Ludwig-Maximilians Universität München
Institut für Statistik

Master thesis

Causal Analysis on SEEG data

Author:
Tobias Riebe

Supervisor:
Prof. Dr. Moritz Grosse-Wentrup

August 6, 2019

Contents

Abstract

Interictal spikes are typical for patients suffering from epilepsy. The relationship between interictal spikes and epileptic seizures is still unclear. In the following thesis I will focus on analysing the patterns of interictal spikes. This might help to understand better how spikes propagate within the brain and might thereby be helpful to localise the epileptogenic zone, what could improve the treatment of patients with epilepsy. The brain activity, was recorded by depth electrodes. I will use Independent Component Analysis (ICA) on the recorded data to unmix the signals and thereby find the true sources of the signals. To examine if and how the spikes from the different components propagate, a causal analysis is conducted to detect possible spatiotemporal properties of interictal spikes. The analysis is conducted on data of two epilepsy patients with implanted depth electrodes. For both patients causal relations between the independent components were found and presented as Directed Acyclic Graphs (DAGs).

1 Introduction

1.1 Epilepsy

Epilepsy is a neurological disorder with a wide range of clinical appearance. It is characterised by epileptic seizures, which are temporary behavioural changes, that can be objective indications or subjective symptoms. Epileptic seizures can be caused by excessive or synchronous neuronal activity in the brain. The onset of such seizures can arise in one or more localised brain regions or in one hemisphere or can be widely spread over both hemispheres.
Epilepsy is defined as having two or more unprovoked seizures. Almost any insult that disturbs the brain activity can result in a seizure, including acquired causes (for example after a stroke), infectious diseases, autoimmune diseases and genetic mutations. 10 percent of all people will have a seizure during their life time and more than 50 million people worldwide are affected by epilepsy. [3]
The first-line treatment for epilepsy are anti-seizure drugs. Even tough there is a wide range of approved drugs on the market, one third of the patients still fail to achieve seizure control. For these patients the most promising chance to become seizure free is by undergoing an epilepsy surgery, though only a small number of patients is eligible for such a procedure. [5] The success rates of a surgery largely depends on what is measured as success and how long the operated patients are observed after the surgery. One estimation shows that around 70% of the patients who have undergone a temporal lobe surgery report that the surgery stopped their seizures and that they became seizure-free. For a further 20% their seizures have been reduced. About 50% of the persons who have temporal lobe surgery are still seizure-free 10 years after their surgery, but most of these people still take their medications for some time [2].
The aim of a surgery is the resection or complete disconnection of the epileptogenic zone, which is a hypothetical construct generated by different results of investigations. It is defined as 'the area of cortex that is necessary and sufficient for initiating seizures and whose removal (or disconnection) is necessary for complete abolition of seizures' [6]. Before the surgery can be conducted, different zones and lesions in the brain of the patient have to be identified that have a variable spatial relationship with the epileptogenic zone. These zones are (see also figure 1): [17]

- the **irritative zone** that generates interictal spikes
- the **seizure onset** or **pacemaker zone** from which seizures arise
- the **ictal symptomatogenic zone** producing ictal symptoms
- the **epileptogenic lesion** as the brain area that initiates seizure disorder
- the **functional deficit zone** as area of cortex that is not functioning normally in the interictal period

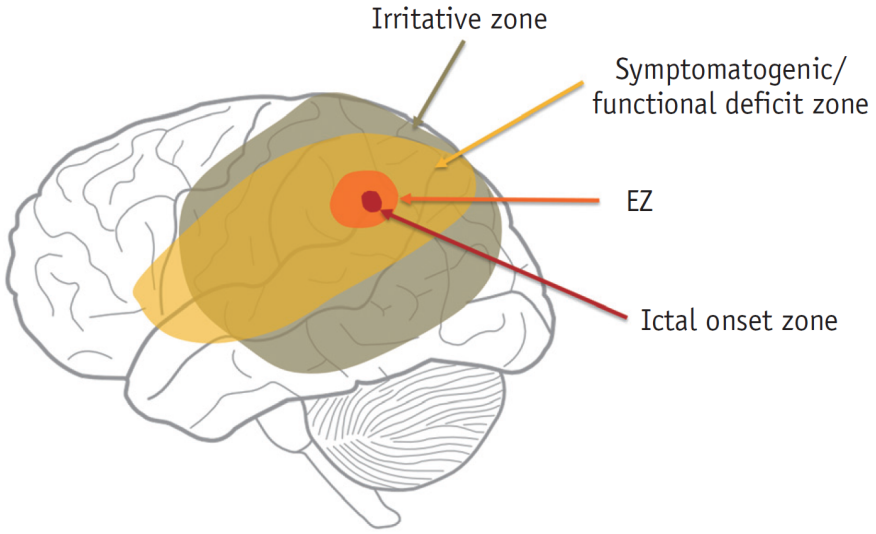

Figure 1: Schematic illustration of areas related to or affected by epileptic discharges. Image taken from [14].

Epileptologists use various non-invasive diagnostic tools like analysis of seizure semiology, electrophysiological recordings, functional testing and neuroimaging techniques to define these five zones. The location and boundaries of the epileptogenic zone is then determined with this gained information. [12]
A problem in the determination of the epileptogenic zone is that it is not possible to define the zones mentioned above with 100% precision, because all the methods used to measure the zones can be faulty. But not only the practical limitations are an issue. It is also not completely clear how all the different zones are associated with each other. In general the seizure onset zone is a subset of the irritative zone, but there is no proof that it has to be. There is evidence that the irritative zone is more extensive than the epileptogenic zone. That means that even the most precise mapping of the irritative zone can be misleading and points to regions that are not included in the epileptogenic zone. The symptomatogenic zone rarely overlaps with the seizure onset zone. But as the seizure originates from the seizure onset zone and then spreads to the eloquent cortex it can be said that the seizure onset zone is in the neighbourhood of the symptomatogenic zone. It still can not localise the exact seizure onset zone or give an indication of the extent of it. [12]

1.2 Propagation of spikes

One example of a typical non-invasive electrophysiological recording to localise the epileptogenic zone mentioned in the section above, is the electroencephalogram (EEG), which records electrical activity in the brain with the electrodes placed along the scalp. When looking on EEG recordings, different types of brain waves can be seen. The brain waves are named according to their frequency. Alpha waves have a frequency of 8 to 13 waves per second and can typically be seen in adults who are relaxed with their eyes closed. When people are awake, beta waves can be observed with frequencies greater than 13 waves per second. Theta waves are also called slow activity and occur while adults are sleeping or in young children (4 to 8 Hz). The slowest type of waves are Delta waves (up to 4 Hz). They appear in children under one year and can also be found in some parts of sleep. Very fast waves are called *spikes*. They clearly stand out from other brain activity and last less than 80 milliseconds [1]. Examples of brainwaves and epileptic spikes can be seen on figure 2.

Spikes that occur between spontaneous recurrent seizures in epileptic patients are called interictal spikes and are one type of interictal epileptiform discharges (IED) [9]. Interictal spikes are typical for epilepsy patients and a lot of research was conducted to understand if and how spikes could potentially influence epileptic seizures and to comprehend how spikes propagate. It is important to mention that 10% of epilepsy patients never show IEDs. Furthermore there are persons that show IEDs, but never develop epilepsy [20]. However, as the majority of epilepsy patients shows interictal spikes, the question of their potential role in epileptic patients remains an interesting question.

One study that aims to find a relationship between spikes and seizures is from Karoly et al. [10], which showed that the spike rate significantly changed prior to seizures in over half of the subjects (for 9 out of 15 subjects) in their study. They linked increased pre-ictal spike rate to improved predictability of an epileptic seizure, but they even so showed that the spike rate also decreased before seizures (in 6 out of the 9 subjects with significantly changed spike rate). Another trial showed that high or low spiking rates do not influence the occurrence of seizures, but increase after them [7]. One further proposal is that interictal spikes and epileptic seizures are not independent events, but the relationship may vary with other influences such as the onset location [4]. Tomlinson et al. [24] showed that the propagation of spikes is more spatially organised in patients, that become seizure free after a surgery than it is the case for patients, that do not become seizure free. With the results of their study they claimed that spike propagation is a valuable and potentially under-utilised tool for assessing the network configuration of the epileptogenic zone. It could therefore be useful to comprehend where and how interictal spikes arise and how they propagate across the brain areas to understand the relation between spikes and seizures better. Sabolek et al. [18] described the propagation of interictal spikes by 3D trajectories with data from an implanted grid of electrodes. They found that paths by which synchronous activity spreads through an epileptic network change with each activation through multiple pathways. In general, network

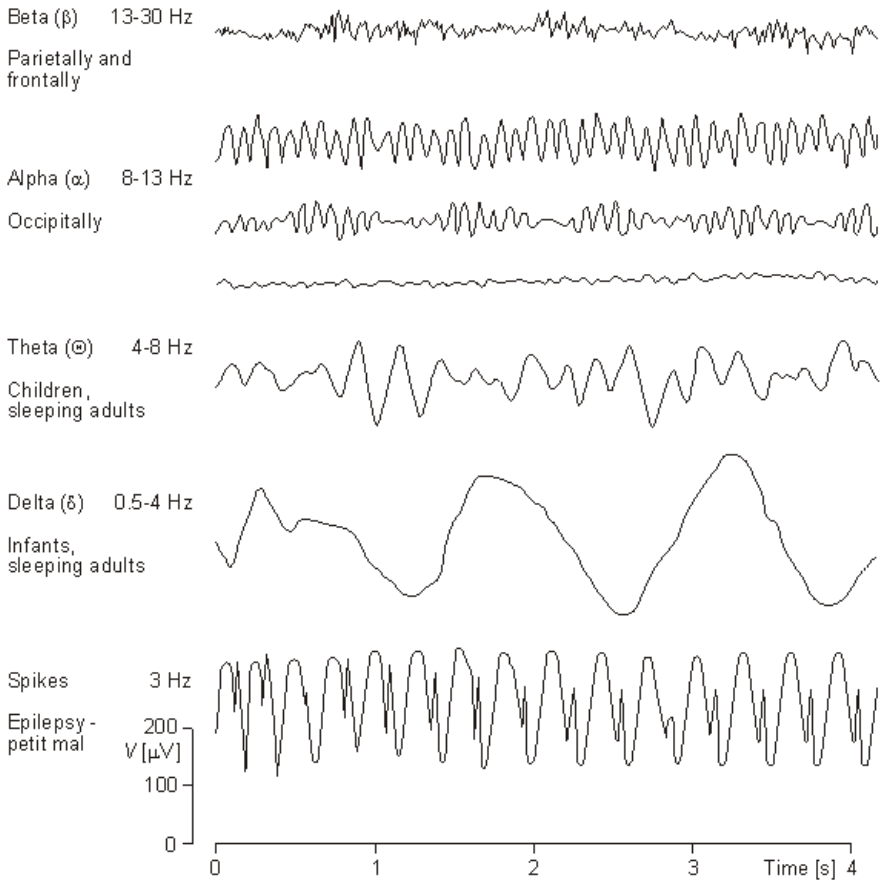

Figure 2: Examples of the different brain waves and epileptic spikes. Image taken from [13].

analysis of EEG recordings got more and more attention in the recent years [27] [26] and produced promising results with respect to find the epileptogenic zone. Also principles from graph theory and network science were used to detect the location of the epileptogenic zone [15]. Van Mierlo et al. [25] recorded signals from depth electrodes and designed such a network, which was based on Granger causality. For all of the eight patients they tested, they were able to detect an electrode, which was among the electrode contacts, that showed the ictal onset on the recordings, which was visually confirmed by an epileptologist.

1.3 Motivation

The approaches described above gave interesting insights in the propagation of spikes and the relation between interictal spikes and seizures, but they suffer from different shortfalls. Some of them work with electrodes that are only able to record the brain activity in the upper layer and therefore brain activity in deeper areas stays uncovered. To tackle these drawback the data analysed in this thesis is recorded by depth electrodes instead of grids of electrodes, as they give highly accurate sampling from sulcal areas and deep brain structures. Some of the studies that were described above used depth electrodes to record brain activity in deeper brain areas as well but these approaches suffered from a poor spatial resolution [26] [25]. It is very challenging to implant the depth electrodes at the locations where the spikes are generated and thus the electrodes might fail to record the spike patterns at the true sources. To overcome the problem of poor spatial resolution, the *Independent Component Analysis* will be used in this thesis to unmix the recorded signals and to therefore recover the true sources of the spikes. Thereby spike patterns might be detected at locations, that were not covered by the recordings of the electrodes. Analysing the data set generated by the *Independent Component Analysis* might give new insights into the patterns of spike propagation. In section 4.1 the mathematical background of the *Independent Component Analysis* and how it was applied to the underlying data will be described.
Causal Analyses will be used to analyse how the spikes propagate. It might be interesting to use this class of models because it is not clear if and how IEDs that appear at different locations in the brain interact with each other. Interictal spikes have the property that they might propagate from one brain area to another within milliseconds. It might as well be that a spike is composed of different components and that the sum of such components produces a spike. As the goal of *Causal Analysis* is to find cause-effect relationships between different variables they might help to reveal spatiotemporal patterns of interictal spikes. Combining *Independent Component Analysis* and *Causal Analysis* could therefore be an attractive approach to analyse and visualise the relationship of hidden components in SEEG data to better understand the propagation of spikes and therefore be able to define the epileptogenic zone more accurately. A short introduction into causal inference and the application on the data will be described in section 4.2.

The data used for this thesis will be presented in section 2. Since it was necessary to detect spikes to be able to carry out valid analyses, an algorithm for spike detection was created and will be shown in chapter 3. Chapter 5 presents the results of the analysis. The thesis ends with a critical discussion about the work and future steps that can be taken. If appropriated, the sections were split into subsections relating to the patients.

2 The data

As mentioned in the chapters above there are invasive and non-invasive methods to record the brain electrical activity. For non-invasive methods no surgery is required in contrast to invasive methods where a surgery is needed. An example for a non-invasive method is the EEG (electroencephalogram). In this case the electrodes are placed along the scalp. EEG is a fast and cheap technique, but it also has disadvantages. One drawback is that the signal picked up is distorted and its amplitude is reduced because the electrical activity is recorded far away from the source. Another downside is that EEG is often contaminated with artefacts. In order to overcome these disadvantages invasive methods could be used, where the electrodes have to be implanted into the patient. As such an implantation is carried out locally and not over the entire brain region the epileptogenic zone, described in chapter 1.1, has to be detected to implant the electrodes close to that zone. One example for an invasive method is the Electrocorticography (ECoG), where a grid of electrodes is implanted in the patient, directly on the exposed surface of the brain. Stereoelectroencephalography (SEEG) is another invasive method of recording electrical activity. SEEG places an electrode at some depth inside the brain and has multiple contacts on it to record the signals in different depths. So it can reach regions ECoG cannot reach.
For this thesis SEEG data sets of two patients with epilepsy were examined. In the following sections the patients and the corresponding data will be described for both patients.

2.1 Patient 1

Patient 1 was male and 23 years old. 11 depth electrodes were implanted in the patient with a total of 125 contacts. The sampling rate of the recorded data was 1024 Hz and four data sets each with a length of 2 hours and 11 minutes were provided. Two of the data sets were recorded while the patient was sleeping, the other two while the patient was awake. The positions of the electrodes, implanted in the patient, can be seen on figure 4. In the two data sets with the patient being awake a total of 57 spikes for the RFf electrode were manually marked by experienced neurologists. The RFf electrode is placed in the centre of the brain and marked with red crosses on the image of figure 4. The raw signals were filtered using a band-pass filter (7-80 Hz). The bivariate representation of the signals in the RFf channel was also calculated, as the neurologists marked the spikes of this data set by examining the bivariate representation of

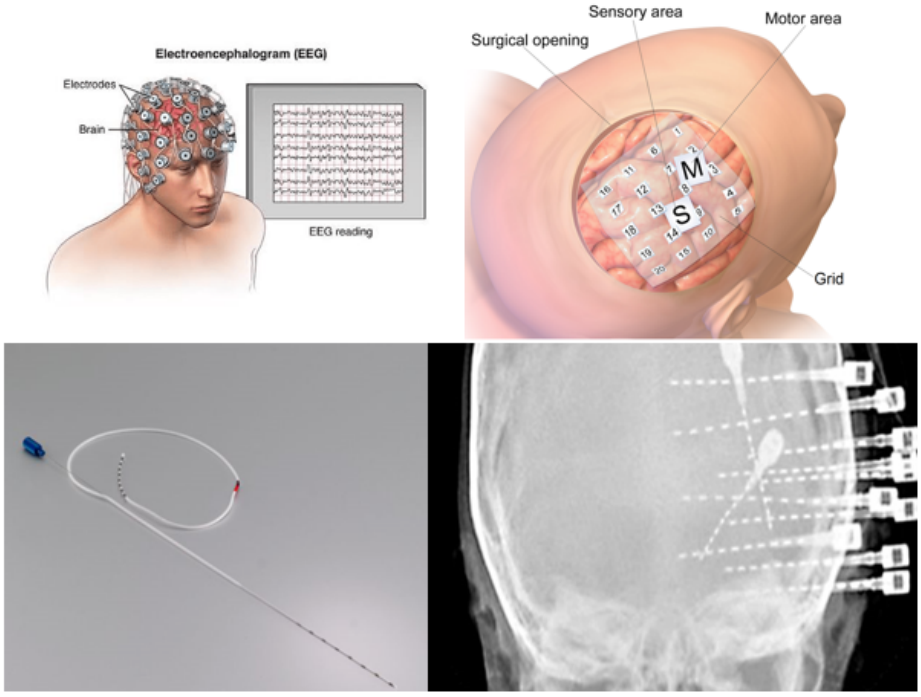

Figure 3: Different methods for recording brain signals. In the upper left image a non-invasive method (EEG) is shown. The other images show invasive methods. The upper right image shows an ECoG. An electrode for SEEG is shown in the bottom left image. To the right of it, it can be seen how the electrodes are implanted into the brain.

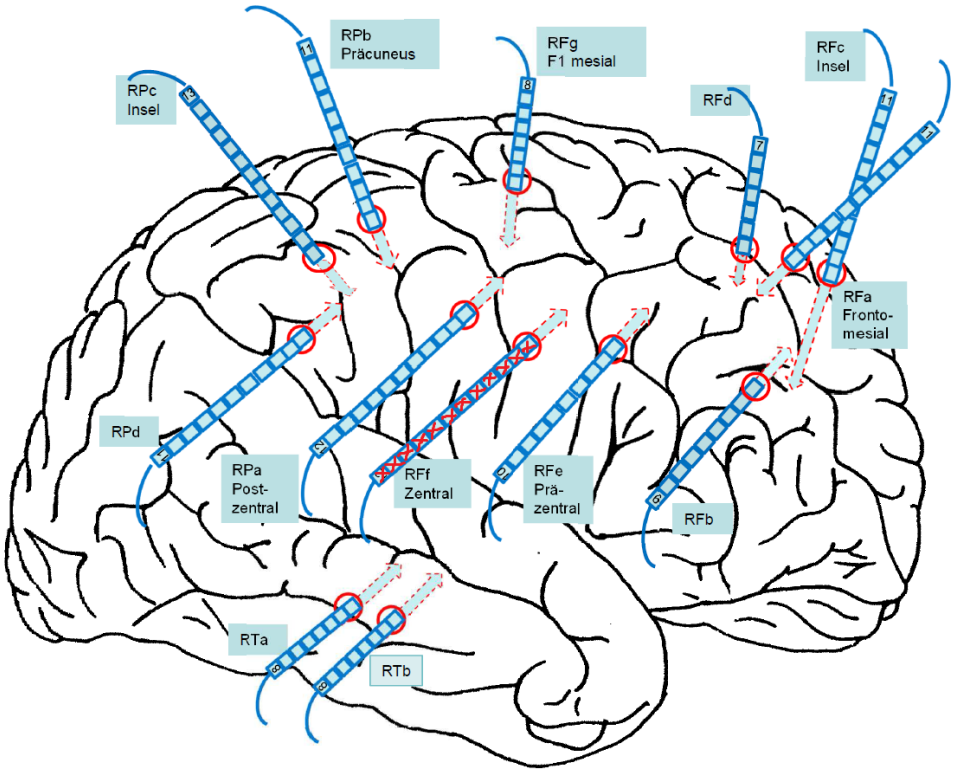

Figure 4: Positions of the electrodes for patient 1. Initially spikes were marked on the RFf electrode (highlighted with red crosses). The lower the number of a contact, the deeper it is inside the brain.

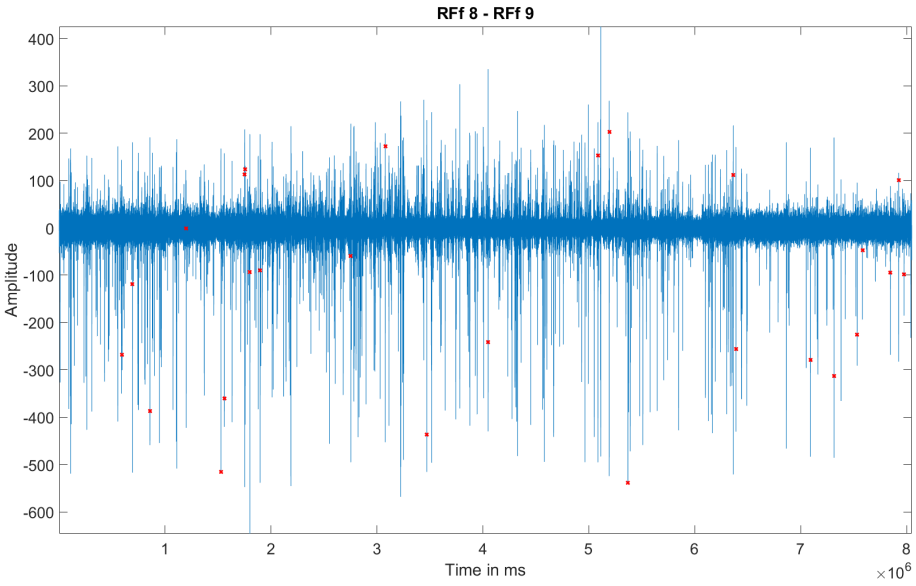

Figure 5: Bivariate representation of the channels RFf 8 and RFf 9. The red dots indicate at which locations spikes were initially marked.

the RFf electrode. The bivariate representation of the signals is calculated by subtracting the time series of two neighbouring signals of an electrode from each other. A signal of one of the bivariate channels can be seen on figure 5. The whole time series of one data set is plotted in blue with the time in milliseconds on the x-axis and the amplitude in microvolt on the y-axis. The red crosses show where a spike was marked initially. An exemplary spike of that channel can be seen by zooming in to one marked location and is shown on figure 6 .

2.2 Patient 2

Patient 2 was female and 33 years old. 11 depth electrodes were implanted in the patient with a total of 98 contacts.The sampling rate was 2048 Hz. Four data sets each with a duration of 1 hour and 6 minutes were provided. Equal to patient 1, two of the data sets were recorded while the patient was sleeping, the other two while the patient was awake. The positions of the electrodes can be seen in figure 8. One data set, where the patient was awake was manually marked with a total of 20 spikes for channel 2 on the electrode RTb. This electrode is placed in the front right part of the brain and is highlighted in figure 8 by a red cross. The raw signals were filtered using a band-pass filter (7-80 Hz). An exemplary spike of the channel RTb 2 is plotted in figure 7

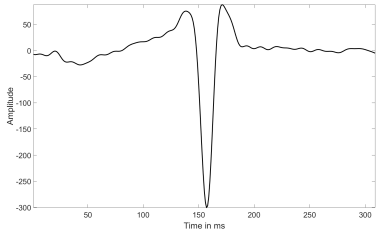

Figure 6: Exemplary spike from channel RFf8-RFf9 for patient 1.

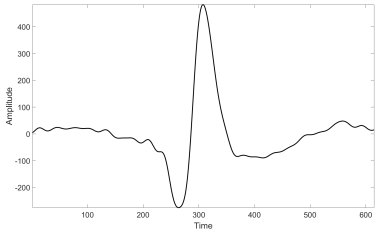

Figure 7: Exemplary spike from channel RTb 2 for patient 2.

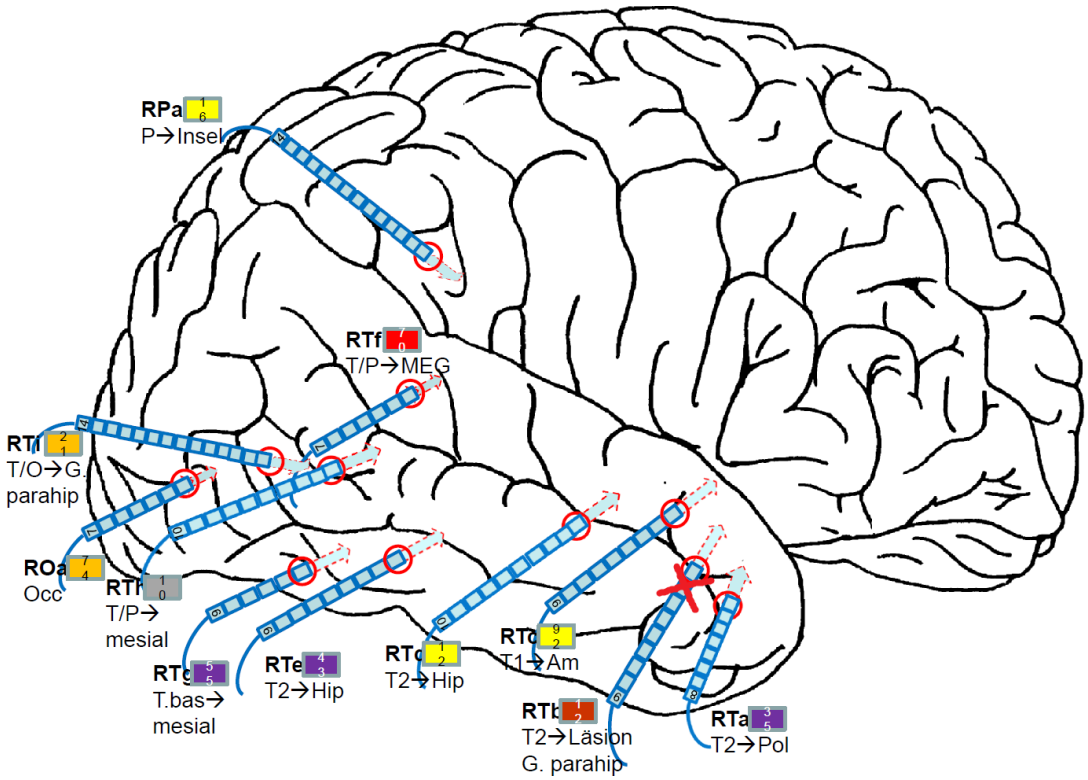

Figure 8: Positions of the electrodes for patient 2. Initially spikes were marked for contact 2 on the RTb electrode (highlighted with red cross). The lower the number of a contact, the deeper it is inside the brain.

3 Detecting spikes in the data

Initially 57 spikes for patient 1 and 20 spikes for patient 2 were marked. In order to carry out valid analyses on the data it was necessary to find more spikes. As it is very time consuming to mark spikes by hand and a specialist is needed for this task, the aim is to automatise the detection of spikes. There are many different approaches for automated spike detection. One idea for instance is to generate a template library by clustering a collection of spikes and background waveforms and to search for spikes with the help of that library [23]. Lodder et al. use a kind of feedback system for their algorithm that consists of two phases: A detection phase that finds all events similar to epileptiform activity by using a large database of template waveforms and a second phase that uses the ten nominations for spikes with highest certainty and presents them to a reviewer one by one for confirmation. Confirmations are used to update certainty values of the remaining nominations. Reviewer feedback is used to update template accuracies globally and improve future detections [11]. These are just two examples of different approaches for spike detection. They all have their advantages and disadvantages, but in general it can be said that there is no gold standard for a method doing spike detection. That is because spikes can look very different depending on the patient, the brain region the spikes appear and the nature of the spike. This makes it very hard to create one common spike detector for all different settings.
For this empiric investigation the spike detector was build by calculating the mean over all marked spikes and taking this mean as template for searching for new spikes in the data. The procedure, evaluation and results of the spike detector will be described in the following, separated by patients.

3.1 Patient 1

As described in section 2.1 the spikes for patient 1 were marked on basis of the bipolar derivation of the signals recorded by the electrode RFf. As the electrode had 12 contacts recording signals, the bipolar derivation had 11 channels. In order to find out, which of these channels is most appropriate for spike detection, they had to be compared.
In a first step snippets with a length of 300 ms around all marked spikes in the first data set were extracted for all RFf channels. An exemplary result of this extraction can be seen in figure 9 for channel RFf8-RFf9, where all snippets around the initially marked spikes of one data set are plotted above each other.
Since the aim is to generate a homogeneous template, all snippets were centred for their minimum or maximum value, whichever absolute value was greater. After shifting the snippets their mean was calculated and taken as template. An example for the centred snippets and the resulting template is shown in figure 10.
A look at figure 11 gives an idea how the templates of all RFf channels look like, compared with the extracted snippets. For the first six channels the snippets

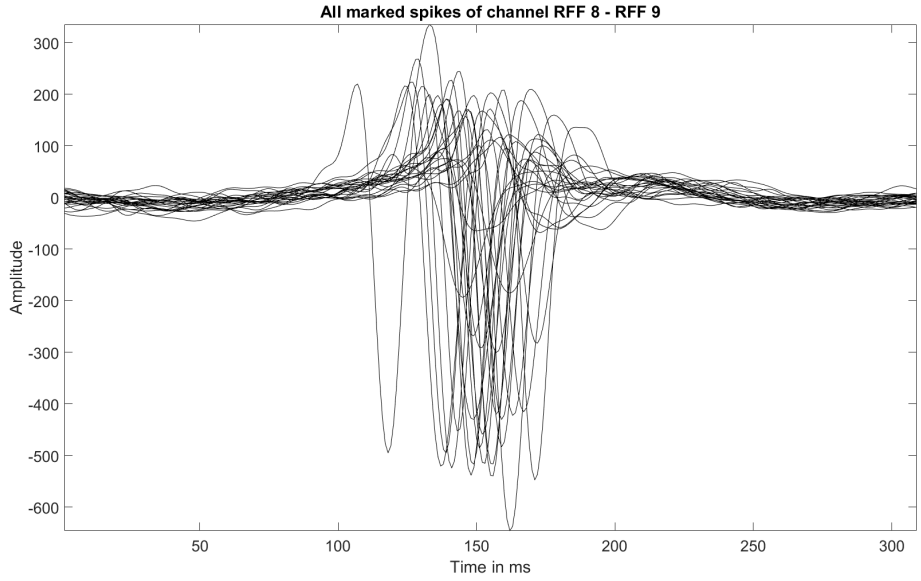

Figure 9: Snippets around all initially marked spikes of channel RFf8-RFf9 with a length of 300 ms.

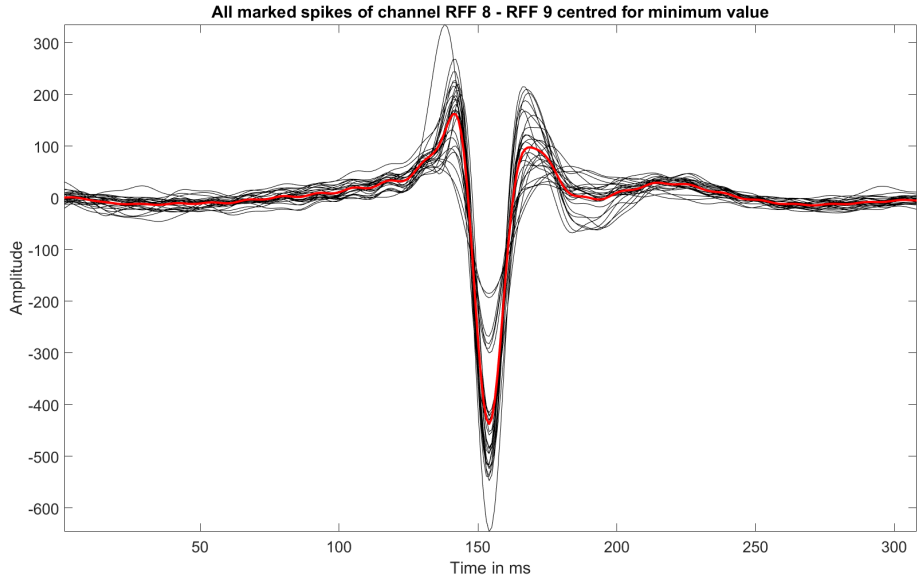

Figure 10: For the exemplary channel RFf8-RFf9 all snippets around the spikes are centred for the minimum and the mean is calculated to obtain a template.

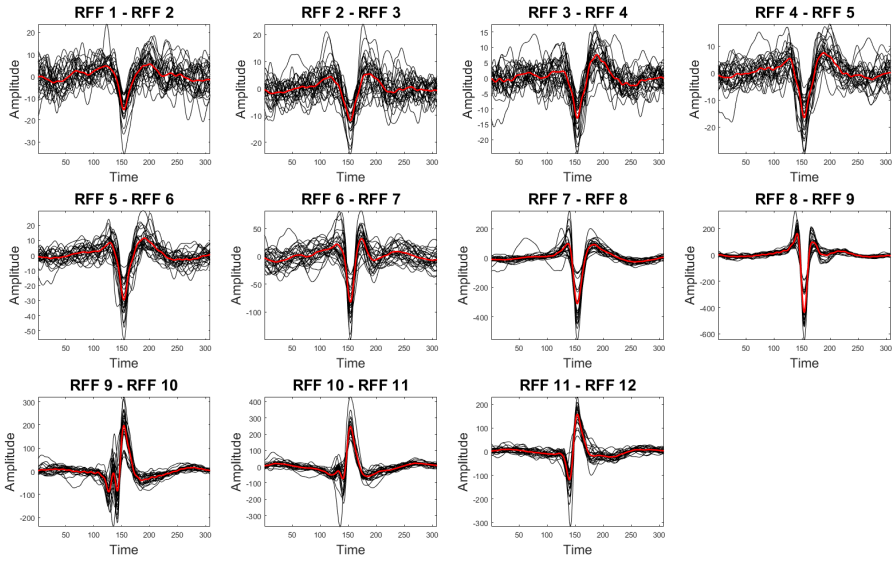

Figure 11: The black 'spike snippets' compared to the red templates. The first six templates seem to be far off the spike snippets. The five remaining templates look more reliable.

(plotted in black) deviate quite a lot from the templates (plotted in red). The five remaining templates look better. The template for channel RFf8-Rff9 seems to be closest to the spike snippets.

In order to evaluate which of the templates does the best job in detecting spikes, they were tested on the second data set with marked spikes. For this purpose each template was tested on the corresponding channel. So for instance the template that was generated on basis of the snippets in the first data set from channel RFf1-RFf2 was tested on channel RFf1-RFf2 on the new data set. For detecting spikes, the new data set was cut into equal pieces, each with the same length as the template (300 ms). All of the pieces were centred for the maximum or minimum value within that snippet depending on the spike in the template. In case the amplitude of the spike in the template was negative, the pieces were centred for the minimum and if it was positive they were centred for the maximum. After the centring, the snippet was compared with the template created for the corresponding channel. If the value of the correlation of the snippet and the template was higher than a certain threshold (here a threshold of 0.9 was chosen), this snippet was considered to contain a spike. (See also figure 12 and algorithm 1.)

Repeating that procedure for all RFf channels resulted in figure 13. Each subplot shows how many spikes in total were found for one specific channel and which proportion of the initially marked spikes was detected. A snippet is cut

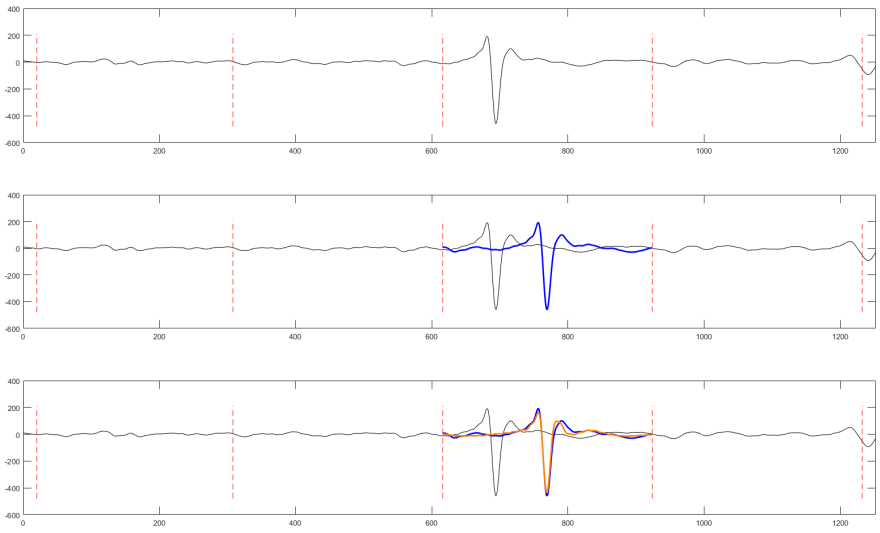

Figure 12: Illustration of the spike detection algorithm. Upper image shows the segmentation of the data set. In the middle plot one exemplary section is centred for the minimum value (plotted in blue). In the bottom image the template is additionally plotted (in orange). If the correlation between the orange and the blue snippet is high this point in time is marked as a spike.

Algorithm 1 Spike Detection

$j = 0$
for $i \leftarrow 1, (length(TestData)/length(template))$ **do**
    $end \leftarrow i * length(template)$
    $start \leftarrow end - length(template)$
    $snippet \leftarrow data(start : end)$
    $absMax \leftarrow max(abs(snippet))$
    $newSnippet \leftarrow data((absMax - snippetLength/2) : (absMax + snippetLength/2))$
    $correlation \leftarrow cor(template, newSnippet)$
    if $correlation > threshold$ **then**
        $j = j + 1$
        $detectedSpikes[j] \leftarrow newSnippet$
return $detectedSpikes$

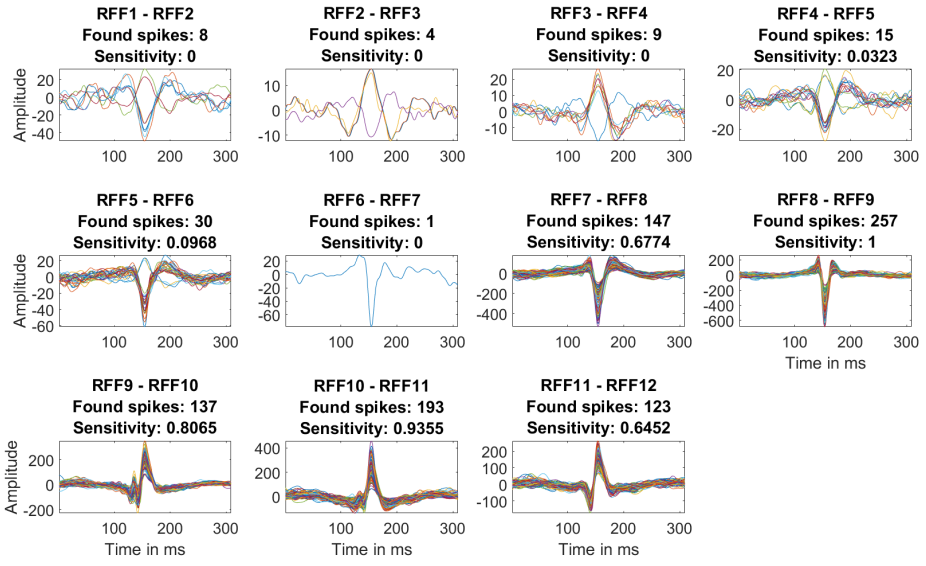

Figure 13: All found spikes that have been detected for each of the channels on the electrode RFf. For each channel the number of found spikes and the true positive rate is listed. The template of the channel RFf8-RFf9 did the best job regarding the sensitivity.

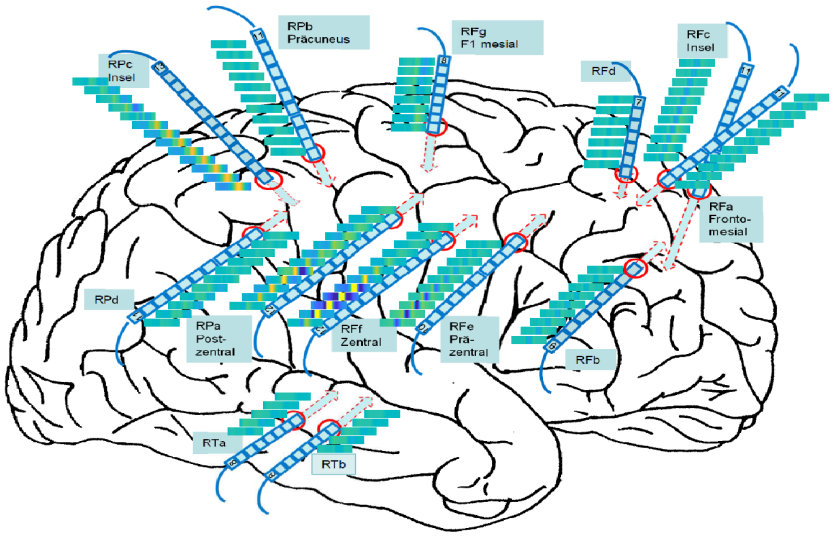

Figure 14: Means of the spike snippets of patient 1 relative to the location of the electrodes. Channels on the RFf electrodes show the clearest spike pattern. Also in electrodes RPa and RPc such patterns can be observed.

around all detected spikes and they are plotted above each other. The templates of the channels RFf1-RFf2 to RFf6-RFf7 did not perform very well, as expected. The other five templates performed way better, with the best being the template of channel RFF8-RFF9 which revealed all previously marked spikes (sensitivity = 1).
The best template was used to detect spikes over the 4 available data sets in the corresponding channel. A total number of 564 spikes, was found by taking the template of channel RFf8-RFf9 and searching for spikes in the channel RFf8-RFf9.
In order to see if there are spike patterns in other electrodes, snippets for all 125 channels over the 11 implanted electrodes were cut at all time points where a spike was detected. The mean over all of these snippets was calculated for each channel and plotted in figure 14 according to the location of the channel. The more coloured the box next to a channel is, the higher is the amplitude of the spike. In channel RFf the spike can be seen the clearest, but also for channels on the RPa and RPc electrodes, spike patterns can be observed.

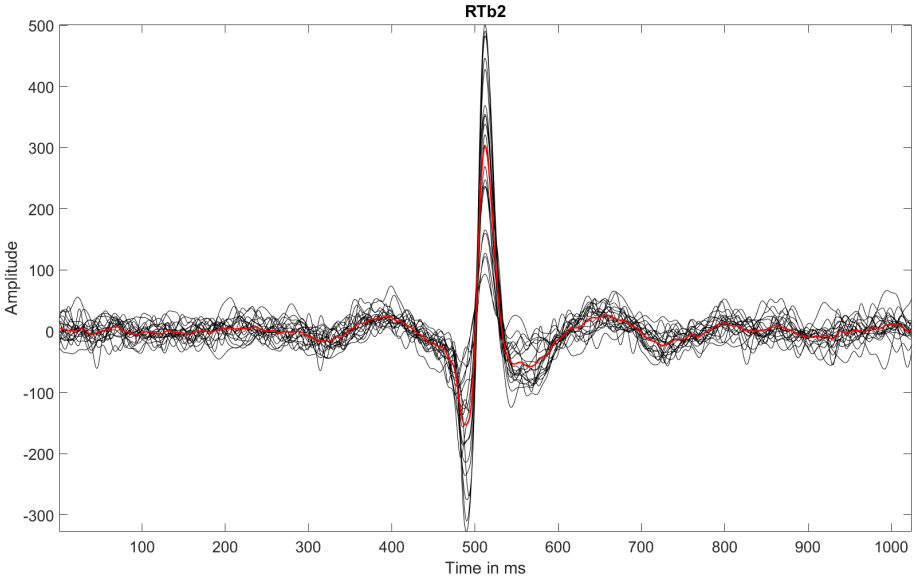

Figure 15: Template and all centred 'spike snippets' for channel RTb 2 in patient 2

3.2 Patient 2

As the spikes for patient 2 were marked on basis of channel RTb 2 (see also section 2.2), a template for that channel was created. Around all 20 marked spikes, snippets were cut and centred for the maximum. Afterwards the mean over these 20 snippets was calculated to create a template (figure 15). Searching for spikes with this template in channel RTb 2 for all available data sets gave a total number of 481 spikes. For each of the 98 channels of patient 2 the snippets around the time points of the detected spikes were collected and the mean was calculated. The resulting image can be seen in figure 16. The clearest spikes can be seen for the channels on electrode RTb. The neighbouring electrodes RTa and RTc also show spike patterns, as well as the electrodes RTe and RTi which are quite far away from the electrode RTb, for which the spikes were initially marked for.

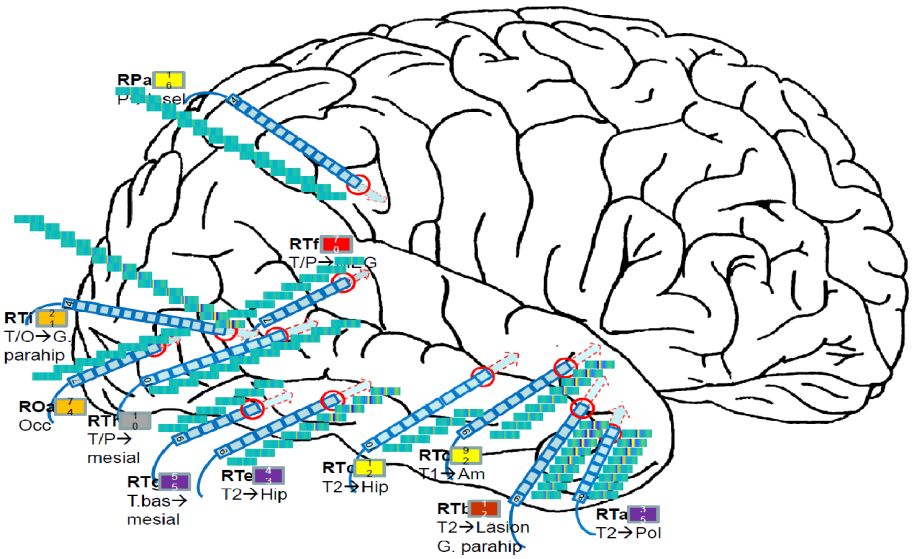

Figure 16: Means of the spike snippets of patient 1 relative to the location of the electrodes. The clearest spikes can be seen for the channels on electrode RTb. The channels on the electrodes RTa, RTc, RTe and RTi also show spike patterns.

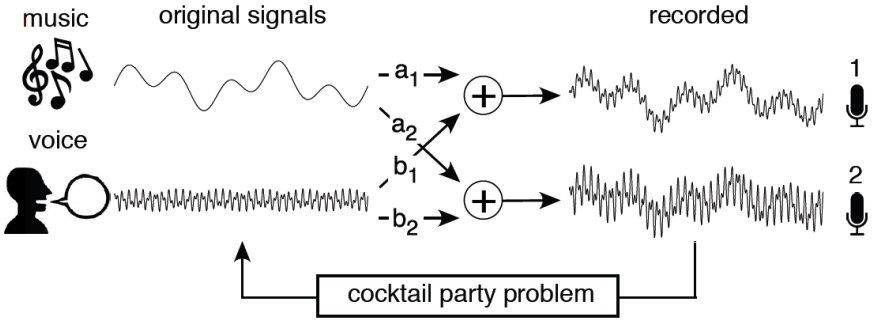

Figure 17: Cocktail Party Problem. Two microphones recording a mix of two sources (music and a voice). The aim is to unmix the recordings and to recover the original signals. Image taken from [19]

4 Methods

4.1 Independent Component Analysis

In the preceding section one could recognise that spikes were detected on multiple channels. It might be interesting to find out if there are common sources that generate the spikes. *Independent Component Analysis* is a tool that is able to find such common sources of which only a mixture might be recorded by the implanted electrodes. Thus the dimensionality of the data could be reduced and also the problem of poor spatial resolution might be solved by finding the true sources that were not recorded by the electrodes because they took place on another location. This might help to create better models of the relationships between the spikes, than using the data of all channels.
The following chapter describes the idea of the *Independent Component Analysis* (ICA), the mathematical background (mainly based on [22] and [19]) and why it can be useful in the context of brain data.
One famous example to understand the principle idea of the ICA is the *cocktail party problem*. [19] Imagine a room where two sounds are generated, one by a person that speaks and the other by music that is played by a musician. Two microphones that are recording the sound in the room are placed at different locations. Each of the microphones records a mixture of the two sounds. After the cocktail party someone is listening to the two recordings and can neither understand clearly what the speaker said because the recording is disturbed by the music, nor can the music that was played at the party be enjoyed again because the voice of the speaker was recorded as well. The aim of the *cocktail party problem* is to filter out the disturbing signal and to recover the original sources (speech and music). For this challenging problem reasonable solutions have been developed based on ideas of ICA.

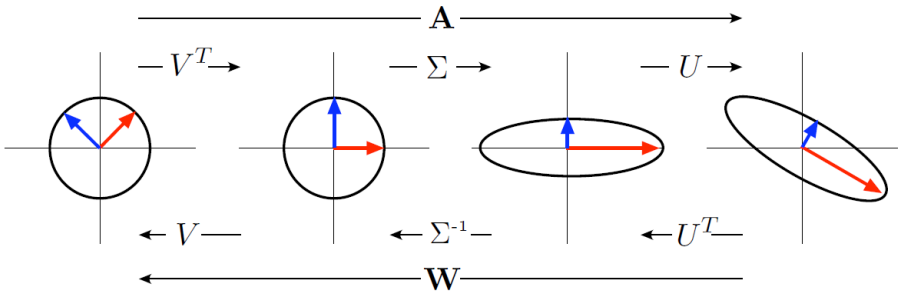

Figure 18: Graphical representation of the SVD. **A** performs three different operations, so it can be decomposed in three matrices (assuming that **A** is invertible). The decomposition is $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$, where **U** and **V** are the rotation matrices and Σ is a diagonal matrices. It can be seen that the basis rotates then stretches and then rotates again during each operation. Image taken from [19]

Writing down the general form of the presented problem would look as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

,where **x** is a matrix of the observed data. In our case **x** corresponds to the recordings of the microphones. Each sample is drawn from an unknown distribution $P(\mathbf{x})$. The matrix **s** contains the values of the underlying sources. In the *cocktail party problem* these are the sounds generated by the speaker and the musician, respectively. For using the ICA, each source s_i has to be independent of all other sources s_j for $i \neq j$. **A** is an unknown invertible square matrix that mixes the components of the sources. It is the goal to find this matrix **A** for recovering **s** from **x**.
The first step in solving this problem is the construction of a new matrix **W**, so that $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$. The linear transformation **Wx** is an estimate of the underlying sources **s**. As equation (1) can be written as $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$ and $\mathbf{s} \approx \hat{\mathbf{s}}$, it can be seen that $\mathbf{W} = \mathbf{A}^{-1}$.
Since in equation (1) the matrix **A**, as well as the true sources **s** are unknown, the problem is under-determined. To tackle that issue one can rewrite the problem by using the *singular value decomposition* (SVD). This leads to the following representation (see also figure 18):

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \tag{2}$$

The matrices **U** and **V** are rotation matrices (which are orthogonal) and Σ is a stretch along the axes. As $\mathbf{W} = \mathbf{A}^{-1}$ and by using the properties that **V** is orthogonal and that Σ^{-1} exists and is well defined (because **A** is invertible) one can write:

$$\mathbf{W} = \mathbf{A}^{-1} = (\mathbf{U}\Sigma\mathbf{V}^T)^{-1} = (\mathbf{V}^T)^{-1}\Sigma^{-1}\mathbf{U}^{-1} = \mathbf{V}\Sigma^{-1}\mathbf{U}^{-1} \tag{3}$$

The aim is to calculate $\mathbf{W}$, wherefore two steps are necessary. The first is to calculate $\mathbf{U}$ and Σ by examining the covariance of $\mathbf{x}$. To solve for $\mathbf{V}$ in the second step, one has to return to the assumption of independence of $\mathbf{s}$. To make things easier let us assume that for $\mathbf{x}$ the mean has been subtracted off for each dimension. This means that the mean of the data is zero. The covariance of the matrix $\mathbf{x}$ can then be written as:

$$\begin{aligned} Cov(\mathbf{X},\mathbf{X}) = \\ E[(\mathbf{x}-\mu_x)(\mathbf{x}-\mu_x)^T] = \\ E[\mathbf{x}\mathbf{x}^T - \mathbf{x}\mu_x^T - \mu_x\mathbf{x}^T + \mu_x\mu_x^T] = \\ E(\mathbf{x}\mathbf{x}^T) - \mu_x\mu_x^T - \mu_x\mu_x^T + \mu_x\mu_x^T = \\ E(\mathbf{x}\mathbf{x}^T) - \mu_x\mu_x^T = \\ E(\mathbf{x}\mathbf{x}^T) \end{aligned} \tag{4}$$

The next step is *whitening* the data. *Whitening* is an operation that removes all linear dependencies in a data set and normalises the variance along all dimensions. For *whitening*, the data has first to be decorrelated by rotating the data and then it has to be normalised. For this purpose one needs to calculate the orthogonal rotation matrix $\mathbf{E}$, that contains the eigenvectors of the data and a diagonal matrix $\mathbf{D}$ with the eigenvalues of $\mathbf{x}$ on the diagonal. The whitened version of the data $\mathbf{x}$ is calculated as follows:

$$\mathbf{x}_w = (\mathbf{D}^{-1/2}\mathbf{E}^T)\mathbf{x} \tag{5}$$

Taking again the matrices $\mathbf{E}$ and $\mathbf{D}$, defined above as the matrices with the eigenvalues and eigenvectors of the data $\mathbf{x}$, it can be proven that (*orthogonal diagonalization*):

$$< \mathbf{x}\mathbf{x}^T >= \mathbf{E}\mathbf{D}\mathbf{E}^T \tag{6}$$

From this it can be seen that $< \mathbf{x}_w\mathbf{x}_w^T >= \mathbf{I}$.
With the equations (1) and (2) and by assuming that $< \mathbf{s}\mathbf{s}^T >= \mathbf{I}$ one can state that the covariance calculated in equation (4) can be written as:

$$\begin{aligned} < \mathbf{x}\mathbf{x}^T >= \\ < (\mathbf{A}\mathbf{s})(\mathbf{A}\mathbf{s})^T >= \\ \mathbf{A} < \mathbf{s}\mathbf{s}^T > \mathbf{A}^T = \\ \mathbf{U}\Sigma\mathbf{V}^T < \mathbf{s}\mathbf{s}^T > (\mathbf{U}\Sigma\mathbf{V}^T)^T = \\ \mathbf{U}\Sigma\mathbf{V}^T\mathbf{V}\Sigma^T\mathbf{U}^T = \\ \mathbf{U}\Sigma^2\mathbf{U}^T \end{aligned} \tag{7}$$

and is thus independent of sources s as well as V. With the results of equations (6) and (7) it follows that $\Sigma^2 = \mathbf{D}$ and $\mathbf{E} = \mathbf{U}$ and the problem is reduced to:

$$\mathbf{W} = \mathbf{V}\mathbf{D}^{-1/2}\mathbf{E}^T \tag{8}$$

Using $\mathbf{x}_w$ from equation (5) and the properties $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ and $\mathbf{W} = \mathbf{V}\Sigma^{-1}\mathbf{U}^{-1}$ from above, results in:

$$\hat{\mathbf{s}} = \mathbf{V}\mathbf{x}_w \tag{9}$$

Whitened data is rotationally symmetric. Therefore $\hat{\mathbf{s}} = \mathbf{V}\mathbf{x}_w$ must be whitened as well. So the assumption $< \mathbf{ss} >= \mathbf{I}$ made further above is reasonable.
As $\mathbf{E}$ and $\mathbf{D}$ can be calculated from the data, the next step is to solve for $\mathbf{V}$. For this a measure of independence is needed. There are various methods for measuring the independence of variables. The method shown here is the so called *mutual information*. It measures the departure of two variables from statistical independence. A generalisation of the *mutual information* is the *multi-information*, which measures the statistical dependence between multiple variables and is derived by:

$$I(y) = \int P(y) log_2 \frac{P(y)}{\prod_i P(y_i)} dy \tag{10}$$

The result of the equation is a non-negative quantity and is 0 if and only if all variables are statistically independent. The goal is to find the rotation matrix $\mathbf{V}$ such that $I(\hat{\mathbf{s}}) = 0$, what means that the estimated sources $\hat{s}$ are statistically independent and can be separated. In practice it is difficult to minimise the *multi-information* but it can be simplified. The multi information is a function of *entropy*, which measures the uncertainty about a distribution $P(y)$ by: $H(y) = -\int P(y) log_2 P(y) dy$. The *multi-information* can be expressed as:

$$I(y) = \sum_i H(y_i) - H(y) \tag{11}$$

Plugging $\hat{\mathbf{s}}$ into that equation and using the property $H(\mathbf{B}\mathbf{x}) = H(\mathbf{x} + log_2|\mathbf{B}|)$ (for a linear transformation $\mathbf{B}$ and a random matrix $\mathbf{x}$) gives:

$$\begin{aligned} I(\hat{\mathbf{s}}) = \sum_i H((\mathbf{V}\mathbf{x})_i) - H(\mathbf{V}\mathbf{x}_w) = \\ \sum_i H((\mathbf{V}\mathbf{x}_w)_i) - (H(\mathbf{x}_w) + log_2|\mathbf{V}|) \end{aligned} \tag{12}$$

Since $log_2|\mathbf{V}| = 0$ as the determinant of a rotation matrix is 1 and because $H(\mathbf{x}_w)$ is independent of $\mathbf{V}$ the problem is reduced to:

$$\mathbf{V} = argmin_v \sum_i H((\mathbf{V}\mathbf{x}_w)_i) \tag{13}$$

This means that the optimisation only consists of finding a rotation matrix which minimises the sum of the marginal entropies of $\hat{\mathbf{s}}$.

After looking at the theoretical explanation, the question arises how all of this could be useful for analysing SEEG-data. Basically we have to solve the

same problem as in the *cocktail party problem*. The microphones that record the sound can be seen as the electrodes, which are recording the brain activity. The sound of the speech and the music can be seen as brain activity. The difference is that we have way more electrodes that record brain activity than just two. Also the occurrence of the brain activity is more complex, than the two sounds coming from the two sources in the *cocktail party problem*. The challenge in both settings is, that the recording devices might not be located at the spot, where the signals are produced and therefore record a mix of the signals. In the *cocktail party problem* this issue could be solved quite easily by bringing the microphones closer to the speaker or the musician, respectively, as it should be possible to find out where each of the sound generating sources is located. In the SEEG setting this is much harder. As described in section 2 the depth electrodes have to be implanted in the brain of the patient for recording the brain activity. It is quite challenging to locate the electrodes ideally in the brain of a patient, meaning close to the region where spikes occur. As the electrodes are quite thin and the structure of the brain is complex it is impossible to record the activity of all locations in the brain with SEEG. Since we want to find out how spikes propagate and where they arise from, it could be of interest to find out what happens at locations in the brain where no electrode is recording the activity of the brain. This is comparable to a situation in the *cocktail party problem*, where the speaker and the musician stand in two different corners of a room and the microphones are located in the middle. ICA might be an useful tool for that challenge. So it could be possible that in the brain there a locations where some activity is happening, but the electrodes can only capture a mixture of different activities. The goal is to reconstruct these activities by using the ICA and to find the signals of the true sources of the brain activity. This would solve the problem of poor spatial resolution, which depth electrodes imply.
As it is a reasonable assumption, that the source signals are independent it is valid to use ICA for SEEG data.

4.2 Causal inference

The goal of causal inference is to find cause-effect relationships between different variables. After uncovering the true sources of the spikes by using ICA, causal inference could help to find relationships between the different components and to represent these relations graphically. This might be of clinical relevance as it is not clear if and how the different ICs interact with each other and thus which parts of the brain of a specific patient are involved in the propagation of spikes and need to be removed for a successful surgery. To find out which different components are responsible for producing spikes and how they interact with each other, in order to be able to determine the epileptogenic zone more precisely and to represent the network graphically, causal inference is carried out.
For the challenging task of finding cause-effect relationships it is necessary to perform randomized experiments to learn about the system and to conclude

about relationships between variables. In medical studies, for instance, where new drugs are tested such randomized experiments are conducted. Various cohorts of patients, that should have similar characteristics (e.g. age or gender), are tested to find out whether a novel drug is better than the present medication or whether it does not bring any benefits to the patients. Therefore one cohort might get the new drug and the other cohorts might be treated with a medication that is already on the market. After a certain time the cohorts are compared and it has to be determined whether the new medication had a positive effect on the patients, who received it compared to the ones who did not receive the new drug. It is important that the cohorts of patients are quite similar, so that possible different outcomes between the groups can be justified by the use of the drug. But in practice it is quite difficult to create such conditions where only one variable is responsible for differences in the various treatment groups. Variables that differ between the treatment groups are called *confounders*. In case all of these *confounders* are measured it is still possible to adjust for these variables and a causal effect can still be estimated. But there might also be situations where the *confounders* can not be measured. In these situations it is impossible to accurately estimate the causal effect.
The problem with the *confounders* is not the only one arising in causal inference. In some situations an experiment, as described above, might not be possible due to ethical or financial reasons. Sometimes experiments are just impossible to conduct. The brain activity is a very complex, not yet fully understood system varying from patient to patient, so that it is not possible to adjust for all possible variables and thus randomized experiments are not possible.
Therefore the SEEG data has to be analysed in another way. In this work the *causal bayesian network* will be used for that task. In this framework cause-effect relations are represented as a graph in which the nodes represent random variables and directed edges show causal influences. These relations are represented by a *directed acyclic graph* (DAG). In the following it will be shown how such DAGs can be created from the data (mainly based on [16]).
Let us assume we have a data set with three variables X, Y and Z. The aim is to construct a DAG out of these variables, for which they have to be tested for independence and conditional independence. At first it is tested if X and Y are independent from each other. If it is found out, that they are independent ($X \perp\!\!\!\perp Y$), then there is no edge between the two variables. If they are not independent ($X \not\perp\!\!\!\perp Y$) an edge between the two variables can be drawn ($X - Y$) but it is not possible to say which of the two variables is the driver and which one is the receiver (i.e. we do not know if $X \leftarrow Y$ or $X \rightarrow Y$ is true). Let us consider that $X \not\perp\!\!\!\perp Y$,$X \not\perp\!\!\!\perp Z$ and $Y \not\perp\!\!\!\perp Z$, which means that we can draw an edge between all of the variables ($X - Y$,$X - Z$ and $Y - Z$). As that is not informative in terms of finding causal structure in the data, it is necessary to test for conditional independence. Imagine that we found the following: $X \not\perp\!\!\!\perp Y|Z$,$X \perp\!\!\!\perp Z|Y$ and $Y \not\perp\!\!\!\perp Z|X$. From this we can conclude that the edge between X and Z can be removed because they are independent conditioned on Y. In other words, if we know Y, X does not provide any additional information regarding Z (or the other way round). That gives us the following representation: $X - Y - Z$.

Still we do not have any information that helps to orient the edges, to see which variable is the driver of another one.
Let us consider that we are in a situation where we found out, that there are edges between X and Y and between Y and Z but not between X and Z, then there are still four possibilities how the edges could be oriented:

- $X \rightarrow Y \rightarrow Z$
- $X \leftarrow Y \leftarrow Z$
- $X \leftarrow Y \rightarrow Z$
- $X \rightarrow Y \leftarrow Z$

The first and the second DAG are called *chains*, the third is a *fork* and the last one is a *collider*. For the first three DAGs we have $X \perp\!\!\!\perp Z|Y$, whereas for the collider it holds that $X \not\perp\!\!\!\perp Z|Y$. By testing for $X \not\perp\!\!\!\perp Z|Y$, we can either orient the edges or rule out an orientation (as $X \rightarrow Y \leftarrow Z$ holds if and only if $X \not\perp\!\!\!\perp Z|Y$ is true). Still one can not distinguish between a fork or one of the chains by just analysing the independence structure of the variables. If more variables are available than just three, more edges can possibly be oriented by finding colliders. Imagine that we found out that $X \rightarrow Y$, for example by figuring out that X and another variable (that is not Z) collide in Y, and we know that there has to be a chain or a fork between $X - Y - Z$ (by having the knowledge that a collider is not a option), than we can follow, that the only remaining possibility is $X \rightarrow Y \rightarrow Z$. Also knowledge that is behind statistical analyses can be used to orient edges. One example for that would be the knowledge of the timing of events. If you know that there is an edge between X and Y and you know that Y appears after X than the orientation of the edge can not go from Y to X.
As one can imagine, checking for all independences and conditional independences can get computational very expensive if the data set is large. For that reason the PC algorithm was introduced to create a DAG in reasonable time, even with large data sets. The algorithm works as follows [16] [21]:

- Connect all variables with an edge
- For each X and Y, check if $X \perp\!\!\!\perp Y$; if so remove their edge
- For all X and Y that are still connected and each third variable Z, check if $X \perp\!\!\!\perp Y|Z$; if so remove the edge between X and Y
- For all X and Y that are still connected and each third and fourth variables Z_1 and Z_2, check if $X \perp\!\!\!\perp Y|Z_1, Z_2$; if so remove the edge between X and Y
- ...
- For all X and Y that are still connected , check if $X \perp\!\!\!\perp Y|$ all the $p-2$ other variables; if so remove the edge between X and Y

- *After these steps one has a skeleton with undirected edges. The following rules have to be applied for orienting the edges. Rule 1 has to be applied first, the order of the remaining rules does not matter.*
  - Rule 1: For all triplet of variables (Z_i, Z_j, Z_k), where Z_i and Z_j as well as Z_j and Z_k are adjacent but Z_i and Z_k are not. Orient $Z_i - Z_j - Z_k$ as $Z_i \rightarrow Z_j \leftarrow Z_k$ if $Z_i \perp\!\!\!\perp Z_k | S$, where S is a set of variables and $Z_j \in S$
  - Rule 2: Orient $Z_j - Z_k$ as $Z_j \rightarrow Z_k$ if there is a directed edge $Z_i \rightarrow Z_j$ such that Z_i and Z_k are not adjacent
  - Rule 3: Orient $Z_i - Z_k$ as $Z_i \rightarrow Z_k$ if there is a directed path $Z_i \rightarrow Z_j \rightarrow Z_k$
  - Rule 4: Orient $Z_i - Z_j$ as $Z_i \rightarrow Z_j$ whenever there are two directed paths $Z_i - Z_k \rightarrow Z_j$ and $Z_i - Z_l \rightarrow Z_j$ such that Z_k and Z_l are not adjacent

5 Results

In the following section the results of the analyses will be presented. First of all it is outlined how the methods were applied to the data and then the findings will be shown, separated by the two patients.
For the ICA the function *runica* from *EEGLAB* in *MATLAB* was used. The number of independent components to compute was set to 64 and was performed for all data sets of the two patients.
For the resulting *independent components* (ICs), that seemed relevant (more in sections 5.1.1 and 5.2.1) a causal analysis was conducted for the snippets around the spikes, that were detected earlier (section 3). To create a DAG out of the components all combinations of three out of the relevant ICs were analysed. For instance, let us consider four ICs seemed relevant for performing a causal analysis, (let us call them IC1 to IC4) then all possible combinations of three:

- IC1, IC2, IC3
- IC1, IC2, IC4
- IC1, IC3, IC4
- IC2, IC3, IC4

were analysed. By checking all possible orders of the ICs within each of these combinations it was tested if delayed causal chains could be found. Assume that t stands for a point in time within the spike snippets, and δ_t expresses a shift in time. Then, for instance, taking the first of the combinations shown above (IC1, IC2, IC3), it was tested for the following chains:

- $\text{IC1}(t) \rightarrow \text{IC2}(t + \delta_t) \rightarrow \text{IC3}(t + 2\delta_t)$

- $\text{IC1}(t) \rightarrow \text{IC3}(t+\delta_t) \rightarrow \text{IC2}(t+2\delta_t)$
- $\text{IC2}(t) \rightarrow \text{IC1}(t+\delta_t) \rightarrow \text{IC3}(t+2\delta_t)$
- $\text{IC2}(t) \rightarrow \text{IC3}(t+\delta_t) \rightarrow \text{IC1}(t+2\delta_t)$
- $\text{IC3}(t) \rightarrow \text{IC1}(t+\delta_t) \rightarrow \text{IC2}(t+2\delta_t)$
- $\text{IC3}(t) \rightarrow \text{IC2}(t+\delta_t) \rightarrow \text{IC1}(t+2\delta_t)$

For computational reasons the time delays between the first and the second and the second and the third component were kept equal (δ_t). The chains were tested for time delays from 1 to 100 samples ($\delta_t = \{1, ..., 100\}$). Each of the possible chains was tested, by checking if the ICs were dependent from each other and by testing if the first IC in the possible chain was independent of the last IC, conditioned on the middle one. For instance, for the chain $\text{IC1}(t) \rightarrow \text{IC2}(t+\delta_t) \rightarrow \text{IC3}(t+2\delta_t)$ it was tested if:

1. $\text{IC1}(t) \not\perp\!\!\!\perp \text{IC2}(t+\delta_t)$
2. $\text{IC1}(t) \not\perp\!\!\!\perp \text{IC3}(t+2\delta_t)$
3. $\text{IC2}(t+\delta_t) \not\perp\!\!\!\perp \text{IC3}(t+2\delta_t)$
4. $\text{IC1}(t) \perp\!\!\!\perp \text{IC3}(t+2\delta_t) \mid \text{IC2}(t+\delta_t)$

To test the first three conditions above a correlation was calculated with the associated p-value for each of the conditions and it was tested if these p-values were below a certain threshold (here $\alpha = 0.05$). The p-value for the partial correlation (fourth condition) was calculated by performing a linear regression from $\text{IC1}(t)$ to $\text{IC2}(t+\delta_t)$ and from $\text{IC3}(t+2\delta_t)$ to $\text{IC2}(t+\delta_t)$ and then determining the correlation between the related residuals. It then was tested if the p-value of this correlation was above a chosen threshold (here $\beta = 0.05$). If all four conditions were met, the underlying test setting was said to be a chain.
For a chain $X \rightarrow Y \rightarrow Z$, where all above conditions were fulfilled it can be proven, that Y is a genuine cause for Z but it is still possible, that a confounder is in between X and Y [8]. Therefore only the relation $Y \rightarrow Z$ can be used in the following parts, to create a DAG.

5.1 Patient 1

5.1.1 ICA

The ICA was performed on all four data sets of patient 1. As described above the number of components, that should be calculated, was set to 64. Around all points in time where spikes were detected (see section 3), snippets with a length of 1 second (1024 samples) were cut for each of the 64 ICs and have then been averaged. The resulting image for one exemplary data set can be seen in figure 19.

In figure 14 in section 3 it was shown, that the spikes of the raw data were

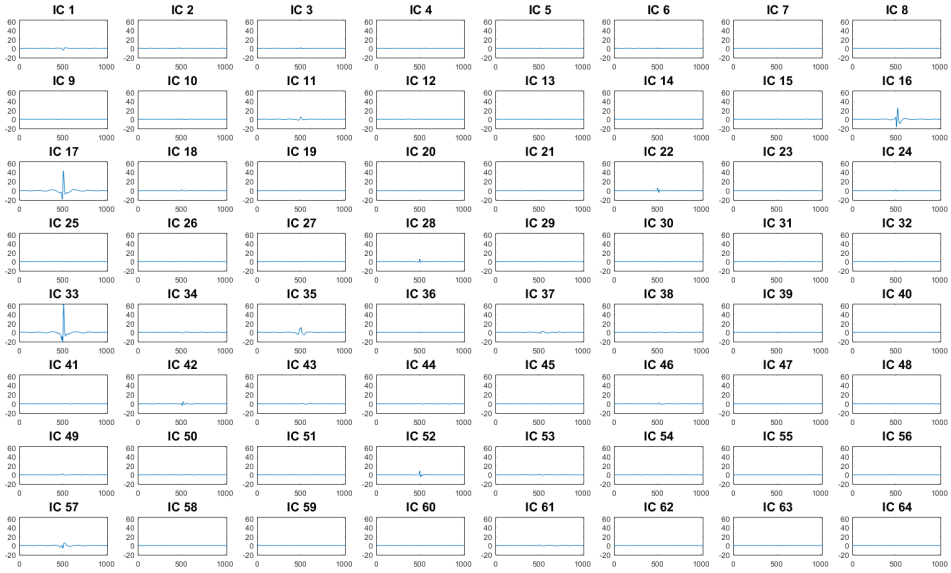

Figure 19: Each subplot represents the mean spike snippets of the *independent component analysis* of one exemplary data set of patient 1. It can be seen that IC16, IC17 and IC33 show significant spikes and might therefore be the true sources of a spike. These ICs were examined in more detail for studying the propagation patterns of the spikes.

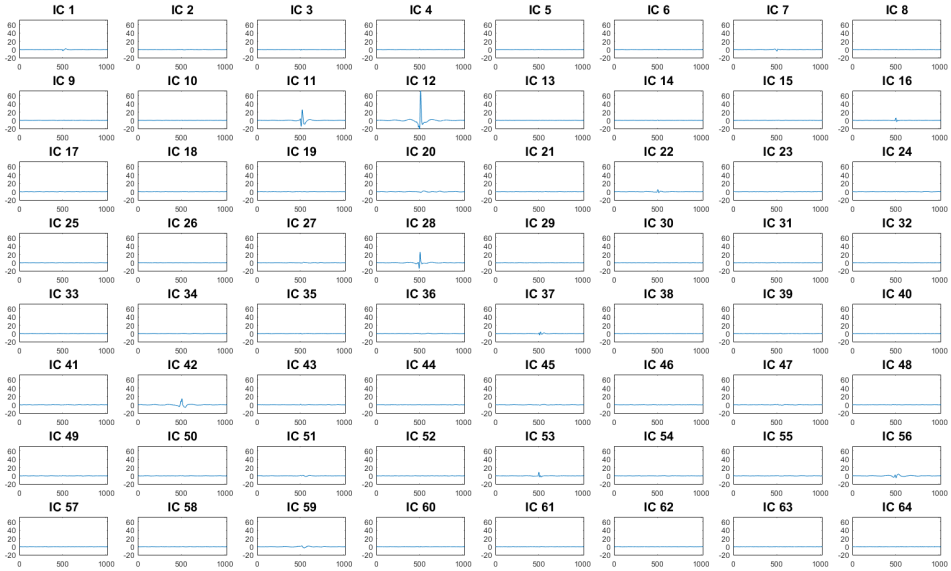

Figure 20: Mean spike snippets of the ICA for a second exemplary data set of patient 1. IC11, IC12 and IC28 seem to be relevant components.

spread over multiple channels. Figure 19 indicates, that the true sources of the spikes are represented by three ICs, which show significant spikes (IC16, IC17, IC33). The same picture can be seen for the other data sets of that patient. For one other data set this is shown in figure 20. Again three components show significant spikes, namely IC11, IC12 and IC 28.

As the ICs are not ordered (like the components in the PCA are, according to their importance) across the different data sets, they had to be matched to the corresponding ICs from the other data sets. For this purpose, the mean spike snippets for the components, that had a significant spike were extracted for all data sets and were then matched. Matching was performed by calculating the correlations between the weight vectors of the ICs of interest and matching those with highly correlated weight vectors. In figure 21 the weight vectors of the corresponding ICs are plotted above each other. Each subplot contains four weight vectors, one from each of the four data sets. The weight vectors that are plotted in one subplot had a high correlation with each other and the corresponding ICs were therefore matched.

The result of that matching process can be seen in figure 22, where the relevant mean IC spike snippets of one data set are plotted in the same column. In each row, the matching IC snippets across the data sets are shown. This image gives the impression, that the spike snippets of the matched ICs look quite similar

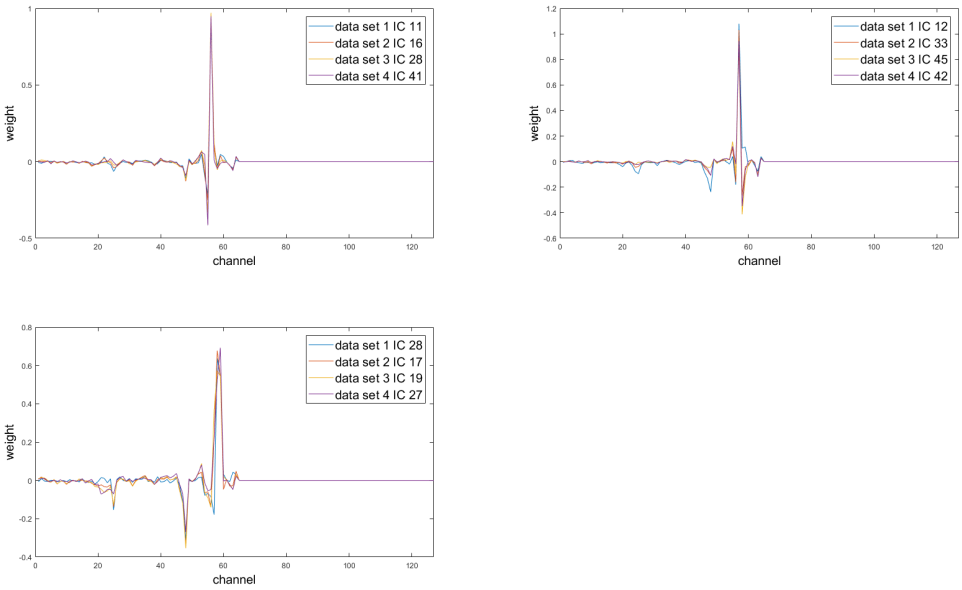

Figure 21: In each subplot four weight vectors are plotted, one for each data set. The weight vectors in one subplot are the ones, that were found to be equal (high correlation) and were therefore matched. The higher a value is, the more a channel contributes to a component.

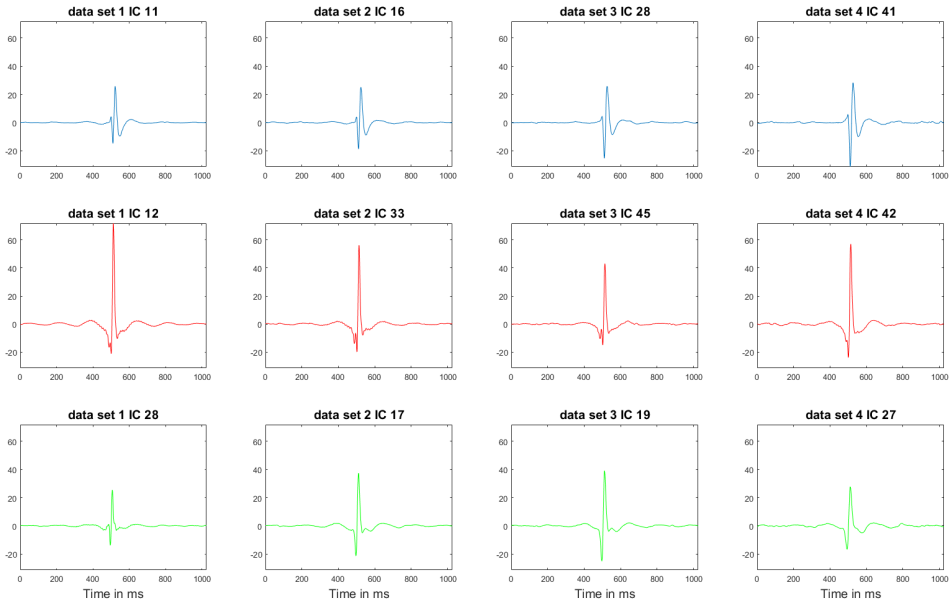

Figure 22: As the ICs are not ordered, they need to be matched over the four available data sets of patient 1. In each column the ICs, that showed significant spikes were extracted for all data sets. The weight vectors of the ICs, that are plotted in the same colour had a high correlation with each other and were therefore matched.

and thus the matching process worked well.

For the further analysis all snippets around the spikes were saved for each component. This results in a data set, containing 564 spike snippets, each with a length of 1 second for each of the three ICs. The names of the relevant ICs have been changed. The IC in the first row in figure 22 is called 'IC1', in the second row 'IC2' and in the third 'IC3'.

For image 23 the mean values over all 564 spike snippets of the three different ICs of interest are plotted. One can see that the red curve (representing IC3) spikes first, then the yellow (IC2) and black curve (IC1) follow. So one could speculate that IC3 drives IC2 and IC1. It is hard to distinguish what happens before the spikes. It can not be seen if one of the ICs gives an impulse to the other ICs.

It is also interesting to take a look at the spatial distribution of the ICs. The projection matrix of the ICs are plotted in figures 24, 25 and 26. The boxes next to each channel indicate the mean value (across all spike snippets) of the projection matrix for that channel. The brighter the colour of such a box, the higher the value for the specific channel and the closer the IC is located to that

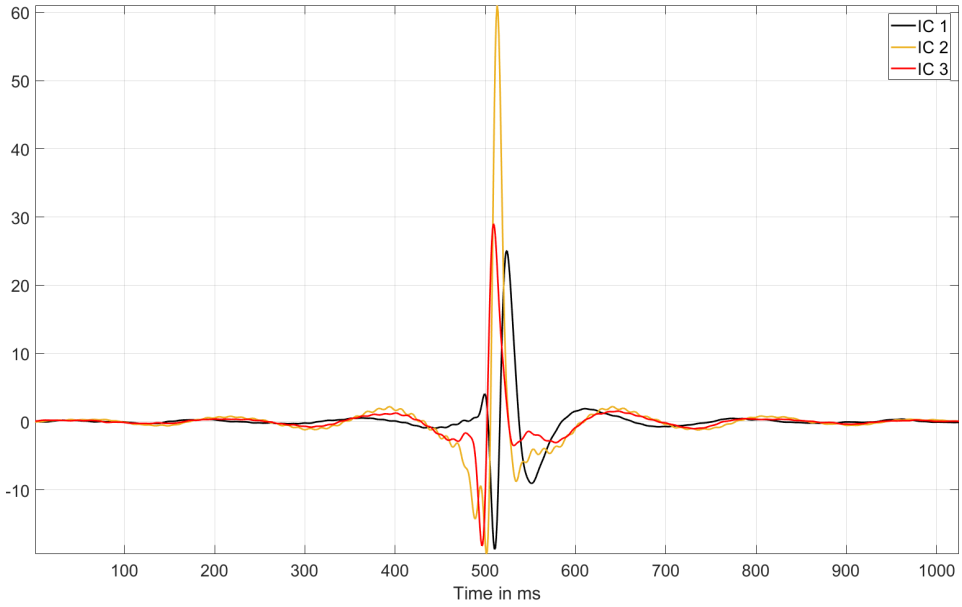

Figure 23: Taking all spike snippets for each of the ICs over all data sets and averaging them shows how the ICs are temporally related. At first IC3 is spiking, then IC2 and IC1 are following.

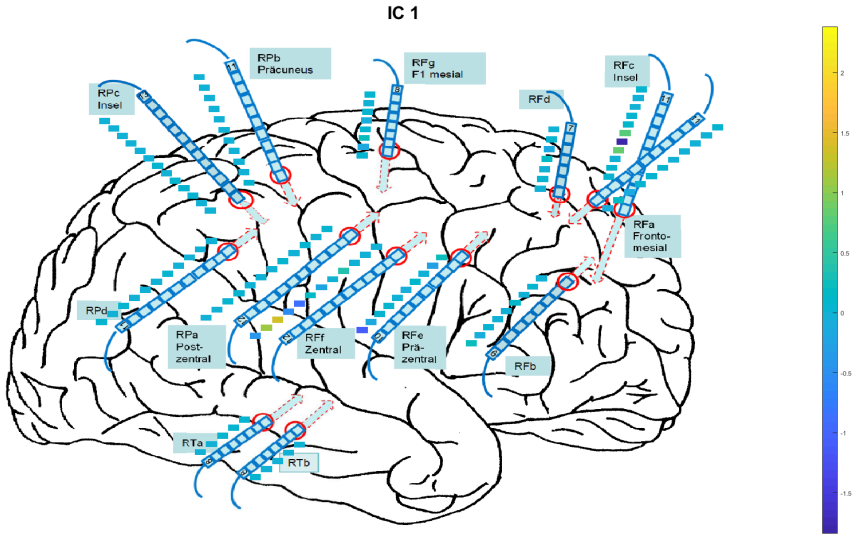

Figure 24: Representing the projection matrix of IC1 spatially. High values can be seen for the channels 'RFf10' and 'RFf11'. The true source might therefore be in between these two channels.

channel.
For IC1 (figure 24) the highest values can be seen for the channels 'RFf10' and 'RFf11'. The other channels do not contribute a lot, so there is reason to believe that the true source lays close to 'RFf10' and 'RFf11'.

For IC2 (figure 25) the channels 'RFf8' and RFf5' stand out from the values of the other channels. The true source might therefore be located anywhere around the RFf-electrode in a depth between the locations of the channels 'RFf8' and RFf5'.

Plotting the values for IC3 (figure 26) shows only one high value, which is located at the channel 'RFf9'. The other values seem to be quiet close to zero. The third component could therefore be close to the channel 'RFf9'.

Even if the electrodes, are spread widely over the brain of the patient, the ICA indicates that the true sources of the spikes lay close to the electrode RFf. This might be a hint, that the spike patterns are quite local for that patient.

5.1.2 Causal Analysis

With the ICs that were found based on the data of patient 1, it is the aim to find potential causal relationships between these components to get a better understanding of how the spikes might propagate.
The causal analysis is based on a data set, that contains 564 snippets around

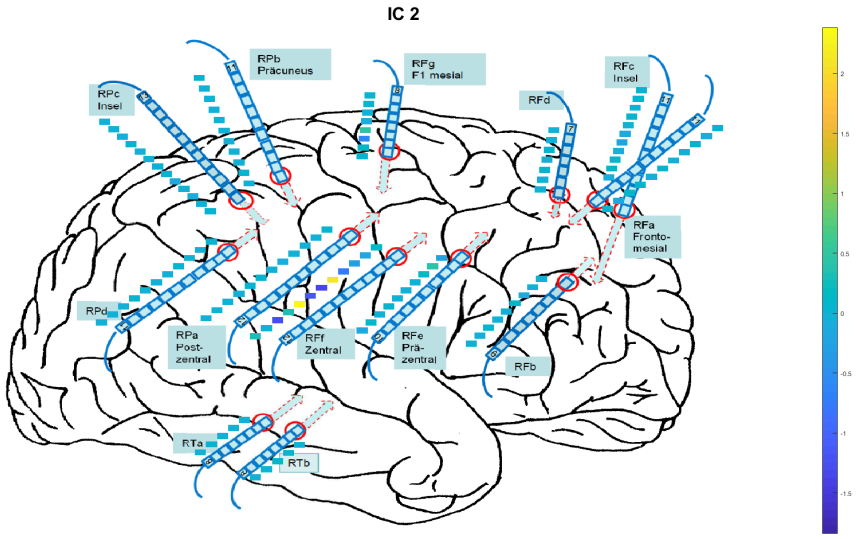

Figure 25: Representing the projection matrix of IC2 spatially. High values can be seen for the channels 'RFf8' and RFf5'. IC2 might therefore be located between the two channels.

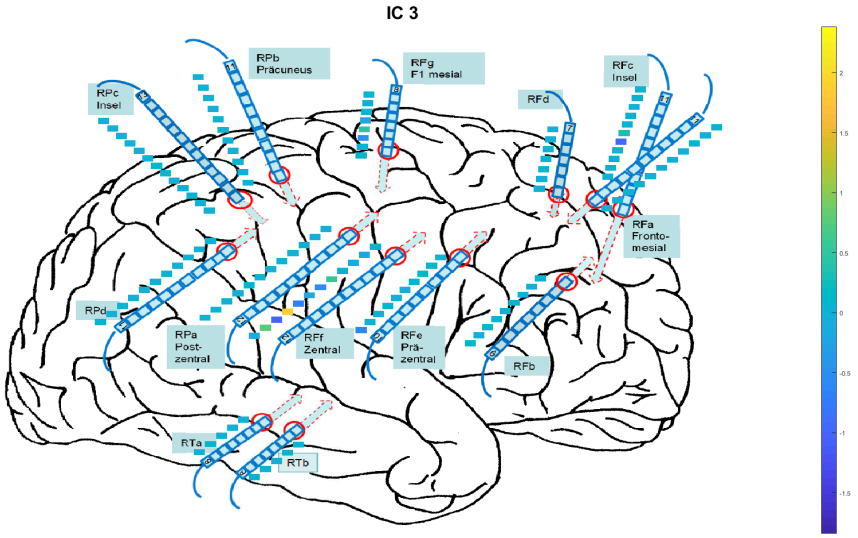

Figure 26: Representing the projection matrix of IC3 spatially. Only one channel ('RFf9') shows a high value. IC3 might be close to that channel.

the spikes, each with a length of 1 second for each of the three ICs that were calculated. As described further above, it is investigated if time delayed causal chains can be found between the ICs. The results of these tests are presented in figure 27. Each row shows the test results of one of the six tested chains. The names of the ICs are shortened and are only represented by the number of the component (e.g. 'IC1' becomes '1'). Also the time lags for which the ICs are tested are not written down in the image, for reasons of transparency. The correct notation of the chain in the top row of figure 27 ($1 \rightarrow 2 \rightarrow 3$) is therefore: $\text{IC1}(t) \rightarrow \text{IC2}(t + \delta_t) \rightarrow \text{IC3}(t + 2\delta_t)$. A test for each chain is performed for all points in time within the spike snippets and considering time delays $\delta_t = \{1, ..., 100\}$. The coloured bar at each point in time shows how many chains were positively tested. The maximum value that could be reached is 100, which would mean that for each time delay (for $\delta_t = \{1, ..., 100\}$) a chain was detected. It can be seen that the most chains were detected in the centre of the one second span. There are some findings slightly to the left and to the right of the centre. The undermost row, representing the chain $\text{IC3}(t) \rightarrow \text{IC2}(t + \delta_t) \rightarrow \text{IC1}(t + 2\delta_t)$, shows the brightest colours, indicating that the most chains were confirmed for that set-up.

Looking at figure 28 shows for which time delays the most causal chains were detected. It can be seen that, generally, the smaller the tested time delays were the more chains were detected. The most positively tested chains were found for delays of about seven to nine milliseconds. This is a reasonable finding as

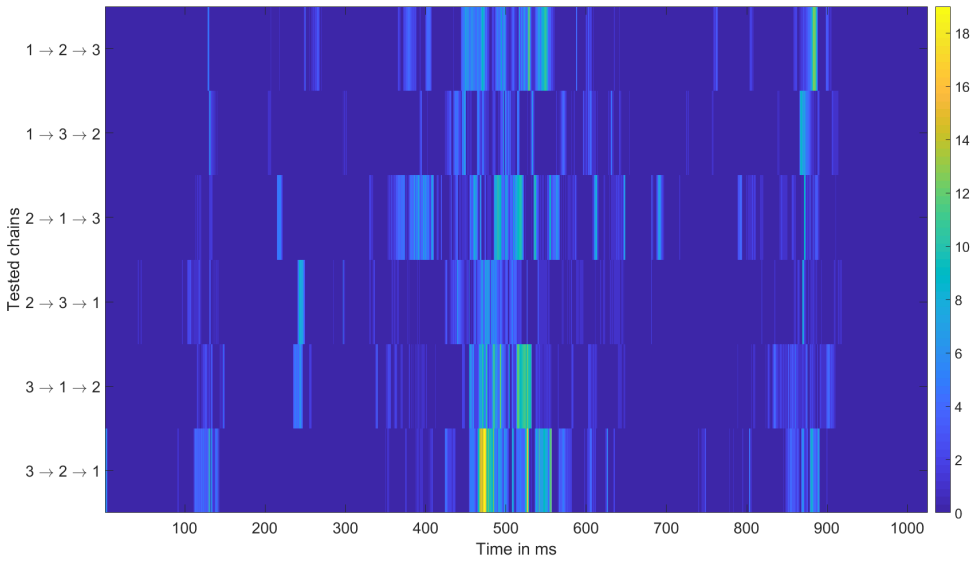

Figure 27: All chains that have been tested. The numbers represent the ICs. Each arrow indicates a lag of δ_t. The brighter the bars get, the more causal chains have been detected for that point in time.

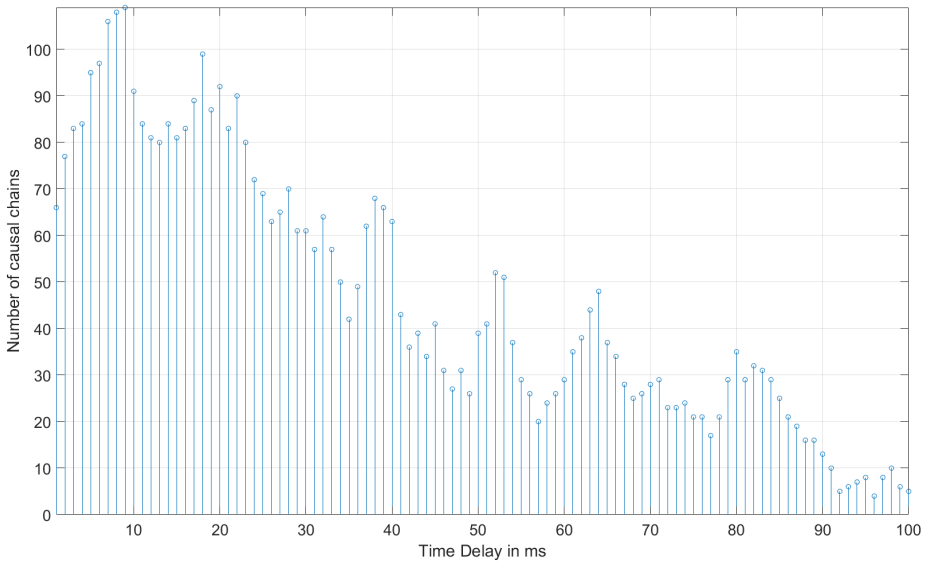

Figure 28: Represents the number of causal chains that have been found over all tested chains for all time delays that were examined. The greater the time delay was, the less causal chains were found. The most causal chains were found for a time delay of about 9 milliseconds.

the ICs are located close to each other (see figures 24 to 26) and thus a potential propagation might happen quite quickly.

To get a better feeling when the assumptions for a chain are fulfilled, relatively to the occurrence of spikes, figure 29 gives a good impression. For each point in time over all tested set-ups, the number of found chains is summed up and plotted (represented by the blue stems). It can be seen that the most chains are found, when the spikes of the ICs occur. But interestingly there are small clusters at time points before and after the spikes appear. So the different ICs seem to already interact before the actual spikes occur.

As already mentioned above, for a chain $X \to Y \to Z$, where all conditions for a chain were met it can be proven that Y is a genuine cause for Z but it is still possible that a confounder is in between X and Y [8]. For that reason the 'endings' of the chains from figure 27 were summarised in figure 30. The figure is splitted into three sections (marked by the horizontal black lines) to explore the relationships between all combinations of two of the ICs. Let us for example consider the upper part of that image. Here the relationship between IC2 and IC3 is represented. Therefore the chains $\mathrm{IC1}(t) \to \mathrm{IC2}(t+\delta_t) \to \mathrm{IC3}(t+2\delta_t)$ and $\mathrm{IC1}(t) \to \mathrm{IC3}(t+\delta_t) \to \mathrm{IC2}(t+2\delta_t)$ from figure 27 are used. For each point in time it was compared if an IC was more often a driver or a receiver compared to the other IC. So for the example with IC2 and IC3, it was analysed for each

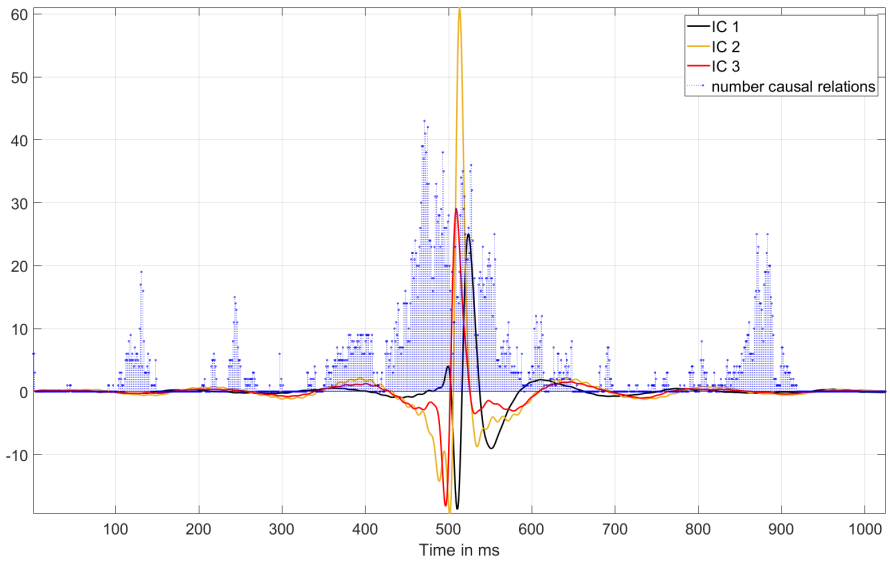

Figure 29: The number of positively tested chains are plotted with blue stems, relative to the mean spike snippets of the ICs. The most causal chains were found close to the spikes. But also some clusters with a high number of detected chains, before and after the spikes can be seen.

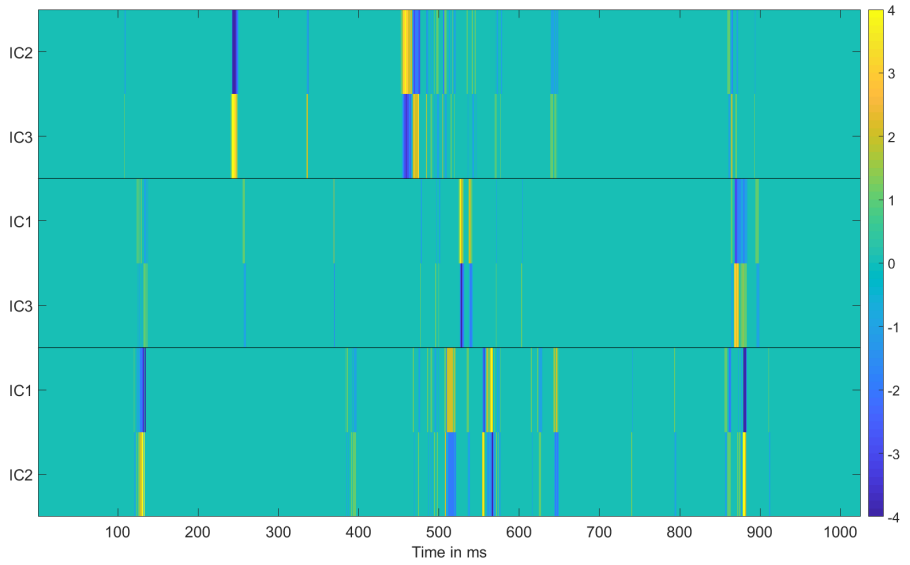

Figure 30: Relation between the different ICs. The figure is splitted into three parts (separated by the black lines) to examine the relationships between all ICs. If a bar is yellow, then an IC more often is a driver than a receiver for that specific point in time, compared to the other IC in the same section. If a bar is blue it is the other way round and the IC is more often a receiver than a driver.

point in time if more chains with the 'ending' IC2$(t+\delta_t) \rightarrow$ IC3$(t+2\delta_t)$ or with the 'ending' IC3$(t+\delta_t) \rightarrow$ IC2$(t+2\delta_t)$ were detected. If an IC was more often a driver at a certain point in time than it was a receiver, the bar was coloured yellow. If it was a receiver more often than a driver, the bar got painted blue. The more intense the yellow or blue the bars are, the more often the IC was a driver or receiver, respectively. To get a more stable result only time delays of $\delta_t = \{9, 10, 11, 12\}$ were used for this plot (these were the time delays where a volley in figure 28 was observed). For none of the combinations it can be said that one IC is always the driver of another IC. It rather looks like the directions of the relations between the ICs are fluctuating over the time course.

To get a better inside, a DAG was created, which is shown on figure 31. The arrows indicate a relation between two ICs, where the direction of an arrow shows if an IC is driver or receiver. The arrows are also colour coded (blue: relation between IC1 and IC2; green: relation between IC1 and IC3; red: relation between IC1 and IC3). The thicker an arrow is, the more causal chains were detected for that point in time. Like in the images above it can be seen that the most interactions take place in the middle, where the spikes occur. Some causal relations are existing before and after the spikes. IC1 and IC3 seem to have the fewest interaction. This plot confirms the finding from before that a

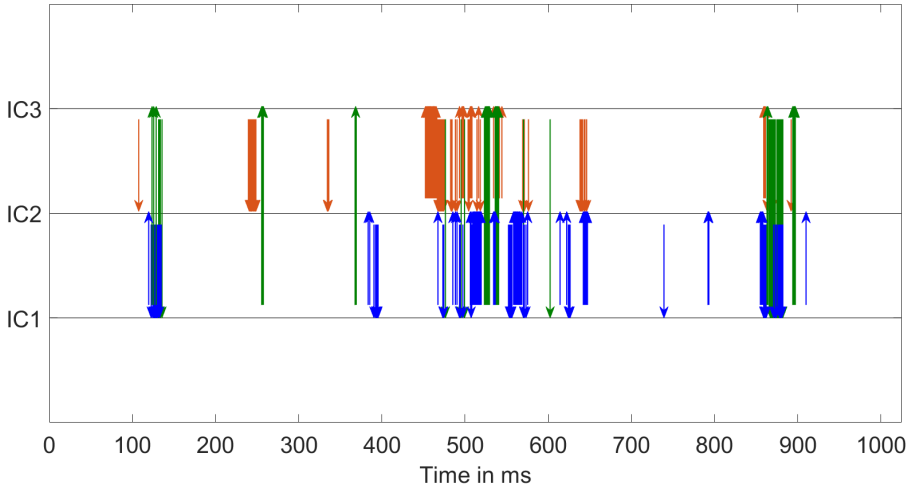

Figure 31: The DAG represents the relations between the ICs. The direction of an arrow indicates which IC is the receiver and which one is the driver. The thicker a line the more causal relations between to ICs were found at a certain point in time.

relation between two variables, such that one IC always drives another IC, can not be found. It is still a very interesting finding that there seems to exist a kind of feedback system between the ICs. Also the outcome that there might be interaction between the ICs before the spikes appear is promising.

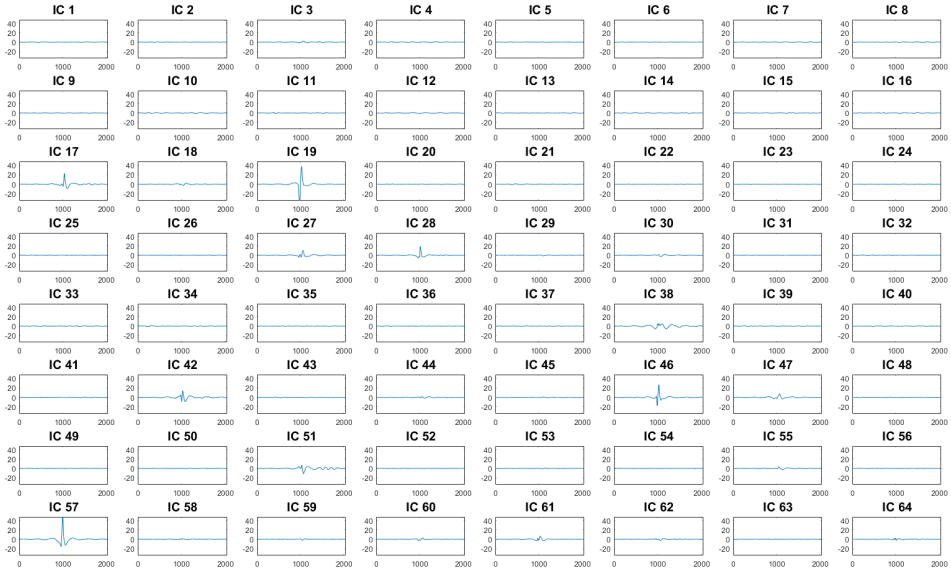

Figure 32: Mean of the spike snippets from the ICs of patient 2 for an exemplary data set. Multiple components show spikes: IC17, IC19, IC28, IC42, IC46, IC51 and IC57

5.2 Patient 2

5.2.1 ICA

Performing the ICA and taking the mean over all spike snippets for the data sets of the second patient gives a result that can be seen in figure 32 for one exemplary data set. The picture is not as clear as it was for patient 1. Multiple components show a spike within the mean snippet (IC17, IC19, IC28, IC42, IC46, IC51, IC57). Comparing the resulting components of this data set to the components found in the other data sets shows that four found components can be matched to ICs, found in the other ones. The ICs that could be matched over the four data sets are shown in figure 34 and the according weight vectors in figure 33. The ICs are renamed after they were matched, according to their row in figure 34 (first row $\hat{=}$ IC1,...,fourth row $\hat{=}$ IC4).

Taking the points in time of all 481 spike snippets and averaging them, for each of the four ICs gives an outcome, represented in figure 35. The green component (IC4) spikes first, afterwards IC2 and IC3 reach their peaks at approximately the same time and IC1 is the last spike occurring. Before the spikes occur, it is hard to distinguish if there are relationships between the different ICs.

Figures 36, 37, 38 and 39 show at which locations high values of the projection

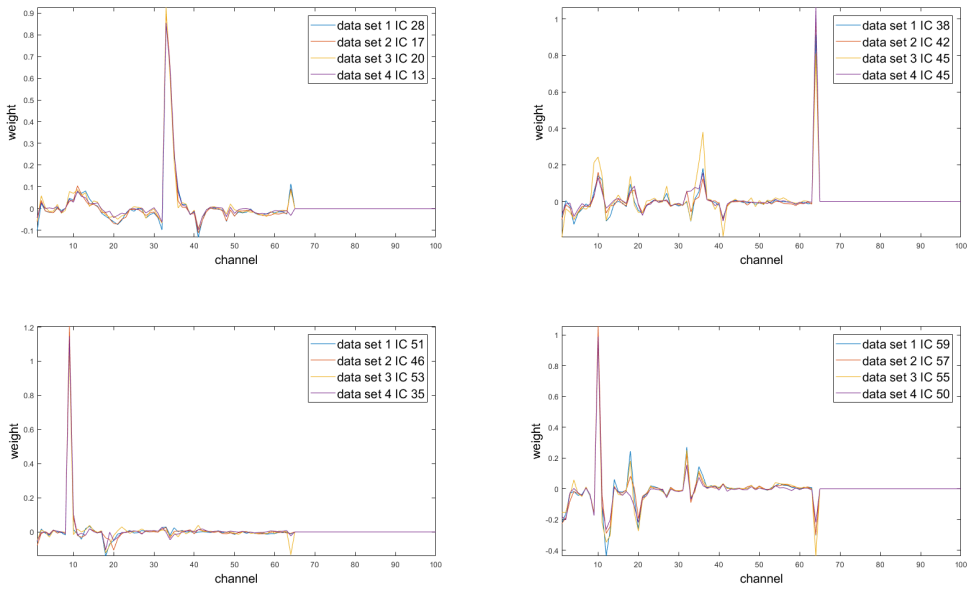

Figure 33: The plotted weight vectors were used to match the ICs over the four data sets. Each of the subplots shows the weight vectors of the four data sets, that had a high correlation with each other.

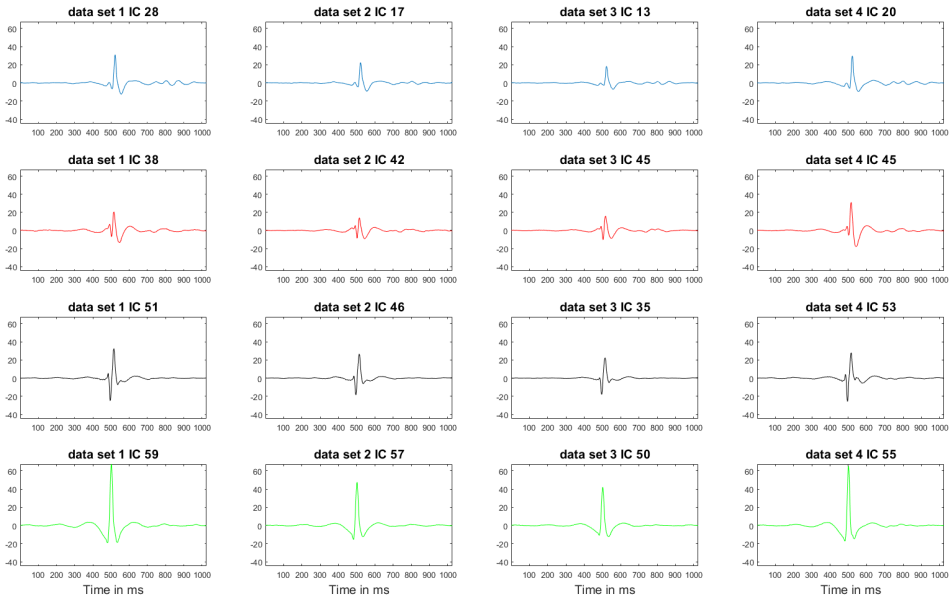

Figure 34: The mean spike snippets over the four data sets, which could be matched due to figure 33. The ICs with the same colour were matched.

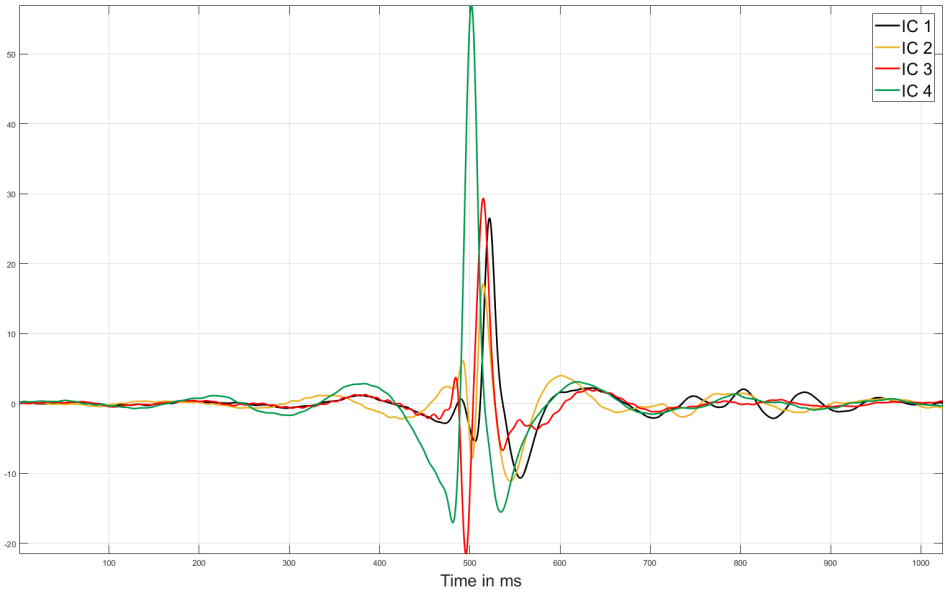

Figure 35: Each curve represents the mean over all spike snippets for one IC. The spike of IC4 occurs first. IC2 and IC3 have their peaks at about the same time. IC1 is the last component that spikes.

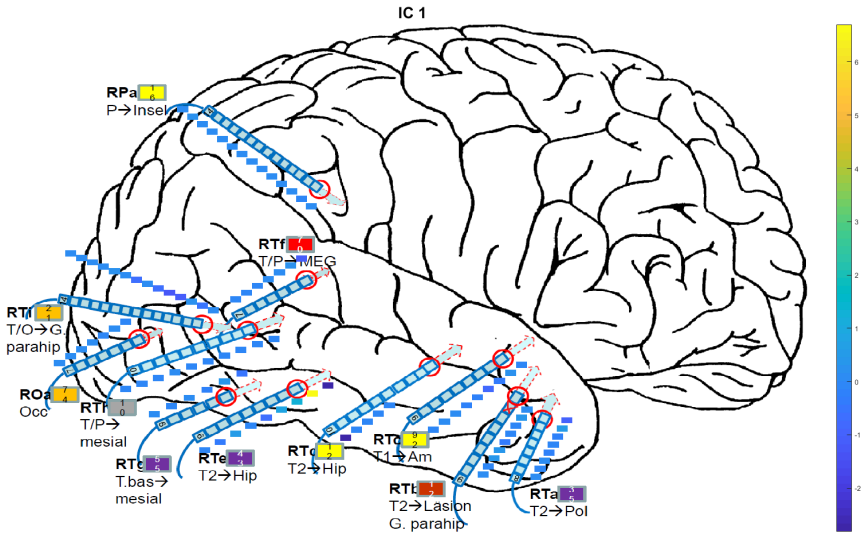

Figure 36: The activation matrix of IC1 shows the highest values for the channel 'RTe2'. The true source might lay close to this channel.

matrix for each component appear.
IC1 shows a high value for the channel 'RTe2', leading to the assumption that the true source lays close to this channel. The spikes were initially marked for channel 'RTb2', from which 'RTe2'is quite a bit away.

The channel that stands out the most for IC2 is 'RTi1', which is even further away from the channel 'RTb2' than the channel 'RTe2'. Another outsatnding value is on the RTe-electrode. The component might therefore be somewhere in between the channels ,'RTi1' and 'RTe5'.

IC3 and IC4 have there maximum values of the projection matrix on the channel 'RTb2' and 'RTb3', respectively. Thus the true sources will be close to the corresponding channels, meaning that the electrode RTb was placed close to two potential true sources.

As the ICs are spread over a wide range in one hemisphere of the brain, this could indicate that spikes are propagated over a wide distance in the brain of the patient.

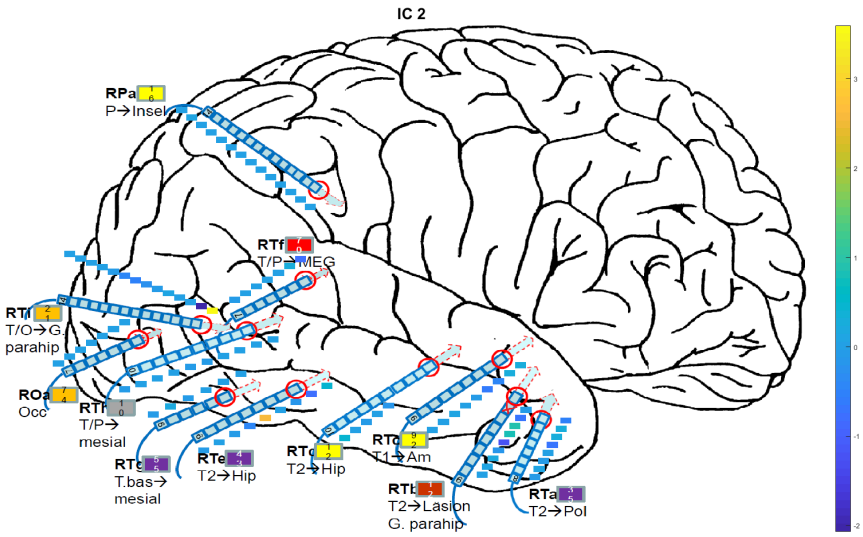

Figure 37: The highest activation value for IC2 is located at channel 'RTi1', which is quite far away from the channel where the spikes were initially marked for ('RTb2').

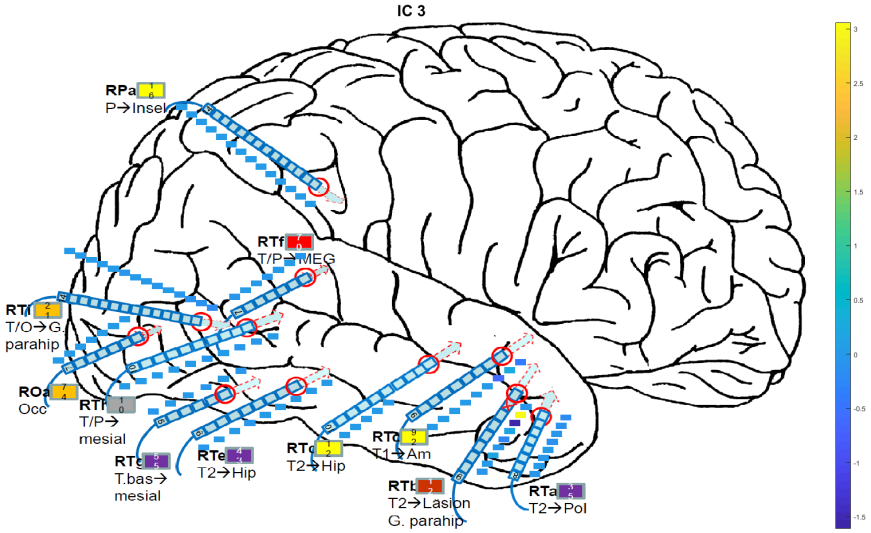

Figure 38: 'RTb2' is the channel that shows the highest value for IC3. This means that the electrode is close to a true source.

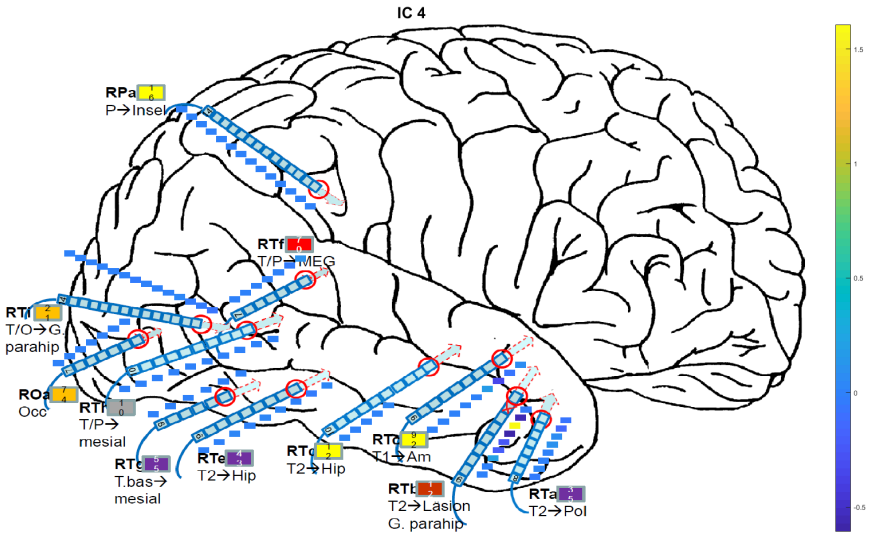

Figure 39: For IC4, the highest value of the activation matrix can be seen for channel 'RTb3'. IC4 will therefore be placed close to this channel.

5.2.2 Causal Analysis

The data set, that was used for the causal analysis of patient 2 consisted of 481 spike snippets, each with a length of 1 second (2048 samples) for each of the four ICs. To understand the causal relations between the ICs and to see how spikes might propagate, the same analyses as for patient 1 were conducted. This time the analyses will get a bit more complex because 4 ICs were found and thus the number of chains, that had to be tested increased.
The number of positively tested chains rose, compared to patient 1 and can be seen on figure 40. The appearance of this figure differs noticeable from the corresponding one of patient 1 (figure 28). Not only the total number of causal chains increased, but also the number of chains with respect to the time delays changed. For patient 1 it was determined that a higher time delay resulted in a lower number of causal chains. Here the number stays at a quite constant level for a delay up to 15 ms and then falls to a lower level but stays fairly the same up to the maximum tested time delay. One possible reason for that different appearance might be that the sampling rate for patient 2 was twice as high as for patient 1 (2048 Hz vs. 1024 Hz). This means that when converting the time delay from sample units into milliseconds, the maximum time delay tested was half as large for patient 2 compared to that of patient 1 ($\approx$ 50 ms vs. $\approx$ 100ms). Looking at the corresponding figure for patient 1, shows that this does not explain the whole difference. For patient 1 the number of causal chains has already dropped a lot for a time delay of 50 ms, whereas the number of causal chains for a time delay of 50 ms is still high for patient 2. Another reason for that contrast might be that the ICs are far more spread for patient 2 than for patient 1 (see figures 36 to 39). So if their exists a causal relationship between ICs, that are far apart from each other, the propagation of spikes takes longer and thus causal relations might be found for greater time delays, than for ICs that are close to each other.

Considering figure 41, all tested chains can be seen and equivalent to the same plot for patient 1, the most chains were detected in the middle of the image but also some chains were tested positively for earlier and later points in time.

To get a feeling for the appearance of the causal chains, with respect to the timing of the spikes, image 42 can be considered. It can be seen that the most detected chains can be found close to the spikes. The number of the causal chains remains at a high level even after the spikes have occurred. The interesting pattern that there are two clusters of causal chains before the spikes start is equal to patient 1. So again there is reason to believe that an interaction between the ICs takes place before the actual onset of the spikes.

Breaking down the results of the analysis to the 'endings' of the chains gives figure 43. For the most combinations it looks like the direction of the interaction between the combinations of the ICs fluctuates in the course of time. Only for the relationship between IC2 and IC4 it looks like IC4 is far more often the driver of IC2 than the other way around.

Condensing figure 43 to a DAG results in figure 44. As there are a lot of chains, the image looks a bit unclear. Still the same patterns that were shown

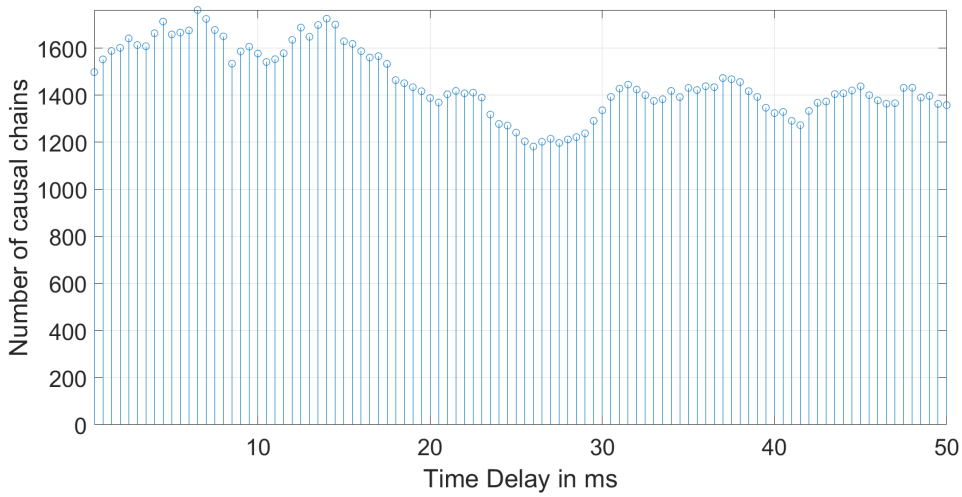

Figure 40: The number of detected causal chains, relative to the time delay. The number stays at a constant level for the first 15 ms, then drops a bit and stays at about the same level until the maximum time delay that was tested.

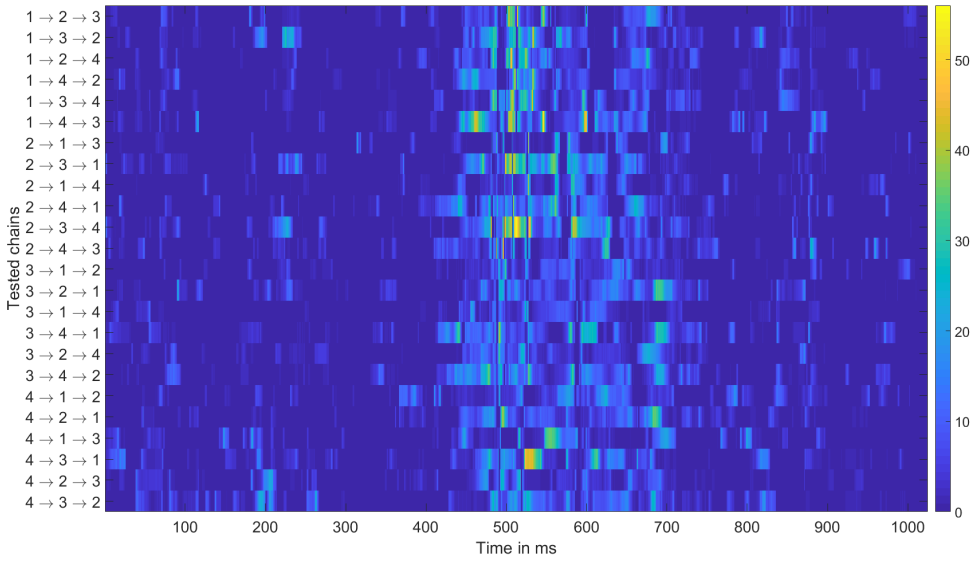

Figure 41: All chains of ICs that were tested. The most positively tested chains can be seen in the middle of the image. Also in the beginning and at the end some chains were detected.

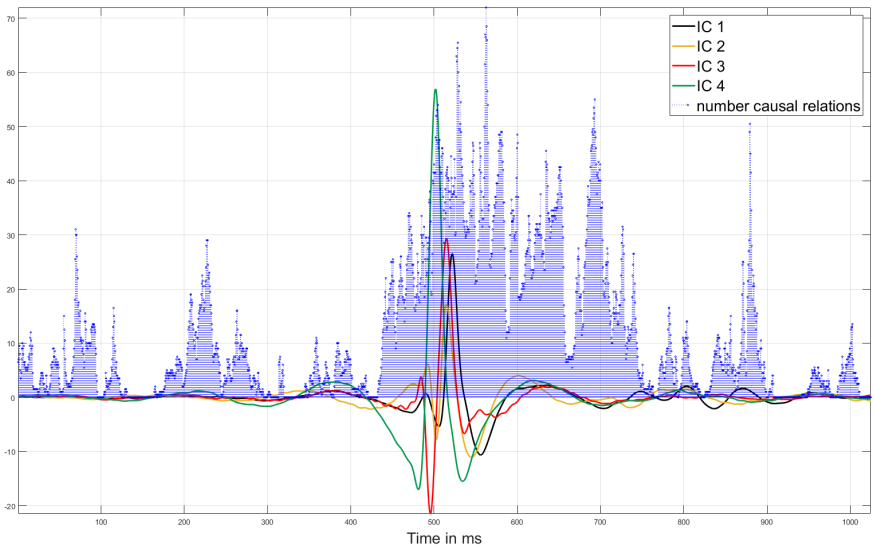

Figure 42: The number of causal relations relative to the appearance of the spikes. The most chains were detected at the time points where the spikes of the ICs appear. Two clusters with a high number of causal chains can be seen before the spikes start.

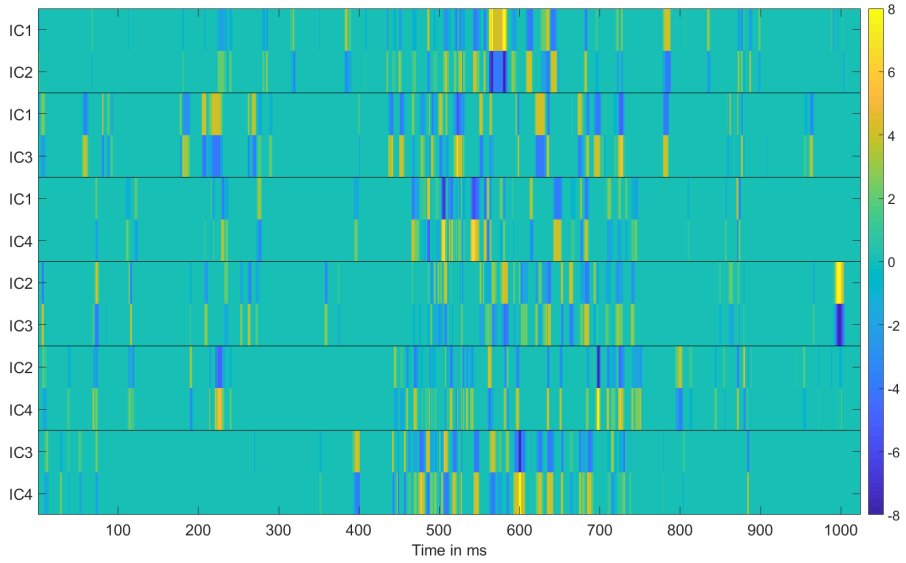

Figure 43: Relations between the ICs. For almost all ICs the direction of the relationship changes. For the relationship between IC2 and IC4 it seems that IC4 is the driver and IC2 the receiver more often.

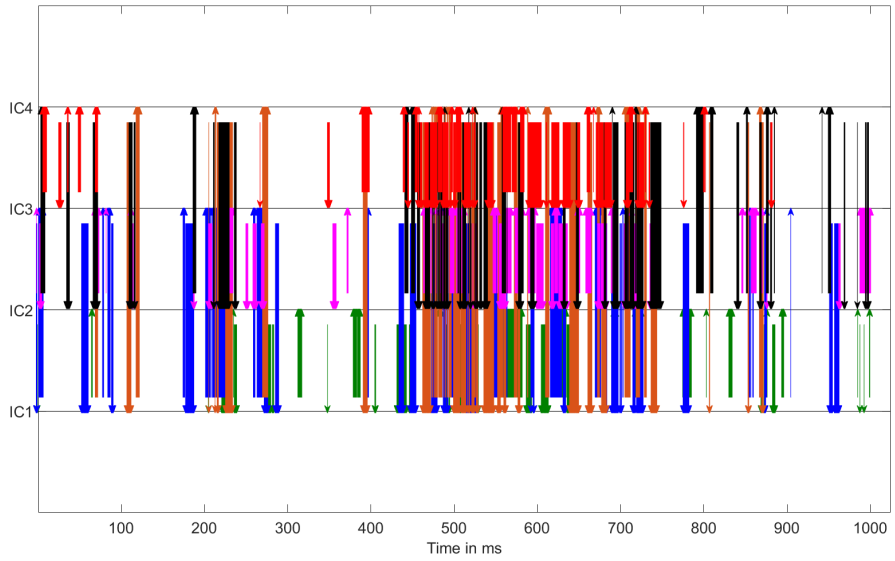

Figure 44: DAG for the four ICs, which confirms the fluctuating relationships between the ICs. Also the relationship between the components before the start of the spikes can be seen.

on the figures before can be recognised here again. There is a heavy interaction between the ICs, especially in the middle of the image at the time where spikes appear. Also before and after the spikes there is a relationship between the ICs.

6 Discussion

The results of the analyses of the data sets of the two patients contained interesting insights and should therefore be examined further. It was interesting to see that the ICA was able to break down the number of true sources to only a couple of components. Even though the causal analysis on the ICs could not find an explicit causal structure, it is promising that an interaction between the independent components was found. A further interesting outcome is the existence of causal relations that show up already before the actual spikes start. The findings can be of clinical interest because they might help localising the epileptogenic zone and could help to understand where spikes start and how they propagate with respect to different brain areas. From these findings there are a couple of critical questions that arise and objectives that might be tackled in future work.

A possible pitfall of this work is that only one form of spikes was used to create a template with which the spikes were detected. It should be considered to use a library that contains multiple templates and performs the detection of spikes. Maybe in the future there will be broadly tested toolboxes that can master this step. It would be interesting to see if and how the analysis would change when another spike detection algorithm is used, as the shape of interictal spikes might change when it is propagated to another location.

It has also to be checked if it is valid to average the data sets of the sleep and awake phases. The findings of the spike detection as well as the results of the ICA seemed quite similar between the sleep and awake data sets, nevertheless a separate analysis might be useful. Therefore different templates for spike detection should be used and it might be reasonable to test if the ICA on the different data sets gives a similar result.

For computational reasons only a small number of time delays from 1 to 100 samples was tested. An open question is if there are also relationships between the components when considering a time delay of more than 100 ms and 50 ms respectively. It should also be thought about enlarging the time window of the spike snippets in order to check if there are causal structures in time windows that have not been observed yet. If the spikes occur in small intervals then there is of course a limitation to the maximum time delay and the length of tested spike snippets that can be tested. Also the time lags between the different components in the test for causal chains should be varied. In this thesis the time delay was always δ_t. Varying the time difference between the first and the second and the second and the third IC might give interesting results as well. It might be sensible to focus on time delays corresponding to the distance of ICs. Which means one should investigate if greater distances of the ICs go hand in hand with greater time delays of causal relations, and if it might be useful to focus on these time delays.

At the current status, running the analysis is quite time consuming because some steps in the analysis are not yet automatised and have therefore been done by hand. This is because the properties of the recordings for each patient differ and the analysis is therefore hard to generalise. For example finding the

relevant ICs and matching them over the different data sets for a patient might be a step that could be automatised as well as the causal analysis, which effort depends strongly on the number of ICs that is found.
Finally it is necessary to conduct the same analysis on a larger number of patients to ensure the outcomes and to make the findings more reliable.

Acknowledgement

I would like to thank Prof. Dr. med. Hajo Hamer, Dr. med. Johannes Lang and Dr. rer. nat. Rüdiger Hopfengärtner from the *Universitätsklinikum Erlangen* for providing and labelling the data and for giving clinical background on the data and possible explanations for the findings.

References

[1] A closer look at eeg. URL https://www.epilepsysociety.org.uk/closer-look-eeg#.XSRNlXduJaR.

[2] Brain surgery for epilepsy. URL https://www.epilepsysociety.org.uk/brain-surgery-epilepsy#.XShG73duJaR.

[3] Epilepsy. URL https://www.who.int/news-room/fact-sheets/detail/epilepsy.

[4] Laetitia Chauvière, Thomas Doublet, Antoine Ghestem, Safia S Siyoucef, Fabrice Wendling, Raoul Huys, Viktor Jirsa, Fabrice Bartolomei, and Christophe Bernard. Changes in interictal spike features precede the onset of temporal lobe epilepsy. *Annals of neurology*, 71(6):805–814, 2012.

[5] O'Brien Terence J. Jette Nathalie Scheffer Ingrid E. de Curtis Marco Perucca Piero Devinsky Orrin, Vezzani Annamaria. Epilepsy. *Nature Reviews Disease Primers*, 4(18024), 2018. URL https://doi.org/10.1038/nrdp.2018.24.

[6] Jerome Engel. Update on surgical treatment of the epilepsies: summary of the second international palm desert conference on the surgical treatment of the epilepsies (1992). *Neurology*, 43(8):1612–1612, 1993.

[7] J Gotman and MG Marciani. Electroencephalographic spiking activity, drug levels, and seizure occurence in epileptic patients. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 17(6):597–603, 1985.

[8] Moritz Grosse-Wentrup, Dominik Janzing, Markus Siegel, and Bernhard Schölkopf. Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach. *NeuroImage*, 125: 825–833, 2016.

[9] Harinder Jaseja and Bhanu Jaseja. Eeg spike versus eeg sharp wave: Differential clinical significance in epilepsy. *Epilepsy & Behavior*, 25(1):137, 2012.

[10] Philippa J Karoly, Dean R Freestone, Ray Boston, David B Grayden, David Himes, Kent Leyde, Udaya Seneviratne, Samuel Berkovic, Terence OBrien, and Mark J Cook. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain*, 139(4):1066–1078, 2016.

[11] Shaun S Lodder and Michel JAM van Putten. A self-adapting system for the automated detection of inter-ictal epileptiform discharges. *PloS one*, 9 (1):e85180, 2014.

[12] Hans O Lüders, Imad Najm, Dileep Nair, Peter Widdess-Walsh, and William Bingman. The epileptogenic zone: general principles. *Epileptic disorders*, 8(2):1–9, 2006.

[13] Plonsey Malmivuo, Jaakko Malmivuo, and Robert Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.

[14] Chinmay Nagesh, Savith Kumar, Ramshekhar Menon, Bejoy Thomas, Ashalatha Radhakrishnan, and Chandrasekharan Kesavadas. The imaging of localization related symptomatic epilepsies: The value of arterial spin labelling based magnetic resonance perfusion. *Korean journal of radiology*, 19(5):965–977, 2018.

[15] Ferruccio Panzica, Giulia Varotto, Fabio Rotondi, Roberto Spreafico, and Silvana Franceschetti. Identification of the epileptogenic zone from stereo-eeg signals: a connectivity-graph theory approach. *Frontiers in neurology*, 4:175, 2013.

[16] Judea Pearl et al. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009.

[17] Felix Rosenow and Hans Lüders. Presurgical evaluation of epilepsy. *Brain*, 124(9):1683–1700, 2001.

[18] Helen R Sabolek, Waldemar B Swiercz, Kyle P Lillis, Sydney S Cash, Gilles Huberfeld, Grace Zhao, Linda Ste Marie, Stéphane Clemenceau, Greg Barsh, Richard Miles, et al. A candidate mechanism underlying the variance of interictal spike propagation. *Journal of Neuroscience*, 32(9): 3009–3021, 2012.

[19] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[20] SJM Smith. Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76 (suppl 2):ii2–ii7, 2005.

[21] Peter Spirtes, Clark N Glymour, Richard Scheines, David Heckerman, Christopher Meek, Gregory Cooper, and Thomas Richardson. *Causation, prediction, and search*. MIT press, 2000.

[22] Alaa Tharwat. Independent component analysis: An introduction. *Applied Computing and Informatics*, 2018.

[23] John Thomas, Jing Jin, Justin Dauwels, Sydney S Cash, and M Brandon Westover. Automated epileptiform spike detection via affinity propagation-based template matching. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3057–3060. IEEE, 2017.

[24] Samuel B Tomlinson, Camilo Bermudez, Chiara Conley, Merritt W Brown, Brenda E Porter, and Eric D Marsh. Spatiotemporal mapping of interictal spike propagation: a novel methodology applied to pediatric intracranial eeg recordings. *Frontiers in neurology*, 7:229, 2016.

[25] Pieter Van Mierlo, Evelien Carrette, Hans Hallez, Robrecht Raedt, Alfred Meurs, Stefaan Vandenberghe, Dirk Van Roost, Paul Boon, Steven Staelens, and Kristl Vonck. Ictal-onset localization through connectivity analysis of intracranial eeg signals in patients with refractory epilepsy. *Epilepsia*, 54(8):1409–1418, 2013.

[26] Pieter Van Mierlo, Margarita Papadopoulou, Evelien Carrette, Paul Boon, Stefaan Vandenberghe, Kristl Vonck, and Daniele Marinazzo. Functional brain connectivity from eeg in epilepsy: Seizure prediction and epileptogenic focus localization. *Progress in neurobiology*, 121:19–35, 2014.

[27] Robert B Yaffe, Philip Borger, Pierre Megevand, David M Groppe, Mark A Kramer, Catherine J Chu, Sabato Santaniello, Christian Meisel, Ashesh D Mehta, and Sridevi V Sarma. Physiology of functional and effective networks in epilepsy. *Clinical Neurophysiology*, 126(2):227–236, 2015.

Erklärung der Urheberschaft

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum	Unterschrift