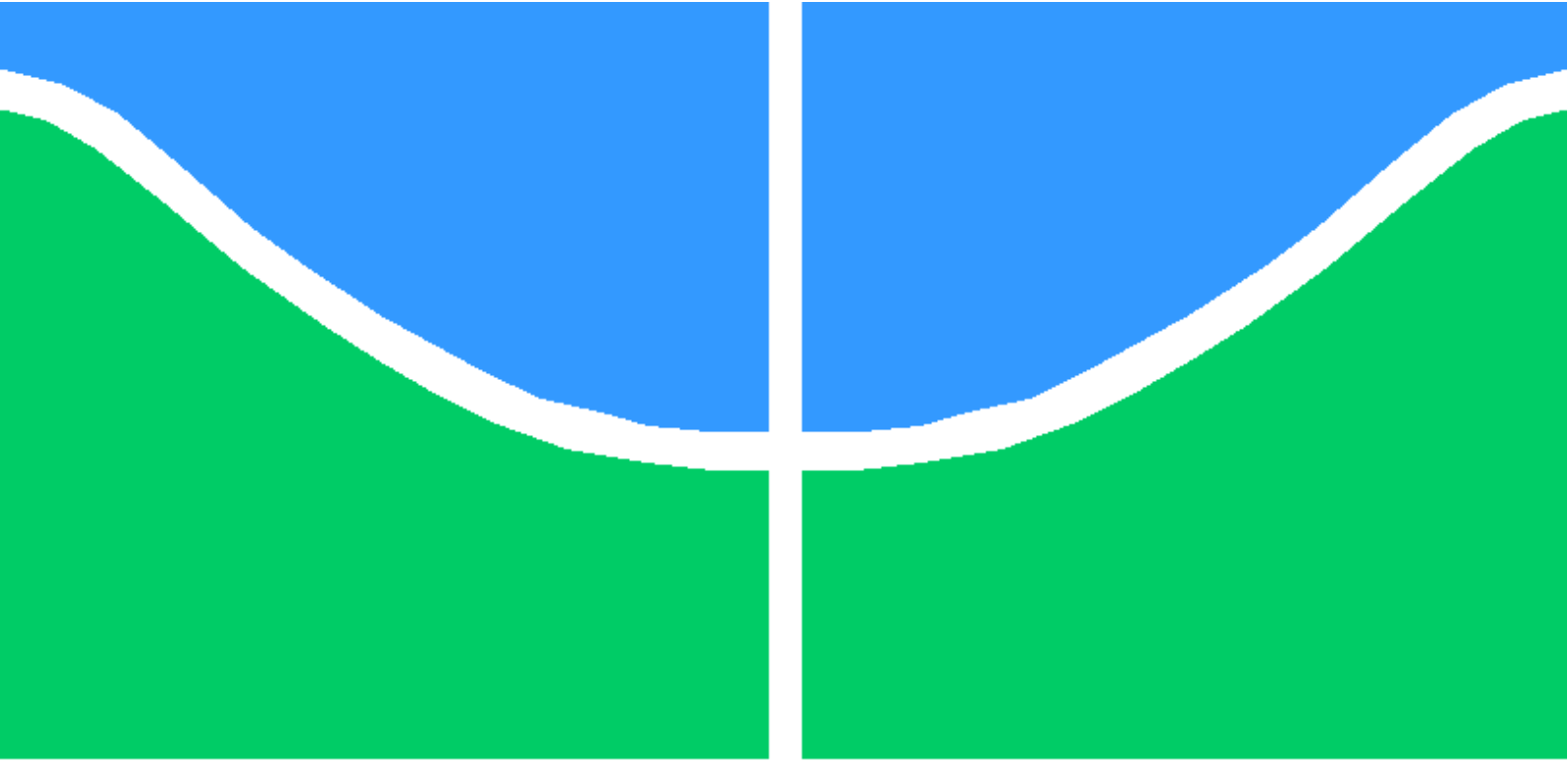


TRABALHO DE GRADUAÇÃO

BLIND SOURCE SEPARATION APPLIED TO THE ANALYSIS OF RAT SOMATOSENSORY EVOKED RESPONSE

Lucas Silva Lopes

Brasília, julho de 2016

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASILIA
Faculdade de Tecnologia

TRABALHO DE GRADUAÇÃO

BLIND SOURCE SEPARATION APPLIED TO THE ANALYSIS OF RAT SOMATOSENSORY EVOKED RESPONSE

Lucas Silva Lopes

Relatório submetido ao Departamento de Engenharia
Elétrica como requisito parcial para obtenção
do grau de Engenheiro Eletricista

Banca Examinadora

Prof. Dr.-Ing. João Paulo Carvalho Lustosa da Costa, ENE/UnB, EMS/Fraunhofer IIS e EMS/TU Ilmenau
Orientador

Prof. Dr. Ricardo Zelenovsky, ENE/UnB
Co-orientador

M. Sc. Ricardo Kehrle Miranda, ENE/UnB
Examinador externo

Prof. Dr. Cristiano Jacques Miosso, FGA/UnB
Examinador interno

Prof. Dr. Leonardo Rodrigues Araújo Xavier de Menezes, ENE/UnB
Suplente

Acknowledgments

First of all, I would like to thank my advisors, João Paulo Carvalho Lustosa da Costa and Ricardo Zelenovsky, for trusting me with such a difficult task and for their valuable teachings during the development of this monograph. Same of these lessons will be decisive in my academic career.

I would also like to thank Dr. Patrik Krieger and Dr. Desire Humanes Valera from Ruhr University Bochum for providing the data analyzed in this work, specially Dr. Patrik Krieger for kindly providing me with some knowledge I could never learn by any other means.

Lucas Silva Lopes

RESUMO

Na última década, o interesse por potenciais de campo local corticais foi renovado. Estes são potenciais elétricos adquiridos de dentro do córtex, geralmente com eletrodos linearmente espaçados inseridos perpendicularmente à superfície do córtex. A ferramenta padrão para a análise de potenciais de campo local de múltiplos eletrodos tem sido análise por densidade de fonte de corrente. Esta permite estimar a corrente por unidade de volume fluindo para dentro (fonte) ou para fora (sumidouro) do espaço extracelular ao redor de um eletrodo. A análise por densidade de fonte de corrente permite estudar conectividade de populações de diferentes camadas do córtex pelo posicionamento das fontes e dos sumidouros de corrente. Neste trabalho, várias técnicas de separação cega de fonte são aplicadas à potenciais evocados adquiridos do córtex somatosensorial de um rato para explicitar associações entre fontes e sumidouros de corrente. As técnicas utilizadas foram análise de componentes principais, análise de componentes independentes, análise de componentes independentes espaço-temporal, análise de componentes independentes triplo-N para misturas convolutivas, e análise de fatores paralelos. Para utilizar com sucesso análise de fatores paralelos com estes dados, uma solução alternativa teve de ser proposta que consiste em aproximar as fases das componentes tridimensionais do tensor de eletrodos por tempo por frequência a serem iguais às do tensor original. Isto permite reconstruir sinais no tempo a partir das componentes e ainda utilizar restrição de não negatividade.

Palavras chave: Córtex Somatossensorial; Potenciais de Campo Local; Análise por Densidade de Fonte de Corrente; Separação Cega de Fontes.

ABSTRACT

In the past decade, the interest on cortical Local Field Potentials (LFPs) has been renewed. These are electric potentials recorded from inside the cortex, generally with linearly spaced electrodes inserted perpendicularly to the cortical surface. The standard tool for the analysis of multielectrode Local Field Potentials has been the Current Source Density (CSD) analysis. This allows to estimate the current per unit volume flowing in (source) or out (sink) the extracellular space around an electrode. The CSD analysis allows to study connectivity of populations from different layers of the cortex by the positioning of the current sources and sinks. In this work several Blind Signal Separation techniques are applied to Evoked Potentials (EP) recorded from the somatosensory cortex of a rat to elicit associations between the current sources and sinks. The techniques used were Principal Component Analysis (PCA), Independent Component Analysis (ICA), Spatiotemporal ICA (stICA), Triple-N ICA for Convolutive Mixtures (TRINICON) and Parallel Factor Analysis (PARAFAC). To successfully use PARAFAC with this data, a workaround had to be proposed

that consists of approximating the phases of the three-way components of the array of electrodes by time by frequency to be equal to the phases of the original array. This allows to reconstruct time signals from each components and still use nonnegativity constraints.

Keywords: Somatosensory Cortex; Local Field Potentials; Current Source Density Analysis; Blind Source Separation

CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Animal experiments have been of great importance for neuroscience. They allow to study the functioning of the brain from much closer than noninvasive methods. To date, it is the best way for studying brain connectivity, how does the numerous neurons in the brain connect to each other to perform complex functions. One of the methods used by the neuroscientists to study the brain are Local Field Potential (LFP) recordings, i.e. electric potentials recorded from the extracellular space of brain tissue. These are particularly suited for bridging the gap between noninvase and single cell measurements [1]. Directly or indirectly, there is always the hope that such experiments can greatly assist in the treatment of human brain diseases.

The first studies using LFP recordings date back to the 1940s [2][3][4], and used no or little signal processing. The two most traditional ways of analyzing the LFP are by means of decomposition in the frequency domain or by Current Source Density (CSD) analysis. CSD analysis was introduced in 1952 by Pitts [5] and was made particularly popular by the works of Nicholson and Freeman [6][7]. Today, CSD analysis is still used as it is in many studies of LFP recordings, for example in [8]. However there is already an acknowledgment by a few authors that the use of Blind Source Separation (BSS) techniques to separate multichannel LFPs or derived CSDs into informative components can greatly assist in the analysis of LFPs.

Several attempts have been made in [9][10][11][12] to separate the contributions of individual neuronal populations by applying Principal Component Analysis (PCA) to CSD waveforms [1]. The constraints assumed by PCA on the obtained components are considered artificial constraints added only for identifiability of the decomposition, and not because of physiological considerations [1][13], requiring posterior rotation of the data to improve physiological interpretation [9]. Independent Component Analysis (ICA), in its turn, has been applied to LFP waveforms in [14][15], although not for the analysis of evoked potentials. Their success in using ICA may be due to the nature of the data analyzed, i.e. spontaneous activity rather than evoked potentials, for which the signals from the different sources of signal may probably be considered statistically independent. In [16] and [17], spatiotemporal ICA was used after CSD analysis to analyze evoked potentials of experimental and simulated data. At the end of their studies, it was concluded that CSD analysis followed by spatiotemporal ICA gave physiologically meaningful results.

A very common object of study by neuroscientists is the rat somatosensory cortex. This work analyzes data from this part of the rat brain. The technology available today allows in vivo voltage recordings at tens or hundreds of contacts across cortex laminae [18][19]. Such recordings can be used to verify which cortex layers are activated by different stimuli. In this work we reinforce that BSS techniques can greatly assist in the analysis of the signals obtained. It helps identify which layers are part of the same activation group for a given stimulus and which are not. To be best of our knowledge, no BSS schemes have been applied to signals of this specific part of the rat somatosensory cortex.

In this work, we propose the analysis of the somatosensory cortex signals of a rat applying BSS schemes. Among the techniques used, two have not been used before in the processing of cortical LFPs and present results consistent with stICA. One of them is an adaptation of Parallel Factor Analysis (PARAFAC) proposed in this work specifically for this problem. PARAFAC is a multi-way analysis technique known to have many advantages over conventional two-way decompositions [20][13][21][22]. For example, it is unique under very mild conditions without the need of additional constraints [20][22]. However, the approach generally used to analyze scalp electroencephalographic signals, which consists of obtaining a third dimension besides electrodes and time by means of Time-Frequency Analysis (TFA), taking only the magnitudes in the time-frequency representations and using PARAFAC with nonnegativity contraints to increase physical significance, is not sufficient in this case because time signals must be recovered from their time frequency representations. We show that complex valued PARAFAC is not an option because it is not able to decompose the LFPs into meaningful components, as attested by their CSDs. We therefore propose a different approach which consists of first using PARAFAC as it is normally used in the analysis of electroencephalographic signals, then use the phases of the original array to reconstruct the time signals.

The other technique used in this work not previously applied to cortical LFP recordings is Triple-N ICA for Convolutive Mixtures (TRINICON). There are indications that, although the mixing of the signals in the brain is certainly instantaneous, the convolutive mixing model can be useful to model the brain activity itself [23]. TRINICON is method for blind separation of convolutive mixtures proposed in [24]. It exploits the nonwhiteness, nonstationarity and nongaussianity of the source signals, each usually exploited alone by other algorithms. To date, it has mostly been used for the analysis of audio signals.

Even though the Electroencephagram (EEG) use a different sensor array type, it shares the same origin, signal processing concepts and techniques as LFPs. Therefore, in Chapter 2, an overal description of Electroencephagraphy is made. Chapter 3 starts by reviewing the most classic BSS techniques, PCA and ICA, then moves to more specific techniques, stICA and TRINICON. Chapter 4 is dedicated to the dataset, it starts with some biological considerations, then moves to the modeling of the dataset, then to the description of the methods used to separate the data and finally to the results. In Chapter 5 we draw our conclusions.

Chapter 2

The Electroencephalogram (EEG)

2.1 Introduction

Even though the Electroencephagram (EEG) use a different type of sensor array, and is recorded from a different perspective, it shares the same origin and signal processing as LFPs. Furthermore, the literature on EEG is far more vast, so a discussion on this topic can greatly improve the understanding of LFPs. For these reasons, the next sections are dedicated to the EEG.

2.2 Measurement

In clinical practice, electrodes are typically placed on the scalp using the 10-20 system, recomended by the International Federation of Clinical Neurophysiology [25][26]. It receives this name because the distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull [27]. Higher electrodes densities are needed specially in research. The 10-20 system was extended in 1991 to the 10-10 system by the American Electroencephalographic Society [28]. In 2001 it was proposed another extension, called the 10-5 system [29]. The 10-20 system comprises 21 electrode, the 10-10 system comprises 74 electrodes, and the 10-5 system comprises 345 electrodes, but these are not the only electrode arrangements used today. Electrode cap manufacturers, for example, produce caps with various electrode arrangements. Figure 2.1 shows the the electrode positions and labels in the 10-20, 10-10 and 10-5 systems. Since the conductivity of the air is almost zero, EEG acquisition systems almost always require electroconductive gel, paste, or salt-water-soaked sponges between the skin and the EEG electrodes [30].

The difference between two electrode potentials is called a channel and is realized by a differential amplifier. The voltage of the person with respect to the amplifier's common is called the common mode voltage, and can be transformed by the amplifier into an interfering differential signal. Therefore it is desirable to attach a third electrode to the person to minimize the common mode voltage [31]. This electrode is commonly called the "ground", but in current designs it is actually isolated from the earth ground [32][33]. This electrode is best placed near the other

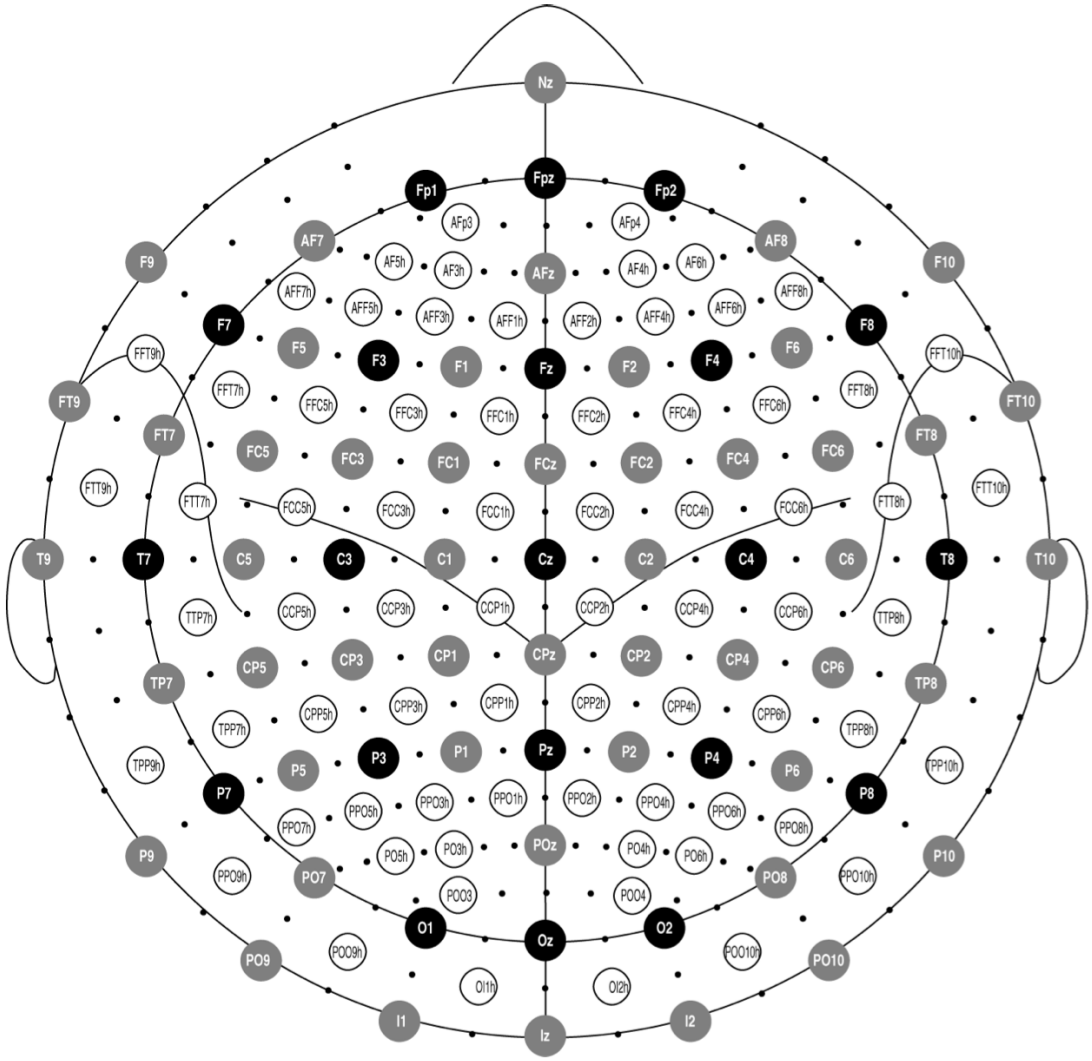
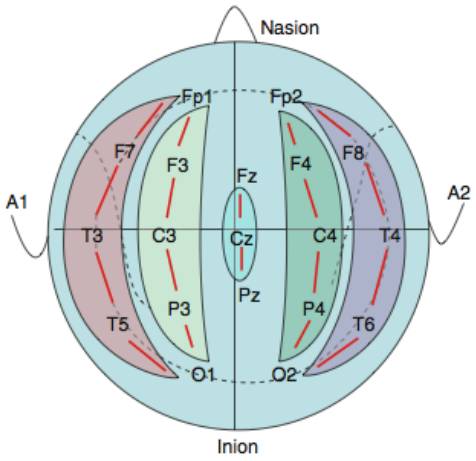


Figure 2.1: Electrodes positions and labels in the 10-5, 10-10 and 10-20 systems. Black circles indicate positions in the original 10-20 system, gray circles indicate positions introduced in the 10-10 system, and dots and open circles in the 10-5 system. Adapted from [29].

recording sites _ a mid-forehead placement is common [34]. An EEG recording system is composed of several differential amplifiers with the same third electrode. The connection of two electrodes to the inputs of an amplifier is called a derivation. A group of derivations is called a montage [35]. There are two types of montages: bipolar, when the electrodes are connected in a chain like fashion, with the reference electrode of one channel being the non reference electrode of another channel; and referential, when all channels have the same reference electrode, or the same reference of technical nature (like in the form of an average) [36]. Any montage can be obtained from any other montage, as long as all electrodes have been referred to each other in the original recording. Because of this, digital EEG systems store the signals in a referential montage containing all electrodes [35]. A great diversity of montages exists among different EEG laboratories [37]. Some standard bipolar and referential montages are proposed by the American Clinical Neurophysiology Society [38]. Figure 2.2 shows an example of bipolar montage: the longitudinal bipolar montage,



Channel No.	LB-18.1	LB-18.2	LB-18.3
1	Fp1-F7	Fz-Cz	Fp1-F7
2	F7-T7 (T3)	Cz-Pz	F7-T7 (T3)
3	T7 (T3)-P7 (T5)	Fp1-F3	T7 (T3)-P7 (T5)
4	P7 (T5)-O1	F3-C3	P7 (T5)-O1
5	Fp1-F3	C3-P3	Fp2-F8
6	F3-C3	P3-O1	F8-T8 (T4)
7	C3-P3	Fp2-F4	T8 (T4)-P8 (T6)
8	P3-O1	F4-C4	P8 (T6)-O2
9	Fz-Cz	C4-P4	Fp1-F3
10	Cz-Pz	P4-O2	F3-C3
11	Fp2-F4	Fp1-F7	C3-P3
12	F4-C4	F7-T7 (T3)	P3-O1
13	C4-P4	T7 (T3)-P7 (T5)	Fp2-F4
14	P4-O2	P7 (T5)-O1	F4-C4
15	Fp2-F8	Fp2-F8	C4-P4
16	F8-T8 (T4)	F8-T8 (T4)	P4-O2
17	T8 (T4)-P8 (T6)	T8 (T4)-P8 (T6)	Fz-Cz
18	P8 (T6)-O2	P8 (T6)-O2	Cz-Pz

Figure 2.2: Longitudinal Bipolar montage or "Double banana". LB-18.1, LB-18.2 and LB-18.3 are just alternative orderings of the same channels. Adapted from [37] and [38].

or "double banana".

Ideally, the reference of the referential montage should be electrophisiologically silent. In practice though, any cephalic or noncephalic site will contribute to the measurements, either with EEG activity or a higher incidence of electrocardiographic (EKG) and electromyographic (EMG) activity. In many conventional EEG studies, a contralateral or (less frequently) unilateral earlobe or mastoid reference is employed. Linked ears or linked mastoids (practices that involve linking electrodes placed on the earlobes or mastoids) are used specially in quantitative EEG studies [34]. Another options used by some investigators are the average reference and laplacian montages. In the first case, the average of all or most electrodes is taken as reference. In the second case, each electrode is referred to the average of the four closest neighbors, which is equivalent to approximating the negative laplacian divided by four [39]. It can be shown that the laplacian is proportional to the current density at that point, and can therefore be used to localize the underlying sources of the EEG [40].

The advantage of the bipolar montages over the referential montages is mainly in the clinical interpretation of the EEG, due to the property of phase reversal: when the same wave appears mirrored in two neighbor channels, it was formed near the common electrode. However, the potential difference between two neighbor electrodes is often very small, which is a disadvantage with respect to the referential montages. In practice, both referential and bipolar montages are used in the clinical interpretation of the EEG [36]. From a signal processing point of view, the average reference can be seen as centering across the electrodes mode [41], while the laplacian and bipolar montages act as spatial filters [34][35].

The EEG, in the most common sense, is the electrical activity of the brain recorded from the scalp. However, the EEG (meaning the summed electrical activities of populations of neurons) may also be measured at a short distance of the sources, in which case it is called Local EEG or Local Field Potentials (LFP), or from the cortical surface, in which case it is called Intracranial EEG or Electrocorticogram (ECoG) [42]. LFPs may be recorded using silicon probes with geometrically

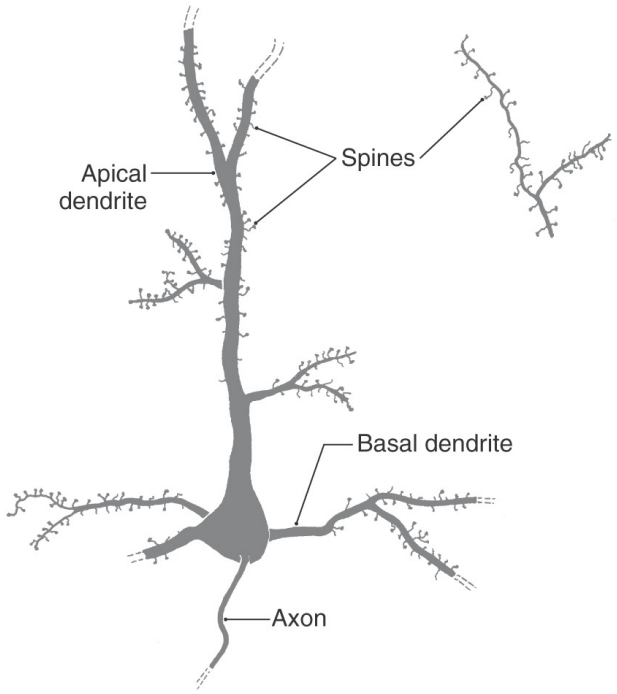


Figure 2.3: Typical neuron: pyramidal cell of the cerebral cortex. Notice the triangular shaped soma of this cell, from which it receives its name. The dendrites of this neuron are divided in one apical dendrite and several basal dendrites. They have membrane specializations called dendritic spines, where most of the synapses take place. Adapted from [45].

arranged electrode contacts [43]. ECoG, in its turn, requires an electrode grid to be surgically implanted on the surface of the cortex. Most of the considerations above still apply for LFP and ECoG, considering that all electrode locations are now determined by the specific measurement device: the silicon probe for the LFP and the electrode grid for the ECoG.

2.3 Origin

The following phrase summarizes the origin of the EEG: "EEG reflects mainly the summation of excitatory and inhibitory postsynaptic potentials at the dendrites of ensembles of neurons with parallel geometric orientation" [30]. In this section, this phrase will be dissected.

Figure 2.3 shows a typical neuron. The neuronal cell body, or soma, presents a great number of ramified projections, the dendrites, from which it receives information from other neurons. The large number of dendrites increase the number of possible afferent information received. There is another longer and thinner extension of the soma, the axon, with few branches in its trajectory and many in its final portion, the telodendron. Each neuron has a single axon, through which it sends efferent information to other neurons. To make many contacts with the dendrites of other cells, the telodendron has many branches [44].

The space between membranes in the contact site between two neurons, the synapse, is called the synaptic cleft. The axon terminal is the presynaptic element, and the dendrite is the postsynap-

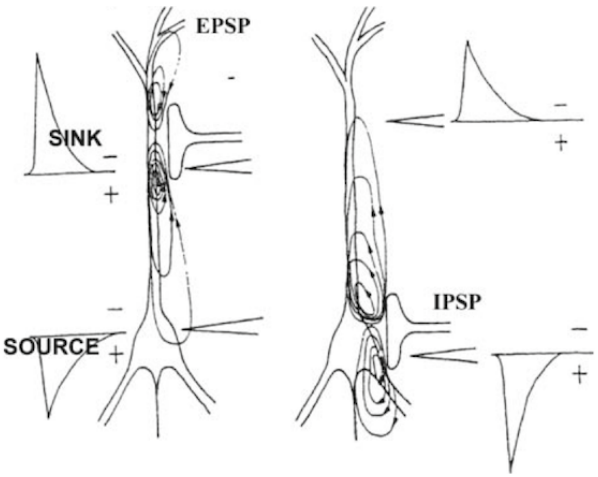


Figure 2.4: To the left, an excitatory synapse creates a current sink at the level of the apical dendrite of a pyramidal cell, which is compensated by distributed current sources along the soma-dendric membrane. To the right, an inhibitory synapse creates a current source at the level of the soma of a pyramidal cell, which is compensated by distributed current sinks along the soma-dendric membrane. At a current sink, the extracellular potential becomes more negative, since positive ions are entering the cell. At a current source, the extracellular potential becomes more positive, since positive ions are leaving (or negative ions are entering) the cell. Adapted from [42].

tic element. When chemical substances sent from the presynaptic element, the neurotransmitters, arrive at the postsynaptic element, changes take place in the membrane potential (the intracellular potential with respect to the extracellular potential), at the location of the postsynaptic terminal. These changes are called postsynaptic potentials and they are caused by the interactions between neurotransmitter, receptors, intracellular second messengers, and ionic channels. Generally speaking, in the Excitatory PostSynaptic Potentials (EPSPs) there are transmembrane currents carried by positive ions inwards (e.g. Na^+). In the case of Inhibitory PostSynaptic Potentials (IPSPs), they are carried by negative ions inwards (e.g. Cl^-) or positive ions outwards (e.g. K^+). Thus, the positive electric current is directed to the inside in the case of an EPSP and to the outside in the case of an IPSP [42].

Since there is no accumulation of charge anywhere in the medium, the transmembrane currents that flow in or out of the neuron at the active synaptic sites are compensated by currents that flow in the opposite direction elsewhere along the neuronal membrane. Consequently, in the case of an EPSP, an active sink is generated in the extracellular medium at the level of the synapse, whereas distributed passive sources appear along the soma-dendritic membrane. In the case of an IPSP an active source occurs at the level of the synapse, whereas distributed passive sinks are formed along the soma-dendritic membrane (Figure 2.4) [42]. In the microscopic level, the electric field associated with these current sources and sinks is quite complicated, but from a distant recording electrode, the distance between the centers of mass of current sources and sinks is small compared to the recording distance, and the field can be approximated by a dipole field [46].

However, the electric field generated by a single neuron cannot be measured from a distant

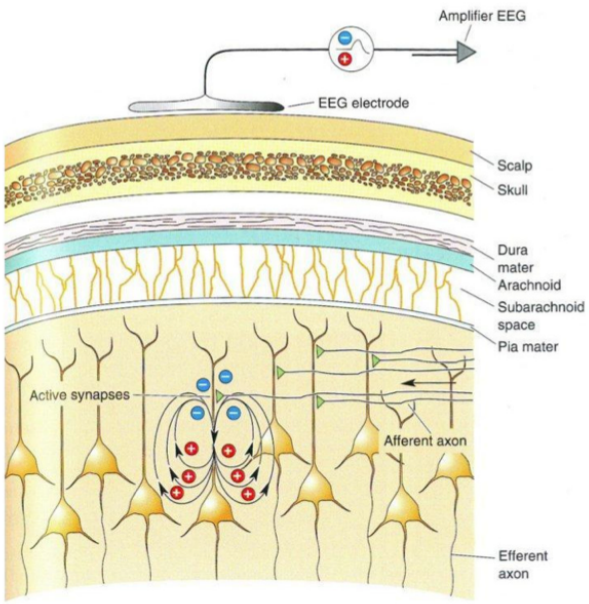


Figure 2.5: Neuronal sources of the EEG: pyramidal neurons of the cortex. The parallel arrangement of these neurons permit their synchronous activity to "add up" and be recorded by an EEG eletrode. Adapted from [49].

electrode. It is necessary that a group of neurons be active at the same time and in the same way in order to generate sufficient voltage to be recorded by the electrode [47]. "At the same time" has to do with synchrony. "In the same way" has to do with geometry. Lorence De Nó [48] identified basically two types of ensemble geometry: closed field and open field. In the closed field geometry, the neurons are arranged radially, either with the cell bodies in the periphery and the dendrites oriented inwards, or with the cell bodies in the center and the dendrites oriented outwards. In the open field geometry, the neurons are aligned parallel to each other, with the cell bodies concentrated in one part and the dendrites concentrated in other part. In the first case current flows entirely within the ensemble with the result that all points outside remain at zero potential, while in the second case the arrangement of current sources and sinks permit the spread of current in the volume of the brain [48].

The neurons that mainly contribute to the EEG are those that form open fields, i.e. the pyramidal neurons of the cortex, since they are arranged in palisades with the apical dendrites aligned perpendicularly to the cortical surface (Figure 2.5) [42]. The fact that the cortex is folded implies that some populations of neurons have apical dendrites that are perpendicular to the overlying skull, whereas others are parallel to the skull. Those that are perpendicular to the skull contribute more to the EEG, while those that are parallel are unlikely to be measured because populations in opposing sides of a sulcus produce electric fields that are likely to cancel each other [30].

The problem of estimating the neural sources given a distribution of scalp potentials is called the inverse problem. It is an ill-posed problem because there are an infinite number of source configurations that can give the same scalp potentials [42]. In this context, the concept of equivalent

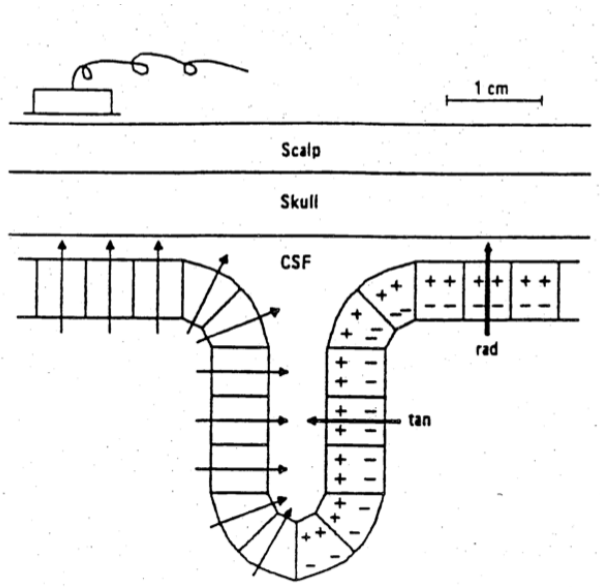


Figure 2.6: General cortical fold. To the left, equivalent dipoles for small cortical segments. To the right, a single tangential dipole for all segments in one side of a sulcus, and a single radial dipole for all segments in a ridge. A single equivalent dipole could be used even for the hole brain [42]. Adapted from [46].

dipole is useful (Figure 2.6). They are called this way because their fields give an equivalent description of the compound activity of all neuronal elements in their vicinity [46]. The simplest source model assumes that the scalp distribution is best represented by a single equivalent dipole current source. This choice has been shown to be useful in certain situations [42].

Chapter 3

Review of Blind Signal Separation (BSS) techniques

3.1 Principal Component Analysis (PCA)

Suppose $\mathbf{X}$ is an I x J matrix containing scores of I subjects on J tests (or I observations described by J variables, or I time samples of J different measurement types, and so on). It is required that the grades are given such that above average scores are positive and below average scores are negative. Each row of $\mathbf{X}$ is a vector that lies in a J-dimensional vector space. The purpose of PCA is to find a new set of F orthonormal basis vectors, $F < J$, that span a subspace of the original space and optimally summarizes the information contained in $\mathbf{X}$ [50].

First suppose $F = J$. The problem can be described as a change of basis. The J x J matrix $\mathbf{B}$ has the new orthonormal basis vectors in its columns such that

$$\mathbf{A}^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}. \tag{3.1}$$

Each row of the I x J matrix $\mathbf{A}$ contains the dot products (i.e. projections) of a specific row of $\mathbf{X}$ on the columns of $\mathbf{B}$. Therefore, the rows of $\mathbf{A}$ contain the vectors expressed in the new orthonormal basis. Equation (3.1) can be rewritten as

$$\mathbf{X} = \mathbf{A}\mathbf{B}^{\mathsf{T}}, \tag{3.2}$$

since $\mathbf{B}$ is an orthogonal matrix, and therefore $\mathbf{B}^{-1} = \mathbf{B}^{\mathsf{T}}$.

The initial goal is to rotate the original basis to minimize redundancy, measured by the covariances between tests, and to maximize the signals, measured by the variances of the tests [51].

This is done by diagonalizing the covariance matrix. The covariance matrix of $\mathbf{X}$ is

$$\mathbf{C_X} = \frac{\mathbf{X}^\mathsf{T}\mathbf{X}}{I} = \frac{\begin{bmatrix} x_{11} & \cdots & x_{I1} \\ \vdots & & \vdots \\ x_{1J} & \cdots & x_{IJ} \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & & \vdots \\ x_{I1} & \cdots & x_{IJ} \end{bmatrix}}{I} = \begin{bmatrix} \frac{\sum_{i=1}^{I} x_{i1}^2}{I} & \cdots & \frac{\sum_{i=1}^{I} x_{i1}x_{iJ}}{I} \\ \vdots & & \vdots \\ \frac{\sum_{i=1}^{I} x_{iJ}x_{i1}}{I} & \cdots & \frac{\sum_{i=1}^{I} x_{iJ}^2}{I} \end{bmatrix}. \tag{3.3}$$

The variances can be seen in the main diagonal and the covariances can be seen in the off diagonal entries. A diagonal matrix has nonzero elements only in its main diagonal, and zero elements in its off diagonal entries. Thus by diagonalizing the covariance matrix we are effectively eliminating the covariances. The covariance matrix of $\mathbf{A}$ is given by

$$\mathbf{C_A} = \frac{1}{I}\mathbf{A}^\mathsf{T}\mathbf{A} = \frac{1}{I}\mathbf{B}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{B} = \mathbf{B}^\mathsf{T}\left(\frac{1}{I}\mathbf{X}^\mathsf{T}\mathbf{X}\right)\mathbf{B} = \mathbf{B}^\mathsf{T}\mathbf{C_X}\mathbf{B}. \tag{3.4}$$

From the Spectral Theorem [52], for $\mathbf{C_A}$ to be diagonal, $\mathbf{B}$ must contain the orthonormal eigenvectors of $\mathbf{C_X}$, the eigenvalues of $\mathbf{C_X}$ being the variances in the main diagonal of $\mathbf{C_A}$. Therefore, $\mathbf{B}$ can be obtained directly from the covariance matrix of $\mathbf{X}$, with the orthonormal eigenvectors ordered according to the size of their eigenvalues, by applying Eigenvalue Decomposition (EVD) to $\mathbf{C_X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$, with $\mathbf{\Lambda}$ diagonal and $\mathbf{Q}$ orthogonal, and making $\mathbf{B} = \mathbf{Q}$. Then, $\mathbf{A}$ could be obtained by Equation (3.1). An alternative approach with the same result would be to apply Singular Value Decomposition (SVD) to $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$, with $\mathbf{\Sigma}$ diagonal and $\mathbf{U}$ and $\mathbf{V}$ orthogonal, and obtain directly $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}$ and $\mathbf{B} = \mathbf{V}$, since $\mathbf{C_X} = \mathbf{X}^\mathsf{T}\mathbf{X}/I = \mathbf{V}\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}/I = \mathbf{V}(\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}/I)\mathbf{V}^\mathsf{T}$, where $\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}/I$ is diagonal.

If only the first F orthonormal eigenvectors are selected, the F-dimensional subspace that optimally summarizes the information contained in $\mathbf{X}$ is selected. Matrix $\mathbf{B}$ is J x F. Matrix $\mathbf{A}$ is I x F (from Equation (3.1)). The matrix $\mathbf{A}$ is called the matrix of factors scores (like if PCA had derived new tests, called factors, that are mixtures of the original tests, and the rows of $\mathbf{A}$ contained the scores of each subject in these new tests). The matrix $\mathbf{B}$ is called the loading matrix (as if each column of $\mathbf{B}$ carried the weights used to compute a new test from the original tests) [53].

With only F components $\mathbf{A}\mathbf{B}^\mathsf{T}$ reconstruct $\mathbf{X}$ just approximately. Each row of $\mathbf{X}$ is a linear combination of the columns of $\mathbf{B}$, but with just F components, $\mathbf{X}$ cannot be reconstructed completely. Other way to see it is that $\mathbf{A}\mathbf{B}^\mathsf{T}$ reconstructs projections of the original vectors in the subspace spanned by the columns of $\mathbf{B}$.

If each element of $\mathbf{X}$ had been divided by $\sqrt{I}$, the analysis would be referred to as a covariance PCA, because in this case the matrix $\mathbf{X}^\mathsf{T}\mathbf{X}$ would be the covariance matrix. If also each column had been normalized, i.e. divided by its variance, the analysis would be referred to correlation PCA, because the matrix $\mathbf{X}^\mathsf{T}\mathbf{X}$ would be a matrix of correlation coeficients [53].

3.2 PCA-based algorithms for Independent Component Analysis (ICA)

In this section, only PCA-based ICA algorithms will be considered. A slight change in the PCA solution from the previous section makes possible to separate statistically independent tests.

First, consider the solution of PCA using SVD

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T} = \mathbf{A}\mathbf{B}^\mathsf{T}. \tag{3.5}$$

Suppose the matrix of factors scores, $\mathbf{A}$, and the loading matrix, $\mathbf{B}$, had been defined as

$$\mathbf{A} = \sqrt{I}\ \mathbf{U} \tag{3.6}$$

and

$$\mathbf{B} = \mathbf{V}\frac{\mathbf{\Sigma}^\mathsf{T}}{\sqrt{I}}, \tag{3.7}$$

instead of $\mathbf{U\Sigma}$ and $\mathbf{V}$ respectively. The factor scores have now become unit variance, since

$$\frac{\mathbf{A}^\mathsf{T}\mathbf{A}}{I} = \mathbf{I}. \tag{3.8}$$

The variance of each factor scores, previously in the diagonal entries of

$$\frac{(\mathbf{U\Sigma})^\mathsf{T}\mathbf{U\Sigma}}{I} = \frac{\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}}{I}, \tag{3.9}$$

had been transferred to the corresponding basis vector by multiplying each eigenvector in the columns of $\mathbf{V}$ by the square root of the corresponding eigenvalue (Equation (3.7)). Now each basis vector have become comparable with the variance in that direction.

For now suppose, without loss of generality, that the number of tests is $J = 2$. Figure 3.1, top right, shows a scatter plot of some example matrix $\mathbf{X}$. The x-axis is the first test, the y-axis is the second test, and each point correspond to a subject. What PCA finds are the orthogonal directions such that the first has the highest variance, then the second has the highest variance, and so on, since by diagonalizing the covariance matrix we are maximizing the variances while minimizing the covariances. Figure 3.1, bottom left, shows a scatter plot of matrix $\mathbf{A}$ as defined in this section. The data has first been rotated then unstretched, which are the effects of the orthogonal matrix $\mathbf{V}^\mathsf{T}$ and the diagonal matrix $\mathbf{\Sigma}/\sqrt{I}$, respectively, when applied to the rows of $\mathbf{X}$,

$$\mathbf{A}^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{X}^\mathsf{T} = \frac{\mathbf{\Sigma}}{\sqrt{I}}\mathbf{V}^\mathsf{T}\mathbf{X}^\mathsf{T}. \tag{3.10}$$

The data in matrix $\mathbf{A}$ is whitened in the sense that it now contains unit variance in all directions, as expressed in Equation (3.8) and as can be seen in Figure 3.1, bottom left.

PCA could be seen as trying to separate independent factors from the data and it would do that by trying to eliminate the covariance between every pair of factors, but it is well known that uncorrelation is a necessary but not sufficient condition for statistical independence. The factors in Figure 3.1, bottom left, are uncorrelated since, if the score of one factor increases, there is no

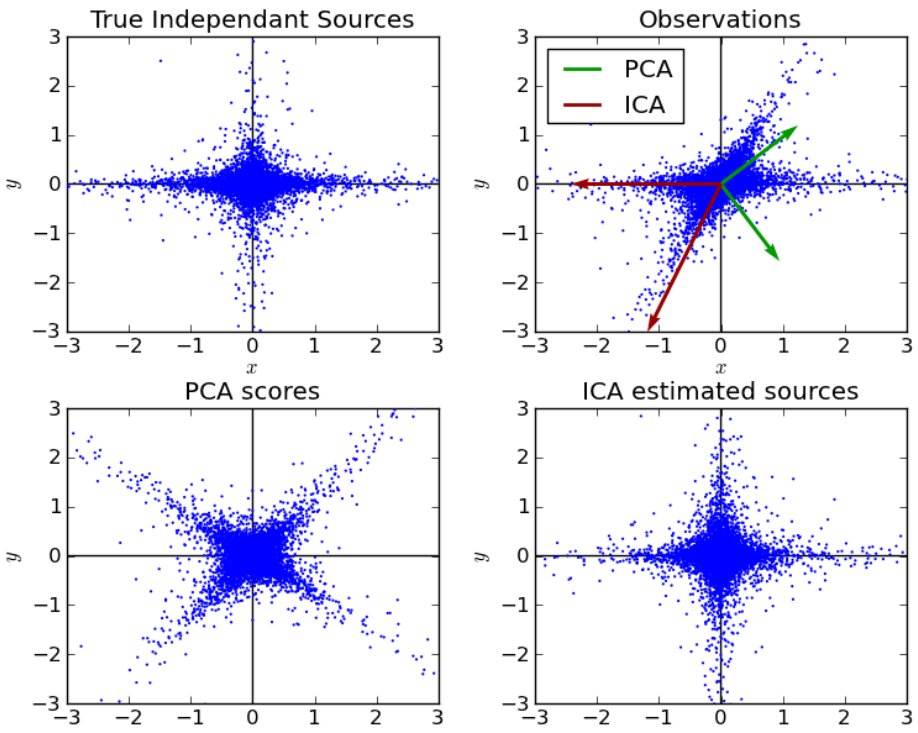


Figure 3.1: Ilustration of Independent Component Analysis (ICA) and Principal Component Analysis (PCA). Adapted from [54].

tendency for the score of the other factor to increase or decrease. However, Figure 3.1, bottom left, does not look like containing two independent factors. If the score of one factor increases, the score of the other factor increases for a group of subjects and decreases for another group of subjects. This is not perceived by the covariance between the two factors. Figure 3.1, top left, looks like containing two independent factors.

It can be seen that PCA was able to find the independent factors up to a rotation. This property of PCA is called rotational invariance [55]. $\mathbf{B}$ and $\mathbf{A}$ could be both rotated by any orthogonal matrix $\mathbf{Q}$ and still give the same fit to the data (although one of the constraints required by PCA would inevitably be violated),

$$(\mathbf{AQ})(\mathbf{BQ})^\mathsf{T} = \mathbf{AQQ}^\mathsf{T}\mathbf{B}^\mathsf{T} = \mathbf{AB}^\mathsf{T}. \tag{3.11}$$

Using the definition of $\mathbf{B}$ and $\mathbf{A}$ from the previous section, $\mathbf{BQ}$ would still be orthogonal, since $(\mathbf{BQ})^\mathsf{T}\mathbf{BQ} = \mathbf{I}$, but $\mathbf{AQ}$ would not be uncorrelated anymore, since $(\mathbf{AQ})^\mathsf{T}\mathbf{AQ}/I = \mathbf{Q}^\mathsf{T}\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma Q}/I$, which is not diagonal. Using the definition of $\mathbf{B}$ and $\mathbf{A}$ from this section, $\mathbf{BQ}$ would not have orthogonal columns, since $(\mathbf{BQ})^\mathsf{T}\mathbf{BQ} = \mathbf{Q}^\mathsf{T}(\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}/I)\mathbf{Q}$, which is not diagonal, but $\mathbf{AQ}$ would remain uncorrelated, since $(\mathbf{AQ})^\mathsf{T}\mathbf{AQ}/I = \mathbf{I}$ [56]. We are interested in the second definition, because now we only need to find a rotation from Figure 3.1, bottom left, to Figure 3.1, bottom right. If the data in Figure 3.1, top right, had just been rotated and not unstretched (like in the previous section), we would never be able to transform it in Figure 3.1, bottom right, by a simple rotation.

To find this rotation we resort to higher order statistics, which basic quantities are higher order moments and higher order cumulants. The N-th order moment of N random variables X_n $(n = 1, 2, ..., N)$ is defined as [55]

$$Mom(X_1, X_2, ..., X_N) = E\left\{\prod_{n=1}^{N} X_n\right\}, \tag{3.12}$$

where $E\{\cdot\}$ represents the expected value operator. One can see that the first order moment is the mean of X_1 and the second order moment is the covariance between X_1 and X_2, if X_1 and X_2 are zero mean. Similarly to the N-th order moment, the N-th order cumulant of N random variables X_n $(n = 1, 2, ..., N)$ is defined as [55]

$$Cum(X_1, X_2, ..., X_N) = \sum(-1)^{K-1}(K-1)!E\left\{\prod_{n\in A_1} X_n\right\} E\left\{\prod_{n\in A_2} X_n\right\} ... \; E\left\{\prod_{n\in A_K} X_n\right\}, \tag{3.13}$$

where the summation involves all possible partitions into K subsets $A_1, A_2, ..., A_K$ $(1 \leq K \leq N)$ of the set of integers $\{1, 2, ..., N\}$. One can see that the first order cumulant is the mean of X_1, since there is only one possible partition of $\{1\}$, into one subset $(K = 1)$, which is $A_1 = \{1\}$. For the second order cumulant, there are two possible partitions of $\{1, 2\}$, namely $\{\{1\}, \{2\}\}$ $(K = 2)$ and $\{\{1, 2\}\}$ $(K = 1)$. So Equation (3.13) becomes

$$Cum(X_1, X_2) = E\{X_1 \cdot X_2\} - E\{X_1\} \cdot E\{X_1\}, \tag{3.14}$$

which is one of the expressions for the covariance, since

$$\begin{aligned} Cov(X_1, X_2) &= E\{(X_1 - E\{X_1\}) \cdot (X_2 - E\{X_2\})\} \\ &= E\{(X_1X_2 - X_1E\{X_2\} - E\{X_1\}X_2 + E\{X_1\} \cdot E\{X_2\})\} \\ &= E\{X_1 \cdot X_2\} - E\{X_1\} \cdot E\{X_2\} - E\{X_1\} \cdot E\{X_2\} + E\{X_1\} \cdot E\{X_2\} \\ &= E\{X_1 \cdot X_2\} - E\{X_1\} \cdot E\{X_2\}. \end{aligned}$$

Therefore we conclude that the second order cumulant is the covariance between X_1 and X_2, regardless of the mean of X_1 or X_2.

A Cumulants is a combination of moments that vanishes when the variables are independent of each other [57]. This explains why the covariance (second order cumulant) vanishes when two variables are independent, and why uncorrelation is a necessary but not sufficient condition for statistical independence. PCA was not able to find the independent factors because it used only lower order cumulants $(N \leq 2)$. ICA uses higher order cumulants $(N \geq 3)$ to finish finding the independent factors.

For the matrix $\mathbf{X}$ under consideration, the Nth-order cumulant tensor is defined by the element-wise equation [55]

$$\left(\mathscr{C}_{\boldsymbol{X}}^{(N)}\right)_{j_1, j_2, ..., j_N} = Cum(X_{j_1}, X_{j_2}, ..., X_{j_N}), \tag{3.15}$$

where $\boldsymbol{X} = (X_1, X_2, ..., X_J)$ represents a J-dimensional random vector from which the rows of $\mathbf{X}$ are samples, and $j_n = 1, 2, ..., J$ $(n = 1, 2, ..., N)$ is the index of the n-th mode of $\mathscr{C}_{\boldsymbol{X}}^{(N)}$. The j_n are

all indexes to the elements $X_1, X_2, ..., X_J$ of the random vector $\boldsymbol{X}$. For example, the second order cumulant tensor is the covariance matrix

$$\mathscr{C}_{\boldsymbol{X}}^{(2)} = \begin{bmatrix} E\{X_1X_1\} - E\{X_1\}E\{X_1\} & \cdots & E\{X_1X_J\} - E\{X_1\}E\{X_J\} \\ \vdots & & \vdots \\ E\{X_JX_1\} - E\{X_J\}E\{X_1\} & \cdots & E\{X_JX_J\} - E\{X_J\}E\{X_J\} \end{bmatrix}. \tag{3.16}$$

Substituting $E\{X_j\}$ by the sample mean $\sum_{i=1}^{I} x_{ij}/I$ one gets the sample covariance matrix, which is equal to Equation (3.3) if the data is zero mean [58]. Since the estimation of a sample cumulant tensor becomes poorer as the order increases (for the same number of samples), one usually uses up to the fourth-order cumulant tensor. Since the third-order cumulant $Cum(X, X, X)$ of a random variable X with symmetric distribution vanishes, the third order cumulant tensor is generally not used [55].

If the random vector $\boldsymbol{X}$ is transformed into the random vector $\tilde{\boldsymbol{X}}$ by some J x J matrix $\mathbf{T}$, i.e. $\tilde{\boldsymbol{X}} = \mathbf{T}\boldsymbol{X}$, then

$$\mathscr{C}_{\tilde{\boldsymbol{X}}}^{(N)} = \mathscr{C}_{\boldsymbol{X}}^{(N)} \times_1 \mathbf{T} \times_2 \mathbf{T} ... \times_N \mathbf{T}, \tag{3.17}$$

where $\times_n$ is the n-mode product of a tensor by a matrix, which is obtained by multiplying all n-mode vectors of the tensors by the matrix on the left. Equation (3.17) is the tensor equivalent of the matrix expression $\mathbf{T}\mathscr{C}_{\boldsymbol{X}}^{(2)}\mathbf{T}^{\mathsf{T}}$. This property is called multilinearity [55]. From Equation (3.10), one can see that

$$\mathscr{C}_{\boldsymbol{A}}^{(4)} = \mathscr{C}_{\boldsymbol{X}}^{(4)} \times_1 (\mathbf{B}^{\mathsf{T}}) \times_2 (\mathbf{B}^{\mathsf{T}}) \times_3 (\mathbf{B}^{\mathsf{T}}) \times_4 (\mathbf{B}^{\mathsf{T}}), \tag{3.18}$$

where $\boldsymbol{A} = (A_1, A_2, ..., A_J)$ is the J-dimensional random vector from which the rows of $\mathbf{A}$ are samples. The desired matrix of independet factors scores $\mathbf{Y}$ is related to $\mathbf{A}$ by the equation

$$\mathbf{Y}^{\mathsf{T}} = \mathbf{Q}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}. \tag{3.19}$$

Therefore

$$\mathscr{C}_{\boldsymbol{Y}}^{(4)} = \mathscr{C}_{\boldsymbol{A}}^{(4)} \times_1 (\mathbf{Q}^{\mathsf{T}}) \times_2 (\mathbf{Q}^{\mathsf{T}}) \times_3 (\mathbf{Q}^{\mathsf{T}}) \times_4 (\mathbf{Q}^{\mathsf{T}}), \tag{3.20}$$

where $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_J)$ is the J-dimensional random vector from which the rows of $\mathbf{Y}$ are samples. The key observation is that $\mathscr{C}_{\boldsymbol{Y}}^{(4)}$ is theoretically a diagonal tensor, i.e. only the entries of which all the indices are equal can be different from zero, otherwise the factors are not independent. Therefore, Equation (3.20) is an EVD-like tensor decomposition. Its result gives the matrix $(\mathbf{Q}^{\mathsf{T}})$. However, in practice, it is generally not possible to fully diagonalize $\mathscr{C}_{\boldsymbol{Y}}^{(4)}$. The way the error is dealt with allows different solution strategies. Some possibilities are ICA by means of: higher order eigenvalue decomposition (HOEVD); maximum diagonalization (MD); joint approximate diagonalization of eigenmatrices (JADE); and simultaneous third order tensor diagonalization (STOTD) [55].

In this Section, PCA with a slightly different definition of matrix of factor scores and loading matrix was used to find the pre-whitened data in $\mathbf{A}$,

$$\mathbf{X} = \mathbf{A}\mathbf{B}^{\mathsf{T}} = \mathbf{A}\left(\mathbf{V}\frac{\mathbf{\Sigma}^{\mathsf{T}}}{\sqrt{I}}\right)^{\mathsf{T}}, \tag{3.21}$$

then the rotation $\mathbf{Q}$ that made the factors approximately independent was found using higher order statistics,

$$\mathbf{X} = (\mathbf{AQ})(\mathbf{BQ})^{\mathsf{T}} = (\mathbf{Y}) \left(\mathbf{V} \frac{\mathbf{\Sigma}^{\mathsf{T}}}{\sqrt{I}} \mathbf{Q} \right)^{\mathsf{T}}. \tag{3.22}$$

3.3 Bell-Sejnowski algorithm for ICA

Consider the problem of finding $\mathbf{s}(n) = [s_1(n) \ ... \ s_Q(n)]^{\mathsf{T}}$ independent sources given a set of measurements $\mathbf{x}(n) = [x_1(n) \ ... \ x_P(n)]^{\mathsf{T}}$ that are instantaneous linear mixtures of the source signals $\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n)$, where $\mathbf{H}$ is called the mixing matrix. In what follows it is assumed that the number of observed signals P is equal to the number of sources Q and that each signal $x_p(n)$ have been mean centered, which implies that the reconstructed signals will be mean centered as well [59]. The sources can be found, up to a permutation and scaling, by $\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n)$, where $\mathbf{W}$ is called the unmixing matrix. This problem is equivalent to the one from the previous Section, i.e. finding statistically independent factors. Define $\mathbf{X}$ as the N x P matrix containing the N observations of the P measurements, and $\mathbf{Y}$ as the N x Q matrix containing the N observations of the Q reconstructed sources. Then $\mathbf{Y}^{\mathsf{T}} = \mathbf{W}\mathbf{X}^{\mathsf{T}}$, which can be rewritten as $\mathbf{X} = \mathbf{Y}(\mathbf{W}^{-1})^{\mathsf{T}}$. This last equation is equivalent to Equation (3.22) with $\mathbf{W}^{-1} = \mathbf{V}(\mathbf{\Sigma}^{\mathsf{T}}/\sqrt{I})\mathbf{Q}$, but in this Section the unmixing matrix will be found in a different way.

In the Bell-Sejnowski algorithm for Independent Component Analysis [60] the unmixing matrix $\mathbf{W}$ is found by maximizing the entropy $H(g(\mathbf{y}(n)))$ of $g(\mathbf{y}(n)) = (g(y_1(n)), ..., g(y_Q(n)))$, where g is a function that should approximate, as closely as possible, the Cumulative Distribution Function (CDF) of each source $s_q(n)$. Plain maximization of $H(\mathbf{y}(n))$ would be inappropriate [61]. The function $H(g(\mathbf{y}(n)))$ is called the contrast function of the Bell-Sejnovski algorithm. The contrast function of a random vector is maximum when its components are statistically independent [62].

To understand the concept of entropy, consider the case for a discrete random variable. The concept for a continuous random variable is similar [63]. The entropy of a discrete random variable X with probability distribution $P(X)$ may be defined as

$$H(X) = -\sum_i P(X = x_i) \log_2 P(X = x_i), \tag{3.23}$$

where x_i are the possible values of X. Entropy is a measure of the unpredictability or randomness of a random variable. If the probability of many outcomes is close to zero, while the probability of one outcome is close to one (low randomness), then the entropy is close to zero. If all probabilities are relatively far from zero or one (high randomness), then the entropy is large. Entropy is closely related to the average number of bits required to code the outputs of X in a sequence. If X has low randomness, one can use short strings to represent outcomes that happen a lot, and longer strings for outcomes that rarely happen, reducing the average number of bits required to code each output of X. The definition of entropy can be extended to a random vector $\boldsymbol{X} = (X, Y, ..., Z)$ with joint probability distribution $P(X, Y, ..., Z)$ if $P(X = x_i)$ is replaced by $P(X = x_i, Y = y_j, ..., Z = z_k)$ and the summation is over all possible combinations $x_i, y_j, ..., z_k$ in Equation (3.23). The entropy $H(\boldsymbol{X})$ gives the code length when all random variables are coded in the same code. If $X, Y, ..., Z$

are statistically dependent, the information they have on each other can be used to reduce the code length. This does not happen when they are statistically independent. This is the idea behind maximizing $H(g(\mathbf{y}(n)))$ to find the matrix $\mathbf{W}$.

3.4 Spatiotemporal ICA

As said before, ICA can be written similarly to PCA as $\mathbf{X} = \mathbf{Y}(\mathbf{W}^{-1})^{\mathsf{T}}$, where $\mathbf{X}$ is the I x J matrix of I observations and J measurements, and $\mathbf{Y}$ is the I x J matrix of I observations and J sources. Contrary to PCA, ICA places constraints only in the matrix $\mathbf{Y}$, i.e. statistically independent columns, while PCA constrained both $\mathbf{B}$ and $\mathbf{A}$ in Equation (3.2) to have orthogonal columns, since $\mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$, which is diagonal, and $\mathbf{B}^{\mathsf{T}}\mathbf{B} = \mathbf{I}$, which is the identity matrix. By analogy to PCA, Spatiotemporal ICA proposes to constrain both the columns of $\mathbf{Y}$ and the columns of $\mathbf{W}^{-\mathbf{1}}$ to be as statistically independent as possible [64]. Suppose now that $\mathbf{X}$ is an I x J matrix containing a sequence of J images of I pixels collapsed to column vectors (could also be J three-dimensional images of I voxels). Suppose also that $\mathbf{X}$ was preprocessed so that each row (i.e. time) and each column (i.e. space) had zero mean [65]. The new decomposition is expressed as $\tilde{\mathbf{X}} = \mathbf{S}\mathbf{\Lambda}\mathbf{T}^{\mathsf{T}}$, where $\mathbf{S}$ is an I x F matrix of mutually independent images, $\mathbf{T}$ is an J x F matrix of mutually independent time-varying image amplitudes, and $\mathbf{\Lambda}$ is a required diagonal scaling matrix. First, $\mathbf{X}$ is decomposed using truncated SVD, giving $\tilde{\mathbf{X}} = \mathbf{U}_F\mathbf{\Sigma}_F\mathbf{V}_F^{\mathsf{T}} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^{\mathsf{T}}$, with $\tilde{\mathbf{U}} = \mathbf{U}_F\mathbf{\Sigma}_F^{1/2}$ and $\tilde{\mathbf{V}} = \mathbf{V}_F\mathbf{\Sigma}_F^{1/2}$, where $\tilde{\mathbf{X}}$ represents an approximation to $\mathbf{X}$ if $F < J$, and is equal to $\mathbf{X}$ if $F = J$. Rewriting ICA as $\mathbf{Y} = \mathbf{X}\mathbf{W}^{\mathsf{T}}$ then applying it to $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ gives $\mathbf{S} = \tilde{\mathbf{U}}\mathbf{W}_S$ and $\mathbf{T} = \tilde{\mathbf{V}}\mathbf{W}_T$, where $\mathbf{W}_S$ and $\mathbf{W}_T$ are transposes of the unmixing matrices. Since $\tilde{\mathbf{X}} = \mathbf{S}\mathbf{\Lambda}\mathbf{T}^{\mathsf{T}}$ and $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^{\mathsf{T}}$, then $\mathbf{W}_S\mathbf{\Lambda}\mathbf{W}_T^{\mathsf{T}} = \mathbf{I}$ and $\mathbf{W}_T = (\mathbf{W}_S^{-1})^{\mathsf{T}}(\mathbf{\Lambda}^{-1})^{\mathsf{T}}$. The matrices $\mathbf{W}_S$ and $\mathbf{W}_T$ can be found simultaneously by maximizing

$$h(\mathbf{W}_S, \mathbf{\Lambda}) = \alpha h_S + (1-\alpha)h_T, \tag{3.24}$$

$$h_S = \ln|\det \mathbf{W}_S| + \frac{1}{I}\sum_{i=1}^{I}\sum_{f=1}^{F}\ln\sigma'(s_{if}), \tag{3.25}$$

$$h_T = \ln|\det \mathbf{W}_T| + \frac{1}{J}\sum_{j=1}^{J}\sum_{f=1}^{F}\ln\tau'(t_{jf}), \tag{3.26}$$

where s_{if} and t_{jf} are the elements of the matrices $\mathbf{S}$ and $\mathbf{T}$, respectively, $0 \leq \alpha \leq 1$ defines the relative weighting afforded to the spatial and temporal entropies, σ approximates the CDF of each spatial signal, and τ approximates the CDF of each temporal signal. In spatiotemporal ICA the distinction between independent components and mixing matrix is completely abolished, since the very same assumptions are made on the mixing matrix and the source signals [63].

3.5 Triple-N ICA for Convolutive Mixtures (TRINICON)

In the instantaneous blind source separation problem it was assumed that the observed signals at a given instant were instantaneous linear mixtures of the source signals, i.e. linear combinations

of the source signals at that instant. In the blind separation of convolutive mixtures, each observed signal at a given instant may also contain delayed versions of the same source [63]. In other words, each observed signal may be expressed as

$$x_p(n) = \sum_{q=1}^{Q} \left(\sum_{k=0}^{M-1} h_{qp}(k) s_q(n-k) \right), \tag{3.27}$$

where $h_{qp}(k)$ is the weight of the k-th delayed version of the source with index q in the observed signal with index p and $M-1$ is the number of delayed versions of each source, assumed the same for all sources. In what follows it is assumed that the number of observed signals P is equal to the number of sources Q. One can notice that the expression between parentheses is the convolution between $s_q(n)$ and $h_{qp}(n)$ and therefore the output of a filter with impulse response $h_{qp}(n)$ [66]. The number of taps, i.e. the number of nonzero values of the impulse response, of this filter is M and therefore it is a Finite Impulse Response (FIR) filter. It can be shown that a set of similar expressions,

$$y_q(n) = \sum_{p=1}^{P} \left(\sum_{k=0}^{L-1} w_{pq}(k) x_p(n-k) \right), \tag{3.28}$$

can reconstruct the sources (up to a permutation and filtering of the original signals) if L is chosen at least equal to M [63] [24], where $y_p(n)$ are the reconstructed signals, $w_{qp}(n)$ is the impulse response of an FIR filter from the p-th observed signal to the q-th source signal, and L is the number of taps of this filter. The overall Multiple Input Multiple Output (MIMO) Linear Time Invariant (LTI) mixing and demixing systems are represented in the block diagram of Figure 3.2.

Several methods exist for estimating the optimum coefficients $w_{pq}(k)$. In this work, the Triple-N ICA for convolutive mixtures (TRINICON) algorithm is used. TRINICON exploits the nonwhiteness, nonstationarity and nongaussianity of the source signals, each usually exploited alone by other algorithms [24]. A white process is a sequence of zero mean uncorrelated same variance random variables [63] [67]. A stationary process is one whose all multivariate densities are translation invariant [68]. A gaussian random variable possesses a gaussian Probability Density Function (PDF).

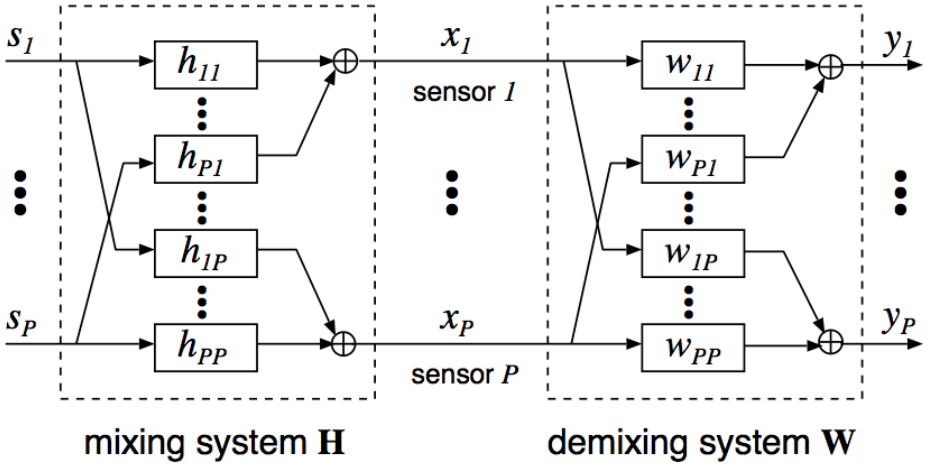


Figure 3.2: Multiple Input Multiple Output (MIMO) Linear Time Invariant (LTI) mixing and demixing systems of the convolutive mixing model. Adapted from [24].

Chapter 4

Application to the Analysis of Rat Somatosensory Evoked Response

4.1 Introduction

In the past decade, the interest on cortical Local Field Potentials (LFPs) has been renewed [18]. These are potentials recorded using small sized electrodes in the extracellular space of cortex brain tissue [43]. More specifically, the LFP correspond to the low frequency part of the recording (below 500Hz) and is associated to the ionic processes around the electrode. The high frequency part (above 500Hz) is called Multiunit activity (MUA) and is related to the action potentials of the surrounding neurons [19]. The renewed interest is due to technological advancements that allowed the recording of potentials at tens or hundreds of contacts across cortex laminae, and the realization that LFPs offer a unique window for studying integrative synaptic processes in cortical neuronal populations [19][18].

The neocortex of the mammalian is generally composed of six well identifiable layers, separated according to the specific types of neuronal cell bodies that are present and the specific connections with other layers and subcortical structures. As an example, Figure 4.1 shows in their characteristic lamina the most frequent types of neurons found in the human cerebral cortex, typical connections between them and typical connections with afferent fibers. Each neuron actually represents several of similar neurons that form populations, and the postsynaptic potentials in several of such populations is what is recorded by an extracellular electrode. Figure 4.2 shows a typical probe with 32 electrodes inserted perpendicularly to a cortical surface for LFP recordings. The neurons shown were reconstructed digitally. Pyramidal neurons are the most abundant in the cerebral cortex and are thought to contribute the most to the signals recorded [69][18].

A standard tool for the analysis of multielectrode LFPs has been the Current Source Density (CSD) analysis, which allows to estimate the net volume density of current entering or leaving the extracellular space at different locations [19]. Consider an arbitrarily small volume of the brain, located at point (x, y, z), where x,y and z are the cartesian coordinates. The net transmembrane current contributed by all cellular elements in this volume, or the CSD, is related to the extracellular

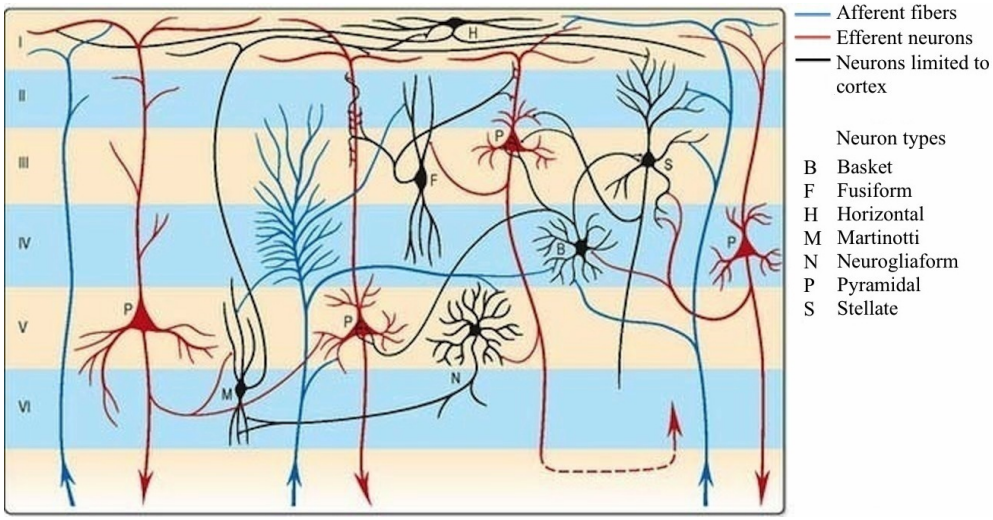


Figure 4.1: Most frequent types of neurons found in the human cerebral cortex, typical connections between them and with afferent fibers. The neurons are shown in their characteristic lamina. Adapted from [70].

field potential $\phi(x, y, z)$ by the Poisson equation $i(x, y, z) = -\sigma\nabla^2\phi(x, y, z)$, where the conductivity σ is considered constant and the same in all directions and ∇^2 is the laplacian operator [6] [18]. It is generally assumed that most variation of neural activity is in the vertical direction, due to the laminar structure of the cortex [75] [18]. Therefore, the CSD is usually estimated by LFP recordings from linearly spaced electrodes inserted perpendicularly to the cortical surface as

$$i(z) \approx -\sigma \frac{\phi(z+\delta) - 2\phi(z) + \phi(z-\delta)}{\delta^2}, \tag{4.1}$$

where δ is the spacing between vertical electrodes. The places where net conventional current is leaving the extracellular space (current sinks) correspond to negative CSDs and the places where net conventional current is entering the extracellular space (current sources) correspond to positive CSDs.

In analysis of EEG and LFPs it is common to differentiate between evoked and spontaneous activity. Evoked Potentials (EPs) refer to potentials elicited by specific sensory stimuli, while spontaneous activity refers to potentials that happen continuously, even in the absence of any stimuli. Averaging over several realizations is the common technique to isolate the EP in one electrode. Several segments of signal with the same length in which a realization of the same stimulus happened at the same time after the beginning of the segment must be added instant by instant then divided by the number of segments to isolate the EP. After this procedure, the instant of occurrence of the stimulus is commonly taken as the time zero. Figure 4.3 illustrates this procedure for continuously recorded data from one electrode. By the end of the procedure, the signal to noise ratio compared to that of a single realization drastically increases.

The data analyzed in this work consists of EPs obtained from the somatosensory cortex of a rat. Figure 4.5 shows the location of the somatosensory cortex in the rat brain and a silicon probe similar to that of Figure 4.2 inserted for LFP recordings. In the experiment under consideration, a

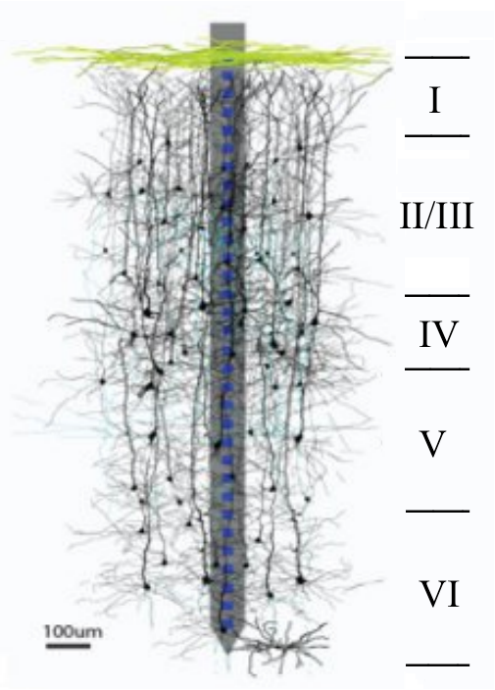


Figure 4.2: Typical probe with 32 electrode inserted perpendicularly to a cortical surface for LFP recordings surrounded by digitally reconstructed neurons. The numbering of layers to the left is merely illustrative. Adapted from [71].

rat received the same somatosensory stimulus every 7 seconds while LFPs were recorded from its somatosensory cortex with 32 electrodes spaced 50 micrometers from each other perpendicularly to the cortical surface. The signals were acquired using a sampling rate of 40kHz and a referential montage. Figure 4.4 shows the EPs obtained for each electrode in this experiment by averaging over 60 realizations. The CSD analysis of this data is shown in Figure 4.6. The CSDs were computed using Equation (4.1) dropping the scaling factor σ/δ^2. Negative CSDs, or current sinks, are represented in blue, and positive CSDs, or current sources, are represented in red. The CSDs were interpolated in the vertical direction for visualization purposes only.

CSD analysis allows to estimate the original current sources and sinks that generated the recorded potentials. When obtained from potentials of linearly spaced electrodes positioned across layers of cerebral cortex, as is the case of Figure 4.6, it allows to study relations between different cortex laminae. It is expected either current sources or sinks in one layer, not both since it is the net activity of all populations in that layer that is recorded. It is believed that a sink is where a cortical signal is generated and that it is then spread to the sources, because sinks are related to excitatory synapses and thus to emission of action potentials by neurons. However, some difficulties arise from the interpretations of the CSD analysis. For example, it is not possible to tell in advance if a spot of negative CSD is due to excitatory synapses or to passive return current from inhibitory synapses [75]. Furthermore, in the analysis of Figure 4.6, it is not explicit what current sources and sinks are part of the same activation group and what are not. This complicates the interpretation of the CSD analysis.

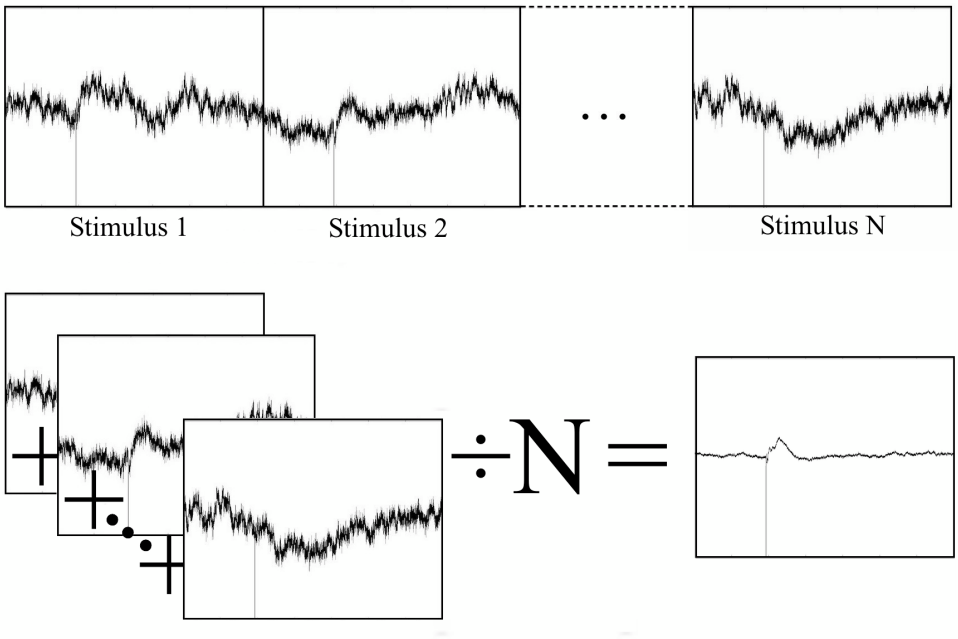


Figure 4.3: Schematic representation of the common procedure used to isolate evoked potentials in one electrode. In this case, data was recorded continuously and a stimulus was given in equally spaced time intervals. Several segments in which the stimulus happened at the same time must be added instant by instant then divided by the number of segments to isolate the EP.

In the following, blind source separation (BSS) techniques will be applied to the experimental set of Figure 4.6 to elicit associations between the current sources and sinks. This may help, for example, identify where is most likely that an activation started. In Section 4.2, it is shown how the data under consideration can be generated according to the instantaneous mixing model and the convolutive mixing model. In Section 4.3 the algorithms used to achieve the separation of the data, that assume either the instantanoues mixing model or the convolutive mixing model, are summarized, and in Section 4.4 the results obtained by each technique are presented.

4.2 Data Model

4.2.1 Instantaneous Mixing Model

Following the description in [20], neglecting inductive effects and for a fixed time interval, the potential at sensor p ($p = 1, ..., P$) at instant n is due to a large number of point current sources and sinks indexed by v,

$$x_p(n) = \sum_{v \in \Gamma} i_v(n) f_{pv}, \tag{4.2}$$

where Γ is a large set of indexes, and f_{pv} is a weight that depends on the location of the vth point current source relative to the pth sensor. A subset of i_v is called an aggregate if their time courses

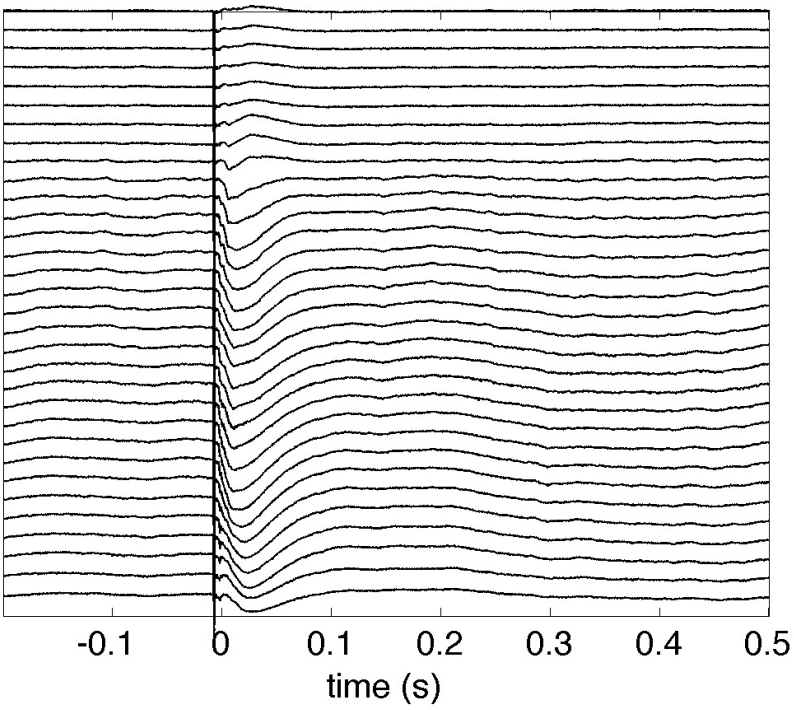


Figure 4.4: Evoked Potential waveforms obtained for each electrode of the experimental set. The EPs are placed such as to reproduce their respective electrode placement during the recording.

differ only by proportionality,

$$i_v(n) = s_q(n)u_v \ , \ v \in \Gamma_q \ , \tag{4.3}$$

where Γ_q corresponds to the indexes of the qth aggregate, $s_q(n)$ is the time course common to the point current sources of the qth aggregate and u_v is the constant of proportionality associated with $i_v(n)$. The compound contribution of aggregate q to the signal recorded at sensor p is

$$\sum_{v\in\Gamma_q} i_v(n)f_{pv} = \sum_{v\in\Gamma_q} s_q(n)u_v f_{pv} = s_q(n)\sum_{v\in\Gamma_q} u_v f_{pv} = s_q(n)h_{pq}, \tag{4.4}$$

where h_{pq} is a weight that depends on the magnitudes of the point current sources belonging to the qth aggregate and their locations with respect to the pth sensor. Finally, the total potential at sensor p can be written as the sum of the contributions from all Q aggregates,

$$x_p(n) = \sum_{q=1}^{Q} s_q(n)h_{pq}. \tag{4.5}$$

Note that the synchronous point current sources of the qth aggregate do not need to be from the same population. According to [9], they can be from spatially distinct neuron populations of separate cortical lamina if their transmembrane currents substantially covary over time.

With a slight change in notation, Equation (4.5) can be rewritten as

$$x_{np} = \sum_{q=1}^{Q} s_{nq}h_{pq}, \tag{4.6}$$

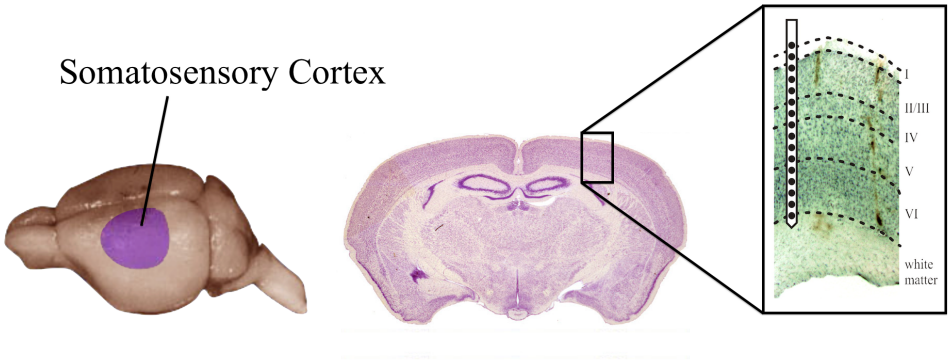


Figure 4.5: To the left, perspective view of the brain of a rat, with the shaded area corresponding to the somatosensory cortex. In the middle, cross section of this same brain. To the right, part of the cortex is zoomed, showing a probe with several electrodes inserted for LFP recordings. Adapted from [72], [73] and [74].

with $n = 1, ..., N$, which allows for the representation

$$\mathbf{X} = \sum_{q=1}^{Q} \mathbf{s}_q \mathbf{h}_q^{\mathsf{T}}, \tag{4.7}$$

where $\mathbf{X}$ is a $N \times P$ matrix with elements x_{np}, $\mathbf{s}_q$ is a $N \times 1$ column vector with elements s_{nq}, and $\mathbf{h}_q$ is a $P \times 1$ column vector with elements h_{pq}. This is equivalent to [22]

$$\mathbf{X} = \mathbf{S}\mathbf{H}^{\mathsf{T}}, \tag{4.8}$$

or

$$\mathbf{X}^{\mathsf{T}} = \mathbf{H}\mathbf{S}^{\mathsf{T}}, \tag{4.9}$$

in which $\mathbf{x}_p$ are the columns of the $N \times P$ matrix $\mathbf{X}$, $\mathbf{h}_q$ are the columns of the $P \times Q$ matrix $\mathbf{H}$ and $\mathbf{s}_q$ are the columns of the $N \times Q$ matrix $\mathbf{S}$. Several approaches exist that try to reconstruct the original signals $\mathbf{s}_q$ only from the set of measurements $\mathbf{x}_p$ in problems that can be described as in Equation (4.9). The general problem consists of finding the matrix $\mathbf{W}$ such that

$$\mathbf{Y}^{\mathsf{T}} = \mathbf{W}\mathbf{X}^{\mathsf{T}}, \tag{4.10}$$

where the columns of $\mathbf{Y}$ consist of signals $\mathbf{y}_q$ such that each $\mathbf{y}_q$ equals one $\mathbf{s}_q$ up to a scaling factor. In Section 4.3 it will be described how to find the matrices $\mathbf{W}_{\text{pca}}$, $\mathbf{W}_{\text{ica}}$, $\mathbf{W}_{\text{stica}}$ using PCA, ICA and spatiotemporal ICA, respectively.

Suppose now that the time courses $s_q(n)$ have a time-frequency representation in which the frequency domain representations differ only by proportionality from one instant to another. This leads to an equation similar to Equation (4.6),

$$x_{pn\omega} = \sum_{q=1}^{Q} a_{pq} b_{nq} c_{\omega q}, \tag{4.11}$$

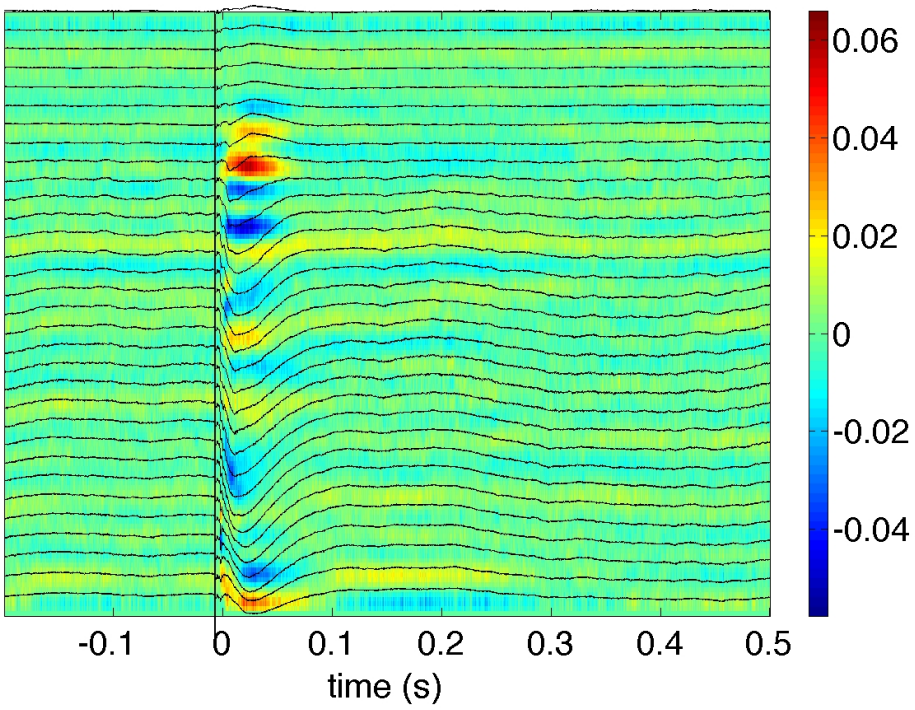


Figure 4.6: Current Source Density analysis of the example experimental set. The computed CSD values are represented in colors. Negative CSDs, or current sinks, are represented in blue, and positive CSDs, or current sources, are represented in red.

where $c_{\omega q}$ is the frequency domain representation, composed of complex numbers, whose amplitude is temporally modulated by the coefficients b_{nq}, and whose resulting time-frequency representation is scaled by a_{pq} for sensor p. Equation (4.11) results in a three way array $\underline{\mathbf{X}}$ of size $P \times N \times \Omega$ with elements $x_{pn\omega}$. In what follows, however, It will be useful to write it in matrix notation [22][13]

$$\begin{aligned}
[\mathbf{X}]_{(1)} = &\left[\sum_{q=1}^{Q} \mathbf{a}_q \mathbf{b}_q^{\mathsf{T}} c_{1q} \ \ldots \ \sum_{q=1}^{Q} \mathbf{a}_q \mathbf{b}_q^{\mathsf{T}} c_{\Omega q}\right] \\
=&[\mathbf{a}_1 \ \ldots \ \mathbf{a}_Q] \begin{bmatrix} \mathbf{b}_1^{\mathsf{T}} c_{11} & \cdots & \mathbf{b}_1^{\mathsf{T}} c_{\Omega 1} \\ \vdots & \ddots & \vdots \\ \mathbf{b}_Q^{\mathsf{T}} c_{1Q} & \cdots & \mathbf{b}_Q^{\mathsf{T}} c_{\Omega Q} \end{bmatrix} \\
=&[\mathbf{a}_1 \ \ldots \ \mathbf{a}_Q] \begin{bmatrix} \mathbf{b}_1 c_{11} & \cdots & \mathbf{b}_Q c_{1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{b}_1 c_{\Omega 1} & \cdots & \mathbf{b}_Q c_{\Omega Q} \end{bmatrix}^{\mathsf{T}} \\
=&\mathbf{A}(\mathbf{C} \diamond \mathbf{B})^{\mathsf{T}},
\end{aligned} \tag{4.12}$$

where $\diamond$ is the Khatri-Rao product and $[.]_{(i)}$ denotes the ith mode unfolding of a multidimensional array as defined in [13]. The matrix $\mathbf{C} \diamond \mathbf{B}$ contains, in its columns, the time-frequency representation of each aggregate signal, such that the first N elements of the column correspond to the first frequency, the next N elements of the column correspond to the second frequency, and so on. Define $\mathbf{A}_{\text{pfac}}$, $\mathbf{B}_{\text{pfac}}$ and $\mathbf{C}_{\text{pfac}}$ the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ found by the PARAFAC algorithm as

described in Section 4.3.5. By analogy to Equation (4.10), define

$$\mathbf{W}_{\text{pfac}} = \mathbf{A}^{+}_{\text{pfac}} = (\mathbf{A}^{\mathsf{H}}_{\text{pfac}}\mathbf{A}_{\text{pfac}})^{-1}\mathbf{A}^{\mathsf{H}}_{\text{pfac}}, \tag{4.13}$$

if $\mathbf{A}_{\text{pfac}}$ is full column rank, where + indicates the Moore-Penrose pseudoinverse and H indicates the conjugate transpose.

4.2.2 Convolutive Mixing Model

No reflection, reverberation or echo take place in the mixing of brain electric signals picked up by extracellular electrodes. However, there may be delayed correlation between different brain regions, in which case the brain activity itself can be modeled by an M-tap convolutive mixing [23]

$$x_p(n) = \sum_{q=1}^{Q}\sum_{k=0}^{M-1} h_{qp}(k)s_q(n-k), \tag{4.14}$$

where the original signals $s_q(n)$, $q = 1, ..., Q$, are filtered by a multiple input multiple output (MIMO) system before they are picked up in the sensor signals $x_q(n)$, $p = 1, ..., P$. In the following, the number of original signals Q will be considered the same as the number of sensor signals P. The $s_q(n)$ will be reconstructed by a similar L-tap MIMO demixing system

$$y_q(n) = \sum_{p=1}^{P}\sum_{k=0}^{L-1} w_{pq}(k)x_p(n-k), \tag{4.15}$$

where $y_q(n)$ are the output signals. It can be shown that, if L is chosen at least equal to M and a proper MIMO demixing system is used, each $y_q(n)$ should equal one of the $s_q(n)$ up to a filtering [63].

For the method described in Section 4.3.4 (TRINICON algorithm) to find the MIMO demixing system, define [76] [77]

$$\mathbf{W}_{pq}(m) = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{4.16}$$

$$\mathbf{W}_{\text{trin}}(m) = \begin{bmatrix} \mathbf{W}_{11}(m) & \cdots & \mathbf{W}_{1P}(m) \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1}(m) & \cdots & \mathbf{W}_{PP}(m) \end{bmatrix}, \tag{4.17}$$

$$\mathbf{x}_p(m,j) = \left[\begin{array}{cccc} x_p(mL+j) & \ldots & x_p(mL+j-2L+1) \end{array} \right], \tag{4.18}$$

$$\mathbf{x}(m,j) = \left[\begin{array}{cccc} \mathbf{x}_1(m,j) & \ldots & \mathbf{x}_P(m,j) \end{array} \right], \tag{4.19}$$

$$\mathbf{y}_q(m,j) = \left[\begin{array}{cccc} y_q(mL+j) & \ldots & y_q(mL+j-D+1) \end{array} \right], \tag{4.20}$$

$$\mathbf{y}(m,j) = \left[\begin{array}{cccc} \mathbf{y}_1(m,j) & \ldots & \mathbf{y}_Q(m,j) \end{array} \right], \tag{4.21}$$

where $\mathbf{W}_{pq}(m)$ is $2L \times D$, $\mathbf{x}_p(m,j)$ is $1 \times 2L$, $\mathbf{y}_q(m,j)$ is $1 \times D$, m is the block index, $j = 0, ..., N-1$ is a time-shift index, and D denotes the number of time lags to exploit the nonwhiteness of the original signals. Then

$$\mathbf{y}(m,j) = \mathbf{x}(m,j)\mathbf{W}_{\text{trin}}(m). \tag{4.22}$$

4.3 Methods

4.3.1 PCA

PCA is closely related to the matrix Singular Value Decomposition (SVD). The truncated SVD of the recorded data matrix $\mathbf{X}$ is

$$\tilde{\mathbf{X}} = \mathbf{U}_R \mathbf{\Sigma}_R \mathbf{V}_R^{\mathsf{T}}, \tag{4.23}$$

where $\tilde{\mathbf{X}}$ is an approximation to $\mathbf{X}$, $\mathbf{U}_R$ have orthonormal columns, $\mathbf{\Sigma}_R$ is diagonal, and $\mathbf{V}_R$ have orthonormal columns. It can be found by first identifying the best rank one approximation [78]

$$\mathbf{X} = \sigma \mathbf{u}\mathbf{v}^{\mathsf{T}} + \mathbf{E} \tag{4.24}$$

in the least squares sense, where σ is a scaling constant term, $\mathbf{u}$ and $\mathbf{v}$ are $N \times 1$ and $P \times 1$ vectors normalized to unit norm, respectively, and $\mathbf{E}$ is a matrix of residual errors. This can be done by solving the classical eigenvalue problems

$$\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{v} = \sigma^2 \mathbf{v}, \tag{4.25}$$

for the maximum eigenvalue and one corresponding unit norm eigenvector, and

$$\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{u} = \sigma^2 \mathbf{u}, \tag{4.26}$$

for the maximum eigenvalue and one corresponding unit norm eigenvector. The nonzero eigenvalues of $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{X}\mathbf{X}^{\mathsf{T}}$ are the same [52]. Call the σ^2, $\mathbf{u}$ and $\mathbf{v}$ found σ_1^2, $\mathbf{u}_1$ and $\mathbf{v}_1$, respectively. Subtract it from the original data,

$$\mathbf{X}_1 = \mathbf{X} - \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{T}}, \tag{4.27}$$

then do the same to $\mathbf{X}_{j-1}$ $(j = 2, .., R)$

$$\mathbf{X}_j = \mathbf{X}_{j-1} - \sigma_j \mathbf{u}_j \mathbf{v}_j^{\mathsf{T}}. \tag{4.28}$$

The $\mathbf{u}_j$ are arranged in the columns of $\mathbf{U}_R$, $\mathbf{v}_j$ in the columns of $\mathbf{V}_R$ and σ_j in the diagonal entries of $\mathbf{\Sigma}_R$. If every column of $\mathbf{X}$ is zero mean, $\mathbf{W}_{\text{pca}}$ can be defined as

$$\mathbf{W}_{\text{pca}} = (\mathbf{V}_R^{\mathsf{T}})^{+} = \mathbf{V}_R^{\mathsf{T}}. \tag{4.29}$$

4.3.2 ICA

The Bell-Sejnowski algorithm [60] is one of the several algorithms for independent component analysis. Suppose the number of original signals Q is equal to the number of sensors signals P, and that each sensor signal $x_p(n)$ have zero mean. Define

$$\mathbf{x}(n) = [x_1(n) \ ... \ x_P(n)]^{\mathsf{T}}, \tag{4.30}$$

$$\mathbf{y}(n) = [y_1(n) \ ... \ y_Q(n)]^{\mathsf{T}} = \mathbf{W}_{\text{ica}}\mathbf{x}(n), \tag{4.31}$$

$$\begin{aligned} \mathbf{z}(n) &= [z_1(n) \ ... \ z_Q(n)]^{\mathsf{T}} \\ &= g(\mathbf{y}(n)) = [g(y_1(n)) \ ... \ g(y_Q(n))]^{\mathsf{T}}, \end{aligned} \tag{4.32}$$

where $g(u) = (1 + e^{-u})^{-1}$ is the sigmoid function, which approximates the form of the cumulative density function of each original signal $s_q(n)$. The algorithm consists of maximizing the multidimensional differential entropy

$$\begin{aligned} H(\mathbf{z}) &= -E[\ln f_{\mathbf{z}}(\mathbf{z})] = -\int f_{\mathbf{z}}(\mathbf{z}) \ln f_{\mathbf{z}}(\mathbf{z}) d\mathbf{z} \\ &= E[\ln |\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|] - E[\ln f_{\mathbf{x}}(\mathbf{x})], \end{aligned} \tag{4.33}$$

where $f_{\mathbf{z}}(\mathbf{z})$ is the multidimensional probability density function of $\mathbf{z}$, $f_{\mathbf{x}}(\mathbf{x})$ is the multidimensional probability density function of $\mathbf{x}$, the term $-E[\ln f_{\mathbf{x}}(\mathbf{x})]$ equals $H(\mathbf{x})$ and $\mathbf{J}_{\mathbf{z}}(\mathbf{x})$ is the Jacobian matrix

$$\mathbf{J}_{\mathbf{z}}(\mathbf{x}) = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_1}{\partial x_P} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_Q}{\partial x_P} & \cdots & \frac{\partial z_Q}{\partial x_P} \end{bmatrix}. \tag{4.34}$$

Since $H(\mathbf{x})$ is unaffected by alterations in $\mathbf{W}_{\text{ica}}$, and the gradient $\nabla|\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|$ with respect to $\mathbf{W}_{\text{ica}}$ is

$$\begin{aligned} \frac{\partial \ln |\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|}{\partial \mathbf{W}_{\text{ica}}} &= \begin{bmatrix} \frac{\partial \ln |\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|}{\partial w_{11}^{\text{ica}}} & \cdots & \frac{\partial \ln |\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|}{\partial w_{1P}^{\text{ica}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln |\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|}{\partial w_{Q1}^{\text{ica}}} & \cdots & \frac{\partial \ln |\det \mathbf{J}_{\mathbf{z}}(\mathbf{x})|}{\partial w_{QP}^{\text{ica}}} \end{bmatrix} \\ &= [\mathbf{W}_{\text{ica}}^{\mathsf{T}}]^{-1} + (\mathbf{1} - 2\mathbf{z})\mathbf{x}^{\mathsf{T}}, \end{aligned} \tag{4.35}$$

one can write the following stochastic gradient ascent rule

$$\Delta \mathbf{W}_{\text{ica}} = \mu \left([\mathbf{W}_{\text{ica}}^{\mathsf{T}}]^{-1} + (\mathbf{1} - 2\mathbf{z})\mathbf{x}^{\mathsf{T}} \right), \tag{4.36}$$

where $\mathbf{1}$ is a vector of ones and μ is the learning rate.

4.3.3 Spatiotemporal ICA

Spatiotemporal ICA maximizes the degree of independence over space and time, without necessarily producing independence in either space or time [65]. It treats both the columns $\mathbf{h}_q$ of $\mathbf{H}$ and the columns $\mathbf{s}_q$ of $\mathbf{S}$ as signals. Suppose all the $\mathbf{h}_q$ and all the $\mathbf{s}_q$ have zero mean. The first step of stICA is to apply truncated SVD to $\mathbf{X}$, resulting in

$$\tilde{\mathbf{X}} = \mathbf{U}_R\mathbf{\Sigma}_R\mathbf{V}_R^\mathsf{T} = \mathbf{U}_R\mathbf{\Sigma}_R^{1/2}(\mathbf{V}_R(\mathbf{\Sigma}_R^{1/2})^\mathsf{T})^\mathsf{T} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\mathsf{T}, \tag{4.37}$$

where $\mathbf{U}_R$ is $N \times R$, $\mathbf{\Sigma}_R$ is $R \times R$ and $\mathbf{V}_R$ is $P \times R$. The next step is to define the $N \times R$ matrix $\mathbf{T}$, the $R \times R$ diagonal matrix $\mathbf{\Lambda}$, the $P \times R$ matrix $\mathbf{E}$, and the $R \times R$ matrices $\mathbf{W}_T$ and $\mathbf{W}_E$ such that

$$\mathbf{T} = \tilde{\mathbf{U}}\mathbf{W}_T, \tag{4.38}$$

$$\mathbf{E} = \tilde{\mathbf{V}}\mathbf{W}_E, \tag{4.39}$$

$$\tilde{\mathbf{X}} = \mathbf{T}\mathbf{\Lambda}\mathbf{E}^\mathsf{T}, \tag{4.40}$$

in which $\mathbf{T}$ contains in its columns maximally statistically independent temporal signals, and $\mathbf{E}$ contains in its columns maximally statistically independent spatial signals. Since, for Equation (4.37) and Equation (4.40) to be valid, $\mathbf{W}_E = (\mathbf{W}_T^{-1})^\mathsf{T}(\mathbf{\Lambda}^{-1})^\mathsf{T}$, the matrices $\mathbf{W}_T$, $\mathbf{W}_E$ and $\mathbf{\Lambda}$ can be found simultaneously by maximizing

$$h(\mathbf{W}_T, \mathbf{\Lambda}) = (1-\alpha)h_T + \alpha h_E, \tag{4.41}$$

$$h_E = \ln|\det \mathbf{W}_E| + \frac{1}{P}\sum_{i=1}^{P}\sum_{j=1}^{R}\ln\sigma'(e_{ij}), \tag{4.42}$$

$$h_T = \ln|\det \mathbf{W}_T| + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{R}\ln\tau'(t_{ij}), \tag{4.43}$$

with respect to $\mathbf{W}_T$ and $\mathbf{\Lambda}$ only, where h_E is the spatial entropy, h_T is the temporal entropy, $0 \leq \alpha \leq 1$ defines the relative weighting afforded to the spatial and temporal entropies, e_{ij} are the elements of the matrix $\mathbf{E}$, t_{ij} are the elements of the matrix $\mathbf{T}$, σ approximates the cumulative density function of each spatial signal, τ approximates the cumulative density function of each temporal signal, and σ' and τ' are their derivatives. This can be done by using gradient ascent, for example, with the following update rules

$$\Delta\mathbf{W}_T = \mu_1\frac{\partial h_{ET}(\mathbf{W}_T, \mathbf{\Lambda})}{\partial \mathbf{W}_T}, \tag{4.44}$$

$$\Delta\mathbf{\Lambda} = \mu_2\frac{\partial h_{ET}(\mathbf{W}_T, \mathbf{\Lambda})}{\partial \mathbf{\Lambda}}, \tag{4.45}$$

where μ_1 and μ_2 are the learning rates of the first and second update rules respectively. Finally, $\mathbf{W}_{\text{stica}}$ can be defined as

$$\mathbf{W}_{\text{stica}} = (\mathbf{E}\mathbf{\Lambda}^\mathsf{T})^+. \tag{4.46}$$

4.3.4 TRINICON

In [24] it was introduced the following cost function to find the demixing filter matrix of Equation (4.17) that takes into account the nongaussianity, nonwhiteness and nonstationarity of the original signals

$$\mathcal{J}(m) = -\sum_{i=0}^{\infty} \beta(i,m) \frac{1}{N} \sum_{j=0}^{N-1} \{\log(\hat{p}_{s,PD}(\mathbf{y}(i,j))) - \log(\hat{p}_{y,PD}(\mathbf{y}(i,j)))\}, \tag{4.47}$$

where $\hat{p}_{s,PD}(.)$ is the desired original multivariate probability density function with dimension PD, $\hat{p}_{y,PD}(.)$ is the output multivariate probability density function with dimension PD, $\beta(i,m)$ is a window function that prioritizes new blocks but still uses previous blocks [79].

The natural gradient $\tilde{\nabla}\mathcal{J}(m)$ with respect to $\mathbf{W}_{\text{trin}}$, in this case defined as

$$\tilde{\nabla}\mathcal{J}(m) = \mathbf{W}_{\text{trin}} \mathbf{W}_{\text{trin}}^{\mathsf{T}} \nabla\mathcal{J}(m), \tag{4.48}$$

where $\nabla\mathcal{J}(m)$ is the gradient with respect to $\mathbf{W}_{\text{trin}}$, is

$$\tilde{\nabla}\mathcal{J}(m) = \frac{2}{N} \sum_{i=0}^{\infty} \beta(i,m) \frac{1}{N} \sum_{j=0}^{N-1} \mathbf{W}_{\text{trin}}(i) \mathbf{y}^{\mathsf{T}}(i,j) \{\mathbf{\Phi}_{s,PD}(\mathbf{y}(i,j)) - \mathbf{\Phi}_{y,PD}(\mathbf{y}(i,j))\}, \tag{4.49}$$

$$\mathbf{\Phi}_{s,PD}(\mathbf{y}(i,j)) = \frac{\partial \log(\hat{p}_{s,PD}(\mathbf{y}(i,j)))}{\partial \mathbf{y}(i,j)}, \tag{4.50}$$

$$\mathbf{\Phi}_{y,PD}(\mathbf{y}(i,j)) = \frac{\partial \log(\hat{p}_{y,PD}(\mathbf{y}(i,j)))}{\partial \mathbf{y}(i,j)}. \tag{4.51}$$

Using only second order statistics, the gradient in Equation (4.49) can be simplified to

$$\tilde{\nabla}\mathcal{J}(m) = 2 \sum_{i=0}^{\infty} \beta(i,m) \mathbf{W}_{\text{trin}}(i) \{\hat{\mathbf{R}}_{yy}(i) - \hat{\mathbf{R}}_{ss}(i)\} \hat{\mathbf{R}}_{ss}^{-1}(i). \tag{4.52}$$

Where $\hat{\mathbf{R}}_{yy}(i)$ and $\hat{\mathbf{R}}_{ss}(i)$ are $PD \times PD$ correlation matrices. This results in the following update rule

$$\mathbf{W}_{\text{trin}}(m) = \mathbf{W}_{\text{trin}}(m-1) - \mu \tilde{\nabla}\mathcal{J}(m). \tag{4.53}$$

Going back to Equation (4.15), the overall Multiple Input Multiple Output (MIMO) Linear Time Invariant (LTI) demixing system can be written in the z-domain as [80]

$$\mathbf{Y}(z) = \mathbf{W}(z)\mathbf{X}(z), \tag{4.54}$$

where $\mathbf{X}(z)$ and $\mathbf{Y}(z)$ are the z-transform of $\mathbf{x}(n) = [x_1(n) \ldots x_P(n)]^{\mathsf{T}}$ and $\mathbf{y}(n) = [y_1(n) \ldots y_Q(n)]^{\mathsf{T}}$ respectively and $\mathbf{W}(z)$ contains the transfer functions $W_{pq}(z) = \mathcal{Z}[w_{pq}(n)]$ of the demixing filters. The fourier transform of a sequence is obtained from its z-transform by setting $z = \exp(j\omega)$ [63], which gives

$$\mathbf{Y}(\omega) = \mathbf{W}(\omega)\mathbf{X}(\omega). \tag{4.55}$$

After obtaining the estimated coefficients $w_{pq}(k)$ by TRINICON, and obtaining the reconstructed signals $\mathbf{y}(n)$, the part of $\mathbf{x}(n)$ generated by a single source or a group of sources will be reconstructed

in this work by setting to zero the signals $y_q(n)$ not of interest, and then, for each frequency ω, making

$$\mathbf{X}(\omega) = \mathbf{W}^{-1}(\omega)\mathbf{Y}(\omega) \tag{4.56}$$

and taking the inverse fourier transform of $\mathbf{X}(\omega)$.

4.3.5 PARAFAC

Consider the complex valued three way array $\underline{\mathbf{X}}$ of size $P \times N \times \Omega$. The following are equivalent ways of writing its R-component PARAFAC model $\underline{\tilde{\mathbf{X}}}$:

$$[\underline{\tilde{\mathbf{X}}}]_{(1)} = \mathbf{A}_{\text{pfac}}(\mathbf{C}_{\text{pfac}} \diamond \mathbf{B}_{\text{pfac}})^{\mathsf{T}}, \tag{4.57}$$

$$[\underline{\tilde{\mathbf{X}}}]_{(2)} = \mathbf{B}_{\text{pfac}}(\mathbf{C}_{\text{pfac}} \diamond \mathbf{A}_{\text{pfac}})^{\mathsf{T}}, \tag{4.58}$$

$$[\underline{\tilde{\mathbf{X}}}]_{(3)} = \mathbf{C}_{\text{pfac}}(\mathbf{B}_{\text{pfac}} \diamond \mathbf{A}_{\text{pfac}})^{\mathsf{T}}, \tag{4.59}$$

where $\mathbf{B}_{\text{pfac}}$ is $N \times R$, $\mathbf{C}_{\text{pfac}}$ is $\Omega \times R$, $\mathbf{A}_{\text{pfac}}$ is $P \times R$. This model can be found by successively assuming two of the matrices $\mathbf{A}_{\text{pfac}}$, $\mathbf{B}_{\text{pfac}}$ and $\mathbf{C}_{\text{pfac}}$ known, and then estimating the unknown matrix by least squares regression [81]. This is called the Alternating Least Squares (ALS) algorithm. First consider that $\mathbf{B}_{\text{pfac}}$ and $\mathbf{C}_{\text{pfac}}$ are known. Then $\mathbf{A}_{\text{pfac}}$ can be estimated as [22][13]

$$\mathbf{A}_{\text{pfac}} = [\underline{\mathbf{X}}]_{(1)}(\mathbf{C}_{\text{pfac}} \diamond \mathbf{B}_{\text{pfac}})^{\mathsf{T}+}. \tag{4.60}$$

Next consider that $\mathbf{A}_{\text{pfac}}$ and $\mathbf{C}_{\text{pfac}}$ are known. Then $\mathbf{B}_{\text{pfac}}$ can be estimated as

$$\mathbf{B}_{\text{pfac}} = [\underline{\mathbf{X}}]_{(2)}(\mathbf{C}_{\text{pfac}} \diamond \mathbf{A}_{\text{pfac}})^{\mathsf{T}+}. \tag{4.61}$$

Finally consider that $\mathbf{A}_{\text{pfac}}$ and $\mathbf{B}_{\text{pfac}}$ are known. Then $\mathbf{C}_{\text{pfac}}$ can be estimated as

$$\mathbf{C}_{\text{pfac}} = [\underline{\mathbf{X}}]_{(3)}(\mathbf{B}_{\text{pfac}} \diamond \mathbf{A}_{\text{pfac}})^{\mathsf{T}+}. \tag{4.62}$$

These steps should be repeated one after the other until convergence. These same equations can be used to find the R-component PARAFAC model of a real valued three way array $\underline{\mathbf{X}}$, since a real number is just a complex number with no imaginary part.

In the complex valued PARAFAC model, there is no guarantee that only the matrix $\mathbf{C}_{\text{pfac}}$ will be composed of complex numbers and that matrices $\mathbf{A}_{\text{pfac}}$ and $\mathbf{B}_{\text{pfac}}$ will be composed of real numbers, but in the model described in the previous section, only the matrix $\mathbf{C}$ is complex valued, while matrices $\mathbf{A}$ and $\mathbf{B}$ are positive. This problem is usually solved by assuming just the amplitude spectrum of the signals instead of the complete frequency representation, with magnitude and phase. This results in a real valued three way array $\underline{\mathbf{X}}$, which can be analyzed using a real valued PARAFAC model, that allows the addition of suitable constraints, like nonnegativity [78]. In many situations this approach is sufficient, since the vectors $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ (columns of the matrices $\mathbf{A}_{\text{pfac}}$, $\mathbf{B}_{\text{pfac}}$ and $\mathbf{C}_{\text{pfac}}$, respectively) already allow to analyze the location, the timing and the power spectrum of each component R [82][13], but they do not allow to reconstruct the original time signals, since the phase information has been discarded. In the problem at hand,

however, the estimates of the original time signals are required to compute the current source densities. For this reason, the following approach is proposed: first, assume just the amplitude spectrum of the signals and compute a real valued R-component PARAFAC model; then assume that the element-wise phase is approximately the same for all components and for the original three way array, which allows for the reconstruction of the original time signals. This can be expressed mathematically as follows. Consider Equation (4.11) again

$$x_{pn\omega} = \sum_{q=1}^{Q} a_{pq} b_{nq} c_{\omega q}. \tag{4.63}$$

Each $x_{pn\omega}$ is a complex number composed of magnitude and phase, $x_{pn\omega} = |x_{pn\omega}|e^{j\angle x_{pn\omega}}$, as well as $c_{\omega q} = |c_{\omega q}|e^{j\angle c_{\omega q}}$, then

$$|x_{pn\omega}|e^{j\angle x_{pn\omega}} = \sum_{q=1}^{Q} a_{pq} b_{nq} |c_{\omega q}| e^{j\angle c_{\omega q}}. \tag{4.64}$$

Instead of assuming each $\angle c_{\omega q}$ dependent only in the frequency ω and the component q, it will be assumed that

$$|x_{pn\omega}|e^{j\angle x_{pn\omega}} = \sum_{q=1}^{Q} \alpha_{pq}\beta_{nq}\gamma_{\omega q} e^{j\angle x_{pn\omega}}, \tag{4.65}$$

where α, β and γ are all positive. The complex exponentials in Equation (4.65) can be disregarded, yielding

$$|x_{pn\omega}| = \sum_{q=1}^{Q} \alpha_{pq}\beta_{nq}\gamma_{\omega q}. \tag{4.66}$$

Therefore one can first estimate the components according to Equation (4.66), and then include the phase as in Equation (4.65).

The model of Equation (4.66) has the same form as the real valued PARAFAC model. It is also involves only positive numbers. To restrict all parameters of the model to be positive, nonnegativity constraints will be imposed in all matrices $\mathbf{A}_{\text{pfac}}$, $\mathbf{B}_{\text{pfac}}$ and $\mathbf{C}_{\text{pfac}}$. Nonnegativity can be imposed in the least squares regression step for each matrix. The reader is referred to [22] for details on the algorithm.

4.4 Results

All techniques were applied to the voltage time courses from -0.005s to 0.095s, except for stICA, which was applied after CSD analysis. Third-party codes used were: the stICA code from J.V. Stone and J. Porrill [83]; the *runica* function, part of the EEGLAB toolbox [84]; the N-WAY toolbox by Rasmus Bro [85], and the complex valued PARAFAC code from Nicholas D. Sidiropoulos and Rasmus Bro [86].

In PCA analysis of EEG signals, each electrode is commonly taken as a variable, and each time point as an observation. We verified that centering the data across observations mode deteriorated the CSD figure. In other words, subtracting each electrode signal by its mean affected the computation of the second derivative across the electrodes mode by creating artificial current sources

and sinks. PCA without centering has been used before and finds application in ecology, chemistry and geology [56]. The technique finds directions of maximum variance through the origin rather than through the centroid of the data set. This is a problem if the center of the observations is a long way from the origin. This is not the case, however, since EEG electrode means tend to be relatively small [87]. Figure 4.7 shows the CSDs of all LFP components extracted by PCA. Figure 4.8 shows the first and second component LFPs together with their CSDs, as well as the recovered signal using only these two components. Figure 4.9 is the same but without the LFPs to better visualize the CSDs. The way to interpret the results from PCA, and the result from any other technique, is that the current sources and sinks that appear in one component are related to each other and are not related to the current sources and sinks that appear in any other component. The difference between the different techniques is the criteria used to state that they are related or not. The criteria used by PCA is statistical uncorrelation of the LFP temporal courses and orthogonality of the LFP spatial profiles.

Figure 4.10 shows the CSDs of all LFP components extracted by ICA. It can be seen that, except for the first few components, ICA basically breaks down the CSD figure in different time instants. A few of such components were selected and ordered in Figure 4.11, which shows the LFPs and the CSDs, as well as the recovered signal using only these components. Figure 4.12 is the same but without the LFPs. The LFP time course of each component is statistically independent from the LFP time course of the other components. However, it can be seen that, for each component corresponding to a specific instant, all current sources and sinks are related. This incapacity of separating the current sources and sinks may be due to the fact that the signals analyzed are evoked potentials, and it is possible that in this case the signals from different underlying sources of signal may not me considered independent. Other possible interpretation is that the signal was indeed generated by a single indivisible underlying source of signal.

The analysis using spatiotemporal ICA was done based on [16], in which stICA was used to analyze a similar dataset. We used

$$\sigma(u) = \int_{-\infty}^{u} \operatorname{sech}^2(u)du, \tag{4.67}$$

to approximate the high kurtosis of the localized spatial components, and

$$\tau(u) = \int_{-\infty}^{u} e^{-u^4}du, \tag{4.68}$$

to approximate the low kurtosis of the oscillatory time courses. As in [16], we verified that stICA with $\alpha = 1$, i.e. purely spatial decomposition, applied after CSD analysis gave the best results. The modeling of CSD waveforms rather than LFP waveforms is justified because the same source in the CSD picture is much more localized than its potential, yielding a higher kurtosis, and the used algorithm searches for high kurtosis in the spatial domain [16]. Additionally, it was verified that changing the number of components of the truncated SVD step affected the stICA components. The solution for a truncated SVD of two components was selected. Figure 4.13 shows the first and second components, as well as the recovered signal using these components. It can be seen that it clearly separated the current source and sink at the bottom of the first component, as well as other current sources and sinks, from the current sources and sinks of the second component.

Previously it was stated, based on [23], that the mixing of the electric activity of the brain is not a convolutive, but the convolutive mixing model could be useful to model the brain activity itself. In fact, it was possible to find a solution very similar to stICA by using TRINICON with $L = 32$. Figure 4.14 shows the CSDs of all LFP components extracted by TRINICON. The third component was very similar to the first component from stICA. The contribution of the third TRINICON component and the contribution of all other components are shown separately in Figure 4.15, which shows the LFPs as well as the CSDs, and the recovered signals using all components. Figure 4.16 is the same but without the LFPs.

In order to use PARAFAC on this data, each electrode signal had to be analyzed using short time fourier transform (other invertible time frequency analysis techniques could be used, but the STFT was chosen in this work). For that, a kaiser window of duration 400 samples was used, with short time analyses being performed every 10 samples. Tests were also made using rectangular, hamming and triangular windows (windows that tapered to zero were avoided since each inverse discrete fourier transform run had to be divided by the window in order to reconstruct the signal). After obtaining a three way complex-valued data of electrodes by time by frequency, the method proposed in the previous section was used to model the data, that means: it was taken the absolute value of each entry of the complex-valued data; a three component PARAFAC model with nonnegativity constraints in all modes was fitted to the resulting array; the phase of each element of each three-way component was estimated by assuming it equal to the phase of the corresponding element in the original three-way array to result in a model like that of Equation (4.65). The contribution of each three-way component to the original data was synthesized by inverting the short time fourier transform on each component, using the conventional overlap add algorithm [88]. The resulting component LFPs together with their CSDs are shown in Figure 4.17, as well as the recovered signal using these components. Figure 4.18 is the same but without the LFPs. The recovered signal is somewhat deteriorated, but it is not possible to tell in advance if this is because of the PARAFAC modeling of the absolute values or the approximation of the phase proposed in this work. The obtained result is again similar to the result of stICA, but with the components reversed. As a matter of comparison Figure 4.19 shows the result of a three component complex valued PARAFAC model. It can be seen that allowing the time courses, the frequency domain representation and the electrode weights to be all complex valued completely deteriorates the component LFPs and CSDs, as well as the recovered LFPs and CSDs.

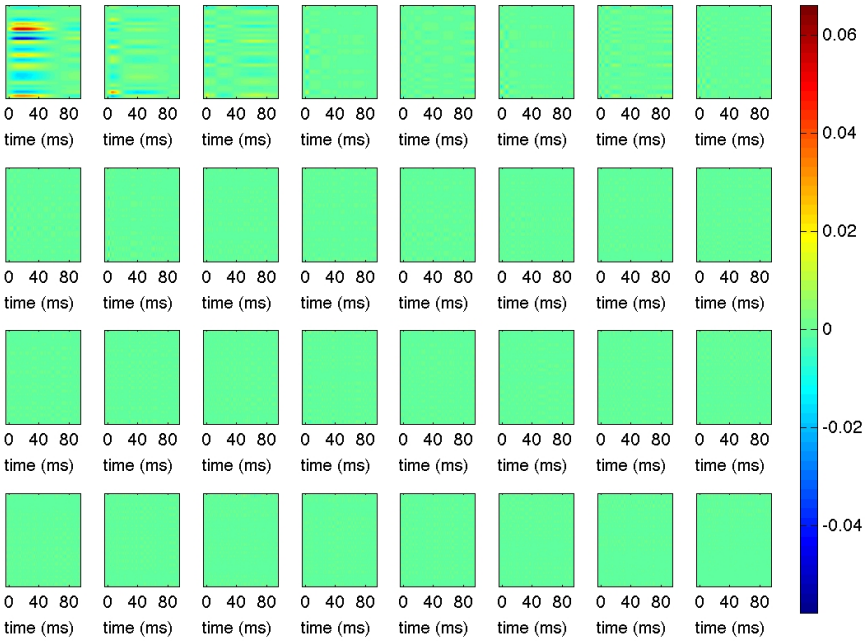


Figure 4.7: CSDs of all LFP components extracted by PCA. The components are ordered according to their variances, such that the first component correspond to the highest variance. First row correspond to the first eight components, second row correspond to the next eight components, and so on.

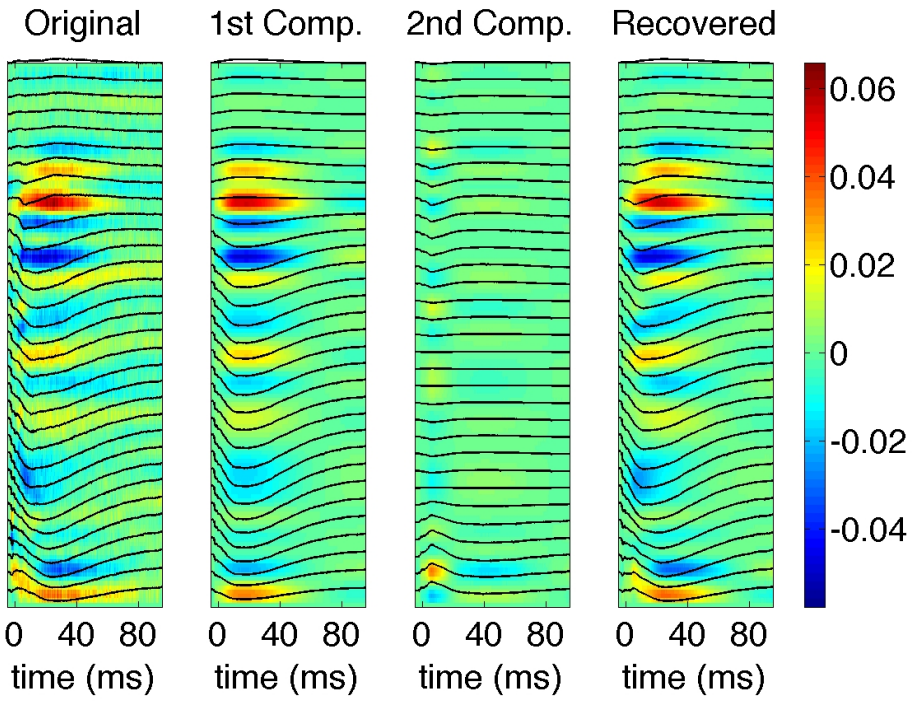


Figure 4.8: LFPs and CSDs corresponding to: the original signal; the first and second components extracted by PCA; and the signal recovered using only these two components.

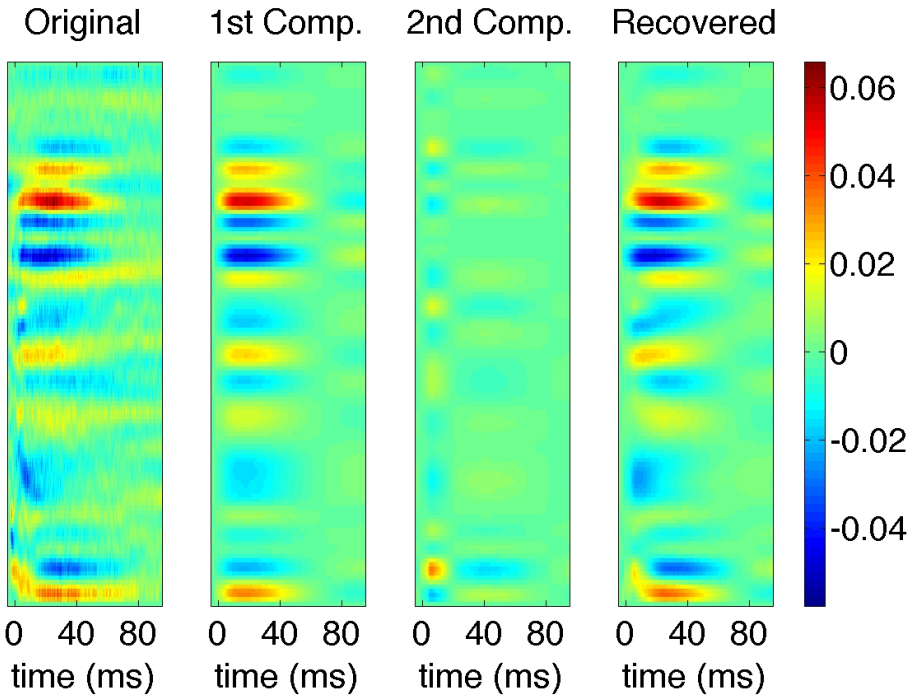


Figure 4.9: CSDs corresponding to: the original signal; the first and second components extracted by PCA; and the signal recovered using only these two components.

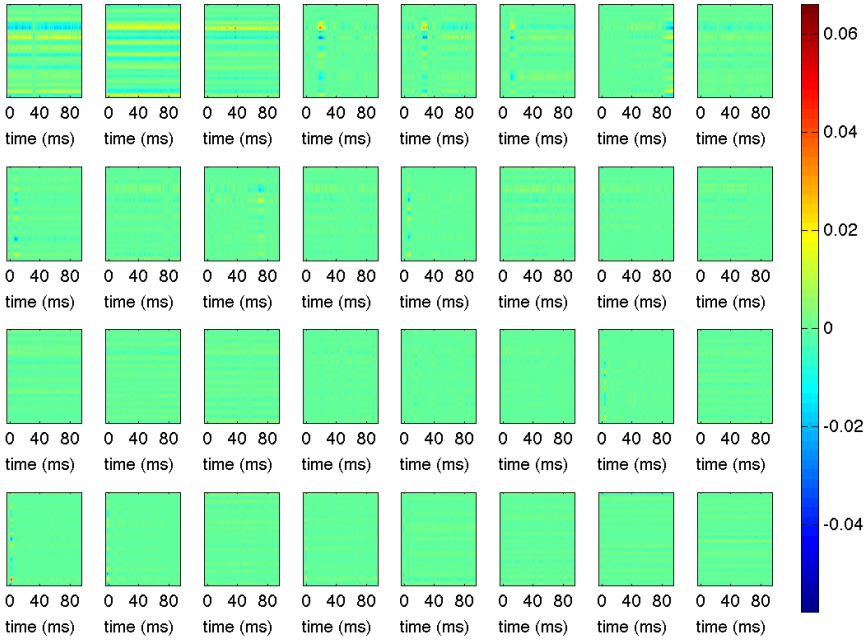


Figure 4.10: CSDs of all LFP components obtained using ICA. Components sorted in descending order of mean projected variance. First row correspond to the first eight components, second row correspond to the next eight components, and so on.

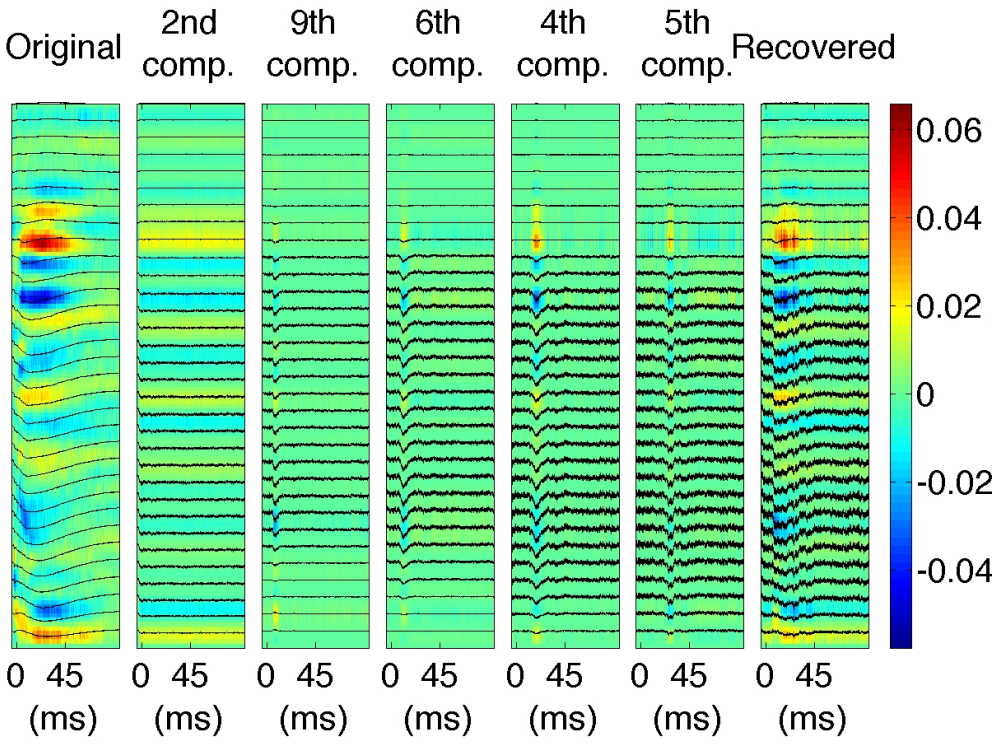


Figure 4.11: LFPs and CSDs corresponding to: the original signal; selected and ordered components extracted by ICA; and the signal recovered using only these components.

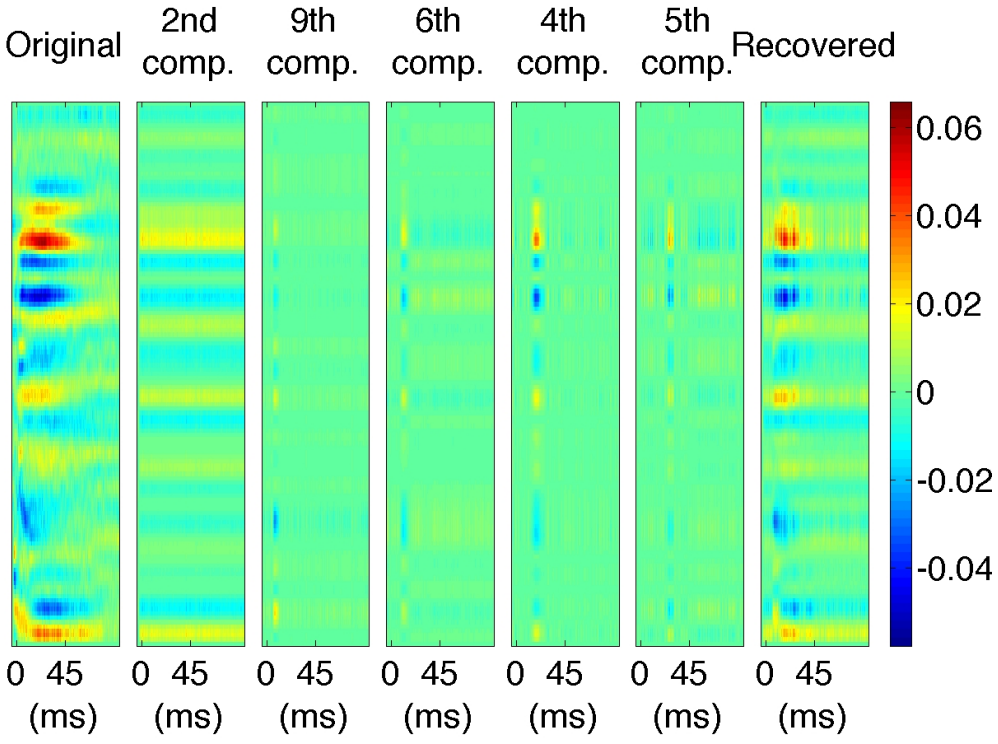


Figure 4.12: CSDs corresponding to: the original signal; selected and ordered components extracted by ICA; and the signal recovered using only these components.

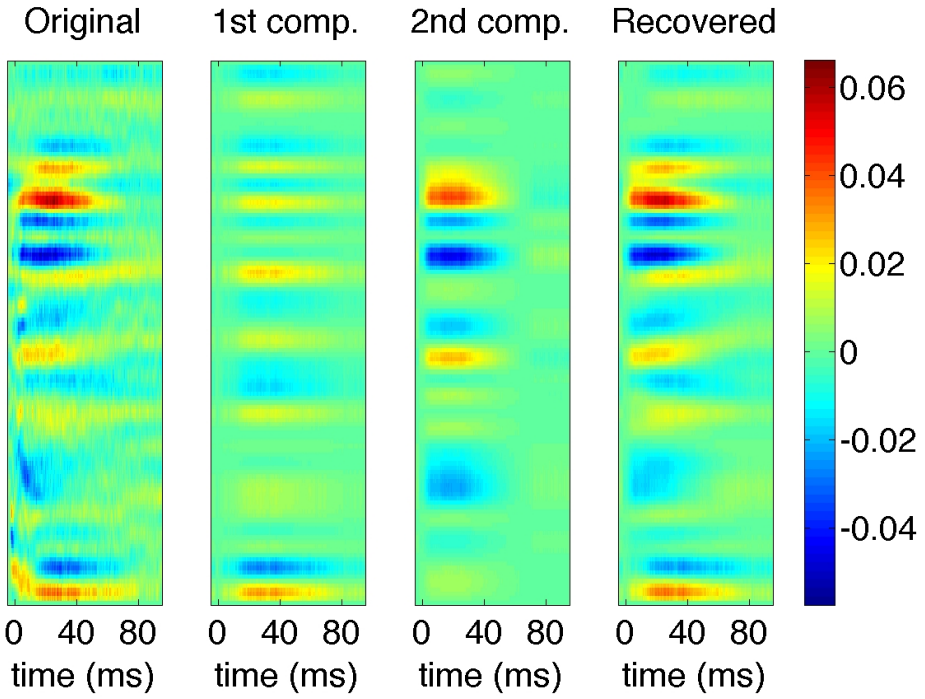


Figure 4.13: CSDs corresponding to: the original signal; components obtained using stICA; and the signal recovered using these components.

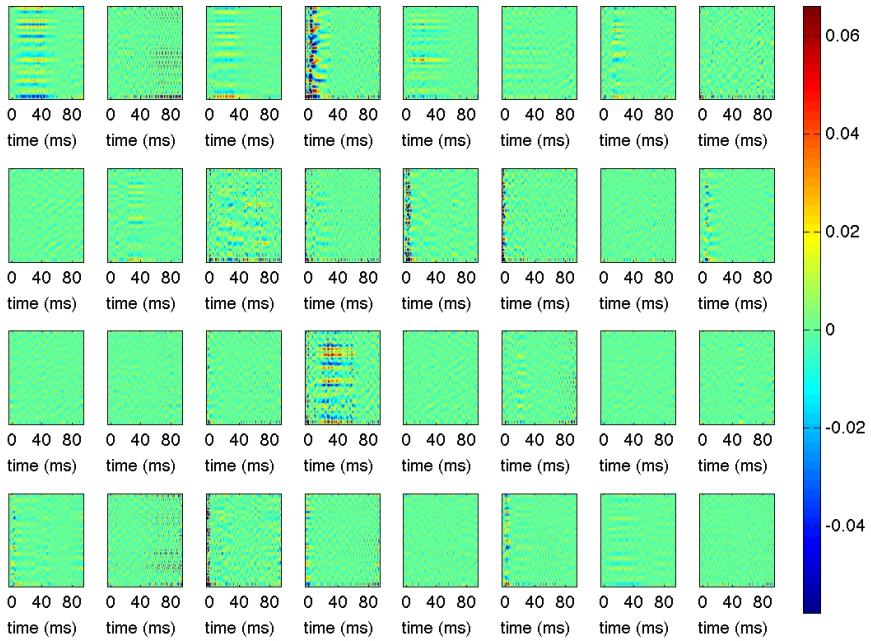


Figure 4.14: CSDs of all LFP components extracted by TRINICON. First row correspond to the first eight components, second row correspond to the next eight components, and so on.

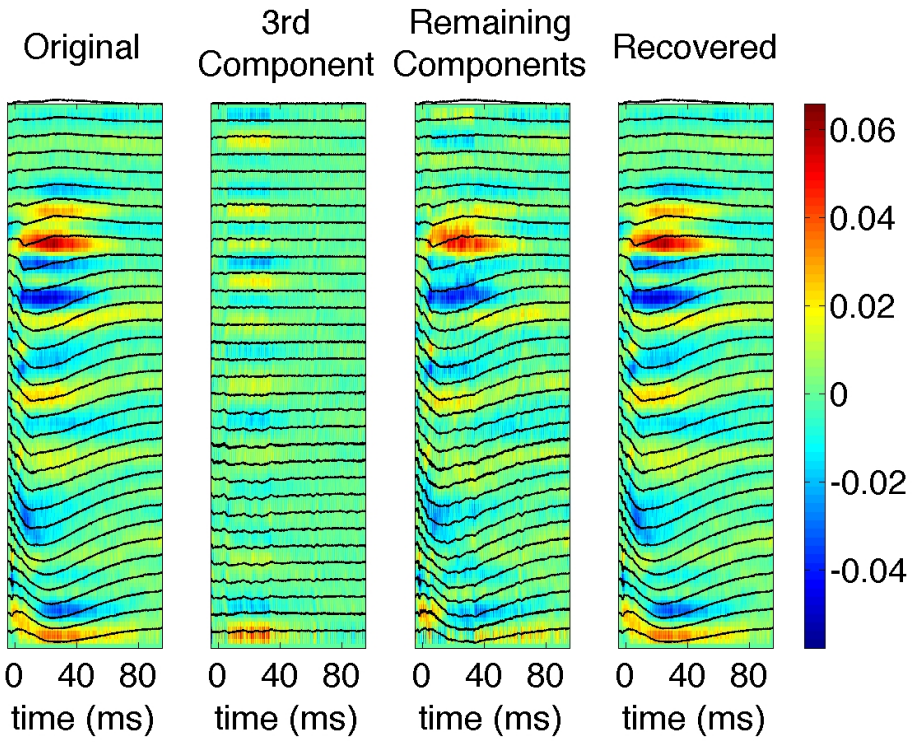


Figure 4.15: LFPs and CSDs corresponding to: the original signal; the third component extracted by TRINICON; the signal recovered using all but the third component and the signal recovered using all components.

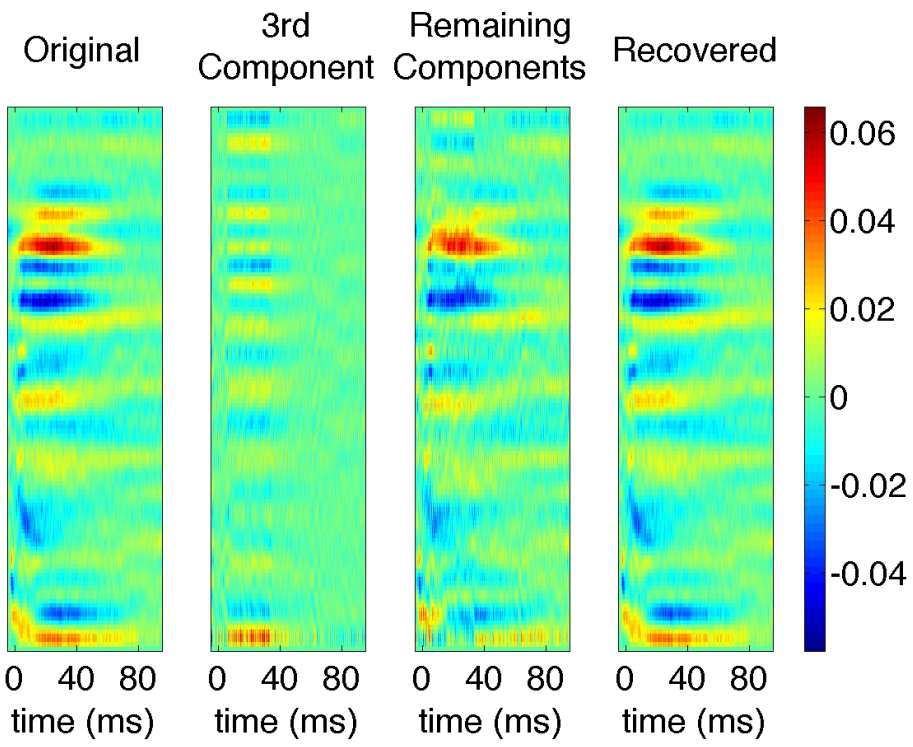


Figure 4.16: CSDs corresponding to: the original signal; the third component extracted by TRINICON; the signal recovered using all but the third component and the signal recovered using all components.

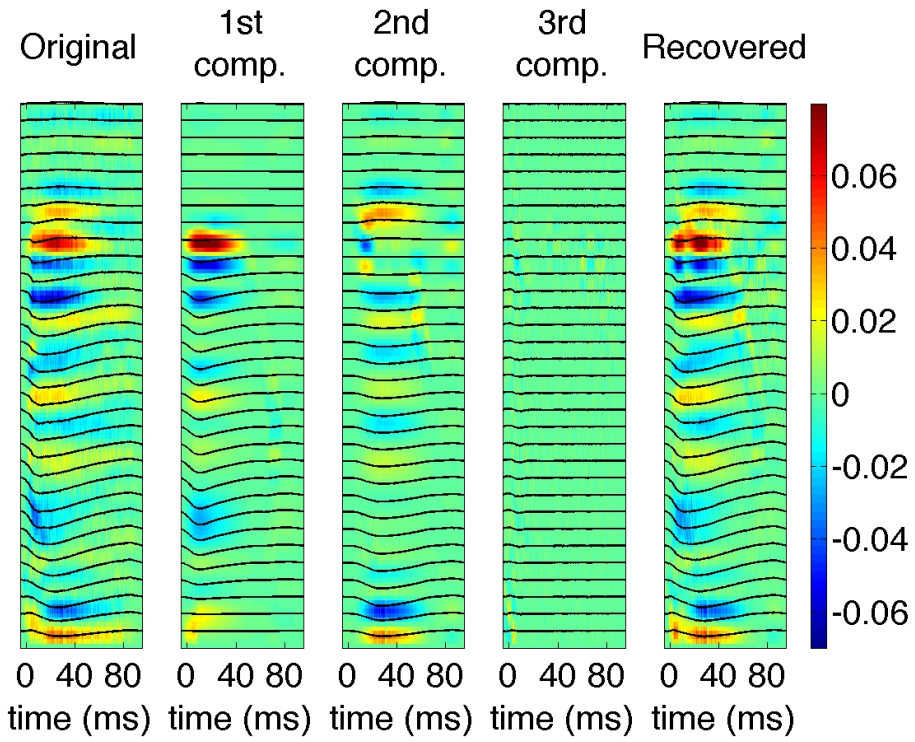


Figure 4.17: LFPs and CSDs corresponding to: the original signal; the first, second and third components extracted by the method proposed; and the signal recovered using these components.

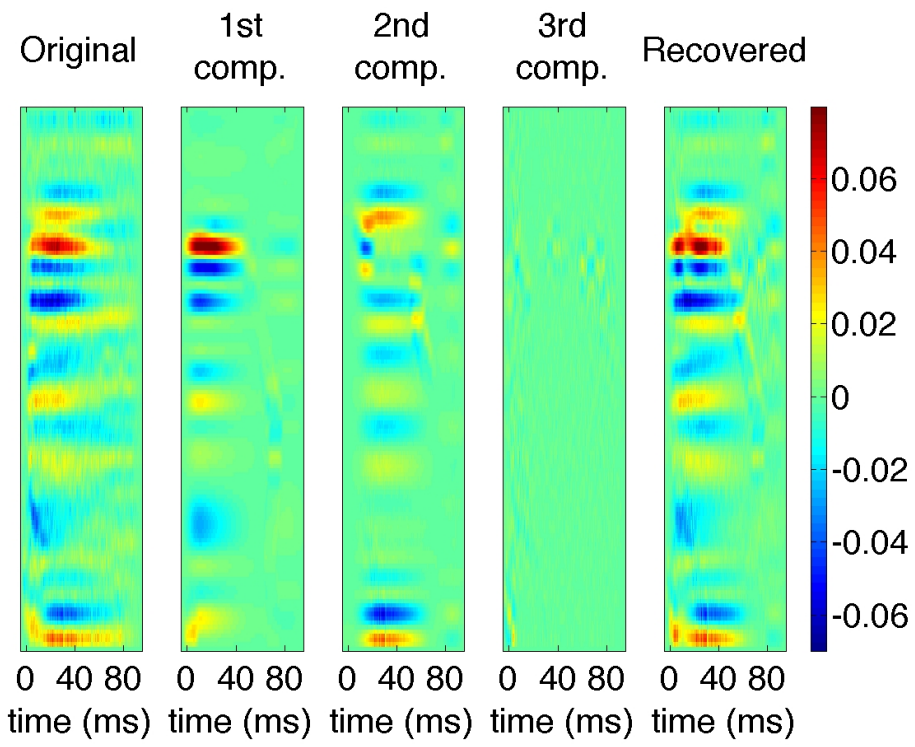


Figure 4.18: CSDs corresponding to: the original signal; the first, second and third components extracted by the method proposed; and the signal recovered using these components.

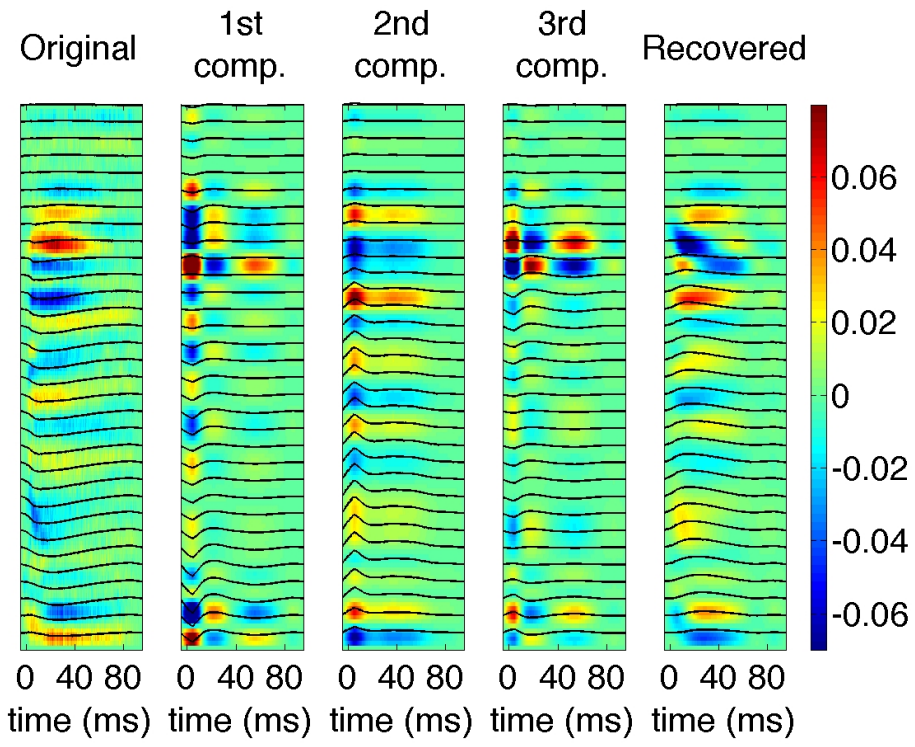


Figure 4.19: LFPs and CSDs corresponding to: the original signal; the first, second and third components extracted by complex valued PARAFAC; and the signal recovered using these components.

Chapter 5

Conclusions

This monograph addresses the problem of the analysis of LFPs, potentials recorded from multi-electrodes in the extracellular space of brain tissue. More specifically it investigates BSS techniques to improve the interpretation of CSD analysis. Up to date, there are few BSS techniques that have been used to the analysis of multichannel LFPs or obtained CSDs, although there is already an acknowledgement by some authors that BSS can assist the interpretation of the recorded signals.

Chapter 2 of this monograph, Section 2.1, presents important concepts related to LFP measurements, most notably the definitions of referential and bipolar montages. Section 2.2 presents the neuronal origin of the LFPs, which are excitatory post synaptic potentials and inhibitory post synaptic potentials. Although the concepts are presented in the context of EEG, there is no loss since the exact same concepts apply to LFPs.

Chapter 3 is dedicated to a review on BSS techniques. Sections 3.1 to 3.3 are dedicated the most classic BSS techniques, PCA and ICA. It is shown how PCA rotates the original basis in which the data was recorded so that the data is uncorrelated in the new basis and how ICA finds an oblique basis from the PCA solution so that the data is actually statistically independent. However, the most frequently used algorithm for ICA is the Bell-Sejnowski algorithm, presented in Section 3.3, which is based on information theory. Section 3.4 is dedicated to a less known BSS techniques, stICA, which is based on the Bell-Sejnowski algorithm and assumes partial statistical independence in time and space. Finally Section 3.5 presents the problem of blind separation of convolutive mixtures, and points TRINICON as a possible solution to this problem.

Chapter 4 is the chapter dedicated to the dataset. Evoked potentials from the somatosensory cortex of a rat are analyzed using several BSS schemes. First, some biological concepts relevant to LFP not provided before are presented. Most notably, the fact that the mammalian cerebral cortex is subdivided into six layers, each with specific cells and connections. The CSD analysis is presented as a method that allows to study connectivity between the cortical layers. However, since it does not explicitly show what current sources and sinks are related and what are not, this motivates the use of BSS techniques to elicit associations between the current sources and sinks. In Section 4.2, it is shown how the data under consideration can be generated according to the instantaneous mixing model and the convolutive mixing model. These are the models assumed by

the techniques described in Section 4.3, which gives details of the algorithms used to decompose the data, including PCA, ICA, stICA and TRINICON. Section 4.3 also describes an adaptation of PARAFAC proposed in this work that allows it to be used in this problem. To use PARAFAC in data of electrodes by time, another dimension is commonly obtained using TFA. Furthermore, only the magnitudes in the time-frequency representations are generally used, which allows for the addition of nonnegativity contraints. However, in this case, time signals must be reconstructed after the decomposition so that the CSDs for each component can be computed. Complex valued PARAFAC is not an option because it does not yield useful components, as attested by their CSDs. The proposed approach consists of using PARAFAC with nonnegativity contraints in the array of magnitudes and then using the phases of the original array to reconstruct the time signals. Section 4.4 presents the results of applying all techniques to the dataset. It is verified a convergence between the results of the method proposed, TRINICON and stICA, which according to previous works yields physiologically meaningful results for evoked potential data.

Much more can be done in the field of BSS applied to LFP data. PARAFAC is only one tensor technique among many others. Also, improvements can be made in the method proposed to yield an even better decomposition.

REFERENCES

[1] EINEVOLL, G. T. et al. Laminar population analysis: Estimating firing rates and evoked synaptic activity from multielectrode recordings in rat barrel cortex. *Journal of Neurophysiology*, v. 97, n. 3, p. 2174–2190, 2007.

[2] BULLOCK, T. H. Problems in the comparative study of brain waves. *Yale J Biol Med*, v. 17, n. 5, p. 657–680.3, 1945.

[3] GALAMBOS, R. Cochlear potentials from the bat. *Science*, v. 93, p. 215, 1941.

[4] MARSHALL, W. H.; WOOLSEY, C. N.; BARD, P. Cortical representation of tactile sensibility as indicated by cortical potentials. *Science*, v. 85, p. 388–390, 1937.

[5] PITTS, W. H. Investigations on synaptic transmission. In: *Cybernetics, Trans. 9th Conf. Josiah Macy Foundation H. von Foerster*. New York: [s.n.], 1952. p. 159–166.

[6] NICHOLSON, C.; FREEMAN, J. A. Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *Journal of Neurophysiology*, v. 38, n. 2, p. 356–368, 1975.

[7] NICHOLSON, C.; FREEMAN, J. A. Experimental optimization of current source-density technique for anuran cerebellum. *Journal of Neurophysiology*, v. 38, n. 2, p. 369–382, 1975.

[8] BELLISTRI, E.; AGUILAR, J.; BROTONS-MAS G. FOFFANI, L. M. d. J. R. Basic properties of somatosensory-evoked responses in the dorsal hippocampus of the rat. *J Physiol.*, v. 591, n. 10, p. 2667–2686, 2013.

[9] BARTH, D. S.; DI, S.; BAUMGARTNER, C. Laminar cortical interactions during epileptic spikes studied with principal component analysis and physiological modeling. *Brain Research*, v. 484, n. 1-2, p. 13–35, 1989.

[10] BARTH, D. S.; DI, S. Laminar excitability cycles in neocortex. *Journal of Neurophysiology*, v. 65, n. 4, p. 891–898, 1991.

[11] BARTH, D. S.; BAUMGARTNER, C.; DI, S. Laminar interactions in rat motor cortex during cyclical excitability changes of the penicillin focus. *Brain Research*, v. 508, p. 105–117, 1990.

[12] DI, S.; BAUMGARTNER, C.; BARTH, D. S. Laminar analysis of extracellular field potentials in rat vibrissa/barrel cortex. *Journal of Neurophysiology*, v. 63, n. 4, p. 832–840, 1990.

[13] WEIS, M. *Multi-Dimensional Signal Decomposition Techniques for Analysis of EEG Data.* Tese (Doutorado) — Ilmenau University of Technology, 2015.

[14] MAKAROV, V. A.; MAKAROVA, J.; HERRERAS, O. Disentanglement of local field potential sources by independent component analysis. *Journal of Computational Neuroscience*, v. 29, n. 3, p. 445–457, 2010.

[15] TANSKANEN, J. M. A.; MIKKONEN, J. E.; PENTTONEN, M. Independent component analysis of neural populations from multielectrode field potential measurements. *Journal of Neuroscience Methods*, v. 145, n. 1-2, p. 213–232, 2005.

[16] LESKI, S. et al. Extracting functional components of neural dynamics with independent component analysis and inverse current source density. *Journal of Computational Neuroscience*, v. 29, n. 3, p. 459–473, 2012.

[17] POTWOROWSKI, J. et al. Extracting activity of individual cell populations from multielectrode recordings. *BMC Neuroscience*, v. 12(Suppl 1), p. P374, 2011.

[18] EINEVOLL, G. T. et al. Modeling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, v. 14, n. 11, p. 770–785, 2013.

[19] GLABSKA, H. T. et al. Generalized laminar population analysis (glpa) for interpretation of multielectrode data from cortex. *Frontiers in Neuroinformatics*, v. 10, p. 1–15, 2016.

[20] MOCKS, J. Decomposing event-related potentials: A new topographic components model. *Biological Psychology*, v. 26, n. 1-3, p. 199–215, June 1988.

[21] LAHAT, D.; ADALI, T.; JUTTEN, C. Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, v. 103, n. 9, p. 1449–1477, August 2015.

[22] BRO, R. *Multiway Analysis in the Food Industry.* Tese (Doutorado) — Universiteit van Amsterdam, 1998.

[23] DYRHOLM, M.; MAKEIG, S.; HANSEN, L. K. Model structure selection in convolutive mixtures. In: ROSCA, J. et al. (Ed.). *Proceedings of the 6th International Conference, ICA 2006.* [S.l.: s.n.], 2006. p. 74–81.

[24] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity. In: *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC).* Kyoto, Japan: [s.n.], 2003. p. 275–278.

[25] SINCLAIR, C. M.; GASPER, M. C.; BLUM., A. S. Basic electronics in clinical neurophysiology. In: BLUM, A. S.; RUTKOVE, S. B. (Ed.). *The Clinical Neurophysiology Primer.* [S.l.]: Humana Press, 2007. p. 3–18.

[26] KLEM, G. H. et al. The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement*, v. 52, p. 3–6, 1999.

[27] TCT. *10/20 System Positioning Manual*. Available online at http://www.trans-cranial.com/local/manuals/10_20_pos_mas_v1_0_pdf.pdf, 2012.

[28] SHARBROUGH, F. W. et al. American electroencephalographic society guidelines for standard electrode position nomenclature. *Clinical Neurophysiology*, v. 8, n. 2, p. 200–202, April 1991.

[29] OOSTENVELD, R.; PRAAMSTRA, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, v. 112, n. 4, p. 713–719, 2001.

[30] COHEN, M. X. *Analyzing Neural Time Series Data: Theory and Practice*. [S.l.]: MIT Press, 2014.

[31] WINTER, B. B.; WEBSTER, J. G. Driven-right-leg circuit design. *IEEE Transactions on Biomedical Engineering*, v. 30, n. 1, p. 62–66, January 1983.

[32] NUNEZ, P. L.; SRINIVASAN, R. *Electric Fields of the Brain: The Neurophysics of EEG*. [S.l.]: Oxford University Press, 2006.

[33] SCHOMER, D. L.; Lopes da Silva, F. H. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. [S.l.]: Lippincott Williams & Wilkins, 2011.

[34] PIVIK, R. T. et al. Committee report: Guidelines for the recording and quantitative analysis of electroencephalo- graphic activity in research context. *Psychophysiology*, v. 30, n. 6, p. 547–558, November 1993.

[35] FISCH, B. J. *Epilepsy and Intensive Care Monitoring: Principles and Practice*. [S.l.]: Demos Medical Publishing, 2009.

[36] STöHR, M.; KRAUS, R. *Introdução à Neurofisiologia Clínica*. [S.l.]: Livraria Santos Editora, 2009.

[37] ALVAREZ, V.; ROSSETTI, A. O. Electroencephalography and evoked potentials: Technical background. In: ROSSETTI, A. O.; LAUREYS, S. (Ed.). *Clinical Neurophysiology in Disorders of Consciousness: Brain Function Monitoring in the ICU and Beyond*. [S.l.]: Springer, 2015.

[38] ACNS. Guideline 6: a proposal for standard montages to be used in clinical EEG. *Journal of Clinical Neurophysiology*, v. 23, n. 2, p. 111–117, April 2006.

[39] LAGERLAND, T. D. Manipulating the magic of digital EEG: Montage reformatting and filtering. *Am. J. END Technol.*, v. 40, p. 121–136, 2000.

[40] WóJCIK, D. K. Current source density (csd) analysis. In: JAEGER, D.; JUNG, R. (Ed.). *Encyclopedia of Computational Neuroscience*. [S.l.]: Springer New York, 2014.

[41] FIELD, A. S.; GRAUPE, D. Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topography*, v. 3, n. 4, p. 407–423, January 1991.

[42] Lopes da Silva, F. EEG: Origin and measurement. In: MULERT, C.; LEMIEUX, L. (Ed.). *EEG - fMRI: Physiological Basis, Technique, and Applications*. [S.l.]: Springer Science and Business Media, 2009. p. 19–38.

[43] SHAROTT, A. Local field potential, methods of recording. In: JAEGER, D.; JUNG, R. (Ed.). *Encyclopedia of Computational Neuroscience*. [S.l.]: Springer New York, 2014. p. 1–3.

[44] LENT, R. *Cem Bilhões de Neurônios*. second. [S.l.]: Atheneu, 2010.

[45] LYNCH, J. *The Cerebral Cortex*. Available online at http://clinicalgate.com/the-cerebral-cortex-2/, 2015.

[46] SCHERG, M. Fundamentals of dipole source potential analysis. In: GRANDORI, F.; HOKE, M.; ROMANI, G. (Ed.). *Auditory evoked magnetic fields and electric potentials*. [S.l.]: S. Karger. Basel, 1990. p. 40–69.

[47] NIEUWENHUYS, R.; DONKELAAR, H. J.; NICHOLSON, C. *The Central Nervous System of Vertebrates: With Posters*. [S.l.]: Springer Science and Business Media, 1998.

[48] Lorente de Nó, R. Action potential of the motoneurons of the hypoglossus nucleus. *Journal of Cellular and Comparative Physiology*, v. 29, n. 3, p. 207–287, June 1947.

[49] BEAR, M. F.; CONNORS, B. W.; PARADISO, M. A. *Neuroscience: Exploring the Brain*. [S.l.]: Lippincott Williams and Wilkins, 2007.

[50] BERGE, J. M. F. T. *Least Squares Optimization in Multivariate Analysis*. [S.l.]: DSWO Press, 1993.

[51] SHLENS, J. *A Tutorial on Principle Component Analysis*. Available online at http://arxiv.org/pdf/1404.1100.pdf, 2014.

[52] STRANG, G. *Linear Algebra and Its Applications*. [S.l.]: Thomsom Brooks/Cole, 2006.

[53] ABDI, H.; WILLIAMS, L. J. Principle component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 2, n. 4, p. 433–459, June 2010.

[54] Scikits-Learn Development Community. *Documentation for the scikits-learn version 0.7.1: FastICA on 2D point clouds*. Available online at http://scikit-learn.sourceforge.net/0.7/auto_examples/plot_ica_vs_pca.html, 2015.

[55] LATHAUWER, L. D.; MOOR, B. D.; VANDEWALLE, J. An introduction to independent component analysis. *Journal of Chemometrics*, v. 14, n. 3, p. 123–149, June 2000.

[56] JOLLIFFE, I. T. *Proncipal Component Analysis*. [S.l.]: Springer, 2002.

[57] BRILLINGER, D. Moments, cumulants and some applications to stationary random process. In: *Proceedings of the Conference on Moments and Signal Processing*. [S.l.: s.n.], 1992. p. 108–126.

[58] FONOLLOSA, J. A. R. Sample cumulants of stationary processes: asymptotic results. *IEEE Transactions on Signal Processing*, v. 43, n. 4, p. 967 – 977, April 1995.

[59] HYVARINEN, A.; OJA, E. Independent component analysis: Algorithms and applications. *Neural Networks*, v. 13, n. 4-5, p. 411–430, 2000.

[60] BELL, A. J.; SEJNOWSKI, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, v. 7, p. 1129–1159, 1995.

[61] CARDOSO, J. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, v. 4, n. 4, p. 112–114, 1997.

[62] COMON, P. Independent component analysis, a new concept? *Signal Processing*, v. 36, n. 3, p. 287–314, 1994.

[63] HYVARINEN, A.; KARHUNEN, J.; OJA, E. *Independent Component Analysis*. [S.l.]: Wiley-Interscience, 2001.

[64] STONE, J. V. et al. Spatial, temporal, and spatiotemporal independent component analysis of fMRI data. In: AYKROYD, R.; MARDIA, K.; DRYDEN, I. (Ed.). *Proceedings of the 18th Leeds Statistical Research Workshop on Spatial-Temporal Modelling and its Applications*. [S.l.: s.n.], 1999. p. 23–28.

[65] STONE, J. V. et al. Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *Neuroimage*, v. 15, p. 407–421, 2002.

[66] PRANDONI, P.; VETTERLI, M. *Signal Processing for Communications*. [S.l.]: CRC Press, 2008.

[67] PALMA, W. *Time Series Analysis*. [S.l.]: John Wiley & Sons, 2016.

[68] GARDNER, W. A. *Introduction to Random Processes, With Applications to Signals & Systems*. [S.l.]: Mcgraw-Hill, 1990.

[69] ROJO, C. et al. Laminar differences in dendritic structure of pyramidal neurons in the juvenile rat somatosensory cortex. *Cereb Cortex.*, v. 26, n. 6, p. 2811–2822, 2016.

[70] CEREBRAL hemisphere. Available online at http://clinicalgate.com/cerebral-hemisphere/, 2015.

[71] SCHJETNAN, A. G. P.; LUCZAK, A. Recording large-scale neuronal ensembles with silicon probes in the anesthetized rat. *J Vis Exp.*, n. 56, p. 3282, 2011.

[72] WATAKABE, A. *What is neocortex?* Available online at http://www.nibb.ac.jp/brish/Gallery/cortexE.html, 2016?

[73] SZYMANSKI, F. D.; GARCIA-LAZARO, J. A.; SCHNUPP, J. W. Current source density profiles of stimulus-specific adaptation in rat auditory cortex. *J Neurophysiol.*, v. 102, n. 3, p. 1483–1490, 2009.

[74] WELKER, W.; JOHNSON, J. I.; NOE, A. *Comparative Mammalian Brain Collections*. Available online at http://www.brainmuseum.org/, 2016?

[75] WOJCIK, D. K. Current source density (csd) analysis. In: JAEGER, D.; JUNG, R. (Ed.). *Encyclopedia of Computational Neuroscience*. New York: Springer, 2014.

[76] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. A generalization of a class of blind source separation algorithms for convolutive mixtures. In: *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*. Nara, Japan: [s.n.], 2003. p. 945–950.

[77] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. Trinicon: a versatile framework for multichannel blind signal processing. In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2004*. Montreal, Canada: [s.n.], 2004. p. 889–892.

[78] CICHOCKI, A. et al. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. [S.l.]: Wiley, 2009.

[79] MIRANDA, R. K. *Metodos Para Melhoria da Qualidade de Separacao Cega de Fontes Sonoras em Angulos Obliquos de Radiacao*. Dissertação (Mestrado) — University of Brasilia, Brazil, 2013.

[80] BESSAI, H. *MIMO Signals and Systems*. [S.l.]: Springer Science & Business Media, 2005.

[81] BRO, R. PARAFAC tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, v. 38, n. 2, p. 149–171, 1997.

[82] MIWAKEICHI, F. et al. Decomposing EEG data into space-time-frequency components using parallel factor analysis. *Neuroimage*, v. 22, n. 3, p. 1035–1045, 2004.

[83] STONE, J. V.; STONE, J.; PORILL, J. *Regularisation Using Spatiotemporal Independence and Predictability*. 1999.

[84] DELORME, A.; MAKEIG, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, v. 134, p. 9–21, 2004.

[85] ANDERSSON, C. A.; BRO, R. The n-way toolbox for matlab. *Chemometrics and Intelligent Laboratory Systems.*, v. 52, n. 1, p. 1–4, 2000.

[86] BRO, R.; SIDIROPOULOS, N.; GIANNAKIS, G. A fast least squares algorithm for separating trilinear mixtures. In: *Proc. ICA99 - Int. Workshop on Independent Component Analysis and Blind Separation*. Aussois, France: [s.n.], 1999. p. 11–15.

[87] FIELD, A. S.; GRAUPE, D. Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topography*, v. 3, n. 4, p. 407–423, 1991.

[88] ALLEN, J. B.; RABINER, L. R. A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, v. 65, n. 11, p. 1558–1564, 1977.