UCLouvain

École polytechnique de Louvain

Estimation of seizure onset zone from ictal EEG using independent component analysis and source imaging

Author: **Aurélie DE BORMAN**
Supervisors: **Pierre-Antoine ABSIL, Riëm EL TAHRY, Arvind KUMAR**
Readers: **Michel VERLEYSEN, Simone VESPA**
Academic year 2020–2021
Master [120] in Mathematical Engineering

Acknowledgements

First, I would like to thank Pierre-Antoine Absil and Riëm El Tahry for providing guidance and feedback throughout the year. It was an enriching experience to be able to work in an interdisciplinary framework and I am very grateful they proposed this fascinating topic.

I would also like to thank Arvind Kumar for being part of this project from abroad. The extra time he spent last year sharing his passion for neuroscience with interested students was a source of inspiration for me.

Special thanks also go to Simone Vespa for his insightful comments and recommendations. I am thankful for his enthusiasm and the time he spent supervising this work. His suggestions were of invaluable help.

Abstract

Epilepsy is the fourth most common neurological disorder. About one third of the patients are drug-resistant, meaning medication does not prevent seizures. In that case, surgery becomes an interesting option and it is then crucial to find from which brain region the seizures originate. In this master thesis, we take advantage of two advanced computational techniques, namely independent component analysis (ICA) and source imaging, to analyze electroencephalogram (EEG) recordings of seizures. While this combination of techniques is not new, we investigated its performance considering low-density EEG and a cohort of 11 patients suffering from extra-temporal lobe epilepsy, known to be more challenging than temporal lobe epilepsy. All patients underwent surgery and became seizure free, suggesting that the seizure onset zone lies within the resection zone which can thus be considered as the ground truth. A pipeline was implemented in order to estimate the seizure onset zone using both ICA and source imaging. While each step of the method required to make some choices, we attempted to justify as rigorously as possible any of the decisions. Quantitative measurements allowed us to objectively analyze the pipeline performance thanks to the co-registration of the method output and the resection zone. Whereas the pipeline is not fully automated, it was designed to be practical and help as much as possible a clinician who would use it.

Contents

Chapter 1

Introduction and outline

Epilepsy is the fourth most common neurological disorder and affects around 50 millions people worldwide [1],[2]. Patients suffering from epilepsy are subject to frequent seizures induced by an abnormal electrical activity in the brain. About one third of the patients are drug-resistant, meaning medication does not prevent the seizures [3]. In that case, surgery becomes an interesting option to treat epilepsy. However, not all patients who suffer from drug resistant epilepsy are eligible for surgery. Three main criteria must be fulfilled to enter a presurgical evaluation [4] : (i) the patient (or his/her legal guardian) needs to be aware of the presurgical evaluation objective and agree with the possibility of surgery (ii) the patient should suffer from disabling seizures that can not be prevented by medication (iii) imaging and electroclinical data should be consistent with the possibility of a successful surgery outcome.

The presurgical evaluation aims at localizing the brain region(s) without which seizures would be avoided, called the epileptogenic zone [5]. The estimation of this zone is critical to the outcome of a surgical epilepsy treatment. The presurgical evaluation may rely on many different procedures and so far no single method is sufficient to estimate the epileptogenic zone with an acceptable level of confidence. Among these procedures, the analysis of electroencephalographic recordings (EEG) and more specifically of ictal EEG (i.e., EEG during a seizure) is often considered as an essential step of the presurgical evaluation [6]. While this master thesis focuses on the extraction of relevant information from ictal EEG, it is worth mentioning some other procedures to illustrate the variety of possible techniques. For instance, the study of interictal EEG (i.e., EEG between seizures) and more specifically of epileptiform discharges that occur some other time than during the seizures sometimes allow to gain knowledge about the location of the epileptogenic zone [7]. A brain magnetic resonance imaging (MRI) can also provide valuable information, especially if it reveals lesions or structural abnormalities. In some cases, it is also of interest to perform an invasive EEG recording. Other methods such as positron-emission tomography (PET), ictal single photon-emission computed tomography (SPECT) and magnetoencephalography (MEG) are also used with the aim of estimating the epileptogenic zone [5]. Recent studies based on EEG also take advantage of functional connectivity analysis to increase the chances to find the epileptogenic zone [8], [9].

For the purpose of EEG analysis, Independent Component Analysis (ICA) is a computational technique that was shown to be a helpful tool in many different contexts [10],[11]. Generally speaking, the ultimate goal of ICA is to recover the independent sources of a linear mixture of signals. One common example to give an intuitive interpretation of the method is the following situation: two microphones placed at different positions in a room capture someone's speech with a song playing in the background. Both microphones catch a mixture of the two independent sound signals: the speech and the song. Under some assumptions, ICA is able to unmix the two signals and recover both separated signals. In the case of EEG, the sound signals correspond to electrical activities while the microphones correspond to the electrodes. In the same way ICA can separate the speech and the song signals, the hope is that it will isolate the seizure activity from the other signals that are captured by the electrodes during a seizure, including artefacts and other non-pathological brain activities. Because ICA returns as many indepedent signals as there are electrodes, a challenging task is to identify which of them are related to the seizure.

Once seizure activity has been isolated by ICA, the next and ultimate goal is to obtain the origins of the so-called ictal components, i.e., independent signals containing mostly seizure activity. Several options are possible to localize the origin of the independent signals obtained by ICA. First of all, scalp maps can be produced for each of the independent components straight from the parameters obtained by ICA. Indeed, ICA not only gives us the estimated independent signals but also a weight matrix that describes the preponderance of each recorded signal in each independent signal. Going back to the example, if the first microphone is closer to the speaker than the second one then this information will probably be noticeable in the weight matrix. Indeed, the coefficients of the matrix will reflect the fact that the first microphone recording has a larger contribution to the reconstructed speech signal. In the case of EEG, the scalp maps illustrate the distribution of one independent source across the different electrodes, meaning how much of the source signal was captured by each electrode. Scalp maps only allow us to position an independent signal with respect to the recording sites, thus on the scalp. In many applications this is not sufficient and that is particularly the case when the goal is to localize the onset zone of seizures.

To overcome the limitations of scalp maps, source imaging becomes an interesting tool in order to localize the origins of independent components. In general terms, source imaging is an emerging technique that aims to recover the brain sources from signals recorded at the scalp [12]. The problem of finding the activation of the brain sources from the scalp potential given a head model is called the inverse problem. This problem is ill-posed as many possible solutions correspond to the same potential field on the scalp. Thus, many techniques of source imaging exist relying on different assumptions. Nevertheless, they all have as purpose to find the anatomical regions responsible for what is observed on the scalp. In order to localize the source of the independent components, we must first back-project the components we are interested in at the level of the electrodes before applying a source imaging technique. This is easily done by using the weight matrix. Doing so, we obtain reconstructed EEG signals that hopefully include only seizure activity. A source

imaging technique can then be applied to obtain an estimated location of the onset zone inside the brain.

To assess the performance of the method, the estimated seizure onset zones are compared to the resection zone of patients who underwent surgery. If the patient became seizure-free, the brain region that was removed or disconnected is considered as the ground truth. The distance between the estimated location and the resection zone allows to assess the reliability and precision of the method. This measure is also useful when it comes to analyzing the ICA decomposition. Indeed, the independent components that lead to a correct location are more likely to isolate seizure activity than others. Thus, this fact can be used in order to help the step where the components that isolate seizure activity must be chosen. Different component features are examined with the purpose of automating the selection of components related to the seizure since this selection is crucial for the success of the method.

This master thesis is divided into seven chapters, including this introduction and a conclusion. We start with Chapter 2 that presents the state of the art and the novelty proposed by this project. Chapter 3 then briefly outlines the material that was used, including both data and software. Next, Chapter 4 introduces the pipeline summarizing the proposed method. Each step is explained in details while some decisions are left to be made afterwards. In Chapter 5, we present the quantitative measures implemented to assess the localization performance before using it to analyze the pipeline and to implement a component selection process. At this stage, final choices are made according to the conducted analysis. Finally, Chapter 6 outlines the practical value of this work giving some concrete examples of seizures where the method was applied.

Chapter 2

State of the art and novelty

During the last two decades, multiple studies have proposed the use of ICA to improve presurgical evaluation in patients with intractable epilepsy. One of the very first studies was conducted in 2002 by Nam et al. who proposed to apply ICA to the seizures of patients with medial temporal lobe epilepsy [13]. The lateralizing power of ictal EEG was proved to be increased thanks to the inspection of the scalp maps of components with ictal nature. In the first place, ICA was mainly used to remove artifacts from ictal EEGs [14],[15] and its localization ability was limited to the scalp maps [13],[16].

The idea to locate more precisely the derived components quickly arose with the emergence of source imaging techniques. Some studies proposed to use methods that compute a single dipole location that best explains a component activity [17], [18]. Other studies suggested the usage of distributed methods to either localize the back-projected activity of a single independent component [19] or multiple components at once [20], [21].

Choosing the independent components with ictal nature is a challenging step. Most of the studies rely on visual inspection of the component signals. If the source imaging consists in a dipole fitting technique, the fact that a dipole is located outside the head allows to eliminate components. In addition, a high residual variance associated to a dipole (meaning the component activity is unlikely to originate from a single dipole) permits to get rid of some more components. To support the visual inspection, Nam et al. reviewed the autocorrelograms and also computed the portion of power in the 2- to 10-Hz bandwidth in order to help the component selection [13]. Other studies also considered event-related spectrum pertubation that allows a visual time-frequency analysis of the components [17], [18].

In the range of studies using distributed methods, Yang et al. [20] proposed a method to select the ictal components where they compute the correlations between time-frequency representations of each independent component and of the electrodes showing the most ictal activity in a reconstructed EEG. The method required some visual inspection to build the noise-deducted EEG (by identifying artefact components) and then to identify the electrodes showing ictal activity in the reconstructed EEG. High-density recordings (76 electrodes) were made in a group of eight patients, six of them suffering from temporal

lobe epilepsy. The method gave an accurate location in 17 of the 19 seizures included in the study. Lu et al. [21] applied the same methodology to nine pediatric epilepsy patients suffering from extra-temporal lobe epilepsy using 32-channel EEG and they correctly localized the onset zone for seven patients. However, five patients were not rendered seizure-free after surgery. Besides, it seems that only one seizure per patient was considered meaning the method was applied to only 9 seizures.

More recently, Habib et al. [22] proposed a recursive ICA-decomposition that requires a clinician's intervention only to specify the frequency of the ictal rhythm. The technique works by repetitively applying ICA, eliminating a component and back-projecting the remaining components until a sufficiently satisfying component is obtained. The method was applied to 21-channel ictal EEG from eight patients suffering from temporal lobe epilepsy. In 17 of the 20 seizures included in the study, the estimated onset zone was concordant with the resection zone.

This master thesis extends the work of these studies and distinguishes itself for several reasons. First, all the patients included in this study suffer from extra-temporal lobe epilepsy. This type of epilepsy is known to be more challenging than temporal lobe epilepsy which often shows a typical sustained rhythmical activity. Most of the mentioned studies considered a majority of temporal lobe epilepsy patients. The patients included in our study suffer from a variety of extra-temporal lobe epilepsy that shows rhythmical activities sometimes lasting for only a few seconds. Moreover, we considered 39 seizures in total which is beyond the number of seizures considered by the cited studies that combine ICA and source imaging.

Second, this study aims at minimizing the technicalities that often discourage the clinical use of both ICA and source imaging. Once the implemented pipeline is installed, its use does not require any technical expertise in the field of signal processing or so. The user must simply give the EEG file with an annotation at the position of the seizure onset time. To keep it simple, we also decided to use a template head model for the source imaging instead of patient specific head models. This reduces the amount of time a clinician would spend on the procedure.

Additionally, this study was conducted with 19-channel EEG recordings. While it is evident that high-density EEG recordings have a better spatial resolution, it remains of interest to show the potential of low-density EEG. In fact, it might not be comfortable for a patient to wear a high-density EEG cap during a long time period (typically several days for epilepsy monitoring). Besides, most epilepsy centers have 19-channel EEG devices while high-density EEG requires more costly devices. Finally, we should also note that more electrodes implies larger datasets, causing longer processing times.

More importantly, the analysis of the pipeline avoids as much as possible any bias that could be caused by visual inspection. To do so, we rigorously implemented a quantitative performance measure, that is the distance between the resection zone and the estimated location. Consequently, it was possible to determine objectively if some components ob-

tained by ICA were localized inside the resection zone and were thus likely to be ictal. This allowed us to compare objectively how to preprocess the data to obtain the most favorable outcome. We compared different type of filters and the length of the time window since these choices differ from one study to another. We also proposed a component selection to help a clinician in his/her choice for ictal components.

Finally, we should note that the method was implemented from zero, gathering the best from different open-source codes. The method relies on a structured pipeline that could easily be modified for future research. Each step of the pipeline could potentially be improved separately allowing a researcher to focus on one aspect of the procedure without the technical burden of other steps.

Chapter 3

Material

3.1 Data

This study was conducted using anonymized retrospective data and was approved by the Institutional Review Board of Saint Luc University Hospital, Brussels, Belgium. The EEG signals were recorded at Saint-Luc University Hospital and at Centre Neurologique William Lennox, Ottignies, Belgium. Between 19 and 23 electrodes were available for each patient but only the 19 electrodes placed according to the International 10-20 system were used for the sake of homogeneity. The sampling rate was 256 Hz.

To be included in this study, patients must satisfy the following inclusion criteria:

1. The patient suffers from drug-resistant extratemporal focal epilepsy.
2. A curative epilepsy surgery was performed.
3. Long-term scalp EEG monitoring and post-operative structural MRI are available.
4. The post-operative follow-up lasted at least one year.
5. The patient became seizure-free after surgery (ILAE class 1 [23]).

According to these criteria, 16 patients were selected from the database of the Centre for Refractory Epilepsy of Saint Luc University Hospital. Because this study relies on the rhythmic nature of seizures, only seizures labelled as "rhythmic" by neurologists were included for the analysis. This left us with 11 patients and a total of 39 seizures. Neurologists identified for each seizure a frequency band characterizing the epileptic activity. Details about the clinical data can be found in Table 3.1.

Pt n°	# seizures	Epilepsy localization	Type of surgery	Frequency band of interest
1	4	Left frontal temporal insular lobe	Lesionectomy	theta
2	3	Right frontal lobe	Disconnection	theta-alpha
3	5	Right temporo-occipital lobe	Disconnection	alpha
4	2	Plurifocal left hemisphere	Hemispherectomy	delta (1), theta (1)
5	5	Plurifocal left hemisphere	Hemispherectomy	theta
6	5	Plurifocal right hemisphere	Hemispherectomy	delta (1), theta (4)
7	4	Right frontal lobe	Disconnection	delta-theta (1), theta-alpha (3)
8	2	Left temporo-occipital lobe	Disconnection	theta
9	1	Plurifocal right hemisphere	Hemispherectomy	beta
10	4	Right parieto-occipital lobe	Disconnection	delta (1), theta (2), delta-theta
11	4	Right frontal lobe	Lesionectomy	alpha

Table 3.1: Clinical data of the patients

3.2 Software

All codes are written in the programming language Matlab. The EEGLAB toolbox [24] provides many useful functions that have been used for this project. EEGLAB is an open source toolbox widely used for processing EEG, magnetoencephalography (MEG) and other electrophysiological data. An interactive graphic user interface (GUI) allows users to visualize and manipulate data in a simple manner that does not require advanced programming skills. EEGLAB is a continuously evolving toolbox as extensions can be downloaded from an open source platform through which developers can propose 'Plug-in' functions that will appear in the EEGLAB menu of users who downloaded them.

The EEGLAB toolbox allows to create Matlab structures from EDF files which is a standard file format for medical time series. The built structure contains a large amount of information about the EEG signal such as the annotations, the sampling rate or the electrode names. The position of the electrodes can be retrieved from the electrode names and a standardized boundary element head model. A function also allows to cut a signal using an event marker which turned out to be useful to delimit the analyzed time window in our case. Finally, we should note that the toolbox comes with a bunch of functions dedicated to the application of ICA. Indeed, several ICA algorithms are available and ready to be applied. The visualization of the components and their properties are also very easily done.

In order to implement the performance measure of the pipeline, two additional software packages were helpful. First, the SPM software package [25] served for the normalization of the MRIs. Second, the MRIcron image viewer [26] allowed to draw the resection zones.

Chapter 4

Methodology

Having in mind the purpose of this master thesis, a general pipeline has been established to summarize the subsequent steps of the method to be implemented. This allowed a global vision of the work to be done and of the potential progress to be made. In this chapter, we first present the general pipeline and then describe how each step was implemented. Some steps required to make some choices and only the different options to be studied are presented in this chapter. The analysis that led to the final choices will be explained in the next chapter.

4.1 Overview of the pipeline

The goal of this master thesis is to localize the onset zone of seizures taking advantage of both ICA and source imaging. From this general idea, a pipeline can be created summarizing the steps to be implemented. This pipeline should therefore have as input the EEG signals including the seizure and as output an estimated position of the seizure onset zone. In this section, we shortly describe each of the pipeline steps but details of the implementation will be found in the next sections. Figure 4.1 shows a graphical version of the pipeline.

STEP 1 : Data acquisition and preprocessing

The very first step consists in opening the EEG data in Matlab. Once it has been done, digital filters are applied at this stage of the pipeline. Bad electrodes are identified and eventually removed. The data is also re-referenced against a common average reference allowing to visualize one signal per electrode. At the end of this step, the signals are loaded in a matrix $\mathbf{\Phi}_f \in \mathbb{R}^{N \times Q}$ where N is the number of electrodes and Q is the number of time points.

STEP 2 : Time window selection

The second step requires to select the time window that will be used for the subsequent analysis. The matrix $\mathbf{\Phi}_f$ is reduced to the matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times T}$ containing only the time interval of interest around the seizure onset. Finding an appropriate time window neither too long nor too short is crucial to the quality of the ICA decomposition.

STEP 3 : ICA decomposition

The third step is the ICA decomposition whose output consists in the independent components to be analyzed. The ICA algorithm computes the weight matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ such that $\mathbf{X} = \mathbf{W}\mathbf{\Phi}$ where the matrix $\mathbf{X} \in \mathbb{R}^{N \times T}$ carries the N independent components.

STEP 4 : Component selection

One of the most challenging steps is the component selection. From the N independent components obtained at the previous step, the hope is that only few of them contain seizure activity. The N_s independent components $\mathbf{X}_s \in \mathbb{R}^{N_s \times T}$ that isolate ictal activity are kept for further analysis.

STEP 5 : Scalp maps

Using the corresponding lines of the weight matrix $\mathbf{W}$, the activities of ictal components are projected to the scalp map. This provides information about how the activity of one component is distributed among the electrodes.

STEP 6 : Source imaging

The final step applies a source imaging technique to obtain an estimated localization inside the brain. Using appropriate columns of the weight matrix inverse, the ictal components can be back-projected to the scalp by computing $\mathbf{\Phi}_s = \mathbf{W}_s^{-1}\mathbf{X}_s$. The activity $\mathbf{\Phi}_s$ is then projected at the brain level using a source imaging method that maps electrodes signal to a distributed source space.

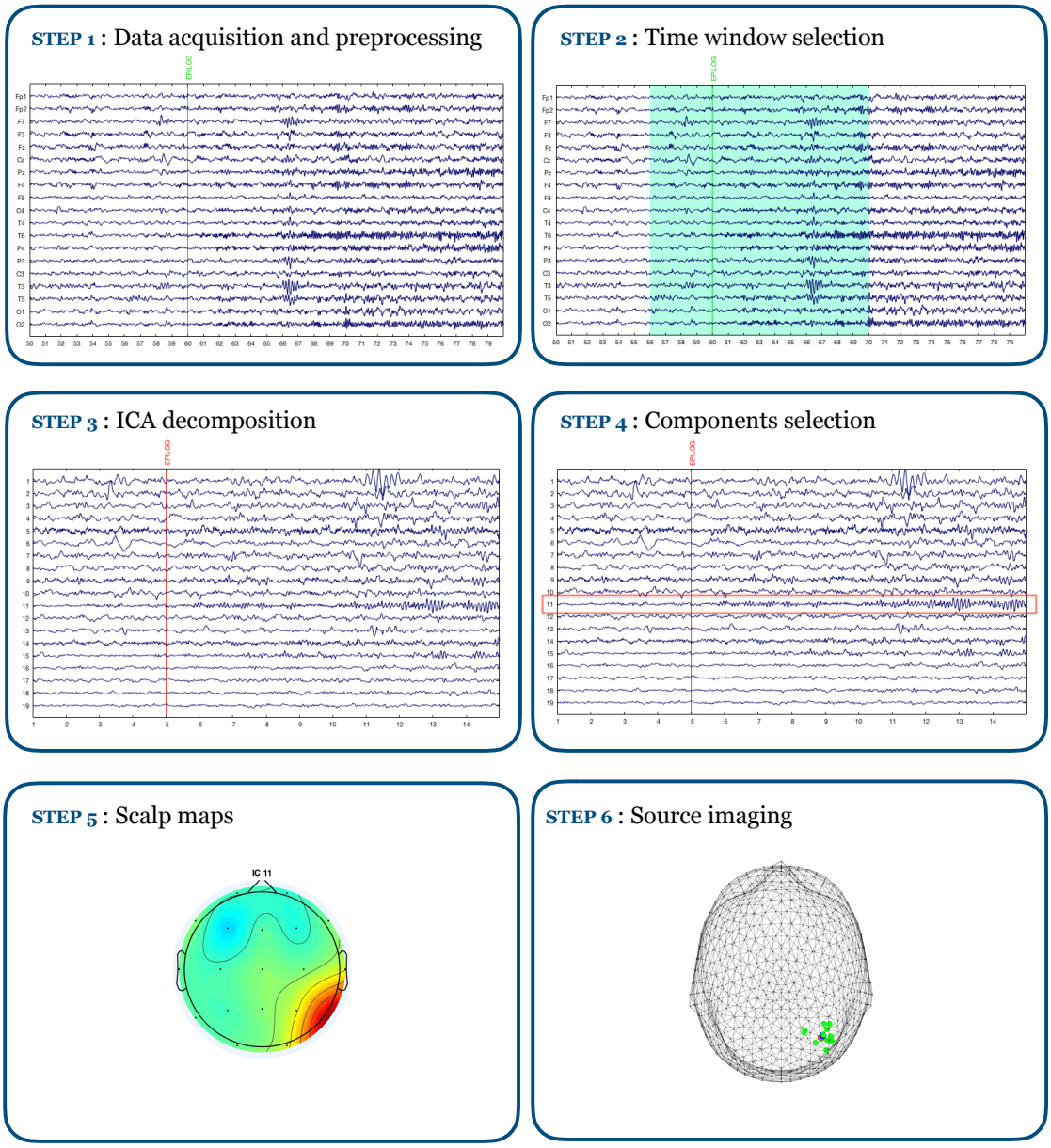


Figure 4.1: Graphical version of the pipeline

4.2 Data acquisition and preprocessing

4.2.1 Data opening in Matlab

The very first step of the pipeline consists in uploading the data. EEG signals were already recorded and thus it was only needed to open the datasets inside Matlab. The files were saved in a .edf file format that can easily be opened thanks to EEGLAB. Each individual file holds one hour of recording with annotations written by a neurologist. Opening it in EEGLAB allows to create a structure containing all the relevant information : annotations, electrode names, sampling rate, etc.

The seizures were marked by experienced neurologists and named according to a specific nomenclature : a letter, a first number and a second number. The letter could be either X or S and was decided from a previous study. The first number indicates the seizure type. This is useful when a single patient experienced several types of seizures. However, a type 1 only means that it is the first type of seizures that the patient experienced but does not indicate a similitude between type 1 seizures across the patients. Then the second number simply identifies the seizure within the type specified by the first number. For instance, seizure X2.03 refers to the third seizure of the second seizure type that the patient had.

As mentioned in section 3.2, the electrode positions could be retrieved thanks to their names and an EEGLAB function. This is the case because the electrodes were placed according to the International 10-20 system. Figure 4.2 illustrates the positions and names of the electrodes. Electrodes A1 and A2 are called the mastoid electrodes and were not used in our case.

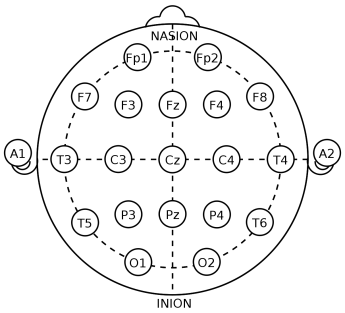


Figure 4.2: International 10-20 system

4.2.2 Filters

At this stage of the pipeline, filters were applied in order to eliminate irrelevant information. To avoid edge artifacts, filtering was done before cutting the portion of signal to be given to ICA. Two types of filters need to be applied : a band-pass filter and a notch filter. Both filters were designed using fourth-order Butterworth filter. Practically, second-order filters were designed using the function `butter` from Matlab. Next, the function `filtfilt` was used to apply the filters. Doing so, the input data is processed in both the forward and reverse directions. This avoids distorting the phase and doubles the order of the initially designed filters.

A first question arises at this point of the pipeline regarding the cutoff frequencies of the band-pass filter to be applied. Getting rid of very low frequencies allows to eliminate

some gross artefacts such as movements or so. Two different lower cutoff frequencies were studied for the band-pass filter: 1 Hz and 2 Hz. Previous studies applying ICA to ictal EEG and cited in Chapter 2 used high-pass filters with a cutoff frequency varying from 0.3 to 1 Hz. While filtering frequencies below 2 Hz seems a bit aggressive, this is motivated by the fact that ICA could be perturbed by very low frequencies. With the purpose of eliminating artefacts using ICA, the influence of high-pass filtering was studied by Winkler et al [27] and they showed that high-pass filtering between 1 and 2 Hz consistently achieved good results in terms of signal-to-noise ratio and percentage of 'near-dipolar' ICA components. While the purpose of this master thesis is different from that of the study, high-pass filtering at 1 and 2 Hz seem rational choices to compare. However, care should be taken for seizure having a dominant activity in the delta band (i.e., 1-4 Hz).

The upper cutoff frequency of the band-pass filter was set to 30 Hz. Regarding the frequency bands of interest marked by the neurologists in this study, this makes sense since no seizure had a dominant rhythm in the gamma band (i.e., above 30 Hz). Hence, there should not be any loss of seizure activity information. Also, this allows to remove some muscles artefacts that are typically characterized by high frequency activity.

The lower and upper cutoff frequencies of the notch filter are 49 and 51 Hz, respectively. That filter eliminates the 50 Hz line noise present in many recordings. One may argue that activity in that band is already eliminated by the band-pass filter. However, because filters are not ideal, a peak is still often visible around 50 Hz in the spectrum of the signals even after band-pass filtering.

4.2.3 Bad electrodes rejection

Another important point to consider at this stage is how to recognize and handle corrupted electrodes. It happened in a few recordings that an electrode had a poor contact with the scalp. Consequently, the electrode severely caught noise from the environment due to a too high impedance. An example is shown in Figure 4.3 where we clearly observe that the electrode Pz behaves abnormally. This figure shows the EEG signal after having applied filtering (2 Hz high-pass) and re-referencing. Figure 4.4 presents the signal before any preprocessing steps were applied. We observe that the electrode Pz catches critically the line noise which is typical for an electrode with bad contact.

A measure that is of interest to detect bad electrodes is the kurtosis [28]. This quantity measures how heavily the tails of a probability distribution differ from those of a normal distribution. The value gets typically high when close-to-discontinuities arise. EEGLAB has a function to detect electrodes having their kurtosis above a certain threshold. For the example shown previously, the electrode with a bad contact has a very high kurtosis value compared to the other electrodes as can be seen in Table 4.1. The kurtosis measurement was used for information but is not 100% reliable. Hence, visual inspection was used to determine electrodes that clearly showed a very low signal-to-noise ratio.

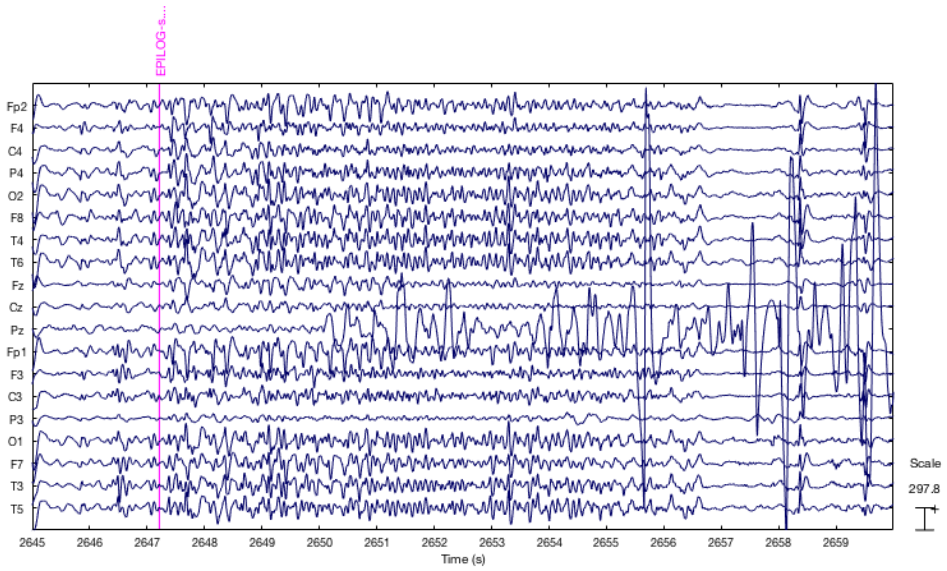


Figure 4.3: EEG channel signals after filtering and re-referencing (patient 7 seizure S1.02)

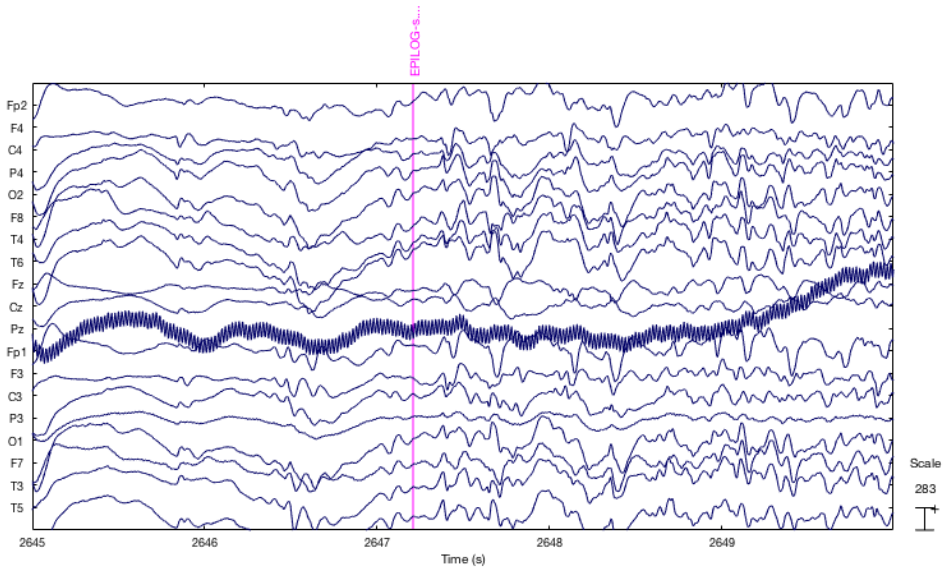


Figure 4.4: Raw EEG channel signals (patient 7 seizure S1.02)

Electrode	Fp2	F4	C4	P4	O2	F8	T4	T6	Fz	Cz	Pz	Fp1	F3	C3	P3	O1	F7	T3	T5
Kurtosis	-0.24	3.40	-1.04	-0.36	1.47	-1.08	-2.01	-0.16	0.77	-0.91	56.71	1.41	-0.56	-0.41	-0.87	1.83	-0.31	-0.82	1.26

Table 4.1: Kurtosis for the EEG signal shown in Figure 4.3

Luckily, the rejection of bad electrodes is not considered as a heavy task for low-density EEG and it usually does not take much time for a clinician to identify bad electrodes when inspecting an EEG. If a doubt persists, it might be reasonable to keep the electrode in

question and then hope that ICA will correctly isolate the artefact. Work could be done in the future to improve that part of the pipeline and incorporate an automatic detection of bad electrodes.

4.2.4 Common average reference

The last preprocessing step consists in re-referencing the data against a common average reference. EEG signals are voltage measurements meaning one signal is the difference in electric potential between an electrode and a reference. This reference is typically an electrode placed on the head with the hope that the electric potential at that position is not influenced by brain activity but captures the same noise as the electrodes of interest. Unfortunately, there is no universally accepted reference position because an ideal reference does not exist [29].

Common average reference is an often performed offline method to modify the reference. In that case, the reference becomes the average of the electrode signals. Mathematically, for a reference montage, it simply implies to remove at each electrode the average of all the electrodes. The theory behind it relies on the fact that the surface integral of the electric potential over a closed volume conductor that contains all the current sources is zero [30]. In fact, the sum of all the electrodes signals after common average re-referencing is equal to zero at any time point. However, the electrodes do not completely and uniformly cover the head so it remains a theoretical assumption. Common average reference has the advantage to be less biased towards any specific location on the head.

Applying common average reference by simply removing the average of all electrodes could cause a problem when ICA is subsequently applied. Indeed, doing so the data matrix (where each row is an electrode signal) becomes rank-deficient as the sum of the rows is zero. This situation is not optimal for ICA because the algorithm supposes independent sources and a full-rank mixing matrix (otherwise the inverse of the mixing matrix that ICA tries to compute would not exist). To overcome that issue, a trick consists in first adding the reference channel (equal to zero all the time), then apply the common average reference and finally discard the added channel [31]. Seen differently, this implies that, if we denote by N the number of electrodes, we divide the sum of the electrodes signals by $N + 1$ instead of N when computing the average.

4.3 Time window selection

Before applying ICA, we must choose the time window that will be given to the algorithm. As it was mentioned before, markers were placed by experienced neurologists to indicate the seizures onsets. Therefore, it remained to choose how many seconds before and after the marker would be given to ICA. Improvements could be done to automate the seizure detection using for instance some advanced machine learning techniques (see a review of

seizure detection using machine learning classifiers [32]). However, EEG recordings are usually annotated by neurologists for other types of analysis and it was therefore not the priority of this study.

Taking into account the markers, what remains to be investigated is the length of the window and its position with respect to the marker. There is a clear tradeoff regarding the length. On the one hand, we do not want to include seizure propagation or large artefacts in the analysis and are tempted to choose a narrow time window. On the other hand, the ICA algorithm requires a large enough dataset to be able to learn the parameters of the weight matrix. Hence, a clever time window length must be chosen to tackle the tradeoff.

In the literature, a rule of thumb suggests that at least $N^2 \times k$ data points should be given to ICA for EEG data where N is the number of electrodes and k is a multiple, typically equal to 20 [33]. In our case there are 19 electrodes and the sampling rate is equal to 256 Hz meaning that a window of around 28 seconds should be given to ICA, taking k equal to 20. However, this rule of thumb has not been studied rigorously and the value of k might also depend on the number of electrodes (being larger for high-density EEG). Therefore, shorter time windows were also studied.

Regarding the position of the window, it is of interest to include some background activity to help the subsequent analysis. Indeed, an activity starting around a marker and not present before it is more likely to contain seizure activity than an activity already present before the marker. A proportion of one third before and two third after onset seemed a rational choice.

The studied time windows are summarized in Table 4.2. Only the last one sticks to the rule of thumb.

	Before (s)	After (s)
1	2.5	5
2	5	10
3	10	20

Table 4.2: Studied time windows

4.4 ICA decomposition

Independent Component Analysis (ICA) is the key element of the proposed pipeline. Generally speaking, ICA aims at separating a linear mixture of signals into their different sources. An intuitive application is the situation where two microphones record simultaneously a speech and a song in a room. Both microphones catch a mixture of the two sound signals. The goal of ICA is to recover the two original signals from the recorded mixtures. This is illustrated in Figure 4.5.

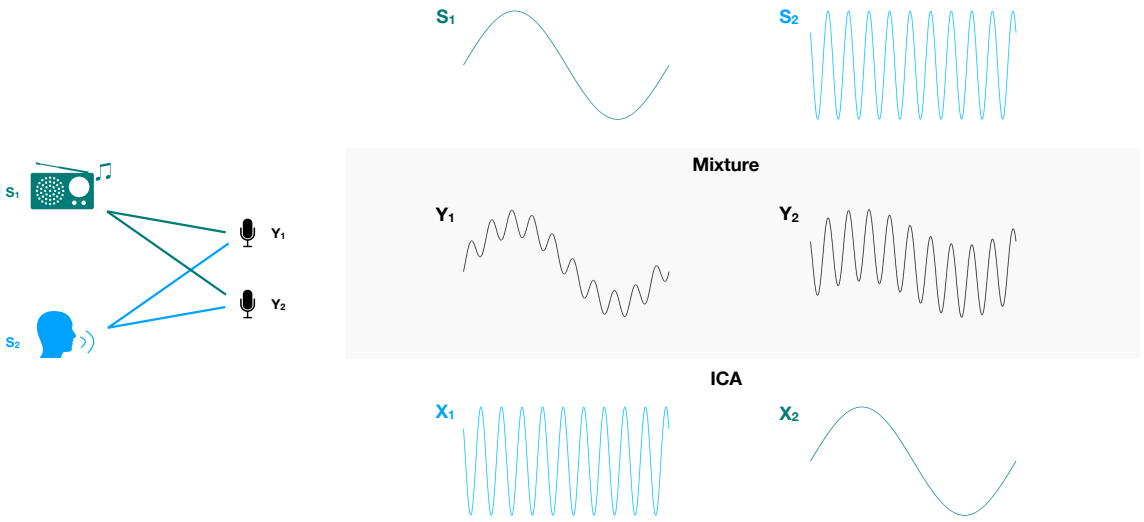


Figure 4.5: ICA illustration - S_1 and S_2 represent the two sources signals, Y_1 and Y_2 represent the recorded signals, X_1 and X_2 represent the independent components recovered by ICA

The concept of ICA was introduced in 1994 by Pierre Comon as an extension of Principal Component Analysis (PCA) [34]. Mathematically, the goal of ICA is to find a linear transformation that minimizes the statistical dependence between the built components. While PCA is limited to second-order statistics, ICA makes use of higher-order information. Since the idea emerged, many algorithms implementing ICA have emerged in the literature.

Over the last decades, ICA has been widely used in the case of EEG analysis [35]. The method is particularly attractive when it comes to remove artefacts [36]. Theoretically, artefacts such as eye blinks are independent from brain signals which suggests that ICA might perform well and separate the artefacts from the activity of interest. Furthermore, ICA is also useful when it comes to disentangle separate brain activities if they are thought independent from each other. In this master thesis, ICA is applied to isolate seizure activity from artefacts and other non-pathological brain activities.

4.4.1 Model and assumptions

ICA assumes that recorded signals come from an instantaneous linear mixture of independent sources. Taking this into account, an EEG recording can be written as :

$$\mathbf{\Phi} = \mathbf{MS} \tag{4.1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{N \times T}$ is the recorded EEG data, $\mathbf{M} \in \mathbb{R}^{N \times N}$ a mixing matrix and $\mathbf{S} \in \mathbb{R}^{N \times T}$ the unknown source data. N denotes the number of channels and T the number of time samples.

First, the model supposes an instantaneous mixture of the sources. While this assumption might be questionable when considering signals travelling at a limited speed (e.g., acoustic signals), this is not an issue in the case of EEG signals regarding the laws of conductance that apply to electrical signals. Similarly, the assumption of linearity is also reasonable since a generated signal has a contribution to a recorded signal proportional to its distance with the recording site. We also note that there are as many independent sources as there are electrodes. While this might seem a doubtful assumption, it does not prohibit the use of ICA for EEG signals in practice since there are commonly a sufficient number of components to isolate independent activities [35].

Another assumption made by ICA is that the source time courses are statistically independent. This might be more arguable regarding EEG data, considering all the work done in the study of connectivity between brain regions. However, multiple studies showed that ICA still perfoms well as long as the sources are not perfectly coupled [35].

Additionally, it follows from the model that the sources remain spatially fixed for the duration of the input data. If that was not the case, the mixing matrix would be time-dependent. In a sense, this seems a rationale assumption for EEG data, considering populations of neurons do not move around. However, signals sometimes propagate across the brain and this could be perceived as a spatial shift. Therefore, care should be taken to avoid such kind of situations and this is particularly true for seizure analysis.

Finally, ICA fails to separate multiple Gaussian signals. Consequently, the last assumption is that the sources are non-Gaussian. This fact comes from the central limit theorem which states that the sum of multiple independent random variables shows a Gaussian distribution. A class of ICA algorithms specifically plays with that law to promote the independence of the sources. In case of EEG signals, this assumption is hard to prove but is supported by the practical value of ICA [35].

4.4.2 Indeterminacies

The goal of ICA is to somehow find the inverse of the mixing matrix $\mathbf{M}$ in equation (4.1). An ICA algorithm therefore returns a weight matrix such that :

$$\mathbf{X} = \mathbf{W}\mathbf{\Phi} \tag{4.2}$$

where $\mathbf{X} \in \mathbb{R}^{N \times T}$ is the matrix whose rows are the independent components time courses and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the weight matrix. Ideally, $\mathbf{X}$ should be equal to the true unknown sources $\mathbf{S}$.

First, we should note that the order of the components is arbitrary. As a matter of fact, by swapping two rows in the weight matrix the order would be changed while the independence criterion of the components would remain unchanged because independence is symmetric. In case of EEG, it might be relevant to order the components according

to the variance of EEG explained by the component. However, the order remains a free choice for the user.

Second, the scaling of the independent components is also not straightforward. Indeed, even if we multiply the weight matrix by a constant α, the independence criterion of the components would stay the same. In a sense, the scaling is distributed between $\mathbf{X}$ and $\mathbf{W}$ so that both matrices should be considered when information about the scaling is needed.

4.4.3 Algorithms

Many algorithms have been developed to solve the ICA problem [37]. Fundamentally, what we need is a function that describes an independence measure of the components, called *contrast function*, and an optimization algorithm to solve the problem that results from the optimization of the contrast function. Different theoretical principles lead to the constructions of various ICA algorithms.

For instance, some algorithms take advantage of non-gaussianity measures. In that case, they heavily rely on the central limit theorem which states that a sum of independent variables tends to show a Gaussian distribution. Therefore, the idea behind these algorithms is to prevent having components exhibiting a Gaussian distribution since it would mean they are still sums of independent signals (under the assumption there is no Gaussian source). Among the variety of non-gaussianity measures that exist, we could for instance cite the kurtosis that we previously mentioned for the bad electrode rejection. Indeed, the kurtosis of a Gaussian distributed variable is zero and thus we want this measure to be maximum for each component to encourage their non-gaussianity.

Other algorithms are inspired by information theory. For instance, the goal of some methods is to minimize mutual information between the components. The mutual information is a measure of dependence between random variables that is zero only for statistically independent variables. Thus, we want to minimize this quantity to promote the independence of the components.

These examples only give a taste of the diversity of techniques that exist. Moreover, the use of different algorithms to solve the underlying optimization problems also give rise to separate methods, generating even more possible ICA algorithms.

4.4.4 Implementation

EEGLAB proposes several ICA algorithms to apply to EEG data. In the context of ICA applied to seizure activity, the most commonly used algorithm is the Extended Infomax algorithm. This latter algorithm is a derivation of the Infomax algorithm that aims at maximizing the mutual information between the input and output of a single-layer feed-forward neural network [38]. The extended version was implemented with the purpose of

separating supergaussian and subgaussian sources [39]. Extended infomax was validated on EEG data in the original paper and has been widely used in the context of seizure activity since then, notably in the majority of the papers cited in the chapter presenting the state of the art. For this reason, the pipeline uses the Extented Infomax algorithm as implemented in the EEGLAB toolbox.

4.5 Component selection

The component selection is for sure the most challenging step of this pipeline. Once the ICA algorithm has computed the N independent components, it remains to decide what are the components that isolate seizure activity. Even with visual inspection, recognizing ictal components requires some expertise in the field of ictal EEG reading. Indeed, what may be obvious to a neurologist's eye is not always obvious to others. As a consequence, the collaboration between clinicians and engineers had a particular importance at this stage of the pipeline.

While completely automating this step is very ambitious, a first goal was to lower the effort of clinicians who would use the pipeline. Reducing the number of components to analyze and ordering them in a clever way is the first step towards an automated pipeline. This is what was achieved through this master thesis. To do so, quantitative features of the components were implemented to identify components likely to contain seizure activity. These features along with new ones could be used in the elaboration of a fully automated pipeline.

At this point, it was chosen that only rhythmic seizures would be included in the study in order to help the design of meaningful component features. However, it should be noted that the pipeline could be used with any type of seizure provided that the selection is then visually performed.

More details about this step of the pipeline and its practical implementation will be found in the next chapter after an analysis of the features that could help to identify ictal components.

4.6 Scalpmaps

Once the ictal components are selected, their corresponding scalp maps may be visualized. A scalp map is an illustration of the prevalence of each electrode in an independent component. It is simply constructed with the weight matrix obtained by the ICA decomposition. Indeed, one row of the weight matrix corresponds to the construction of one independent component and each element of that row corresponds to an electrode. Colors are given to each element of the row according to its value : blue for large negative value and red

for large positive values. Interpolation is then used to find the colors in the area that the electrodes cover.

EEGLAB allows to display the scalp maps in a 2D or 3D frame. Examples are shown on Figures 4.6 and 4.7. The black dots correspond to the electrodes.

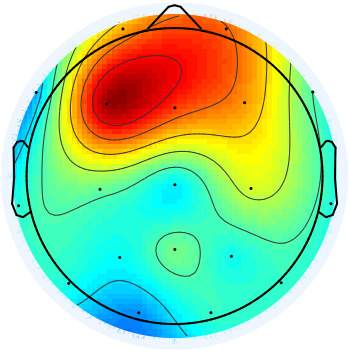


Figure 4.6: 2D scalp map

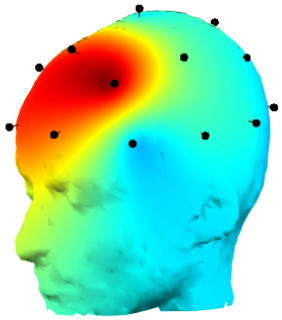


Figure 4.7: 3D scalp map

4.7 Source imaging

While scalp maps could give a first insight into the onset location, they only provide information at the level of the scalp. As the goal is to help a presurgical evaluation, it is more relevant to propose a position inside the brain rather than at its surface. For this reason, a source imaging method was applied once the seizure activity was isolated.

Source imaging regroups techniques that aim at localizing the neural sources from activities recorded at the scalp. Whereas EEG is known for its good temporal resolution, its spatial resolution remains a major limitation that these methods attempt to tackle. Among the diversity of techniques that exists, two categories can be distinguished. On the one hand, some methods assume a limited number of generating dipoles. The solution is thus a few 3D coordinates inside the brain. On the other hand, some methods obtain a distributed solution that is spread across thousands of dipoles. In that latter case, an activation level is available for each dipole of the source space.

Whereas ICA and source imaging both aim at recovering sources in a sense, the two approaches should not be confused since they rely on very different principles. On the one side, ICA aims at unmixing a linear mixture of signals to obtain independent components. The method does not require more information than the signals, it is a so-called *blind source separation* technique. The localization of the components is not precise and can only be roughly approximated by looking at the weight matrix (scalp maps in case of EEG). On the other side, a source imaging technique intends to specifically localize the sources of EEG signals within the brain. The technique requires some additional information

compared to ICA. In fact, details are needed about the head morphology (by the means of a head model constructed thanks to MRI) to determine how signals propagate inside the brain to the scalp. In addition, source imaging does not rely on any sort of independence assumption. In fact, it was initially designed to localize the brain activity at a precise time point. Therefore, it does not need signals but only a measure of potential for each electrode at a fixed time point. In summary, in this pipeline, ICA is used to isolate seizure activity whereas source imaging is used to localize the resulting activity.

4.7.1 eLORETA

For this master thesis, we decided to opt for a distributed source method because it allows to have an idea of the activity zone spread. The chosen method is called exact low-resolution electromagnetic tomography (eLORETA) [40],[41]. This method was derived from the low-resolution electromagnetic tomography (LORETA) introduced in 1994 by Pascal-Marqui [42]. LORETA was the first developed source imaging method to obtain a distributed 3D solution and is still widely used nowadays. A smooth solution is constructed motivated by the fact that neighbouring neurons are likely to be active simultaneously. Instead of obtaining a few dipoles inside the brain like the other techniques used at the time of its creation, a smeared solution is obtained trading some spatial resolution (hence "low-resolution") for a more realistic solution. Based on the same principle of smooth solution, eLORETA method achieves an exact localization under ideal (noiseless) conditions while accomplishing outperforming results on real EEG data [41].

4.7.2 General formulation

Fundamentally, what is of interest for source imaging methods is the relation between the scalp potential and the activity level at each brain source point. This relation can be examined in two directions. In the first direction, we obtain what is called the forward equation:

$$\boldsymbol{\varphi} = \mathbf{K}\mathbf{j} \tag{4.3}$$

where $\boldsymbol{\varphi} \in \mathbb{R}^{N\times 1}$ denotes the potential recorded at the N scalp electrodes, $\mathbf{K} \in \mathbb{R}^{N\times M}$ is the lead field matrix and $\mathbf{j} \in \mathbb{R}^{M\times 1}$ is the activity level at M (typically thousands) fixed cortical voxels.

From known activity levels of the sources $\boldsymbol{j}$, equation (4.3) permits to compute the potential recorded at the electrodes on the scalp. This is possible because the lead field matrix captures how the signals propagate through the different tissues inside the head. This matrix can be built using information from MRI and thus each patient has a specific lead field matrix.

For this master thesis, we used a standard lead field matrix from the New York head (ICBM-NY) [43]. This template was constructed by non-linearly averaging 512 human

heads and is hence less biased than models, such as the commonly used Colin27 [44], that are based on a single man head. The New York head comes with a mesh using MNI coordinates[1] that allows to easily visualize the source space. Using a standard head model instead of a patient specific model reduces the computational load of the procedure at the expense of loosing some spatial resolution. However, showing that the pipeline performs well without the burden of a patient specific head model could encourage its use in clinical practice.

In the reverse direction, obtaining the activity level of the sources from known scalp potential is called the inverse problem and is much more challenging. The problem is ill-posed as M is typically much larger than N: for given scalp potential $\boldsymbol{\varphi}$, many activity levels $\mathbf{j}$ could satisfy (4.3). Consequently, a priori assumptions must be made to select the best solution. That is why we observe such a diversity of source imaging methods due to the variety of possible assumptions.

The eLORETA method is a linear method that looks for a solution to the regularized problem:

$$\hat{\mathbf{j}} = \arg\min_{\mathbf{j}} ||\boldsymbol{\varphi} - \mathbf{K}\mathbf{j}||^2 + \alpha \mathbf{j}^T \mathbf{Z} \mathbf{j} \tag{4.4}$$

where $\alpha \geq 0$ is the Tikhonov regularization parameter and $\mathbf{Z} \in \mathbb{R}^{M \times M}$ is a symmetric positive definite matrix. The vector $\hat{\mathbf{j}} \in \mathbb{R}^{M \times 1}$ is the estimated activity level of the sources.

Depending on the choice of $\mathbf{Z}$, different inverse methods can be derived. Setting $\mathbf{Z} = \mathbf{I}$ gives the classical non-weighted minimum norm solution that yields poor localization performance in practice. The LORETA method uses a matrix that implements the squared three-dimensional spatial Laplacian operator. The eLORETA method uses a matrix such to obtain a zero localization error when tested with point sources anywhere in the brain under noiseless conditions.

The solution to problem (4.4) is given by

$$\hat{\mathbf{j}} = \mathbf{Z}^{-1}\mathbf{K}^\top (\mathbf{K}\mathbf{Z}^{-1}\mathbf{K}^\top + \alpha \mathbf{H})^+ \boldsymbol{\varphi}$$

where the superscript '+' denotes the Moore-Penrose pseudoinverse and $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{\mathbf{1}^\top \mathbf{1}}$. Essentially, what these linear methods compute is a spatial filter $\mathbf{P} \in \mathbb{R}^{M \times N}$ such that $\hat{\mathbf{j}} = \mathbf{P}\boldsymbol{\varphi}$. In a sense, they compute an intelligent inverse of the non-square lead field matrix $\mathbf{K}$. Going from the electrode space to the source space then consists in applying the matrix $\mathbf{P}$ to the scalp potential $\boldsymbol{\varphi}$.

So far, we considered the case where we want to compute the activity level of the sources at one time point. However, it is often of interest to look at the activity level over a specific time period. In that case, the matrix $\mathbf{P}$ could simply be applied to the EEG signals $\boldsymbol{\Phi} \in \mathbb{R}^{N \times T}$ where T is the number of time samples. Doing so, we obtain one time series per point in the source space and it remains to identify which source points are the most

[1] MNI stands for "Montreal Neurological Institute". MNI coordinates refer to a coordinate system that is widely used for source imaging and MRI processing [45].

active. A common procedure consists in computing the power of the source time series. By normalizing the power, we obtain a value between 0 and 1 indicating how active a source point is. It is also possible to compute the spectral power in a certain frequency band and see what source points are then the most active. To monitor a temporal evolution, one could also compute the power over consecutive time intervals.

4.7.3 Implementation

The implementation of the source imaging step relies on codes made available on GitHub in the context of a study about functional connectivity in simulated EEG [46], [47]. The spatial filter $\mathbf{P}$ associated to eLORETA could be computed thanks to a function that takes as input the lead field matrix and the regularization parameter. Similarly to what is done in the cited study [46], the lead field matrix of the New York head was used and the regularization parameter was set to 0.01.

In the specific context of the pipeline, we want to localize the activity of independent components obtained by ICA. Two options are possible: we can either localize the activity of a single component or the activity of multiple components at once. In any case, we first need to back-project the activity of the component(s) at the level of the electrodes before applying the spatial filter.

This back-projection can simply be done using the inverse of the weight matrix $\mathbf{W}$ given by ICA. Indeed, if the purpose was to back-project all the independent components at the level of the electrodes, it is evident that we would compute: $\mathbf{\Phi} = \mathbf{W}^{-1}\mathbf{X}$. In a similar manner, if we want to back-project the activity from the i-th component alone, we will only use the i-th column of $\mathbf{W}^{-1}$ and the i-th row of $\mathbf{X}$. If we want to back-project the activity of multiple components at once, we would then use the corresponding columns of $\mathbf{W}^{-1}$ and corresponding rows of $\mathbf{X}$. Figure 4.8 summarizes the approach where N_s is the number of selected components to back-project, $\mathbf{W}_s^{-1}$ refers to the columns of $\mathbf{W}^{-1}$ corresponding to the selected components and $\mathbf{X}_s$ refers to the corresponding rows of $\mathbf{X}$.

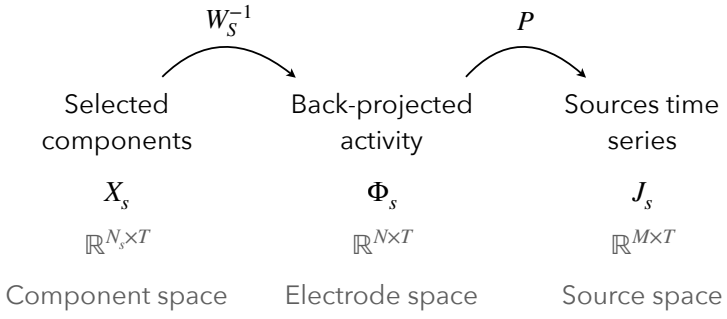


Figure 4.8: Back-projection scheme

While the pipeline allows to localize the activity of multiple components at once, the subsequent analysis of the pipeline will require to localize each component activity separately. In that case, both $\mathbf{W}_s^{-1}$ and $\mathbf{X}_s$ are vectors. We will denote them as $\mathbf{w}_i^{-1}$ and $\mathbf{x}_i$, respectively, considering we want to localize the activity of the i-th component. An important thing to note here is that the back-projected activity of the i-th component $\mathbf{\Phi}_i = \mathbf{w}_i^{-1}\mathbf{x}_i$ is a rank-one matrix since it results from an outer product of two vectors. Consequently, the matrix $\mathbf{J}_i = \mathbf{P}\mathbf{\Phi}_i$ is also of rank one. This means that the source time series are identical up to a multiplicative constant. Therefore, there is no need to go through the power computation to know what the most active sources are. Indeed, the level of source activation is completely captured by the multiplicative constants.

These multiplicative constants are stored in the vector $\mathbf{P}\mathbf{w}_i^{-1}$. In fact, we have that:

$$
\begin{aligned}
\mathbf{J}_i &= \mathbf{P}\mathbf{w}_i^{-1}\mathbf{x}_i \\
&= \begin{pmatrix} p_{11} & \dots & p_{1N} \\ \vdots & & \vdots \\ p_{M1} & \dots & p_{MN} \end{pmatrix} \begin{pmatrix} | \\ \mathbf{w}_i^{-1} \\ | \end{pmatrix} \begin{pmatrix} — & \mathbf{x}_i & — \end{pmatrix} \\
&= \begin{pmatrix} p_{11}(\mathbf{w}_i^{-1})_1 + \dots + p_{1N}(\mathbf{w}_i^{-1})_N \\ \vdots \\ p_{M1}(\mathbf{w}_i^{-1})_1 + \dots + p_{MN}(\mathbf{w}_i^{-1})_N \end{pmatrix} \begin{pmatrix} — & \mathbf{x}_i & — \end{pmatrix}
\end{aligned}
$$

We observe that the time course of the i-th independent component does not come into play to determine what source points are the most active. Only the corresponding column in the inverse of the weight matrix plays a role. Due to this, there is no point at comparing the activity of the sources for a specified frequency band since the most active sources are fully specified by $\mathbf{P}\mathbf{w}_i^{-1}$ that does not depend on the component time course $\mathbf{x}_i$. For the same reason, computing the power of consecutive intervals would always lead to the same active brain region. This conforms with the ICA assumption stating that the sources are spatially fixed.

To accommodate with what would be obtained computing the power of the sources, the coefficients of $\mathbf{P}\mathbf{w}_i^{-1}$ are squared. Subsequently, they are normalized between 0 and 1 by dividing by the largest coefficient. It is then possible to plot the solution in a 3D space where the darkness of a source point depends on its level of activity. A threshold of 0.8 was chosen to determine what we call the *active cloud* of a component. Accordingly, the active cloud outlines the most active source points of a component. An example of component activation is shown in Figure 4.9 and 4.10. In Figure 4.10, the average of the active cloud refers to its centroid, i. e., the position computed by taking the average of each coordinates.

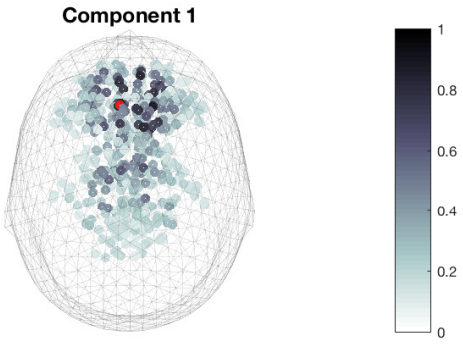


Figure 4.9: Source activity normalized between 0 and 1; the sources with an activity smaller than 0.1 were not displayed for readability; the red dot is the most active source

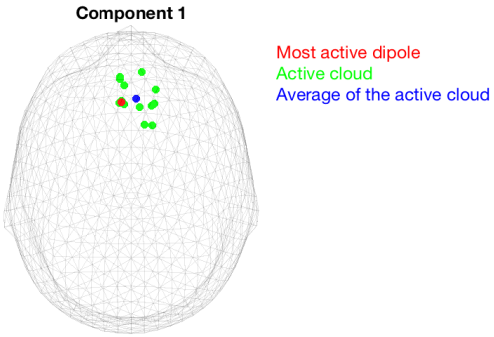


Figure 4.10: Source activity summarized by the most active dipole (red), the active cloud (green) and the average of the active cloud (blue)

Chapter 5

Pipeline analysis

In this chapter, we first present how we implemented quantitative measures to assess the pipeline quality. Next, we study different setups, using these measures, in order to choose the optimal preprocessing parameters. Afterwards, we describe and analyze component features that could possibly help to recognize components that isolate seizure activity. Finally, we present the component selection of the final proposed pipeline and discuss the achieved performance.

5.1 Performance measures

All patients included in this study became seizure-free after surgery. Accordingly, the seizure onset zone is assumed to be located inside the removed or disconnected zone (called *resection zone*). It is then of interest to measure the distance and overlap between the resection zone and the estimated zone given by the pipeline. In order to objectively assess the pipeline performance, three quantitative measures were implemented based on the resection zone.

5.1.1 Resection zone drawing

For the purpose of drawing a resection zone, the available MRI must first be normalized in order to obtain an image in a reference brain space. This is necessary since we do not use individual head models to perform source imaging. Therefore, the pipeline gives us a location inside a standard brain model using MNI coordinates and we must consequently define the resection zone in that same space. The effect of spatial normalization is shown in Figure 5.1 and 5.2.

The next step consists in drawing the brain regions that were removed or disconnected. The image viewer MRIcron allows to draw 3D volumes on MRI. More specifically, there is a tool that draws 3D shapes that fit specific regions by detecting the brightness changes. This was helpful when the resection zone simply consisted in removed regions that are typically

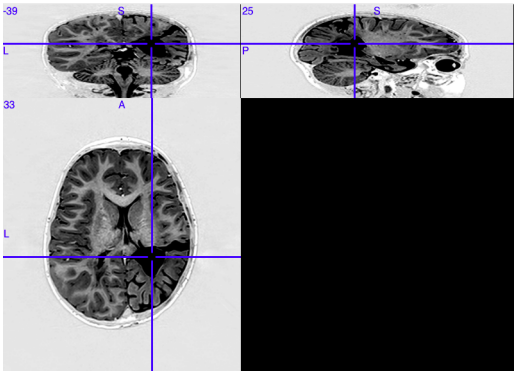


Figure 5.1: Before normalization

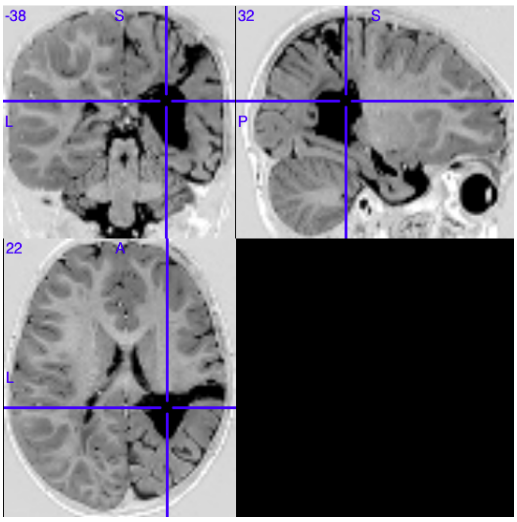


Figure 5.2: After normalization

contrasting. However, the tool could not detect regions that were only disconnected. To tackle this issue, the additional regions were drawn manually by adding up small spheres. Sometimes, a region could be specified by a simple condition on the coordinates and thus it did not require to draw manually. This was the case for hemispherectomies where a whole hemisphere is disconnected. Another example is illustrated in Figure 5.3 where the patient underwent a temporal-parietal-occipital disconnection. First, the boundary of the resection zone was drawn thanks to the brightness detection tool. Second, additional regions were drawn manually so that disconnected parts were also selected. Finally, the region on the back of the brain was specified by a simple condition on the coordinates (see dashed green line in the right part of Figure 5.3) and so did not need to be drawn manually.

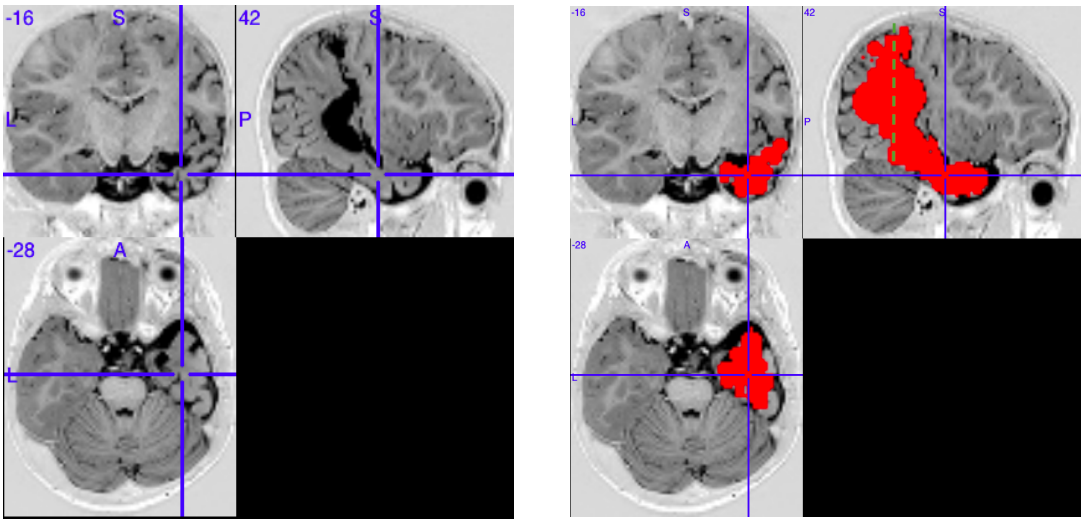


Figure 5.3: Image before (left) and after (right) drawing part of the disconnected region (red: selected region, dashed green line: delimitation behind what everything is disconnected)

Finally, the drawn region could be uploaded in Matlab. From MRIcron, the volume of interest could be saved as a NIfTI file which is a typical format for MRI that EEGLAB can handle and build structures from. The corresponding structure contains a 3 dimensional matrix containing the intensity of each voxel. In this case, the matrix takes only binary values depending if the voxel belongs to the drawing or not. Using a change of coordinate, it is possible to retrieve the MNI coordinates of the drawing.

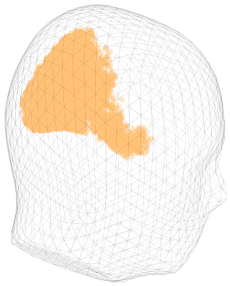


Figure 5.4: Resection zone opened in Matlab

5.1.2 Quantitative measures

Once the resection zone was drawn and uploaded in Matlab, it became possible to compare it with the pipeline output. A first measure that was implemented is the distance between the most active source point and the resection zone. This is done by calculating the Euclidean distance between the most active source point and the closest coordinate of the resection zone. In other words, we compute the distance to the border of the resection zone.

However, the source imaging method that we use not only gives us one 3D coordinate but a distributed solution across the source space. Therefore, it is informative to consider not only the most active source point but also the source points having an activity above a fixed threshold. The concept of active cloud (introduced in subsection 4.7.3) allowed us to compute two additional quality measures.

First, we computed the distance between the resection zone and the centroid of the active cloud. This latter measure is often more robust than the distance with the most active source. Indeed, the source imaging method we use does not claim to find one dipole explaining the scalp activity as it would be the case with a single dipole fitting method. Therefore, judging the quality of the pipeline based on a single dipole could bias its true potential. This is illustrated in the left part of Figure 5.5 where the centroid (blue dot) is more representative of the estimated activity than the most active source point (red dot).

Second, we also computed the percentage of the active cloud located inside the resection zone. That measure informs about the volume overlapping between the estimated zone and the resection zone. To decide whether a source point is located inside the resection zone, a tolerance of one centimeter was set. First, this is justified by the fact that we compute the distance between a point and a cloud of points. Even if the point is located inside the cloud, the distance is not likely to be exactly zero since the cloud is not a continuous space. Second, a certain tolerance should be accepted regarding the drawing of the resection zones that might be imperfect. Finally, we note that we consider low-density EEG with inter-electrode distances ranging from 5 to 7 centimeters. Consequently, we

do not expect a resolution in the order of millimeters. However, a potential improvement could be to calculate the distance to the convex hull of the resection zone to overcome the first issue.

All the quantitative measures are visible in Figure 5.5 for two components coming from different patients.

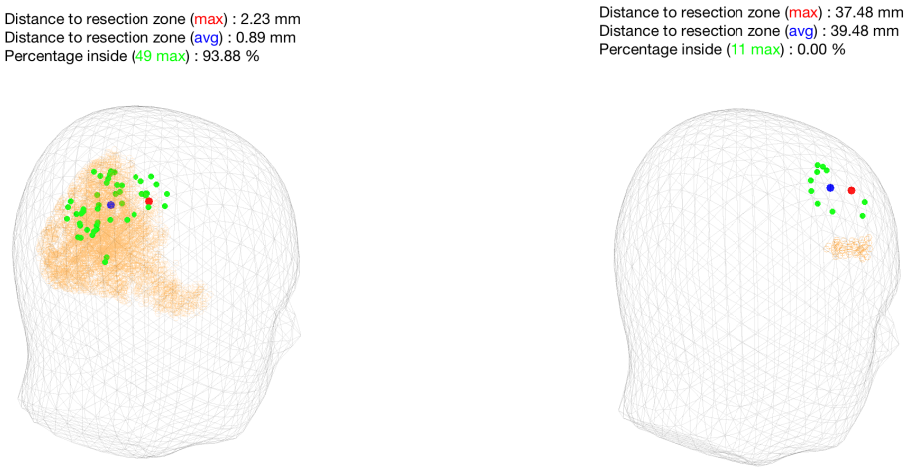


Figure 5.5: The red dot, green dots and blue dot represent the most active source point, the active cloud and its centroid, respectively. The light orange cloud represents the resection zone.

5.1.3 Types of resection zone

To assess optimally the pipeline, the resection zones of the patients were divided into three categories depending on their sizes:

- Type 1 : hemispherectomy
- Type 2 : disconnection
- Type 3 : lesionectomy

An example for each category is presented in Figures 5.6, 5.7 and 5.8. In the Material chapter, Table 3.1 gives the type of each patient included in this study.

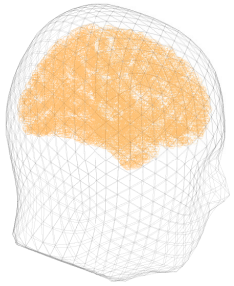


Figure 5.6: Type 1 Hemispherectomy

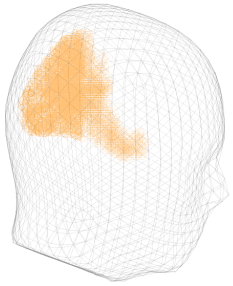


Figure 5.7: Type 2 Disconnection

Figure 5.8: Type 3 Lesionectomy

The division in categories was necessary in order to assess in a reliable manner the components quality. Indeed, stricter criteria are applied to hemispherectomies compared to lesionectomies. Table 5.1 shows the thresholds chosen to consider a component as localized within the resection zone and thus likely to be ictal. For types 1 and 2, a tolerance of 10 mm was set for the reasons explained in the previous subsection. For type 3, a bit larger tolerance was chosen regarding the small size of the resection zones. The percentage threshold was also chosen according to the resection zone extent. We should note that the 0 associated to type 3 means at least one dipole must be inside the resection zone since we consider a strict inequality.

	Distance to maximum (mm)	Distance to average (mm)	Percentage inside resection zone (%)
Type 1	10	10	90
Type 2	10	10	70
Type 3	20	20	0

Table 5.1: Criteria that defines if a component is localized within the resection zone

5.2 Setup analysis

As it was mentioned in the previous chapter, there is no consensus about either the filters to apply or the length of the time window to select when ICA is applied to seizure signals. Therefore, in this section we attempt to answer the following questions:

1. Which high-pass filter is optimal ?
2. What time window length is optimal ?

To do so, we established a series of setups to be analyzed according to the options proposed in the previous chapter. The studied setups are summarized in Table 5.2.

	High-pass filter (Hz)	Before (s)	After (s)
Setup 1	1	2.5	5
Setup 2	2	2.5	5
Setup 3	1	5	10
Setup 4	2	5	10
Setup 5	1	10	20
Setup 6	2	10	20

Table 5.2: Setups to be analyzed

In order to avoid the bias of visual inspection, we compared the different setups using quantitative measures. For each seizure, we counted the number of components having their origin in the resection zone using the quality criteria established in the previous section. Nevertheless, we should keep in mind that having numerous components lying in the resection zone is not necessarily better than having a single component that is

correctly localized. In fact, in the second case the seizure activity was probably isolated in one single component which is more desirable than having the activity spread over multiple components. What is essentially needed is that at least one component is located within the resection zone for each seizure. Therefore, a setup quality was evaluated according to the number of seizures for which at least one component was located within the resection zone rather than the number of components having their origin in the resection zone. This way, setups that could not find any component lying in the resection zone were penalized.

The strategy adopted to select the optimal setup implied to gradually restrict what was considered as an *accurate* component. First, a component was considered as accurate only if it satisfied the quality criteria presented in subsection 5.1.3. Namely, the centroid and maximum of the component active cloud must be sufficiently close to the resection zone and the percentage of active cloud inside the resection zone must be sufficiently high. As described earlier, the thresholds used depend on the type of resection.

Motivated by the fact that no significant difference was observed, two additional features were used to discern what setup should be chosen. The first feature measures the dispersion of the active cloud. Indeed, it is not desirable to consider as accurate a component that is diffuse within the brain. The second feature measures how much power of the signal can be found in the frequency band identified by a neurologist as being representative of the ictal rhythm. This latter measure therefore anticipates the subsequent analysis in the sense that it promotes setups for which well located components present a rhythmic pattern. These two features will be explained in more detail in the next section of this chapter.

Table 5.3 presents the percentage of seizures for which at least one component was accurate according to the different sets of quality criteria. Overall, we observe that setup 6 performs slightly better than the other setups. For any set of quality criteria, setup 6 achieves the best performance even though it is not necessarily the only one to reach that performance. Consequently, it seemed reasonable to continue the analysis of the pipeline using that setup. Nonetheless, the difference between the setups is not considerable, meaning that the pipeline performs usually well and is probably not too sensitive regarding the preprocessing steps that are applied.

Quality criteria					Setups					
Distance to maximum (mm)	Distance to average (mm)	Percentage inside (%)	Dispersion	Percentage of power of interest (%)	1	2	3	4	5	6
10,10,20	90,70,0	10,10,20			**89.74**%	87.18%	84.62%	84.62%	84.62%	**89.74**%
10,10,20	90,70,0	10,10,20	500		87.18%	84.62%	84.62%	82.05%	84.62%	**89.74**%
10,10,20	90,70,0	10,10,20	500	20	79.49%	79.49%	82.05%	76.92%	82.05%	**87.18**%

Table 5.3: Setup comparison : in white (left), quality criteria used, if three values are given then it relates to seizures of type 1, 2 and 3 respectively; in blue (right), percentage of seizures having at least one accurate component (bold font means it yields the best performance across the setups); "percentage inside" (third column) refers to the percentage of active cloud inside the resection zone; 39 seizures were analyzed in total

According to these results, setup 6 was chosen for the rest of the analysis. This setup corresponds to the case with the high-pass cutoff frequency equal to 2 Hz and the largest time window (10 seconds before the marker, 20 seconds after). Regarding the high-pass filter, this confirms the fact that ICA could benefit from a relatively more aggressive filter that removes a series of low frequency artefacts. Concerning the time window length, choosing the longest time window is in accordance with the rule of thumb that can be found in the literature and that was mentioned in section 4.3.

Looking more carefully at the results, we noticed that the seizures of a patient consistently achieved poor results for each of his/her seizures. Indeed, four seizures were available for patient 1 and they did not lead to any accurate components except for one of his/her seizures with setup 1. If the seizures from that patient were removed, setup 6 is the only setup to achieve a 100% performance for the first two sets of quality criteria and fails to find an accurate component for only one seizure (out of the 35 remaining seizures) with the last set of quality criteria.

Patient 1 was excluded for the rest of the analysis because his/her seizures did not yield any accurate components with the chosen setup. Keeping that patient would have biased the analysis to identify features that distinguish accurate and non-accurate components because in that case the seizure activity is spread across non-accurate components. Furthermore, we should note that the seizures experienced by this patient already raised doubts when the neurologists had to choose the beginning of seizure activity. Two potential time markers were placed 20 seconds apart and the second was finally chosen for this study. Therefore, we might also suspect that seizure propagation was included for the analysis. One of the seizures can be seen in Figure A.2 in the appendix.

Other hypotheses could also explain why the pipeline failed to find accurate components for patient 1. First of all, the resection zone is quite deep inside the brain (see Figure A.1 in Appendix) and so low-density EEG might not be sufficient to localize such type of sources. Second, as it can be seen in Figure A.2 in the Appendix, high frequency activity starts at the first seizure marker. This activity might give a hard time to ICA that would have difficulty isolating it. Finally, the patient is also a special case as he underwent a double resection surgery.

5.3 Feature analysis

Now that the setup choice is done, we are interested in knowing whether the components that correctly localize the seizure onset are somehow identifiable. Knowing there are components that would lead to a correct location is a proof-of-concept that the pipeline has the potential to correctly localize the seizure onset zone. However, it would not be of great help if these components are not recognizable. Now, we would like to find features that distinguish components that are well located from the other components. To do so, different features were implemented and analyzed. We considered as accurate

the components satisfying the quality criteria related to the position of the active cloud (first set of criteria in Table 5.3) without any constraints on the dispersion or rhythmical activity because these are features to be analyzed.

For each feature, we compared the mean, median and standard deviation for the accurate and non-accurate components separately. This gives a first idea whether the feature could potentially help to distinguish the accurate components. We displayed box plots[1] for both accurate and non-accurate components to gain insight into the distributions. Next, we also computed the correlation between the feature and the distance from the centroid of the active cloud to the resection zone. This would help to know if there is a linear trend that could be useful for a component selection. Finally, we conducted statistical tests to further assess the difference between accurate and non-accurate components. Depending on the feature distribution, either a t-test or Wilcoxon rank sum test was performed.

Additionally, receiver operating characteristic (ROC) curves were created to assess whether a threshold on the feature would allow to discriminate between accurate and non-accurate components. Note that the curves were plotted with the specificity on the x-axis and the sensitivity on the y-axis. For each curve, the area under the curve (AUC) was computed. Even if we do not expect a single feature to classify perfectly accurate and non-accurate components, this gives an idea whether setting a threshold on the feature would enable to get rid of some non-accurate components while keeping accurate components.

At this point, we should remind that ICA cannot recover the scaling of the independent components as it was mentioned in section 4.4. The scaling is not completely lost but rather spread between the weight matrix and the component signals. To stay cautious, implementation was made in such a way that the scaling would not impact the features.

The field of seizure detection inspired the design of features. Seizure detection aims at classifying seizure and non-seizure EEG recordings using machine learning models [49], [32]. The purpose is to identify the time windows that present ictal activity in long EEG recordings. Clearly, there is a similarity between the component selection of our pipeline and the classification task addressed by seizure detection. The main difference relies in the fact that we consider a single component signal whereas seizure detection considers EEG recordings with multiple electrodes. Moreover, as explained in the previous paragraph we must be cautious about the scaling of the components signals.

Dispersion measure

A first feature that was implemented is the dispersion of the active cloud. This feature is the only one that relies on the geometry of the component active cloud and not the time

[1] "On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol." [48]

signal. In theory, an ictal component should isolate an activity generated by a specific population of neurons. If a component leads to a diffuse active cloud, it is unlikely that the component isolates seizure activity because the seizure onset zone is supposed to be focal.

Several implementations are possible to compute the dispersion a cloud of points. We decided to compute the Frobenius norm of the covariance matrix of the 3D coordinates. Examples are shown in Figures 5.9, 5.10 and 5.11 where we observe different spreads of the sources confirmed by contrasting dispersion measures.

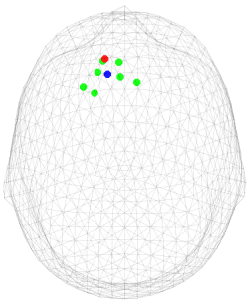


Figure 5.9: Dispersion: 185.25

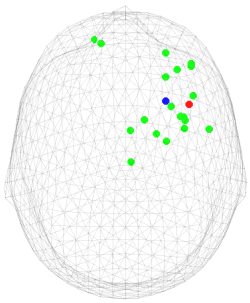


Figure 5.10: Dispersion: 1403.18

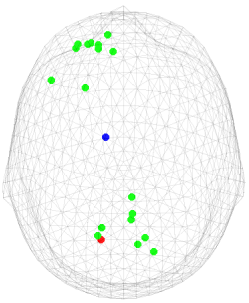


Figure 5.11: Dispersion: 5894.55

Table 5.4 summarizes the statistical analysis for the dispersion. Figure 5.12 shows box plots for both accurate and non-accurate components. We observe that the mean and the median of the feature are smaller for accurate components than for non-accurate components. That is in accordance with the expectations because well located components are likely to isolate seizure activity originating from a limited brain region. The very small correlation suggests that there is no simple linear relation between the dispersion of a component and its distance to the resection zone. A Wilcoxon rank sum test was performed since the distribution is unlikely to be normal, preventing the use of a classical t-test. The test was significant, suggesting that it is unlikely that both accurate and non-accurate groups come from the same distribution.

Figure 5.13 shows the ROC curve for the dispersion feature. As we could expect, setting a threshold on the feature does not allow to classify correctly the components. However, the curve detaches from the random classification meaning that it could benefit a classification process.

	Accurate	Non-accurate
Mean	416.63	569.89
Std	463.19	673.53
Median	308.32	358.95
Correlation	-0.0403	
Wilcoxon rank sum test p-value	3.47e-04	

Table 5.4: Statistics: dispersion

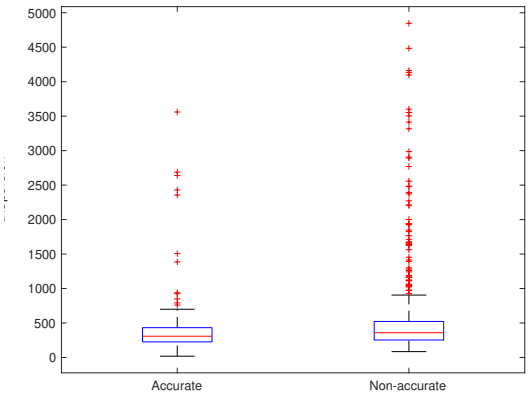


Figure 5.12: Box plots: dispersion

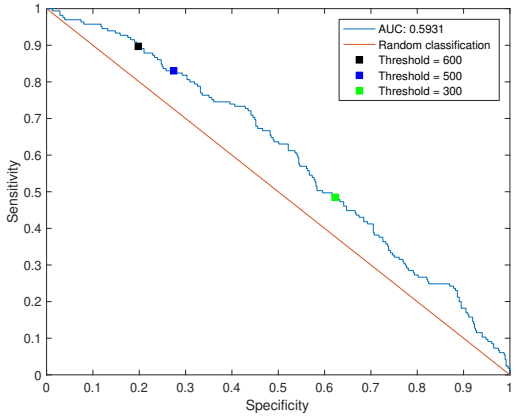


Figure 5.13: ROC curve: dispersion

Power in the frequency band of interest

The next feature was also already mentioned in the previous section: the percentage of power in the frequency band of interest. The motivation behind this feature lies in the fact that the seizures are rhythmic and therefore the components that isolate seizure activity should be sufficiently rhythmic. This feature required a neurologist to identify the frequency band of the ictal rhythm. Typical bands of brain activity are summarized in Table 5.5 with some intermediary bands as well. Information about what frequency band was chosen for each seizure can be found in Table 3.1 in the Material chapter.

	Frequency band (Hz)
delta	1-4
theta	4-8
alpha	8-12
beta	12-20
delta-theta	2-6
theta-alpha	6-10
alpha-beta	10-15

Table 5.5: Typical frequency ranges

Instead of simply looking at the average power in the frequency band of interest, it is the percentage of power that was computed by calculating the ratio of the average power in the frequency band of interest and the total power of the component signal. Doing so, we avoid any troubles with the scaling of the components. Indeed, if a component time course is multiplied by a factor α, its average power in the frequency band is multiplied by α while the percentage remains the same.

Table 5.6 summarizes the statistics for the feature and Figure 5.14 shows the box plots. We notice that as expected the mean and median of the feature are larger for accurate components than for non-accurate components. The correlation is also higher than for the previous feature indicating a weak linear trend. The negative sign of the correlation

corresponds to the intuition since we expected that a large power in the frequency band of interest would lead to a small distance to the resection zone. The t-test being significant confirms that there is indeed a difference in mean between accurate and non-accurate groups.

Figure 5.15 presents the ROC curve for the power in the frequency band of interest. We observe that the AUC is greater than for the dispersion feature but still does not suggest that a simple threshold could discriminate between the accurate and non-accurate components. Nevertheless, we could take once again advantage of a threshold on the feature to eliminate non-accurate components.

	Accurate	Non-accurate
Mean	0.3663	0.3050
Std	0.1207	0.1342
Median	0.3601	0.2986
Correlation	-0.2457	
T-test p-value	2.7729e-07	

Table 5.6: Statistics: power in the frequency band of interest

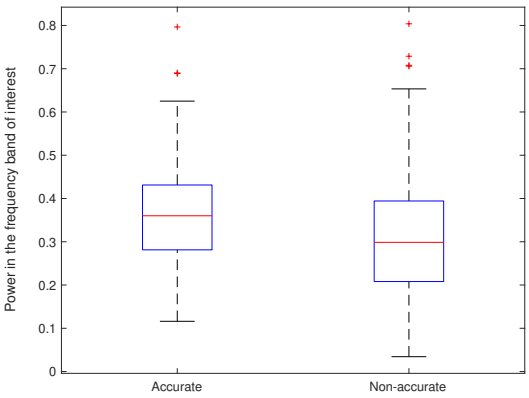


Figure 5.14: Box plots: power in the frequency band of interest

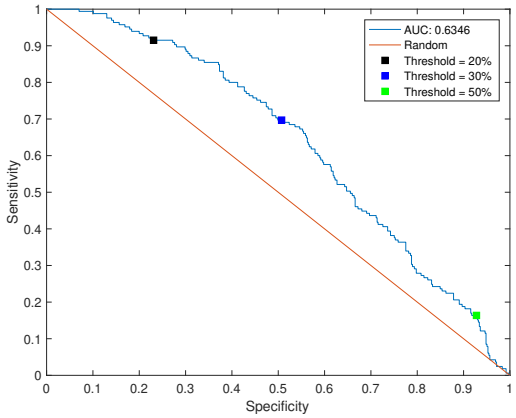


Figure 5.15: ROC curve: power in the frequency band of interest

Power in the frequency band 2-20 Hz

The power in the frequency band from 2 to 20 Hz is somewhat similar to the previous feature. The motivation behind that feature is that seizures included in this study, and more generally typical rhythmic extra-temporal seizures, have their activity in that frequency range. This feature is more robust than the previous one in the sense that it does not rely on a clinician's choice. However, this broader frequency range could also correspond to non-pathological brain activities. To avoid troubles with the scaling, the feature was implemented similarly to the previous one as a percentage of power in the frequency band 2-20 Hz.

Table 5.7 summarizes the different statistical values computed for the feature and Figure 5.16 shows the box plots. Overall, the difference between accurate and non-accurate components is not striking for this feature. The mean and the median are slightly higher for accurate components as expected but the variation is rather small. The sign of the correlation is in line with the expectation that the more power, the closer to the resection zone. The statistical test shows that there is a difference between accurate and non-accurate components yet the p-value is higher than for the previous features.

Figure 5.17 shows the ROC curve of the feature. We observe that if we want to keep a high sensitivity (≥ 0.8) setting a threshold would not help to eliminate non-accurate components better than with a random classification. Therefore, setting a threshold on that feature does not seem to be the optimal way to extract information from it.

	Accurate	Non-accurate
Mean	0.9156	0.9040
Std	0.0737	0.0747
Median	0.9351	0.9216
Correlation	-0.1837	
Wilcoxon rank sum test p-value	0.0035	

Table 5.7: Statistics: power in the frequency band 2-20 Hz

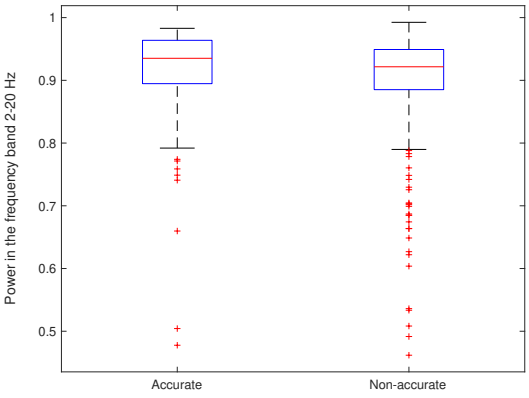


Figure 5.16: Box plots: power in the frequency band 2-20 Hz

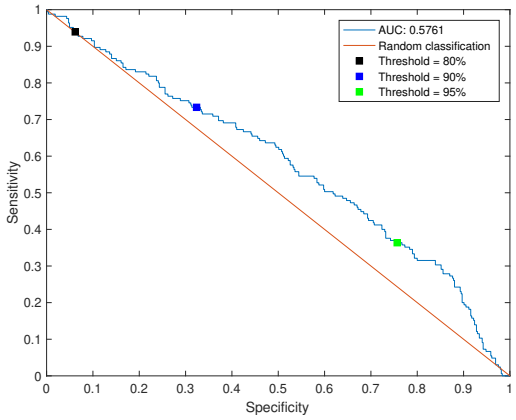


Figure 5.17: ROC curve: power in the frequency band 2-20 Hz

EEG variance contribution

Another interesting feature is how much of the EEG variance a component is responsible for. We expect that a component accounting for seizure activity should contribute significantly to the EEG variance. However, gross artefacts typically explain much of the EEG variance and thus the feature is not expected to be sufficient on its own. This feature distinguishes from others because it also uses information from the EEG and the weight matrix. It allows to gain knowledge about how predominant a component is without be-

ing tricked by the scaling indeterminacy. In a way, it combines the information of the component signal and the weight matrix to learn about the relative scale of a component.

Its implementation was inspired from the `eeg_pvaf` function from the EEGLAB toolbox. This function computes what they call the *percent variance accounted for*. However, we will talk about *EEG variance contribution* as the value could be negative. The feature needs to compute the ratio of the variance of the EEG without the component and the variance of the initial EEG. To remove a component contribution, its signal is back-projected at the level of the electrodes using the inverse of the weight matrix (see subsection 4.7.3 for more details about the back-projection) and then removed from the EEG. The equation behind the feature is presented hereunder:

$$feat = 1 - \frac{\text{var}(EEG - \text{back\_proj}(\boldsymbol{x}_i))}{\text{var}(EEG)} = 1 - \frac{\text{var}(EEG - \boldsymbol{w}_i^{-1}\boldsymbol{x}_i)}{\text{var}(EEG)} \tag{5.1}$$

While it might be counter-intuitive to obtain negative values, this could and did happen. A value close to zero means that with or without the component the EEG variance remains the same. A negative value means that by removing the component activity the EEG variance increases. A possible explanation is that the component annihilates the activity of another component. Therefore, once the component is removed, the other component contributes to the EEG variance. We expect this kind of situation to be very unlikely for components that isolate seizure activity since it is not in accordance with the concept of a component isolating a real physiological phenomenon.

Table 5.8 shows the statistical properties of the EEG variance contribution and Figure 5.18 presents the box plots. The mean and the median of the feature are larger for accurate components than they are for non-accurate components. That corresponds to our expectations. Despite that, the correlation is not significant as it is very small. Yet the statistical test yielded the conclusion that both groups come from different distributions. Finally, the box plots confirm that no accurate components reach a negative value while this happens for non-accurate components.

Figure 5.19 presents the ROC curve for the EEG variance contribution. The curve detaches from the random classification which indicates that we might take advantage of the feature to help the classification.

	Accurate	Non-accurate
Mean	0.0623	0.0529
Std	0.0590	0.0639
Median	0.0421	0.0333
Correlation	-0.0612	
Wilcoxon rank sum test p-value	0.0082	

Table 5.8: Statistics: EEG variance contribution

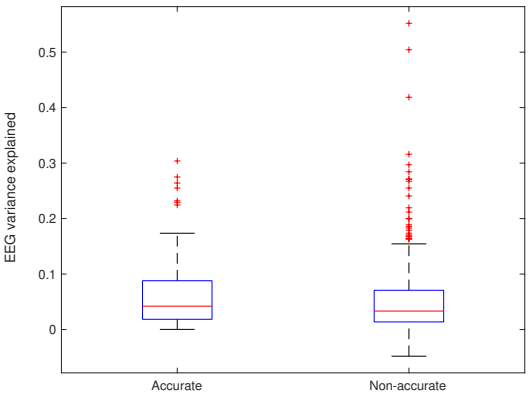


Figure 5.18: Box plots: EEG variance contribution

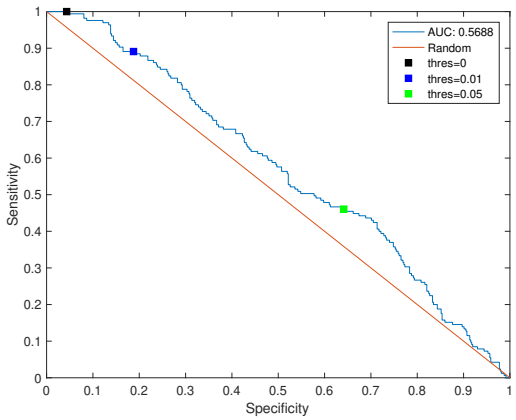


Figure 5.19: ROC curve: EEG variance contribution

Change of variance

Another proposed feature is the change of variance observed at the seizure onset. So far, the features were computed over a period of 10 seconds after the seizure marker. This feature attempts to take advantage of the fact that the seizure should start at the marker. A component that isolates seizure activity should have its variance that increases during the period after the onset in comparison with the period before the onset.

To avoid any issues with the scaling, the feature was implemented by taking a ratio: the variance after the onset is divided by the variance before the onset. Doing so, we expect a larger value for accurate than non-accurate components.

As for other features, Table 5.9 shows the statistical properties and Figure 5.20 the box plots. Contrary to our expectations, the mean and median were slightly larger for non-accurate than for accurate components. The statistical tests and box plots also did not suggest that the feature was relevant to analyze further, so we did not compute the ROC curve.

The fact that the mean and the median are larger for non-accurate than for accurate components could be explained by the fact that components isolating artefacts could also show a large change of variance. A typical component isolating a movement artefact is flat except during the movement. Therefore, if the movement occurs after the seizure onset (which is likely) then the feature will be very large even though the component is not accurate.

Another possible explanation why this feature does not seem to help for the component selection is that it highly depends on the position of the onset marker. This feature is indeed not robust if the marker is misplaced by a few seconds.

	Accurate	Non-accurate
Mean	3.114	4.548
Std	7.793	13.212
Median	1.437	1.676
Correlation	0.0983	
Wilcoxon rank sum test p-value	0.0878	

Table 5.9: Statistics: Change of variance

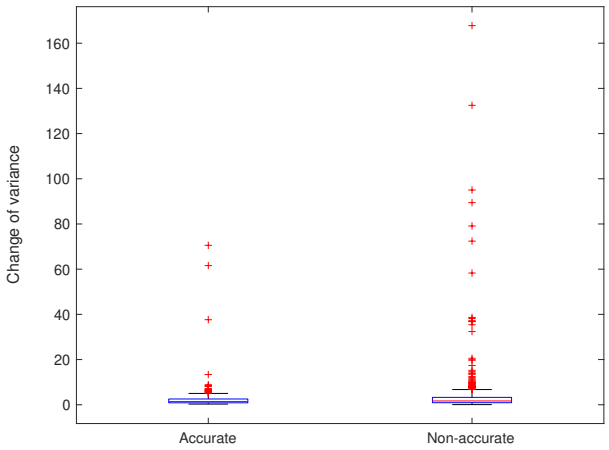


Figure 5.20: Box plots: change of variance

Change of entropy

A last feature that was analyzed is the change of entropy at the onset. The entropy of a signal is a measure of its uncertainty. A high entropy typically means that the signal is hardly predictable. The motivation behind the feature lies in the fact that brain activity is thought to be more predictable during a seizure than it usually is [49]. Thus a drop of entropy could be observed at the onset marker in components isolating seizure activity.

The spectral entropy was computed before and after the marker onset. The entropy of a signal does not depend on its scale. Therefore, the feature was simply implemented as a difference of entropy after and before the marker. A negative value would therefore mean that the entropy has dropped.

Table 5.10 and Figure 5.21 summarize the statistical analysis and box plots for the feature. We observe that both the mean and the median are close to zero, suggesting that there is no significant change in entropy at the onset. The p-value of the t-test also confirms that there is no significant difference between accurate and non-accurate components. Regarding these facts, we did not compute the ROC curve for the feature.

While entropy has been shown to be relevant in the field of EEG seizure detection [49], the feature does not seem beneficial in the context of our pipeline. There is a potential

explanation to this. Indeed, a component that perfectly isolates seizure activity should show little activity before the onset. Therefore, the signal should be almost constant before the onset meaning it has a very low entropy. Consequently, it makes sense that a component that isolates seizure activity will not necessarily show a drop of entropy. In conclusion, while this feature is suited for seizure detection it is not applicable to the component selection of our pipeline.

	Accurate	Non-accurate
Mean	0.017	0.036
Std	0.446	0.445
Median	0.003	0.032
Correlation	0.1677	
T-test p-value	0.6381	

Table 5.10: Statistics: Change of entropy

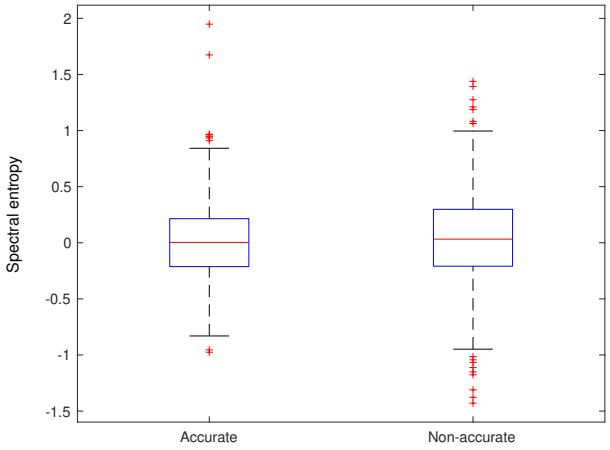


Figure 5.21: Box plots: change of entropy

5.4 Component selection

In light of the feature analysis, a component selection was constructed taking into account the most relevant features. In the framework of the pipeline, the component selection can be seen as a typical machine learning task: either of regression if the goal is to predict the distance to the resection zone for each component or of classification if we only want to know whether a component is accurate or not.

The approach that was implemented in this master thesis is nevertheless a bit different from usual machine learning models. Regarding the small number of features, a more transparent method was used to obtain a component selection that corresponds to intuition in the way it works. In some sense, we implemented manually a model so that each choice made by the model is intuitive and justified.

The component selection consists of two steps:

1. Reject components having features that do not satisfy some thresholds
2. Sort the accepted components according to a feature

The first step resembles how a decision tree would work. Depending on some thresholds fixed for each feature, only a subset of components is kept. Next, the remaining components are sorted according to a well-chosen feature. Consequently, the selection does not operate in a fully automated manner but rather helps the clinician to choose ictal components by eliminating irrelevant components and by sorting the accepted components in a way that the components which are the most likely to isolate seizure activity are presented at the top.

From the analyzed features, only three of them were used for the component selection: the dispersion, the power in the frequency band of interest and the EEG variance contribution. The thresholds were fixed separately for each feature according to the ROC curves. For each feature, the threshold that yielded a 90% sensitivity was chosen. In this particular context, a high sensitivity is desirable because we want to avoid eliminating accurate components. Thresholds are summarized in Table 5.11 and were visible in each of the ROC curves presented in section 5.3.

	Dispersion	Power in the frequency band of interest	EEG variance contribution
Threshold to reject	> 600	$< 20\%$	< 0.01

Table 5.11: Thresholds used to reject components

The second step of the selection sorts the accepted components according to a feature. In consideration of the previous analysis, the feature that showed the largest correlation was chosen: namely the power in the frequency band of interest. This is a very natural choice that conforms with the intuition that the components that are the most likely to isolate seizure activity should show a predominant rhythm.

Two types of analysis were conducted to assess the performance of the component selection. On the one hand, a seizure-based analysis takes into account to which seizure a component belongs. On the other hand, we also performed a component-based analysis that considers the components regardless of their associated seizures. From a clinical point of view, the first analysis is the most relevant because what we want in the end is to have at least one accurate component per seizure. However, the component-based analysis gives a general picture of what the selection process achieved.

Table 5.12 summarizes the results for both seizure- and component-based analysis. The results are presented for each type of resection.

	Type 1	Type 2	Type 2,3	Type 1,2,3
	Seizure-based			
First component is accurate	61.54%	50.00%	45.45%	51.43%
At least one accurate component:				
• within the 3 first components	84.62%	72.22%	63.64%	71.43%
• within the 5 first components	84.62%	88.89%	81.82%	82.86%
• within the accepted components	92.31%	94.44%	95.45%	94.29%
	Component-based			
Total number of accepted components	68.29%	50.30%	47.77%	55.54%
Percentage of accurate components:				
• within all components	38.62%	19.39%	17.33%	25.38%
• within accepted components	40.48%	28.92%	27.46%	33.52%

Table 5.12: Component selection performance

There is approximately a 50% chance that the first component is accurate. In comparison, this probability drops to 16.68% for type 2 and 3 if components are sorted randomly and without rejecting components (the experiment was repeated 100 times) proving that the selection process is efficient. While 50% is not high enough to consider that the component selection is automated, we observe that the probability that a component is accurate within the 5 first components exceeds 80% for any type. Consequently, the process guides the clinician to analyze the first few components rather than ensuring the first component is accurate.

Besides, we observe that there is almost a 95% chance that an accurate component was kept. In terms of number of seizures, only 2 out of 35 seizures (one of type 1 and one of type 2) had no accurate component among the accepted components. This is reasonable considering a patient usually has multiple seizures to analyze. Seen differently, a patient-based analysis that takes into account to which patient is associated which seizure would show that for every patient all the seizures except at most one have an accurate component.

The component-based analysis shows how many components were accepted overall. We observe that the percentage is higher for patients of type 1 than for the other types. This could be explained by the fact that a patient who underwent a hemispherectomy had typically multiple seizure foci dispatched across the pathological hemisphere. Therefore, these multiple concurrent activities are associated to multiple components. Accordingly, it makes total sense that more components are accepted since it is likely that there are more ictal components. For type 2 and 3, approximately 50% of the components are kept. This has a practical value as it means a clinician would have twice less components to analyze.

Figure 5.22 shows in a graphical manner the number of components that are accepted per seizure. For instance, the leftmost bar of the graph means there is one seizure of type 2 for which only two components were accepted. From the figure, it is clear that more components are accepted for type 1 patients than for the two other types.

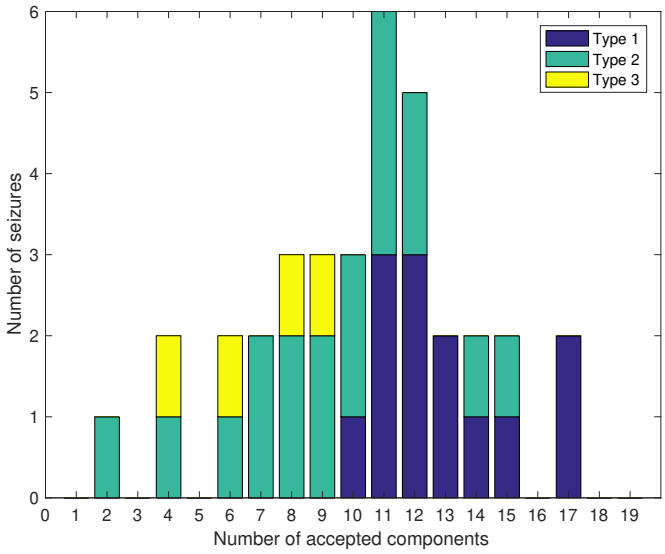


Figure 5.22: Histogram of the number of accepted components per seizure

Another point of the component-based analysis presents what percentage of components was accurate before and after the selection process. Initially, the percentage is the highest for patients of type 1. This is explained both by the extent of the resection zone and the multifocal nature of these seizures. The gain of the selection process is not large for that type of patients. However, there is a substantial increase of the percentage after the selection process for type 2 and 3.

In order to assess the benefits of the sorting alone, Table 5.13 presents the seizure-based analysis if no components were rejected and only the sorting was performed. Interestingly, the chance that the first component is accurate remains the same. This means that components showing a high power in the frequency band of interest are likely to be accepted and even accurate. The benefit of the component rejection appears when we observe that it increases by approximately 10% the chance that at least one component is accurate within the 5 first components for patients of type 2 and 3. Yet the main advantage of the rejection is to reduce the number of components to analyze.

	Type 1	Type 2	Type 2,3	Type 1,2,3
	Seizure-based			
First component is accurate	61.54%	50.00%	45.45%	51.43%
At least one accurate component:				
• within the 3 first components	84.62%	61.11%	54.55%	65.71%
• within the 5 first components	84.62%	77.78%	72.73%	77.14%

Table 5.13: Sorting performance

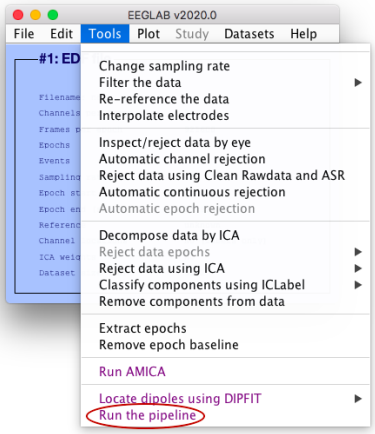
Chapter 6

Practical value

So far, this master thesis focused on the theoretical aspect of the work. However, it is important to note that behind the theory the pipeline was implemented in a way that would encourage clinicians to use it. In this chapter, we first present what an EEGLAB extension applying the pipeline would look like. Finally, we provide some practical examples of seizures on which the pipeline was applied.

6.1 EEGLAB plugin

EEGLAB regroups a variety of so-called *plugins* that were implemented by users and are available to download on the website [50]. These plugins are fairly easy to install facilitating code sharing between users. Our pipeline could easily be integrated into a plugin that would allow users that are not familiar with Matlab scripts to apply the pipeline. More than that, the plugin could provide a good basis for subsequent works related to the combination of ICA and source imaging in the context of estimation of seizure onset zone.



Not much information is needed to run the implemented pipeline. Figure 6.1 shows what would be necessary considering that the EEG data file was imported in EEGLAB and that a time marker was placed at the seizure onset. Leaving blank the last text fields would choose the default values that are set according to what was discussed in this master thesis. However, the user would remain free to make other choices. Once the 'Ok' button is pressed, two windows will appear within less than a minute displaying both the channel and component signals. The user would then be able to select one or more components to localize in brain space. This latter choice would be assisted by the component selection proposed in this master thesis.

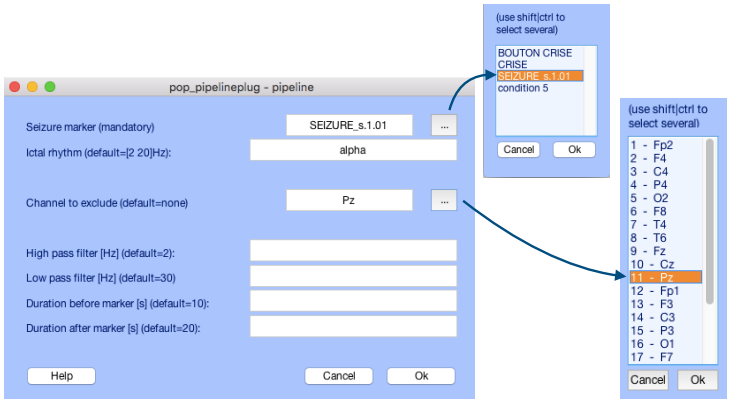


Figure 6.1: Pop-up window for applying the pipeline in EEGLAB

6.2 Examples

In this section, concrete examples of the pipeline application are shown and briefly commented. For each seizure, we show the EEG recording as well as the ICA decomposition. The scale is in seconds and 15 seconds are displayed. Components that are accepted by the selection are displayed in a darker colour. This way, rejected components are not completely thrown away and a clinician could still be able to select one of them to know where it originates from. This could be useful for instance if one of the rejected components isolate seizure propagation or any other phenomenon that might be interesting to localize. As explained in the previous chapter, the accepted components are sorted according to the power in the frequency band of interest.

Patient 2 : seizure S2.02

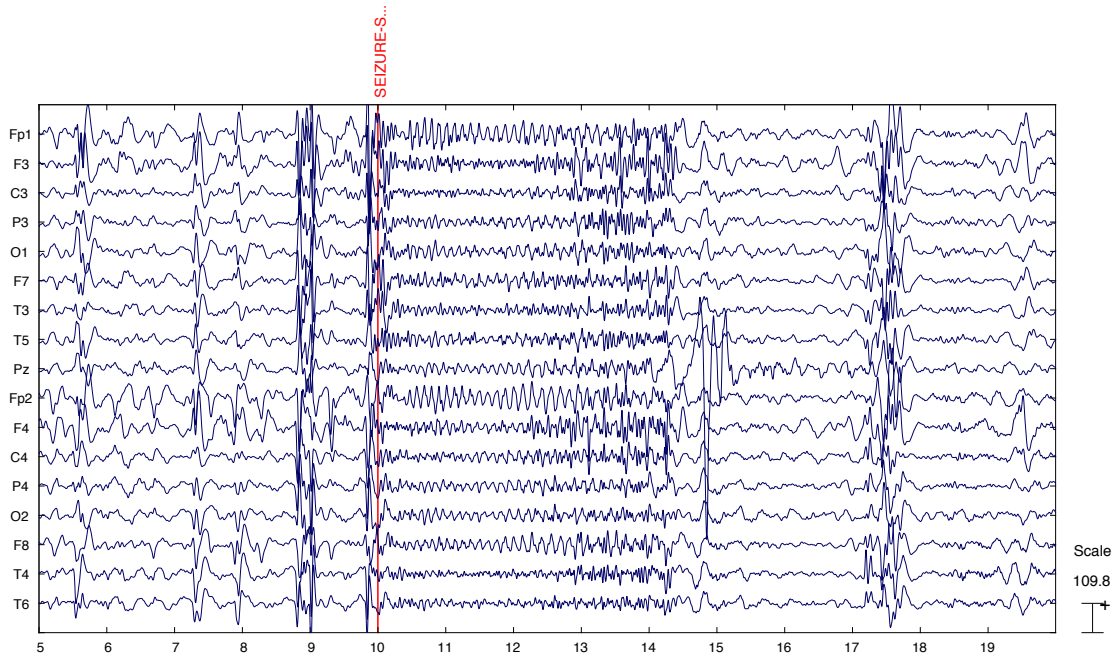


Figure 6.2: Patient 2 - Seizure S2.02 - Channels

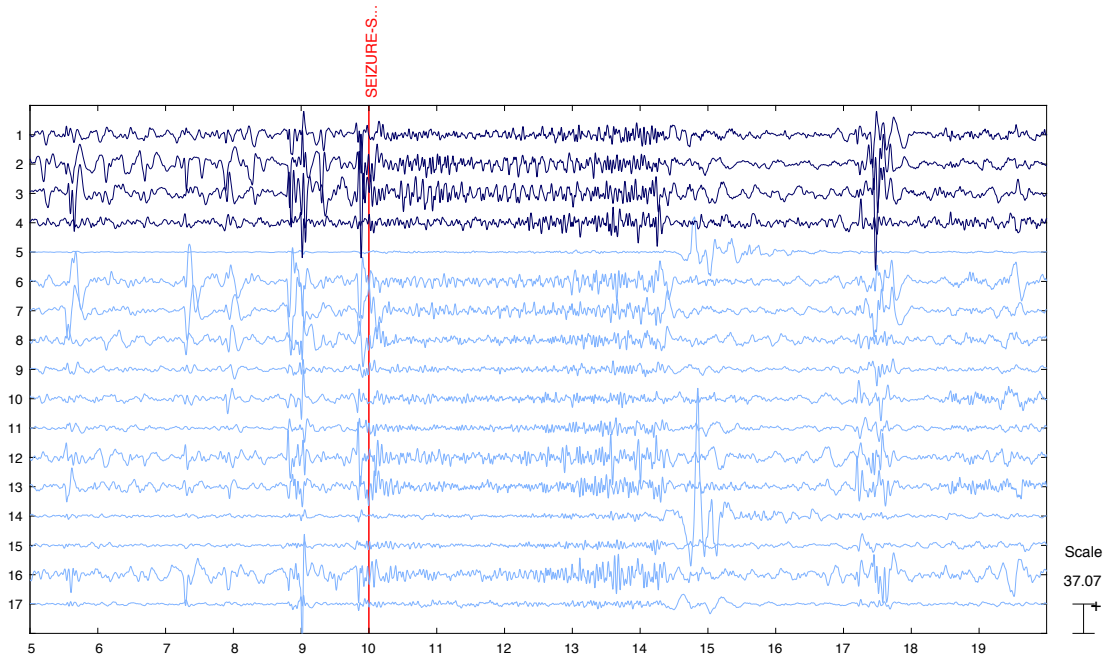


Figure 6.3: Patient 2 - Seizure S2.02 - Components

For the first example, two electrodes (Fz and Pz) were rejected due to loose contact. Electrode Pz was kept despite some artefacts around 15 seconds. The seizure activity is relatively short as it lasts only a few seconds. Figure 6.3 shows the ICA decomposition. We observe that the artefact present in Pz is isolated into two components (5 and 14). Only four components were retained by the component selection and ordered accordingly. The localization of each component is visible in Figures 6.4, 6.5, 6.6 and 6.7. As it might be hard to choose a component among the selected ones, we also show in Figure 6.8 what happens if all the selected components were back-projected. While this last functionality was not assessed in the last chapter and only mentioned in 4.7.3, it remains an interesting feature of the pipeline. The ICA decomposition is quite good since we note that the rhythmic activity appears in a few components only. The localization performance is also satisfying, especially regarding the fact that we rejected the electrode Fz that was close to the resection zone.

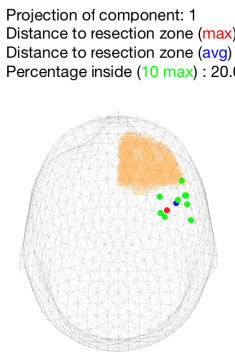


Figure 6.4: Patient 2 - Seizure S2.02 - Component 1

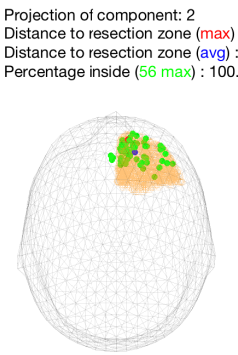


Figure 6.5: Patient 2 - Seizure S2.02 - Component 2

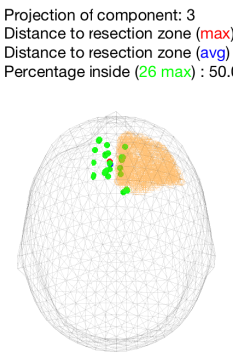


Figure 6.6: Patient 2 - Seizure S2.02 - Component 3

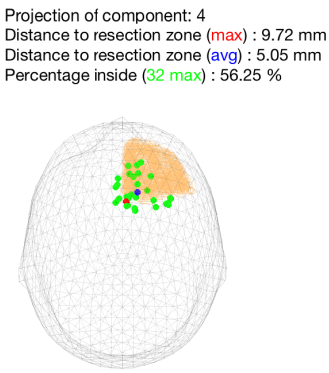


Figure 6.7: Patient 2 - Seizure S2.02 - Component 4

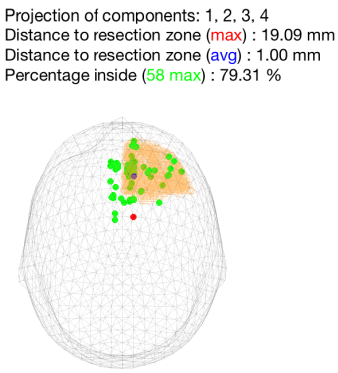


Figure 6.8: Patient 2 - Seizure S2.02 - Components 1 to 4

Patient 3 : seizure S1.01

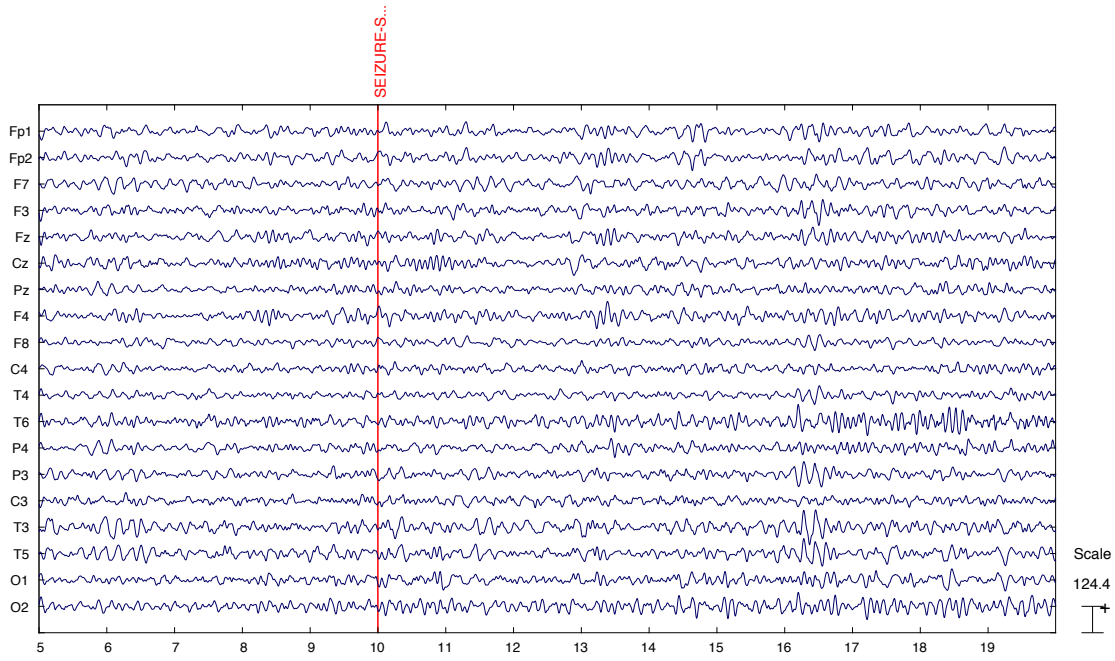


Figure 6.9: Patient 3 - Seizure S1.01 - Channels

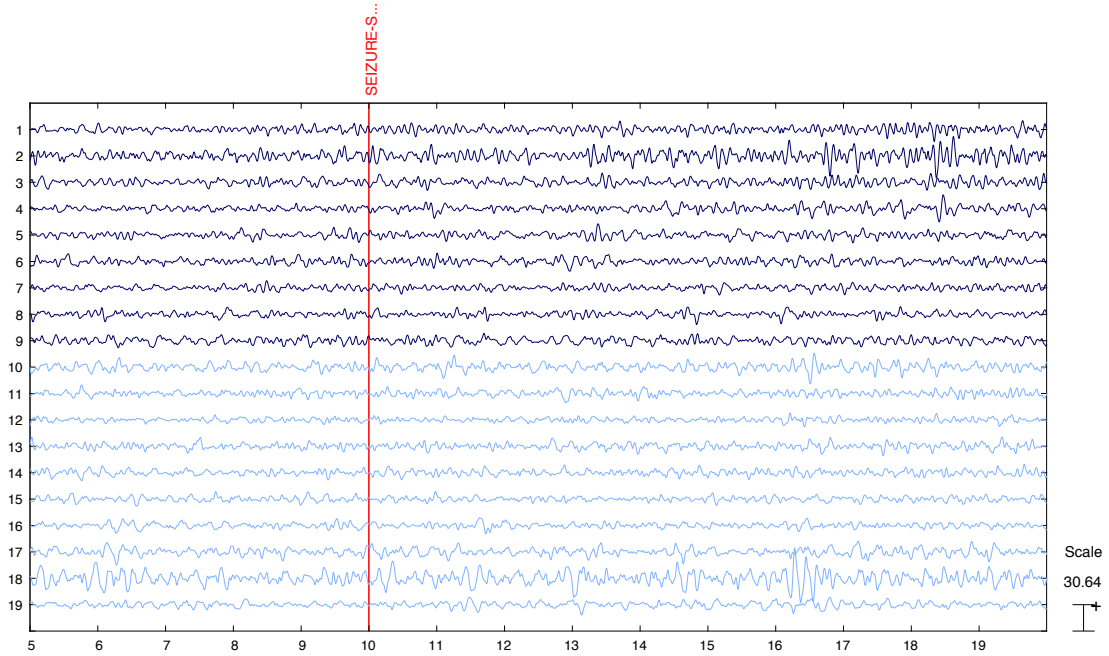


Figure 6.10: Patient 3 - Seizure S1.01 - Components

For this seizure, component 2 draws attention because it shows a typical ictal activity, which is rhythmic and gradually evolving in intensity. As can be seen in Figure 6.11, this component leads indeed to a correct location meaning it probably isolates genuine seizure. However, choosing the first component would also have led to a correct location as it is shown in Figure 6.12. Furthermore, it is interesting to see how ICA isolated slow waves in component 18 while this activity was initially spread across several electrodes.

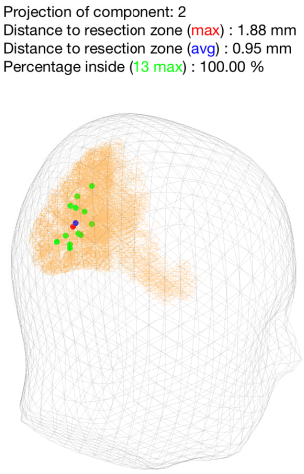


Figure 6.11: Patient 3 - Seizure S1.01 - Component 2

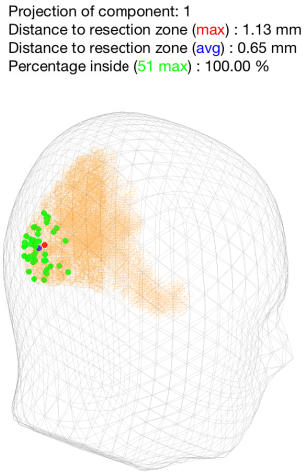


Figure 6.12: Patient 3 - Seizure S1.01 - Component 1

Patient 7 - seizure S1.02

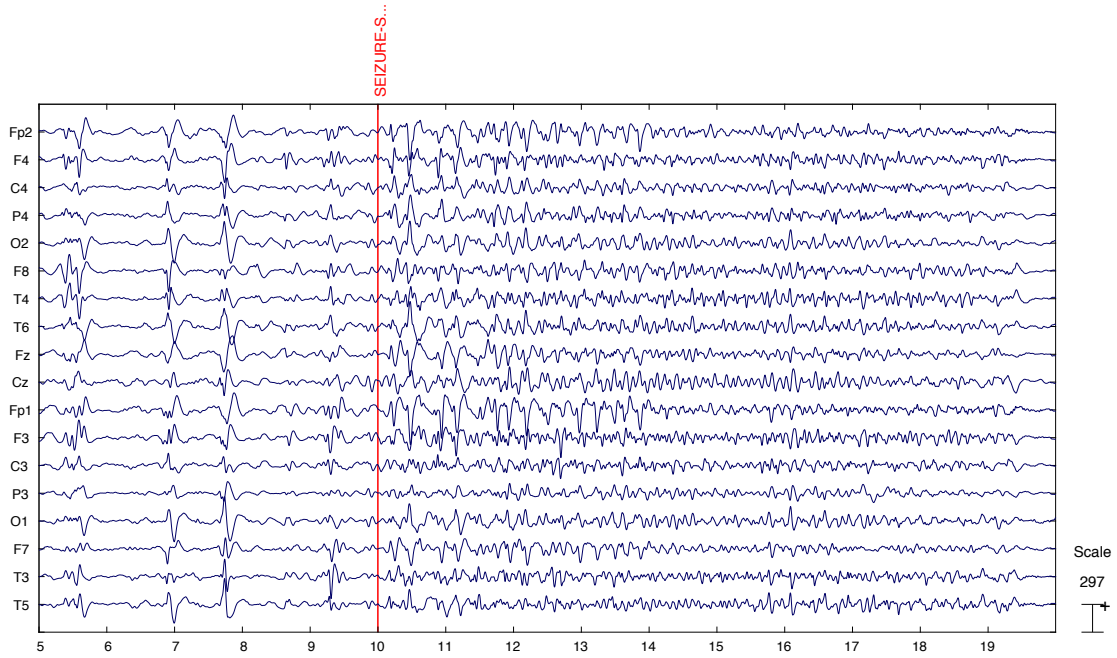


Figure 6.13: Patient 7 - Seizure S1.02 - Channels

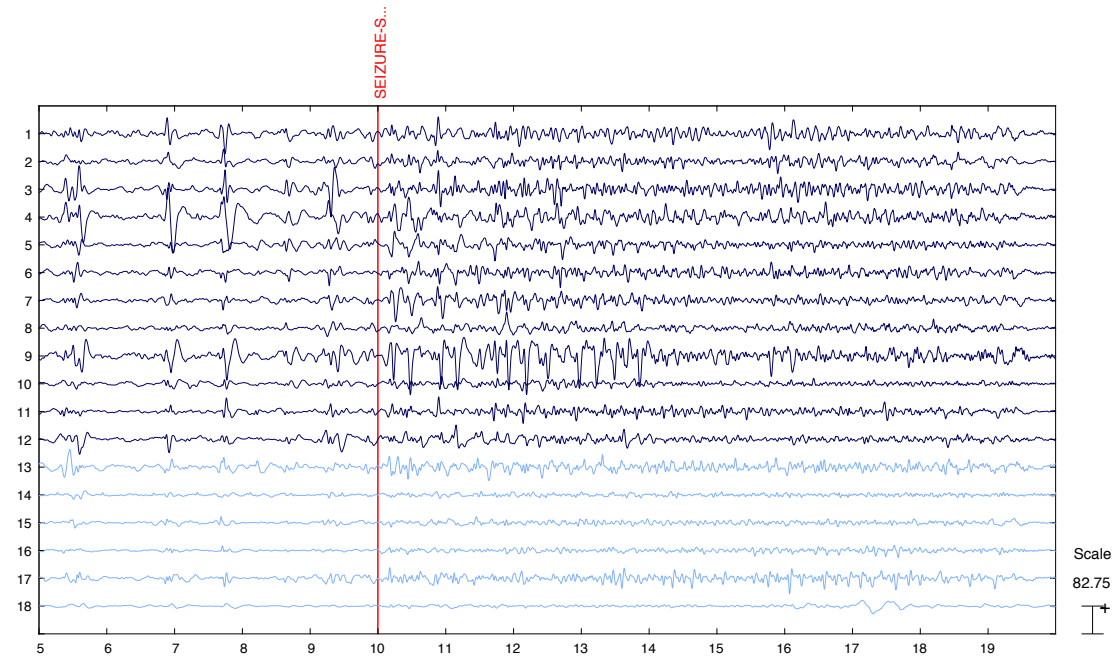


Figure 6.14: Patient 7 - Seizure S1.02 - Components

For this example, component 9 seems particularly interesting and could isolate some abnormal brain activity. Looking at its localization (see Figure 6.15) confirms the intuition that the activity is probably ictal. In contrast, the first-ranked component is located in the left occipital lobe as it is shown in Figure 6.16. The failure to order correctly the

components could be explained by the fact that the frequency band of interest for this seizure is the theta-alpha band and that physiological alpha activity is commonly found in the occipital lobes as a background rhythm. Consequently, the ordering is misled by a concurrent physiological alpha activity. This shows a limitation of the components ordering when an overlap with background physiological rhythms might be suspected. However, visual inspection might have led a clinician to choose component 9 anyway.

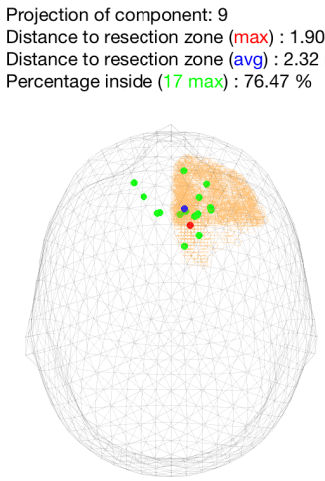


Figure 6.15: Patient 7 - Seizure S1.02 - Component 9

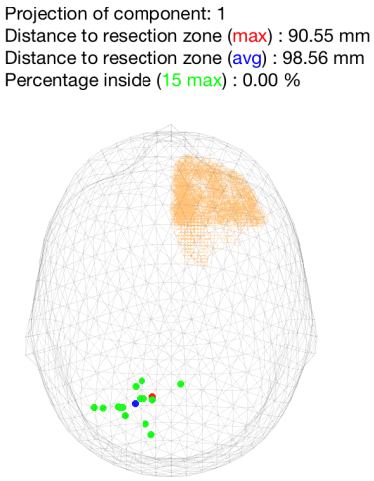


Figure 6.16: Patient 7 - Seizure S1.02 - Component 1

Patient 10 - S1.02

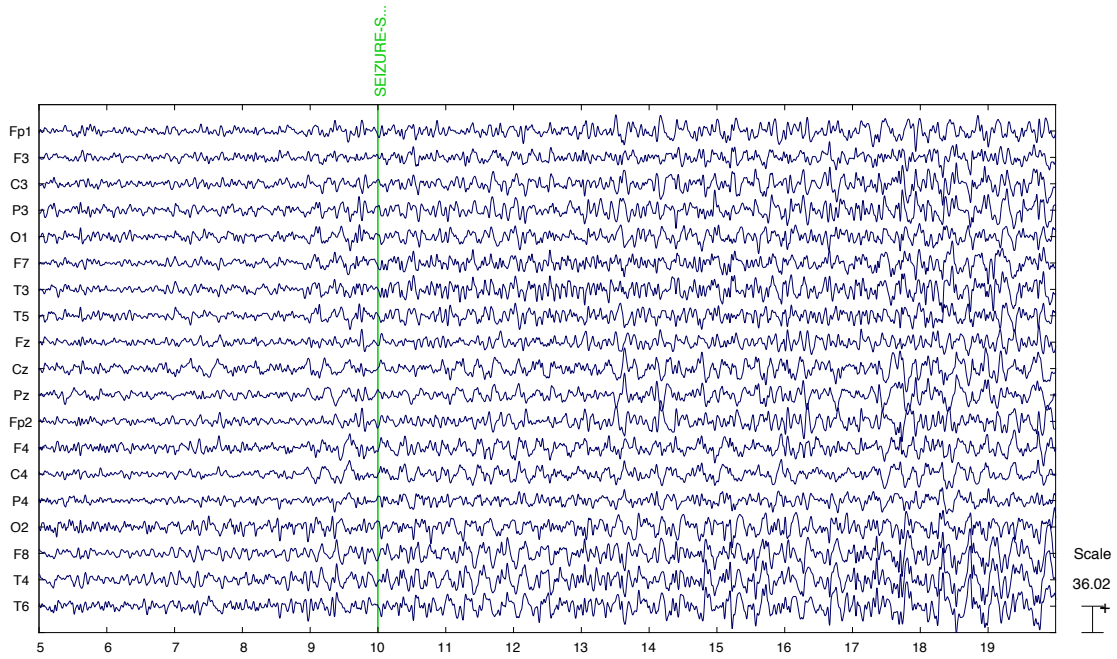


Figure 6.17: Patient 10 - Seizure S1.02 - Channels

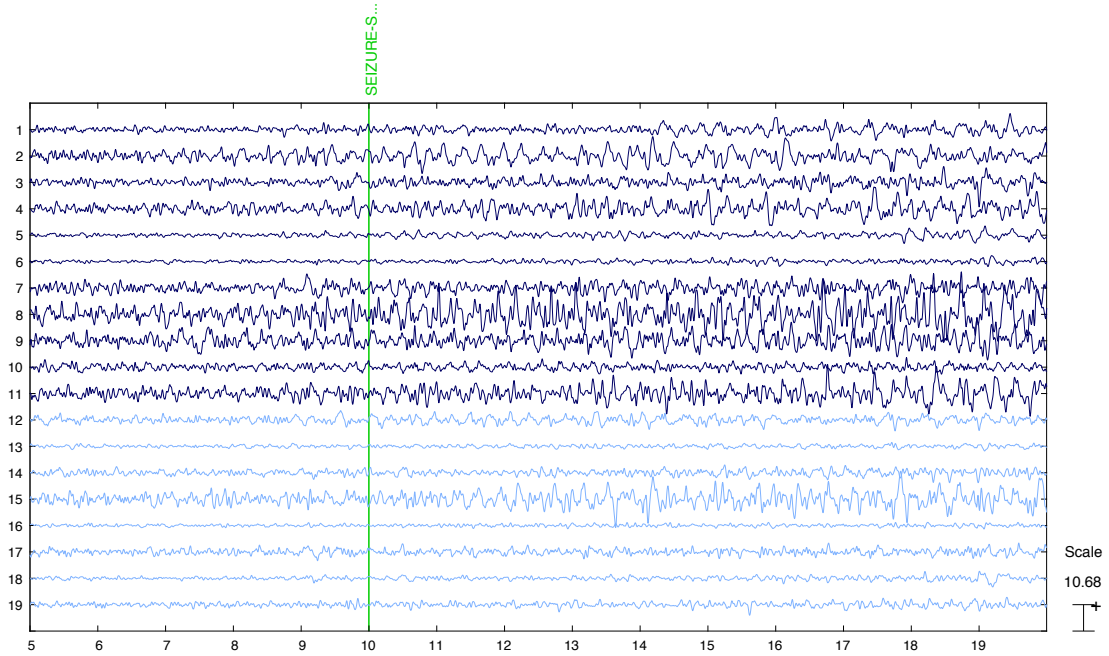


Figure 6.18: Patient 10 - Seizure S1.02 - Components

The ICA decomposition of this seizure is quite interesting. In fact, it might be unclear what rhythm characterized the seizure based on the EEG recording for someone without experience in ictal EEG reading. Nonetheless, the ICA decomposition leads to a component showing a clear rhythm (component 2). Moreover, this component correctly localize the seizure onset as can be seen in Figure 6.19. Finally, we should note that in this case the seizure activity was identified in the delta band and despite the 2 Hz high-pass filter the ICA decomposition achieves a good performance.

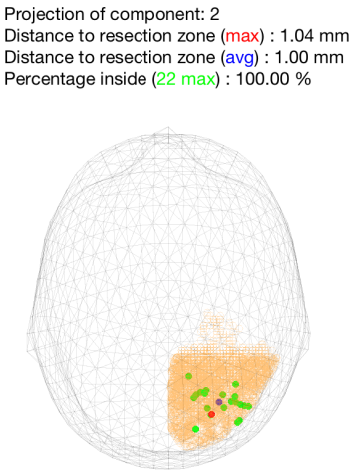


Figure 6.19: Patient 10 - Seizure S1.02 - Component 2

Patient 11 - S1.01

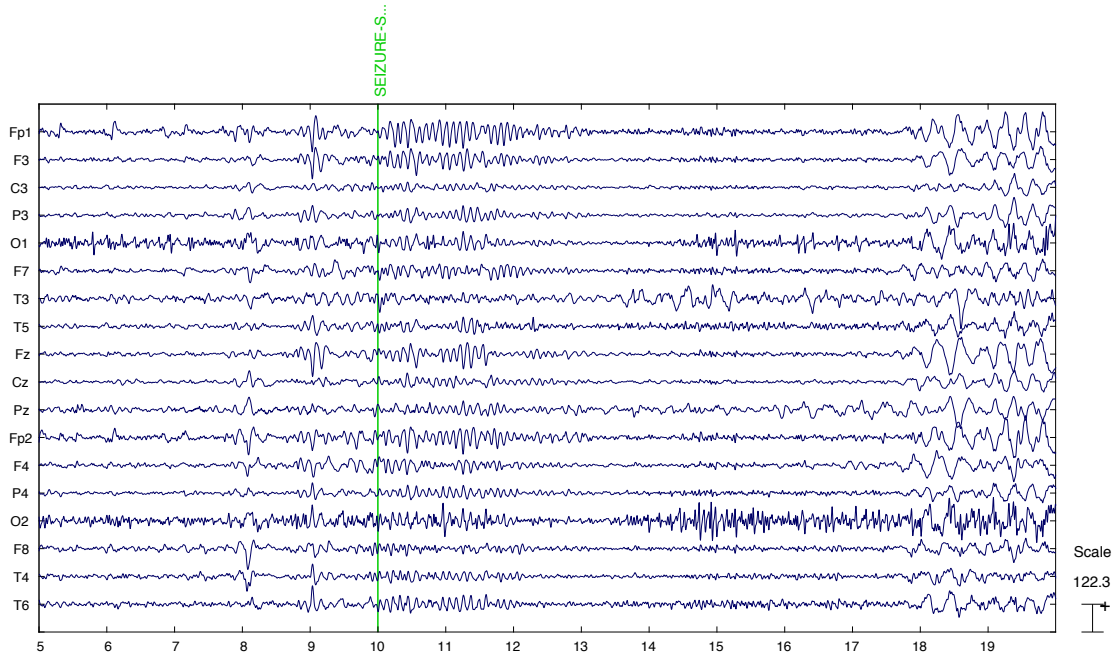


Figure 6.20: Patient 11 - Seizure S1.01 - Channels

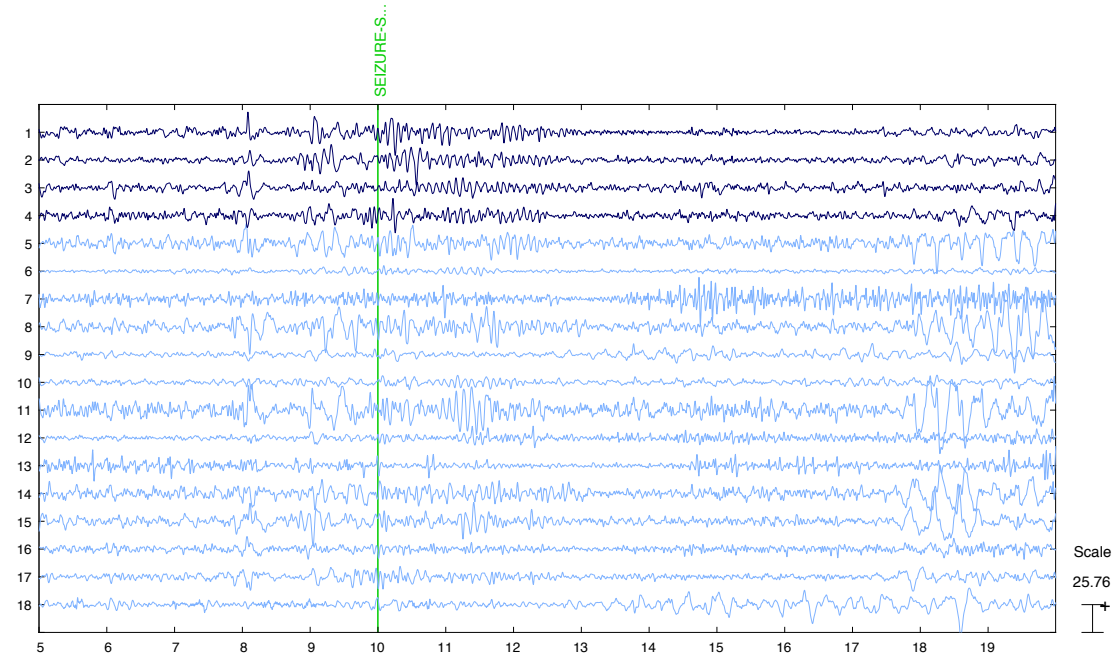


Figure 6.21: Patient 11 - Seizure S1.01 - Components

In this last example, the component selection is particularly efficient since only four components were accepted. The first component is very close to the small region that was removed as it is shown in Figure 6.22. Once again, it might also be interesting to look at what happens if we back-project the activity of the four accepted components together. Figure 6.23 demonstrates that it also correctly localizes the seizure onset.

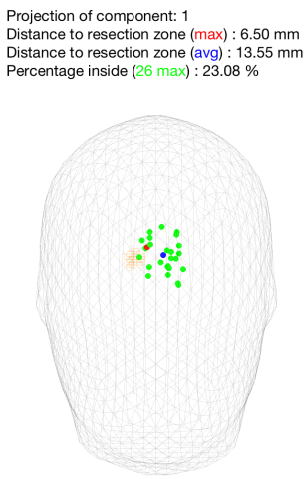


Figure 6.22: Patient 11 - Seizure S1.01 - Component 1

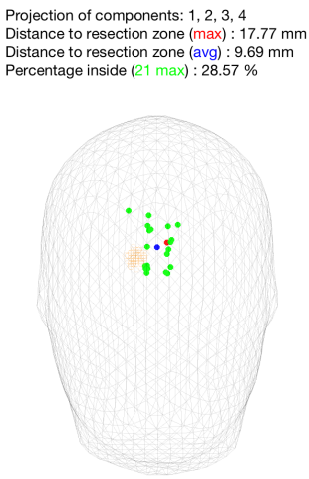


Figure 6.23: Patient 11 - Seizure S1.01 - Component 1 to 4

Chapter 7

Conclusion

This master thesis aimed at estimating seizure onset zones using advanced computational techniques: namely, ICA and source imaging. A pipeline was implemented to reach that goal and structure the work. While each step of the pipeline required to make some choices, we attempted to justify as rigorously as possible any of the decisions. Quantitative measures allowed us to objectively analyze the pipeline performance while many studies in the field of ICA applied to ictal EEG rely on visual inspection to validate their results. This was possible thanks to the co-registration of the pipeline output and the resection zone.

While the pipeline is not completely automated, its use remains practical for clinicians. In a way, it has more freedom than an automated pipeline since the clinician is free to choose the signals he wants to localize that might be something else than pure seizure activity. Indeed, ICA has more to offer than just localizing seizure onset zones. The technique allows to take a new look at EEG signals. In a similar way that time-frequency representations are now commonly used for EEG analysis, ICA offers one more perspective to inspect signals. Through this master thesis, we hope we demonstrated that advanced techniques such as ICA or source imaging do not require a high level of expertise to be used.

We attempted to keep the methodology of our pipeline understandable, regardless of background since this master thesis was the result of an interdisciplinary work between clinicians and engineers. The first step of the pipeline consists in the data preprocessing. We showed that using a 2 Hz high-pass filter might be beneficial when ICA is subsequently applied. Once data is preprocessed, the window to be analyzed must be chosen. We demonstrated that long windows seem preferable. Next, the ICA decomposition is performed and we proposed a process to help the clinicians to choose components isolating seizure activity. This process is based on some component features that were thoroughly analyzed. Each component has a scalp map that gives a first idea of the activation at the scalp level. Finally, to obtain more precision, a source imaging method was employed to localize the activity of the selected components at the brain level.

It is important to note that the pipeline might be improved in many ways. Almost every step could lead to potential future work. First, the preprocessing could be improved by integrating an automatic rejection of corrupted electrodes. Next, the time window to analyze could be automatically detected by incorporating one of the seizure detection models mentioned in section 5.3. The component selection can also definitely be enhanced to increase the chance that it would automatically lead to a correct estimation of the seizure onset zone. This could be done by generating even more features and using some supervised machine learning models such as support vector machines or random forests. While these suggested improvements were out of the scope of this master thesis due to limited time, the implemented pipeline could certainly serves as a basis for these future researches.

Chapter A

Appendix

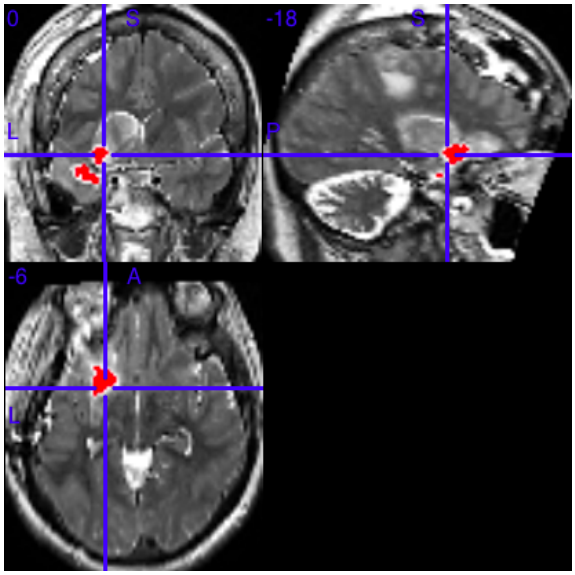


Figure A.1: MRI patient 1 - the resection zone is in red

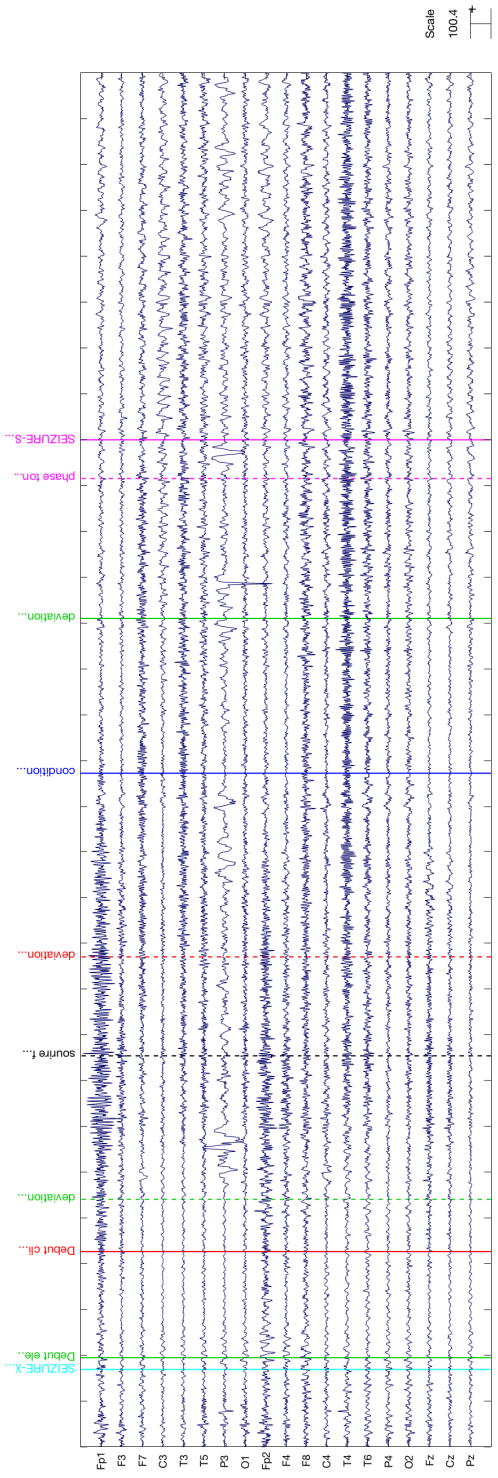


Figure A.2: Patient 1 - seizure X/S1.01

Bibliography

[1] *What is Epilepsy? Disease or Disorder?* URL: https://www.epilepsy.com/learn/about-epilepsy-basics/what-epilepsy (visited on 02/07/2021).

[2] *Epilepsy*. URL: https://www.who.int/news-room/fact-sheets/detail/epilepsy (visited on 02/07/2021).

[3] W. Xue-Ping et al. "Risk factors for drug-resistant epilepsy: A systematic review and meta-analysis". In: *Medicine* 98.30 (July 2019), e16402. ISSN: 0025-7974. DOI: 10.1097/MD.0000000000016402.

[4] P. Ryvlin and S. Rheims. "Epilepsy surgery: eligibility criteria and presurgical evaluation". In: *Dialogues in Clinical Neuroscience* 10.1 (Mar. 2008), pp. 91–103. ISSN: 26083477, 12948322. DOI: 10.31887/DCNS.2008.10.1/pryvlin.

[5] F. Rosenow. "Presurgical evaluation of epilepsy". In: *Brain* 124.9 (Sept. 2001), pp. 1683–1700. ISSN: 14602156. DOI: 10.1093/brain/124.9.1683.

[6] S. Noachtar and J. Rémi. "The role of EEG in epilepsy: A critical review". In: *Epilepsy & Behavior* 15.1 (May 2009), pp. 22–33. ISSN: 15255050. DOI: 10.1016/j.yebeh.2009.02.035.

[7] A. G. Baroumand et al. "Automated EEG source imaging: A retrospective, blinded clinical validation study". In: *Clinical Neurophysiology* 129.11 (Nov. 2018), pp. 2403–2410. ISSN: 13882457. DOI: 10.1016/j.clinph.2018.09.015.

[8] W. Staljanssens et al. "EEG source connectivity to localize the seizure onset zone in patients with drug resistant epilepsy". In: *NeuroImage: Clinical* 16 (2017), pp. 689–698. ISSN: 22131582. DOI: 10.1016/j.nicl.2017.09.011.

[9] S. Vespa et al. "Ictal EEG source imaging and connectivity to localize the seizure onset zone in extratemporal lobe epilepsy". In: *Seizure* 78 (May 2020), pp. 18–30. ISSN: 10591311. DOI: 10.1016/j.seizure.2020.03.001.

[10] Lisha Sun, Ying Liu, and P. Beadle. "Independent component analysis of EEG signals". In: *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005.* Suzhou, China: IEEE, 2005, pp. 219–222. ISBN: 9780780390058. DOI: 10.1109/IWVDVT.2005.1504590.

[11] R. Vigario et al. “Independent component approach to the analysis of EEG and MEG recordings”. In: *IEEE Transactions on Biomedical Engineering* 47.5 (May 2000), pp. 589–593. ISSN: 00189294. DOI: 10.1109/10.841330.

[12] C. M. Michel and D. Brunet. “EEG Source Imaging: A Practical Review of the Analysis Steps”. In: *Frontiers in Neurology* 10 (Apr. 2019), p. 325. ISSN: 1664-2295. DOI: 10.3389/fneur.2019.00325.

[13] H. Nam et al. “Independent Component Analysis of Ictal EEG in Medial Temporal Lobe Epilepsy”. In: *Epilepsia* 43.2 (Mar. 2002), pp. 160–164. ISSN: 00139580, 15281167. DOI: 10.1046/j.1528-1157.2002.23501.x.

[14] E. Urrestarazu et al. “Independent Component Analysis Removing Artifacts in Ictal Recordings”. In: *Epilepsia* 45.9 (Sept. 2004), pp. 1071–1078. ISSN: 0013-9580, 1528-1167. DOI: 10.1111/j.0013-9580.2004.12104.x.

[15] P. LeVan, E. Urrestarazu, and J. Gotman. “A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification”. In: *Clinical Neurophysiology* 117.4 (Apr. 2006), pp. 912–927. ISSN: 13882457. DOI: 10.1016/j.clinph.2005.12.013.

[16] J. Iriarte et al. “Independent Component Analysis in the Study of Focal Seizures:” in: *Journal of Clinical Neurophysiology* 23.6 (Dec. 2006), pp. 551–558. ISSN: 0736-0258. DOI: 10.1097/01.wnp.0000236579.08698.23.

[17] A. J. Leal, A. I. Dias, and J. P. Vieira. “Analysis of the EEG dynamics of epileptic activity in gelastic seizures using decomposition in independent components”. In: *Clinical Neurophysiology* 117.7 (July 2006), pp. 1595–1601. ISSN: 13882457. DOI: 10.1016/j.clinph.2006.03.020.

[18] K.-Y. Jung et al. “Spatiotemporospectral characteristics of scalp ictal EEG in mesial temporal lobe epilepsy with hippocampal sclerosis”. In: *Brain Research* 1287 (Sept. 2009), pp. 206–219. ISSN: 00068993. DOI: 10.1016/j.brainres.2009.06.071.

[19] A. J. Leal et al. “Analysis of the dynamics and origin of epileptic activity in patients with tuberous sclerosis evaluated for surgery of epilepsy”. In: *Clinical Neurophysiology* 119.4 (Apr. 2008), pp. 853–861. ISSN: 13882457. DOI: 10.1016/j.clinph.2007.11.176.

[20] L. Yang et al. “Dynamic imaging of ictal oscillations using non-invasive high-resolution EEG”. In: *NeuroImage* 56.4 (June 2011), pp. 1908–1917. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.03.043.

[21] Y. Lu et al. “Dynamic imaging of seizure activity in pediatric epilepsy patients”. In: *Clinical Neurophysiology* 123.11 (Nov. 2012), pp. 2122–2129. ISSN: 13882457. DOI: 10.1016/j.clinph.2012.04.021.

[22] M. A. Habib et al. “Recursive independent component analysis (ICA)-decomposition of ictal EEG to select the best ictal component for EEG source imaging”. In: *Clinical Neurophysiology* 131.3 (Mar. 2020), pp. 642–654. ISSN: 13882457. DOI: 10.1016/j.clinph.2019.11.058.

[23] H. G. Wieser et al. "Proposal for a New Classification of Outcome with Respect to Epileptic Seizures Following Epilepsy Surgery". In: *Epilepsia* 42 (May 2003), pp. 282–286. ISSN: 00139580, 15281167. DOI: 10.1046/j.1528-1157.2001.4220282.x.

[24] *EEGLAB*. URL: https://sccn.ucsd.edu/eeglab/index.php (visited on 02/07/2021).

[25] *SPM - Statistical Parametric Mapping*. URL: https://www.fil.ion.ucl.ac.uk/spm/ (visited on 03/11/2021).

[26] *NITRC: MRIcron: Tool/Resource Info*. URL: https://www.nitrc.org/projects/mricron (visited on 03/11/2021).

[27] I. Winkler et al. "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan: IEEE, Aug. 2015, pp. 4101–4105. ISBN: 9781424492718. DOI: 10.1109/EMBC.2015.7319296.

[28] A. Delorme, T. Sejnowski, and S. Makeig. "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis". In: *NeuroImage* 34.4 (Feb. 2007), pp. 1443–1449. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2006.11.004.

[29] L. Leuchs. "Choosing your reference – and why it matters". In: *Brain Products Press Release* (May 2019). URL: https://www.brainproducts.com/files/public/products/brochures_material/pr_articles/1901_Referencing.pdf.

[30] O. Bertrand, F. Perrin, and J. Pernier. "A theoretical justification of the average reference in topographic evoked potential studies". In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 62.6 (Nov. 1985), pp. 462–464. ISSN: 01685597. DOI: 10.1016/0168-5597(85)90058-9.

[31] *Makoto's preprocessing pipeline*. URL: https://sccn.ucsd.edu/wiki/Makoto%5C%27s_preprocessing_pipeline (visited on 04/05/2021).

[32] M. K. Siddiqui et al. "A review of epileptic seizure detection using machine learning classifiers". In: *Brain Informatics* 7.1 (Dec. 2020), p. 5. ISSN: 2198-4018, 2198-4026. DOI: 10.1186/s40708-020-00105-1.

[33] J. Onton et al. "Imaging human EEG dynamics using independent component analysis". In: *Neuroscience & Biobehavioral Reviews* 30.6 (Jan. 2006), pp. 808–822. ISSN: 01497634. DOI: 10.1016/j.neubiorev.2006.06.007.

[34] P. Comon. "Independent component analysis, A new concept?" In: *Signal Processing* 36.3 (1994). Higher Order Statistics, pp. 287–314. ISSN: 0165-1684. DOI: https://doi.org/10.1016/0165-1684(94)90029-9.

[35] M. Ullsperger and S. Debener, eds. *Simultaneous EEG and fMRI: recording, analysis, and application*. OCLC: ocn428770799. New York: Oxford University Press, 2010. ISBN: 9780195372731.

[36] J. Iriarte et al. "Independent Component Analysis as a Tool to Eliminate Artifacts in EEG: A Quantitative Study:" in: *Journal of Clinical Neurophysiology* 20.4 (July 2003), pp. 249–257. ISSN: 0736-0258. DOI: 10.1097/00004691-200307000-00004.

[37] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications". en. In: *Neural Networks* 13.4-5 (June 2000), pp. 411–430. ISSN: 08936080. DOI: 10.1016/S0893-6080(00)00026-5.

[38] A. J. Bell and T. J. Sejnowski. "An Information-Maximization Approach to Blind Separation and Blind Deconvolution". In: *Neural Computation* 7.6 (Nov. 1995), pp. 1129–1159. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco.1995.7.6.1129.

[39] T.-W. Lee, M. Girolami, and T. J. Sejnowski. "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources". In: *Neural Computation* 11.2 (Feb. 1999), pp. 417–441. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/089976699300016719.

[40] R. D. Pascual-Marqui. *Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization.* 2007. arXiv: 0710.3341 [math-ph].

[41] R. D. Pascual-Marqui et al. "Assessing interactions in the brain with exact low-resolution electromagnetic tomography". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1952 (Oct. 2011), pp. 3768–3784. ISSN: 1364-503X, 1471-2962. DOI: 10.1098/rsta.2011.0081.

[42] R. Pascual-Marqui, C. Michel, and D. Lehmann. "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain". In: *International Journal of Psychophysiology* 18.1 (Oct. 1994), pp. 49–65. ISSN: 01678760. DOI: 10.1016/0167-8760(84)90014-X.

[43] Y. Huang, L. C. Parra, and S. Haufe. "The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting". In: *NeuroImage* 140 (Oct. 2016), pp. 150–162. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2015.12.019.

[44] C. J. Holmes et al. "Enhancement of MR Images Using Registration for Signal Averaging:" in: *Journal of Computer Assisted Tomography* 22.2 (Mar. 1998), pp. 324–333. ISSN: 0363-8715. DOI: 10.1097/00004728-199803000-00032.

[45] *A short history of stereotaxic data volumes at the MNI.* URL: http://www.bic.mni.mcgill.ca/~louis/stx_history.html (visited on 06/01/2021).

[46] A. Anzolin et al. "Quantifying the Effect of Demixing Approaches on Directed Connectivity Estimated Between Reconstructed EEG Sources". In: *Brain Topography* 32.4 (July 2019), pp. 655–674. ISSN: 0896-0267, 1573-6792. DOI: 10.1007/s10548-019-00705-z.

[47] P. Presti. *paolop21/simulation_source_connectivity.* original-date: 2018-01-19T17:12:41Z. July 2018. URL: https://github.com/paolop21/simulation_source_connectivity (visited on 05/17/2021).

[48] *Visualize summary statistics with box plot - MATLAB boxplot - MathWorks Benelux.* URL: https://nl.mathworks.com/help/stats/boxplot.html (visited on 05/23/2021).

[49] P. Fergus et al. "Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques". In: *BioMed Research International* 2015 (2015), pp. 1–17. ISSN: 2314-6133, 2314-6141. DOI: 10.1155/2015/986736.

[50] *List of plug-ins available for download in EEGLAB 2019.1 and later versions*. URL: https://sccn.ucsd.edu/eeglab/plugin_uploader/plugin_list_all.php (visited on 05/17/2021).

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**