    b. Tuberous sclerosis
    c. Progressive myoclonic epilepsy
    d. Metabolic syndrome
    e. Channelopathy

f. Mitochondrial disorder
g. Chromosomal aberration
h. Other

10. Structural brain abnormality of unknown cause
11. Other
12. Unknown etiology

Tier 4: Seizure frequency

Important for the management of the patient, which is focused mainly on reducing the frequency and intensity of epileptic seizures.

1. Daily seizures:
   One or more seizures per day
2. Persistent seizures:
   Less than one seizure per day but at least one seizure/6 months
3. Rare or no seizures:
   Fewer than one seizure every 6 months
4. Undefined:
   Seizure frequency cannot be specified because of unknown seizure frequency, recent onset of epilepsy or recent surgery of epilepsy

Tier 5: Related medical conditions

Provides information regarding non-etiological conditions the patients may be suffering. For example: Previous surgical procedures, depression, mental retardation, hemianopia, etc.

Below two examples of classification according to the 5-tiers system are presented:

Case (Patient X):

Tier 1 (Epilepsy Location): Right frontal

Tier 2 (Semiological Seizure Type): Left hand clonic seizures ⇨generalized tonic-clonic seizure

Tier 3 (Etiology): Right frontal malformation of cortical development

Tier 4 (Seizure frequency): Undefined

Tier 5 (Related medical conditions): Postsurgical left hand paralysis, right frontal lobectomy (January, 2008)

Case (Patient Y):

Tier 1 (Epilepsy Location): Generalized (Lennox-Gastaut syndrome)

Tier 2 (Semiological Seizure Type): Generalized astatic seizures, Dialeptic seizures, Generalized tonic seizures, Generalized tonic-clonic seizures

Tier 3 (Etiology): Diffuse hypoxic-ischemic injury

Tier 4 (Seizure frequency): Daily

Tier 5 (Related medical conditions): Severe mental retardation, left hemiparesis

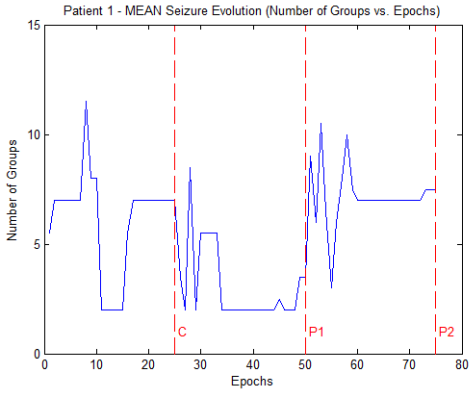
APPENDIX B

Average Results for the Population under study

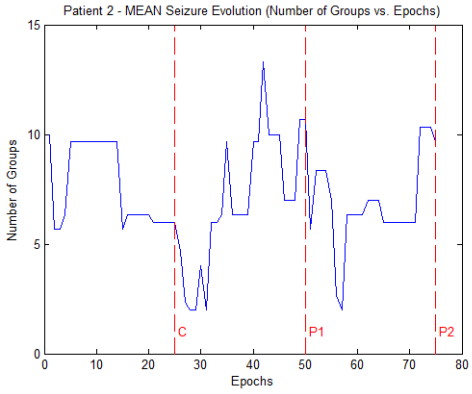
In this Appendix are shown the figures corresponding to the average results obtained for each one of the 14 patients analyzed. In particular, the appendix shows the evolution of the K values and the changes in the average number of observations (SEEG channels) per cluster and in the number of principal components explaining the 90% of the variance in the original data along the epochs.

- **Evolution of K along the epochs:**

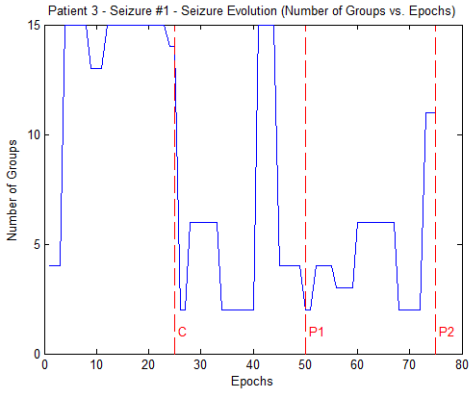
a)



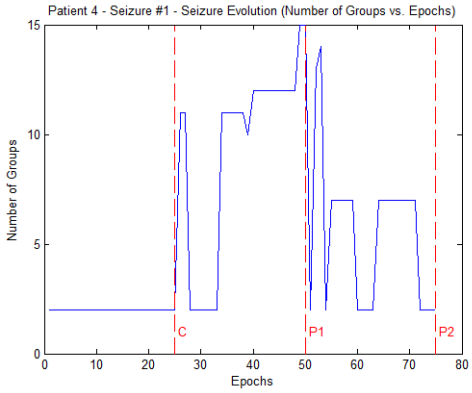
b)



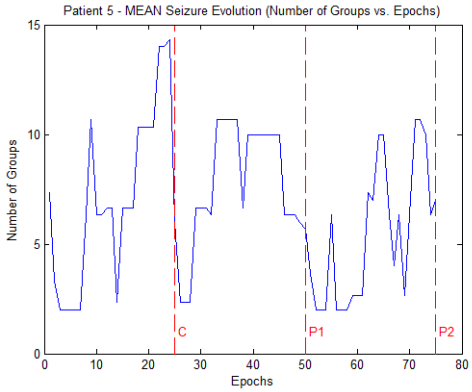
c)



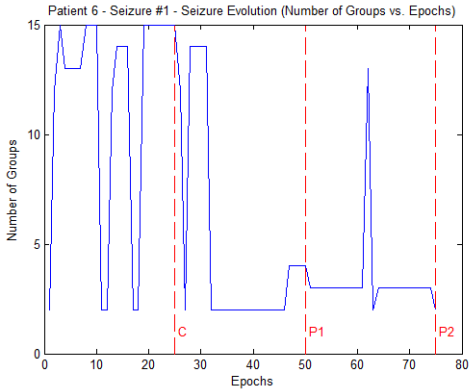
d)



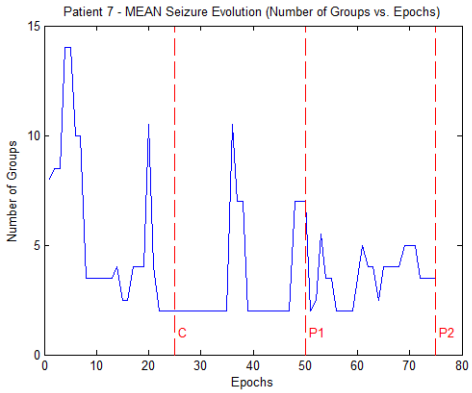
e)


Patient 5 - MEAN Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

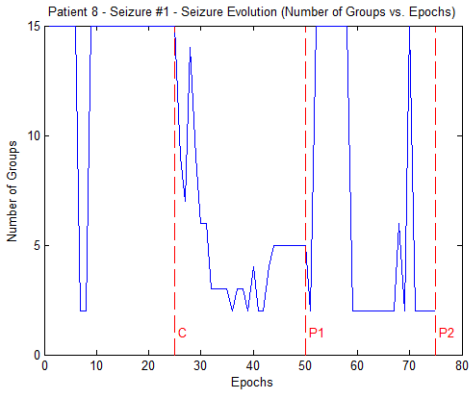
f)


Patient 6 - Seizure #1 - Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

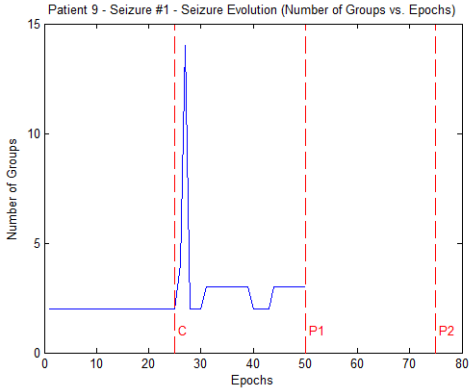
g)


Patient 7 - MEAN Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

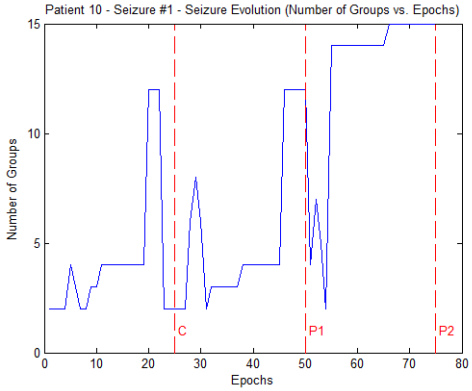
h)


Patient 8 - Seizure #1 - Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

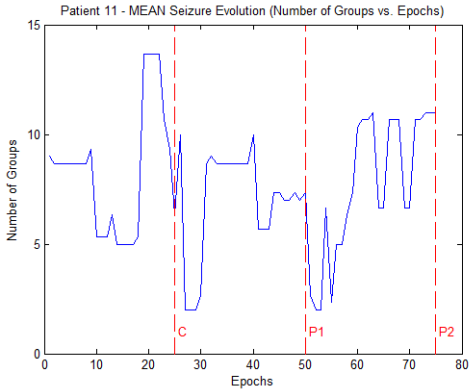
i)


Patient 9 - Seizure #1 - Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

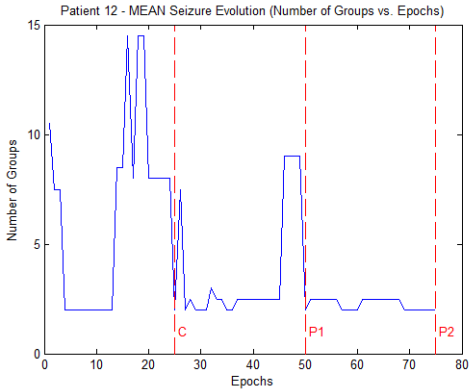
j)


Patient 10 - Seizure #1 - Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

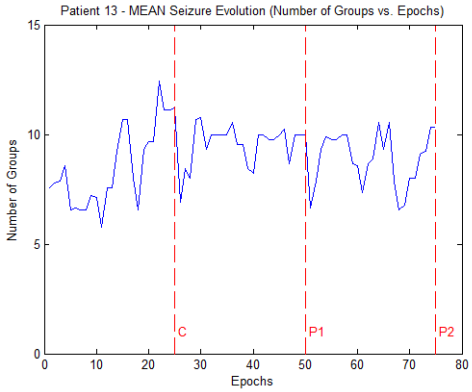
k)


Patient 11 - MEAN Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

l)


Patient 12 - MEAN Seizure Evolution (Number of Groups vs. Epochs)
Number of Groups
Epochs
C
P1
P2

m)



n)

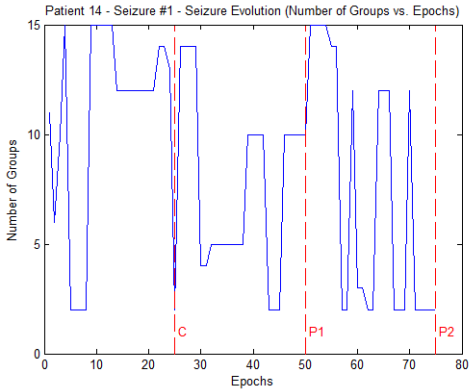
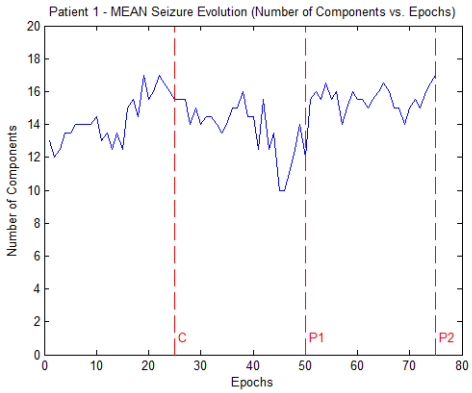


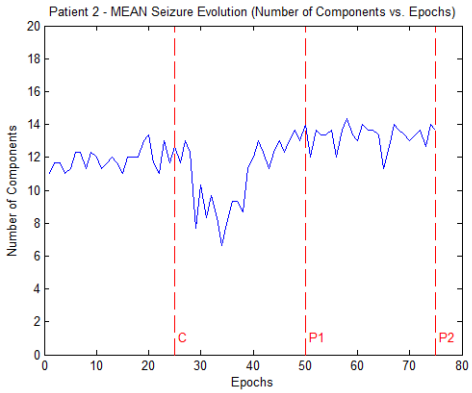
Figure 28. Patients' average evolution of the number of groups found by the algorithm along the epochs. (a-n) correspond to patients 1-14 respectively. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist (or the end of the "pre" time interval). While P1 and P2 constitute the end of the "seizure" and "post" intervals respectively.

- **Number of principal components needed to explain the 90% of the variance along the epochs:**

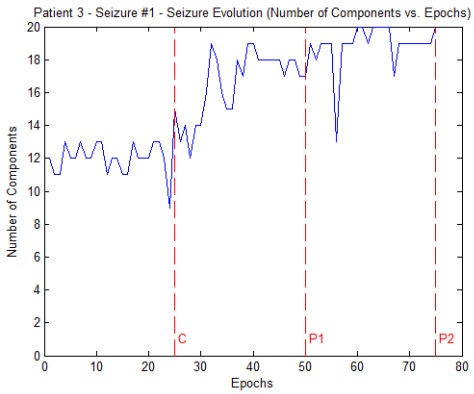
a)



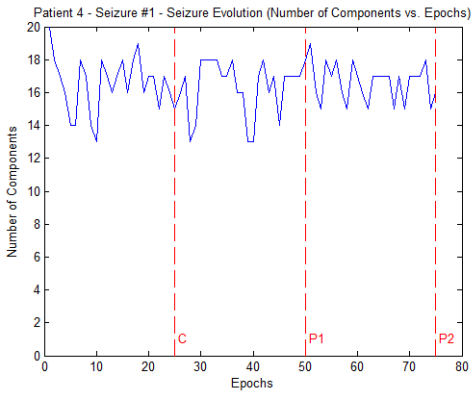
b)



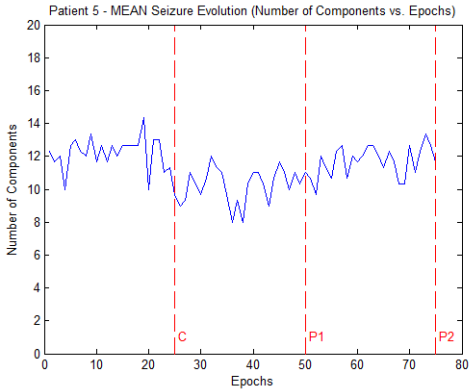
c)



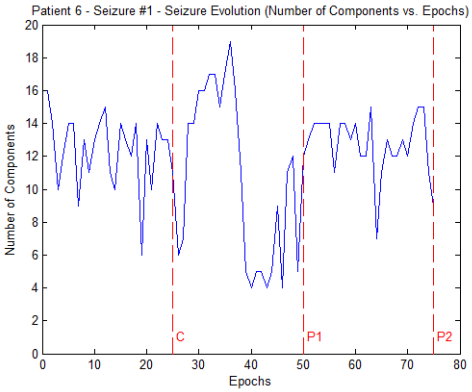
d)



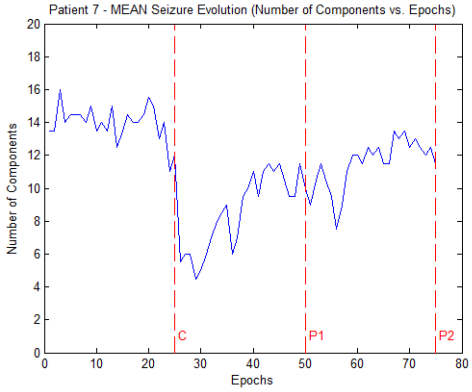
e)


Patient 5 - MEAN Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

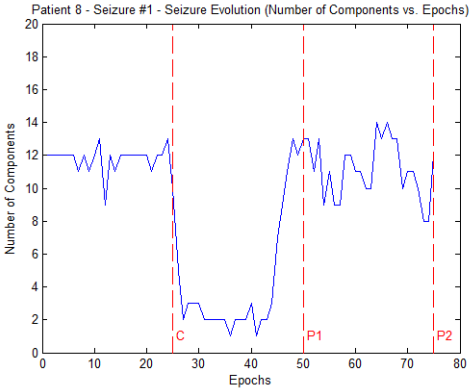
f)


Patient 6 - Seizure #1 - Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

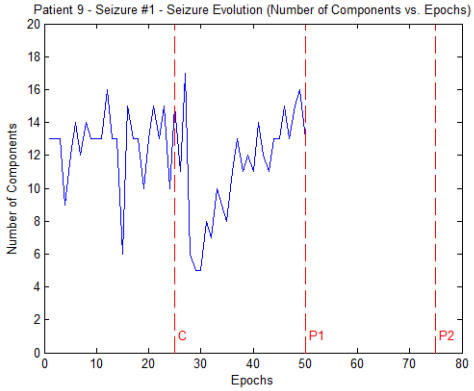
g)


Patient 7 - MEAN Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

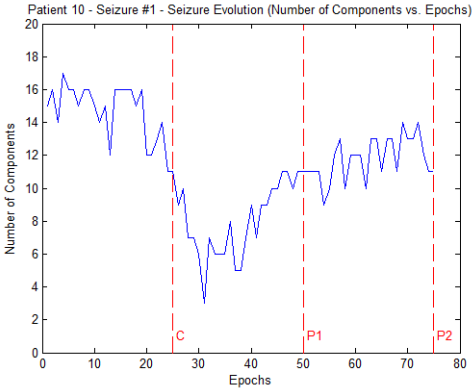
h)


Patient 8 - Seizure #1 - Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

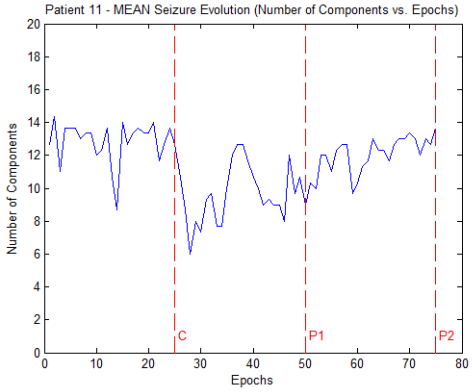
i)


Patient 9 - Seizure #1 - Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

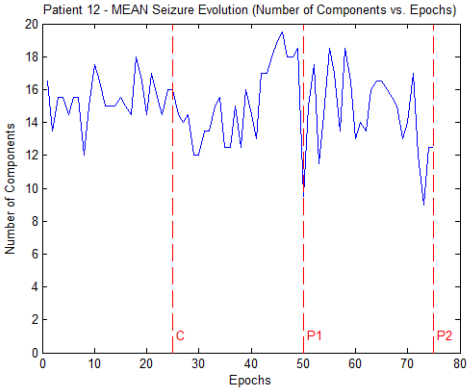
j)


Patient 10 - Seizure #1 - Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

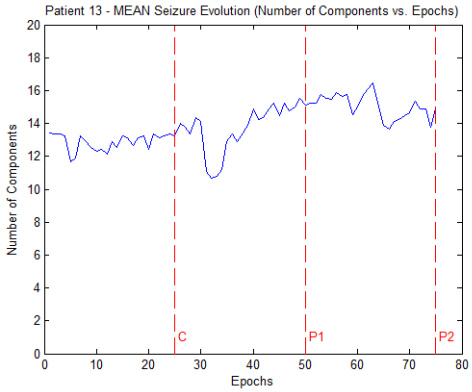
k)


Patient 11 - MEAN Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

l)


Patient 12 - MEAN Seizure Evolution (Number of Components vs. Epochs)
Number of Components
Epochs
C
P1
P2

m)



n)

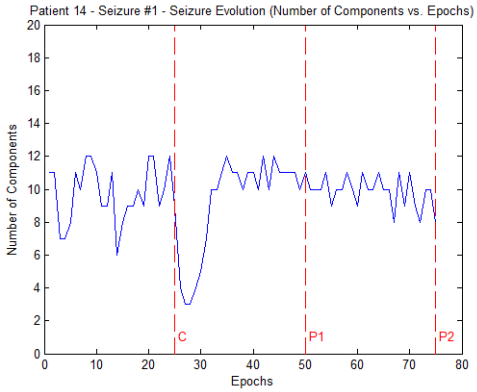
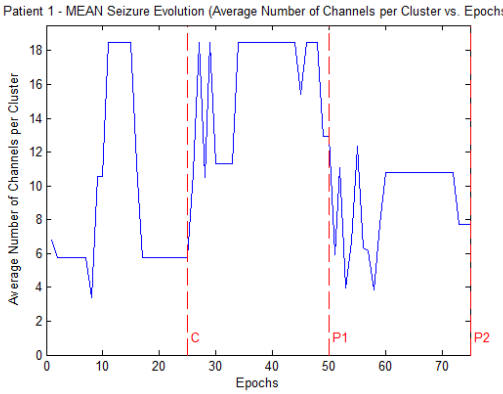


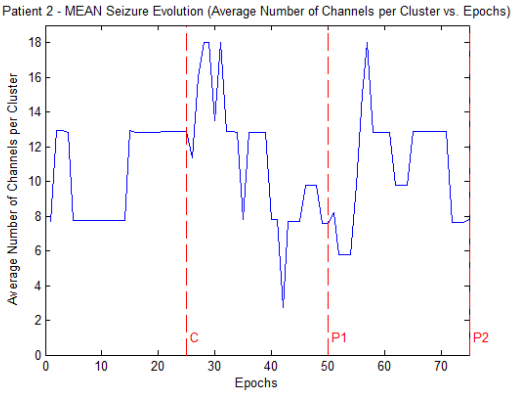
Figure 29. Patients' average evolution of the number of principal components needed to explain the 90% of the variance for each one of the different time epochs. (a-n) correspond to patients 1-14 respectively. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist (or the end of the "pre" time interval). While P1 and P2 constitute the end of the "seizure" and "post" intervals respectively.

- **Average number of channels per cluster along the epochs:**

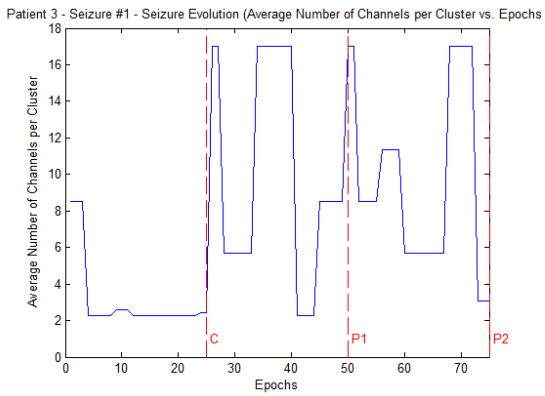
a)



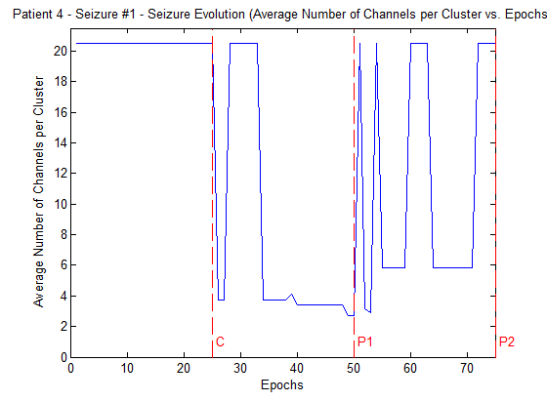
b)



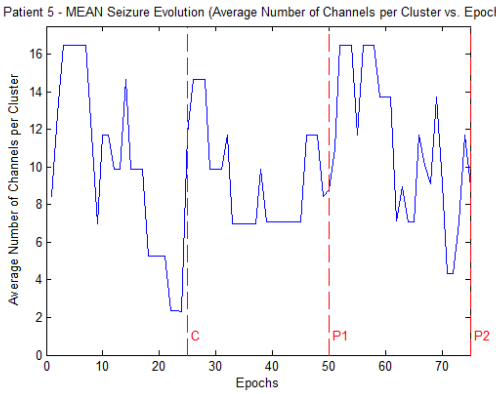
c)



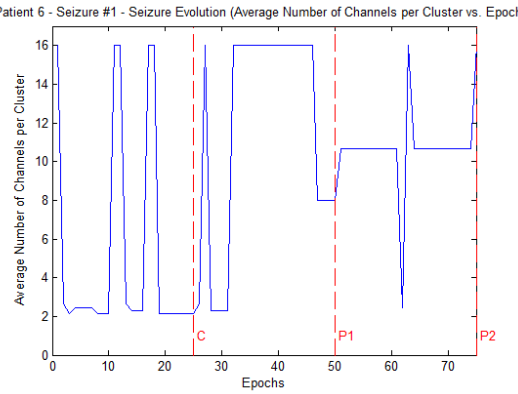
d)



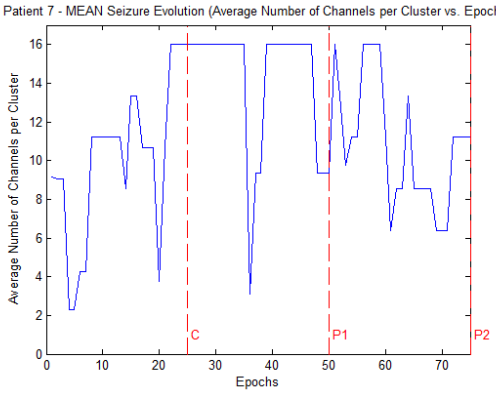
e)


Patient 5 - MEAN Seizure Evolution (Average Number of Channels per Cluster vs. Epochs)
Average Number of Channels per Cluster
Epochs
C
P1
P2

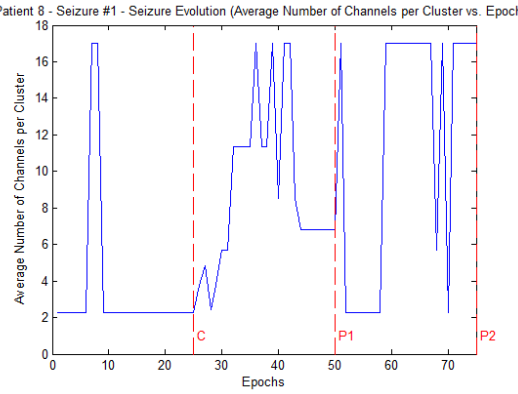
f)


Patient 6 - Seizure #1 - Seizure Evolution (Average Number of Channels per Cluster vs. Epochs
Average Number of Channels per Cluster
Epochs
C
P1
P2

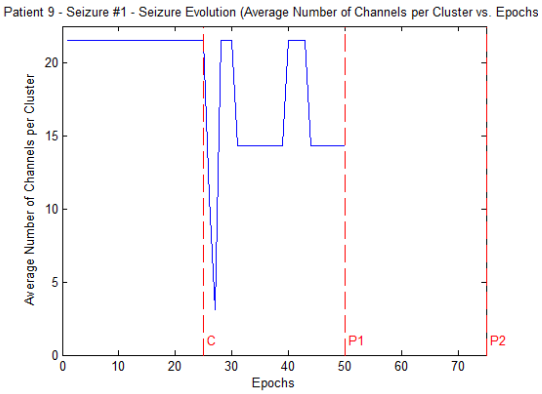
g)


Patient 7 - MEAN Seizure Evolution (Average Number of Channels per Cluster vs. Epochs)
Average Number of Channels per Cluster
Epochs
C
P1
P2

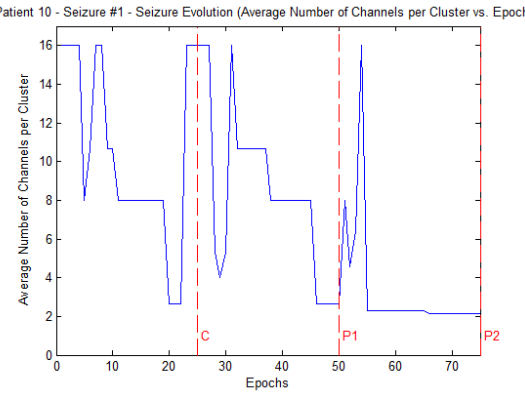
h)


Patient 8 - Seizure #1 - Seizure Evolution (Average Number of Channels per Cluster vs. Epochs
Average Number of Channels per Cluster
Epochs
C
P1
P2

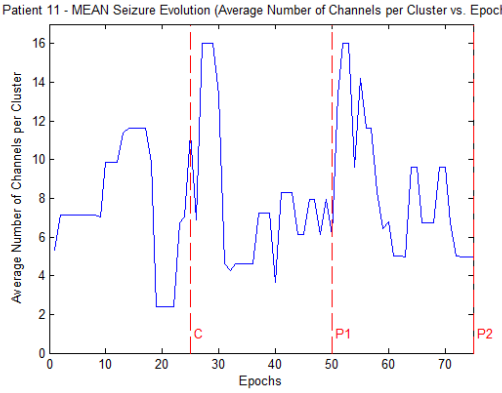
i)


Patient 9 - Seizure #1 - Seizure Evolution (Average Number of Channels per Cluster vs. Epochs
Average Number of Channels per Cluster
Epochs
C
P1
P2

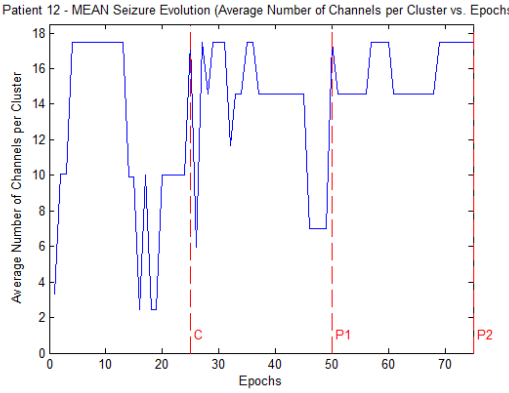
j)


Patient 10 - Seizure #1 - Seizure Evolution (Average Number of Channels per Cluster vs. Epochs
Average Number of Channels per Cluster
Epochs
C
P1
P2

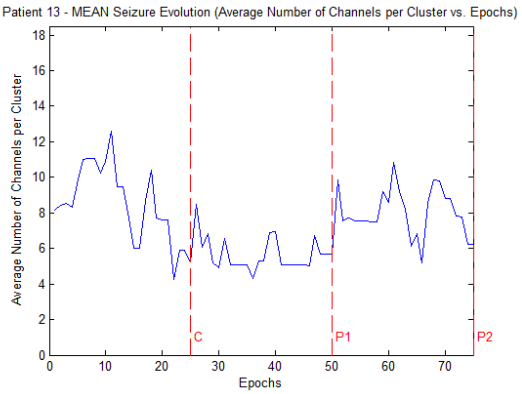
k)


Patient 11 - MEAN Seizure Evolution (Average Number of Channels per Cluster vs. Epochs)
Average Number of Channels per Cluster
Epochs
C
P1
P2

l)


Patient 12 - MEAN Seizure Evolution (Average Number of Channels per Cluster vs. Epochs)
Average Number of Channels per Cluster
Epochs
C
P1
P2

m)



n)

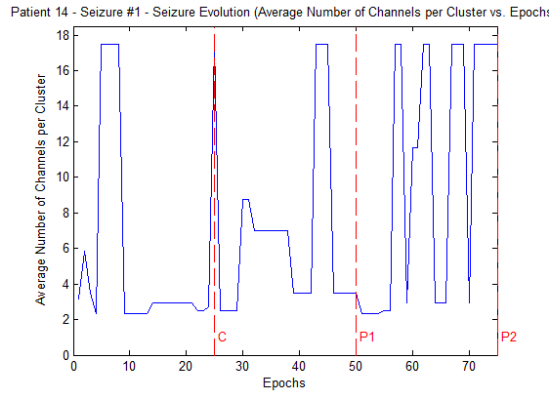


Figure 30. Patients' mean evolution of the average number of channels per cluster along the epochs. (a-n) correspond to patients 1-14 respectively. The dotted red lines constitute the end of each of the three time intervals under study. The red line named C represents the beginning of the seizure according to the information provided by the neurologist (or the end of the "pre" time interval). While P1 and P2 constitute the end of the "seizure" and "post" intervals respectively.