TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

D I P L O M A R B E I T

EEG Microstate Analysis with Respect to the Severity of Alzheimer's Disease

Ausgeführt am Institut für
Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von
Ao.Univ.Prof. Wolfgang Scherrer

durch

Lisa-Marie Mohr
F.Woltrongasse 21
2630 Ternitz

Wien, 22. September 2014 _____

Kurzfassung

Während der letzten Jahrzehnte ist die Anzahl von Betroffenen der Alzheimer-Krankheit, einer der häufigsten Gründe für Aktivitätseinschränkungen im fortgeschrittenen Alter, rasch angestiegen. Immer öfter wird die Elektroenzephalographie (EEG) verwendet, um die Änderungen in der Informationsverarbeitung des Gehirns im Laufe der fortschreitenden Alzheimer-Erkrankung zu untersuchen und Ergänzungen zur klinischen Diagnostik zu liefern.

Im Rahmen dieser Arbeit wird das EEG-Signal als Sequenz von elektrischen Potentiallandschaften mit gleichbleibender Topographie aufgefasst. Diese sogenannten Mikrozustände werden bezüglich ihrer Dauer, ihrer Topographie, ihrem Anteil und ihrem Auftreten analysiert. Diese Auswertung wird für 96 Alzheimer-Patienten der PRODEM-AUSTRIA Datenbank durchgeführt. Dabei wird der Schweregrad der Alzheimer-Krankheit mithilfe des Mini-Mental-State-Examination (MMSE) Score gemessen.

Innerhalb dieser Arbeit werden zwei Vorgehensweisen für die Bestimmung der EEG-Mikrozustände vorgestellt. Die erste Herangehensweise bestimmt die Zentren der positiven und negativen Potentialgebiete, um verschiedene Mikrozustände zu unterscheiden und dadurch die EEG-Sequenz zu segmentieren. Die andere verwendet einen modifizierten K-Means Algorithmus, um alle vorkommenden Mikrozustände in eine zuvor bestimmte Anzahl an Klassen zu gruppieren.

Die statistische Auswertung besteht ebenfalls aus zwei voneinander unanbhängigen Teilen. Im ersten werden die Daten in Bezug auf den MMSE Score mithilfe der Methode der kleinsten Quadrate analysiert. Im anderen werden die Daten für 79 Patienten bezüglich des MMSE Scores ausgewertet, indem für die Methode der kleinsten Quadrate auch die demographischen Variablen Geschlecht, Alter, Dauer der Alzheimer-Erkrankung und Ausbildungsgrad miteinbezogen werden. Für beide statistischen Vorgehensweisen kann sowohl bei der Segmentierung als auch der Gruppierung keine signifikante Änderung der Charakteristiken der EEG-Mikrozustände im Laufe der fortschreitenden Alzheimer-Erkrankung festgestellt werden.

Für die zukünftige Forschung ist die Auswirkung von Longitudinalstudien interessant, da die Grundeigenschaften von EEG-Mikrozuständen von Alzheimer-Patienten sehr stark variieren. Daher könnte das Vergleichen von vergangenen und aktuellen EEG-Aufzeichnungen in Bezug auf die Verschlechterung der Symptome der Alzheimer-Erkrankung und die stärkere Einschränkung eines Patienten neue Erkenntnisse liefern.

Abstract

The last several decades have witnessed a rapid increase of people suffering from Alzheimer's disease (AD), one of the most popular causes of disability in late-life. More and more frequently, electroencephalography (EEG) is used to investigate the changes in the brain's information processing in the course of AD in order to find supplements for the clinical diagnostics.

In this thesis, the EEG signal is considered as a sequence of electric potential landscapes with stable topography. These so-called microstates are analyzed with respect to their duration, topography, ratio and occurrences. The analysis is performed for 96 AD patients of the PRODEM-AUSTRIA database. Thereby, the severity of AD is measured by using the Mini-Mental-State-Examination (MMSE) score.

Within this thesis, two approaches for the determination of EEG microstates are presented. The first strategy determines the centroids of the positive and negative potential areas to distinguish different microstates and thereby segment the EEG sequence. The other approach uses a modified K-means algorithm to cluster all occurring microstates into a predetermined number of classes.

The statistical evaluation also consists of two independent parts. In the former, the data are analyzed versus MMSE score by using a least squares regression. In the other, the data of 79 patients are evaluated versus MMSE score by using a least squares regression including demographic variables as sex, age, duration of AD and degree of education. For both statistical approaches and the segmentation as well as the clustering procedure, the results could not document a significant change in the characteristics of EEG microstates in the course of AD.

For future research, the impact of longitudinal studies is interesting since the basic characteristics of EEG microstates of AD patients vary a lot. The comparing of past and more actual EEG recordings in relation to the worsening of the AD symptoms and the more severe impairment of one patient could therefore be revealing.

Danksagung

Hiermit möchte ich mich bei all jenen bedanken, die entweder mit fachlichem Wissen oder persönlicher Unterstützung das Zustandekommen dieser Diplomarbeit ermöglicht haben.

Mein Dank gilt Herrn Heinrich Garn, der mir im Rahmen eines Forschungsprojektes des Austrian Institute of Technology die Möglichkeit gegeben hat, diese Diplomarbeit zu verfassen und mich während dieser Zeit fachlich unterstützt hat. Ebenso bedanke ich mich bei meinem Betreuer an der Technischen Universität Wien, Herrn Wolfgang Scherrer, der mir bei der Umsetzung und der Verbesserung meiner Vorgehensweise behilflich gewesen ist. Zudem gilt ein besonderer Dank meinem Arbeitskollegen Markus Waser, der immer versucht hat, all meine Fragen zu beantworten und mir sowohl bei der mathematischen Analyse und der Programmierung als auch der sprachlichen und formativen Umsetzung meiner Diplomarbeit eine große Hilfe gewesen ist.

Der größte Dank geht an meine Familie, insbesondere an meine Eltern, die mich während meiner gesamten Studienzeit unterstützt hat und ohne die die Fertigstellung dieser Arbeit und der Abschluss meines Studiums nicht so schnell vonstattengegangen wäre. Dies gilt auch für meine Studienkollegen, welche über die Zeit zu wahren Freunden geworden sind und meine Studienzeit unvergesslich gemacht haben. Ein letzter Dank geht an meine Freunde, die mich immer unterstützt haben und denen ich so manchen lustigen Abend verdanke, der das Durchhalten vereinfacht hat.

Lisa-Marie Mohr

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter is written to provide a basic knowledge of the topic for the reader. It includes information and facts about Alzheimer's disease, electroencephalography and the use of latter in the diagnostics of Alzheimer's disease. Previous studies reporting EEG alterations of Alzheimer's patients are presented. The research question of this work and a thesis outline conclude this chapter.

1.1 Alzheimer's Disease

Dementia (taken from Latin, meaning "without mind") is a disorder due to disease of the brain which is characterized by a progressive impairment of cognitive abilities. It affects predominantly elderly people and is one of the most common causes of disability in late-life. There are different forms of dementia, namely vascular dementia, frontotemporal dementia, dementia with Lewy bodies and *Alzheimer's disease* (AD) which is the most common [1]. The disease is named after the German psychiatrist and neuropathologist Alois Alzheimer (1864-1915) who published first findings in 1906 [2].

The neuropathology of AD is associated with a loss of cells in the brain and the formation of cortical amyloid plaques and neurofibrillary tangles. The brain changes over at least 20 to 30 years while the concretions spread up to the cortical regions [3].

AD symptoms include the loss of cognitive abilities like memory, learning, orientation, comprehension, judgment and, in a late stage, also the loss of motor functions [2].

In 2010, there were about 35.6 million people with all forms of dementia in

the world. It is supposed that the number of affected people is going to increase to 65.7 million by 2030 and 115.4 million by 2050 [1]. In Austria, the estimated number of patients with dementia in 2010 was $112\ 600$. Among them, $74\ 300$ were suffering from AD. New research and statistical expansions suggest that the number increases more rapidly than first expected [4].

At the moment, there exist neither treatment nor drugs which cure AD. For improving the patient's life, there is symptomatic treatment for particular core symptoms but not for stopping the progress of AD. So the principal goal is an early diagnosis and detecting symptoms [3].

Up to this moment, the only way to attest AD is a post-mortem analysis of the patient's brain. In one's lifetime, the disease is classified as *possible* or *probable* AD according to the NINCDS-ADRDA Alzheimer's Criteria [5]. This classification includes a clinical and neuropsychological examination which analyzes the neurological, internistic and psychiatric status and includes neuropsychological tests, a complete blood count, and cerebral magnetic resonance imaging (MRI). There are additional experimental examinations like genotyping, serology, liquor analysis, positron emission tomography (PET), functional MRI (fMRI) and the electroencephalography (EEG) which can be performed [4].

1.2 Electroencephalography and Alzheimer's Disease

The words *encephalo* and *graphy* themselves are taken from Greek meaning "in-head", which refers to the brain, and "write", which denotes the act of recording [6]. An EEG signal is a measurement of currents flowing during synaptic excitations in the central cortex. These currents generate an electric field over the scalp which can be measured by EEG systems, usually with scalp electrodes [7]. The human EEG signal was first measured by Hans Berger (1873-1941) who experimented by placing electrodes on his daughter's scalp in 1929 [6]. The advantages of the EEG in clinical diagnosis of AD are that it is non-invasive, a low-cost examination and it is easier accessible than other imaging procedures.

Since each human scalp has a different form, there is a conventional electrode setting, the so-called *international 10-20 system* [8]. The numbers 10 and 20 refer to the fact that the distances between electrodes are either 10 or 20% of a measured distance - nasion to inion - of the patients head. The electrodes are named in a way that enables an easy assignment to a specific lobe and position. F stands for frontal, C for central, P for parietal, T for temporal and O for occipital. Even numbers are associated with the right hemisphere, odd with the left and z with the middle line [6], [8]. Figure 1.1 explains this electrode placement on the patient's scalp.

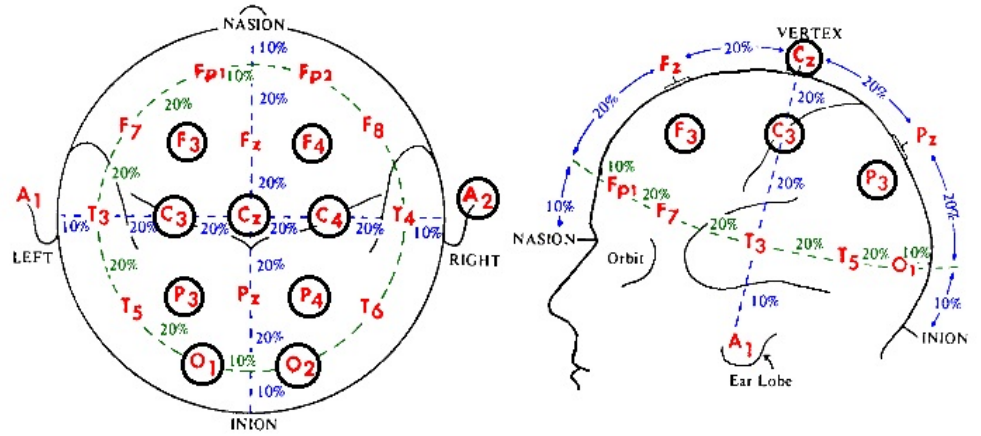

Figure 1.1: Electrode setting given by the international 10-20 system. This image was taken from http://gerstner.felk.cvut.cz/biolab/bionika2004/cepek/bci.html, accessed 05 November 2013.

There have already been several studies about changes in the EEG of AD patients. Up to the present, three major alterations have been reported. First, there is the EEG slowing which means an increase of power in low and a decrease of power in high frequency range. By applying different measures, studies have also shown that the complexity within the EEG is reduced when suffering of AD. A last finding is that perturbations of EEG synchrony occur when comparing healthy controls with AD patients [9]. There are only few studies concerned with EEG microstate alterations in AD patients. All of them following the method of [10] revealed a shortening of the microstate duration for more cognitive impairment [11], [12], [13]. Two of these studies reported an anteriorisation of the centroid locations, i.e. of the brain electric fields [11], [13]. The usage of the clustering algorithm presented in [14] to investigate EEG microstate changes of AD patients is, to the best of the author's knowledge, a novelty in this field.

1.3 Research Question

The data for this work were provided by the Austrian Alzheimer Society in form of the multi-centric study PRODEM-AUSTRIA [15]. Within a subproject *"Advanced" EEG in der Vorhersage des Verlaufs der Alzheimerdemenz* (project no. 827462), EEG samples of patients suffering from AD were recorded. The database includes recordings of 96 subjects with probable AD.

This work's purpose is to examine whether changes in the parameters of EEG microstates, e.g. duration or topography, relate with the severity of AD measured

by the Mini-Mental State Examination score [16]. Two different approaches of determining microstates within an EEG epoch will be introduced. The first is based on the segmentation of an EEG epoch into microstates and the second on the clustering into a finite number of microstate classes which represent different states of the human brain. Particular attention is paid to the changes in duration of EEG microstates in the course of AD.

The main hypothesis of this thesis are the following:

- H_0^1: The EEG microstate duration shortens in the course of AD.
- H_0^2: The topography of the microstates' electric potential landscapes changes in the course of AD.
- H_0^3: The characteristics, e.g. occurrences, duration or ratio, of specific microstate classes change in the course of AD.

1.4 Thesis Outline

The thesis is structured as followed: Chapter 2 provides a basic knowledge about the sample data and the preprocessing. Two methods used to determine the EEG microstates are explained. First, the segmentation of epochs into microstates and afterwards the clustering of epochs into microstate classes. Chapter 3 presents the results and the analysis by applying the methods discussed in Chapter 2. The last part, Chapter 4 and Chapter 5, discusses the findings, completes the thesis by comparing the results to the scientific status quo in the research field of EEG microstates and provides some critical thoughts.

Chapter 2

Material and Methods

The purpose of this chapter is to describe the sample data used in this work, the EEG preprocessing and the main mathematical tools for the signal processing. Later on, the segmentation and clustering of EEG recordings into microstates is introduced and the main characteristics of such short brain states are explained.

2.1 Sample Data

The data used in this work were provided by the Austrian Alzheimer Society in form of the multi-centric study PRODEM-AUSTRIA [15]. The EEG samples were recorded at the Medical Universities of Graz, Innsbruck, Vienna, and the General Hospital Linz. The participants of this study have to fulfill the following criteria [17]:

- diagnosis of AD according to NINCDS-ADRDA Alzheimer's Criteria [5]
- family caregiving
- older than 40 years
- consent of patient and caregiver.

The database includes EEG recordings of 96 subjects (57 female, 39 male) with probable AD. The average age of the subjects is 73.39 years ($\sigma = 8.75$). Additionally, the *Mini-Mental State Examination score* (MMSE score) is provided for each participating subject [16]. This score quantifies the severity of AD on the basis of cognitive impairment. The test is divided into two parts, consisting of 21 and nine questions respectively, each worth one point. The first part covers orientation, memory and attention and the participants have to give oral response.

The second part tests the ability to name, follow verbal and written commands, write a sentence spontaneously and copy a complex polygon. The results of the test range from 0 to 30 points where lower scores are associated with more severe impairment [16]. The subjects in this study have MMSE scores between 15 and 26 with an average of 22.16 ($\sigma = 3.14$). These empirical distributions are shown in the histograms of Figures 2.1 and 2.2.

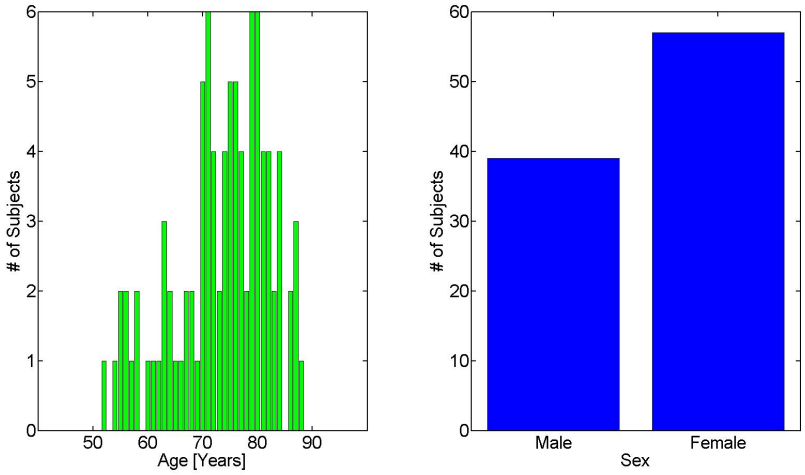

Figure 2.1: Empirical distributions of age (left) and sex (right) of the study's subjects.

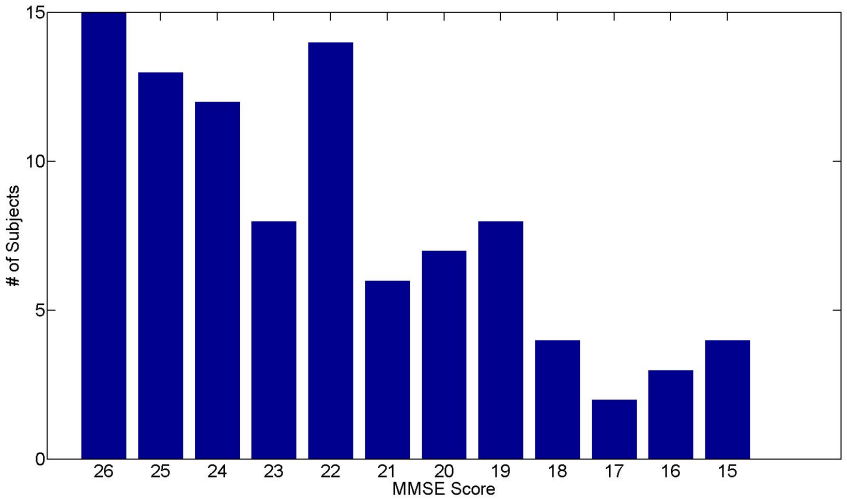

Figure 2.2: Empirical distribution of the MMSE score of the study's subjects. Lower MMSE score is associated with more cognitive impairment.

For 79 patients the following information is provided by the database and is later on included in the analysis: additionally to age and sex, the duration of AD and the degree of education are stated. The average age of these 79 subjects is 73.57 years ($\sigma = 9.22$), ranging from 52 to 88 years. The duration of AD ranges from 2 to 120 months with a mean of 25.54 ($\sigma = 22.08$). The degree of education is scaled from 1 to 6 where 6 indicates the highest education level. Using this scale, the mean degree of education is 2.37 ($\sigma = 1.58$). The levels are characterized as follows: 1 stands for Volks/Hauptschule (primary school), 2 for Lehre (apprenticeship), 3 for AHS (grammar school), 4 for BHS (higher vocational school), 5 Lehrerbildungsanstalt (teacher training school) and 6 for Hochschule (tertiary institution). The empirical distributions of age, sex, duration of AD and degree of education of the 79 patients are given in Figure 2.3. The selected 79 subjects have MMSE scores between 15 and 26 with an average of 22 ($\sigma = 3.12$). This is presented in Figure 2.4.

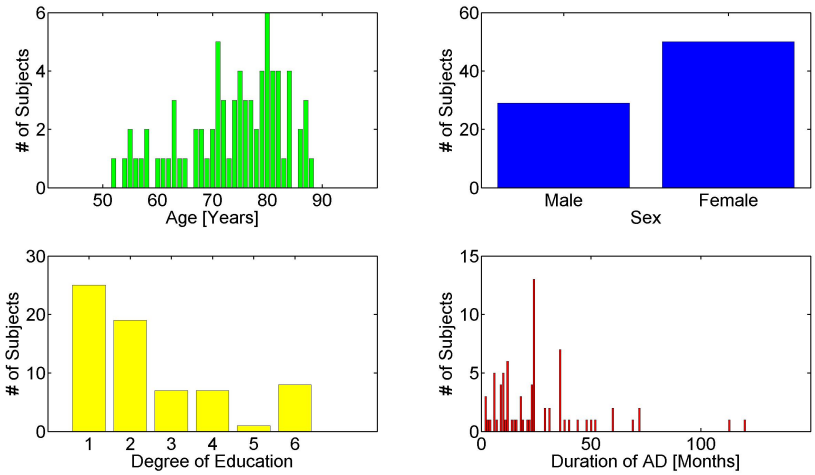

Figure 2.3: Empirical distributions of age (top left), sex (top right), degree of education (bottom left) and duration of AD (bottom right) of the 79 subjects with more information given.

The PRODEM-AUSTRIA study uses the *Neurospeed* software of the *alpha-trace digitalEEG System* with sampling rate of 256 Hz to digitalize the samples [18]. For the EEG recordings, 19 gold cup electrodes are positioned on the scalp according to the international 10-20 system, cf. Figure 2.5.

Additionally to the EEG, horizontal and vertical electrooculogram (EOG) channels and an electrocardiogram (ECG) channel are recorded. The EOG measures the eye's position by placing a pair of electrodes either above and below (vertical) or to the left and right of the eye (horizontal). The recording starts with a resting state sequence of 180 seconds in which the subjects are sitting in a resting but awake condition with eyes closed. After that they open their eyes and are asked to

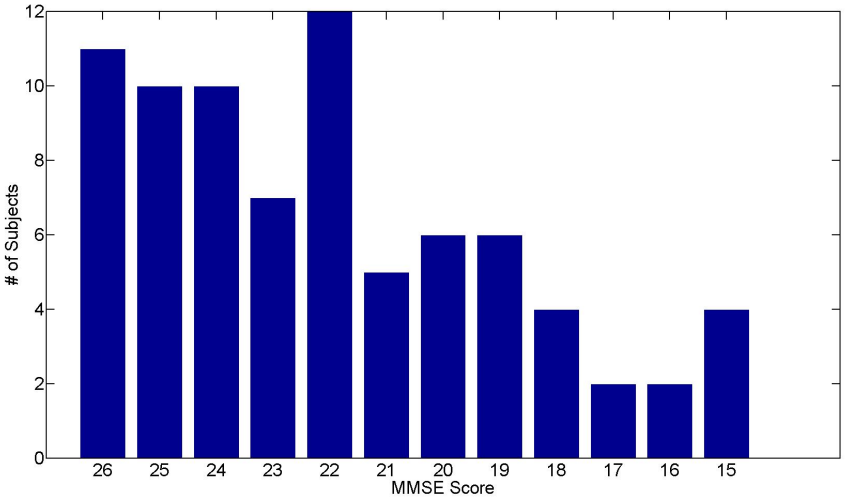

Figure 2.4: Empirical distribution of the MMSE score of the 79 subjects with more information given. Lower MMSE score is associated with more cognitive impairment.

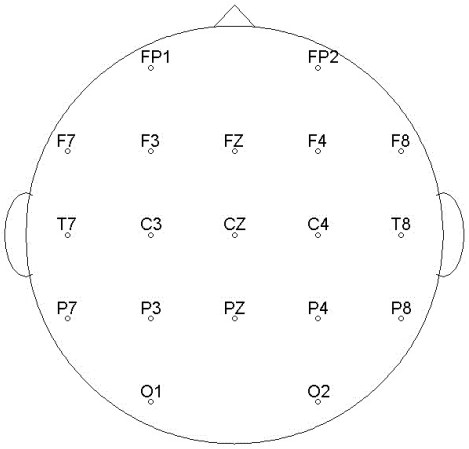

Figure 2.5: Positions of the 19 electrodes in this study according to the international 10-20 system. The image was taken from [17].

perform the following tasks [17]:

- remember faces and corresponding names (50 seconds)
- recall names while only faces are shown
- memorize faces and names again (50 seconds) .

2.2 Data Preprocessing and Artifact Removal

Due to physiological and technical reasons, it is very common that EEG recordings are altered by so-called *artifacts*. Physiological interferences include eye movement, muscular activity, sweating and cardiac electric fields [17]. Figures 2.6 and 2.7 show both eye and cardiac artifacts in EEG recordings. For a better analysis, it is important to determine and remove these artifacts to obtain the true neuronal signals. This was done as presented in [17] and summarized in this section. All further computations of this work have been implemented in MATLAB® 7.10.0 (R2010a) [19].

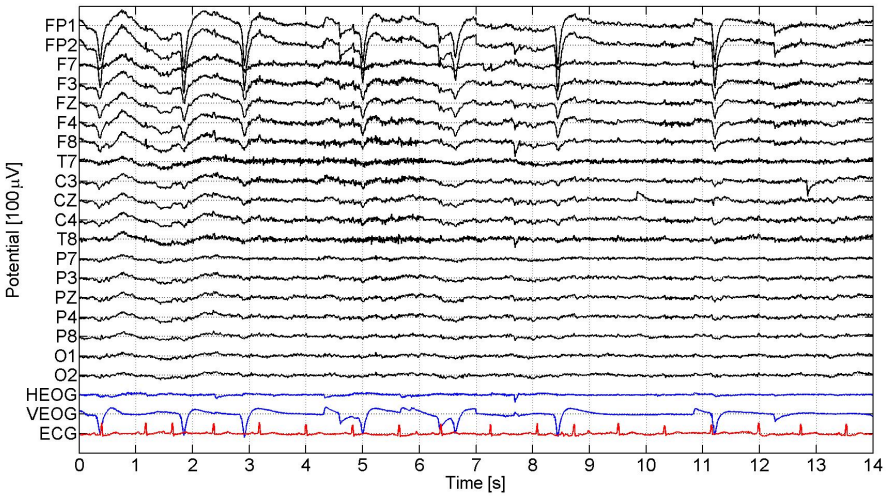

Figure 2.6: A 14 seconds long EEG recording with vertical and horizontal EOG and ECG signals corrupted by eye movement. Like in the majority of cases, the frontal electrodes (indicated with F) are more affected.

In order to achieve the original signals, the EEG, ECG and EOG signals were first high-pass filtered to remove slow variations. This was done by using a finite impulse response (FIR) filter with order 340 and border frequency of 2 Hz.

Subsequently, the EEG signals were examined in order to determine a possible alteration due to cardiac fields. A visual inspection and the so-called energy interval histogram (EIH) method were applied to remove these artifacts [20], [21]. This method uses the fact that cardiac artifacts are spike shaped and nearly periodically.

The also very common eye artifacts were corrected by applying a static linear regression of each EEG signal on the EOG signals. Prior to that, the EOG signals were low-pass filtered using a FIR filter with order 340 and border frequency 12 Hz. Since the EOG also measures electric fields from neuronal activities, this filtering

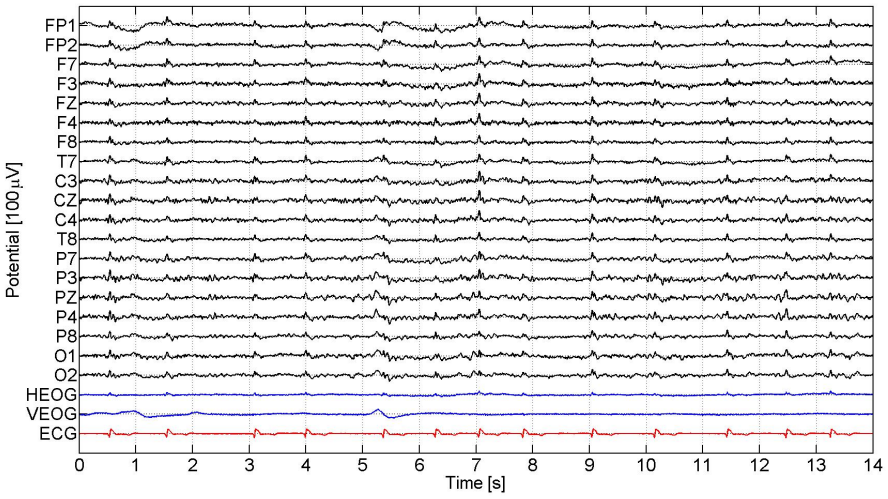

Figure 2.7: A 14 seconds long EEG recording with vertical and horizontal EOG and ECG signals corrupted by cardiac activity.

was used to remove these high-frequency interferences.

The last step in this preprocessing was to low-pass filter (same filter properties as above) the whole EEG recording at 15 Hz to remove high-frequency components caused, for example, by muscle tension.

Figure 2.8 and Figure 2.9 demonstrate these preprocessing steps. The first figure shows a corrupted signal which was then corrected and finally appears as in Figure 2.9.

The further computations and investigations are carried out on a artifact-free EEG epoch of the resting state which has a duration of five seconds.

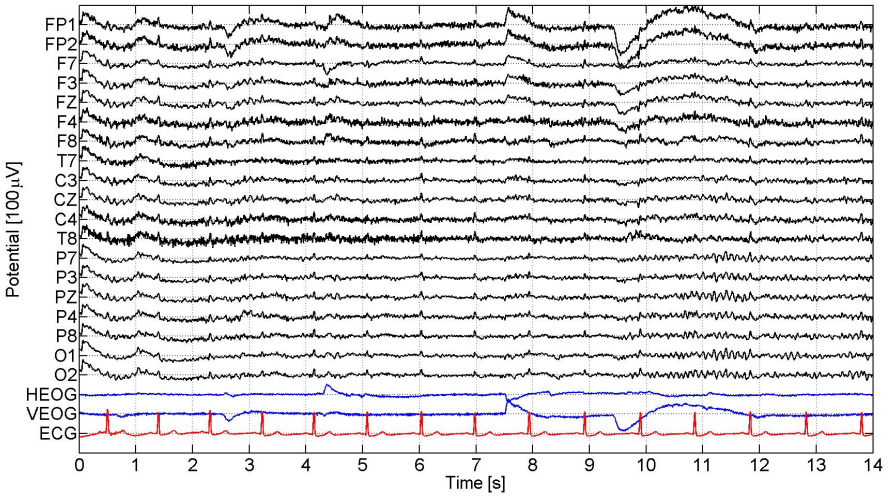

Figure 2.8: A 14 seconds long EEG recording with vertical and horizontal EOG and ECG signals corrupted by cardiac activity and eye movement.

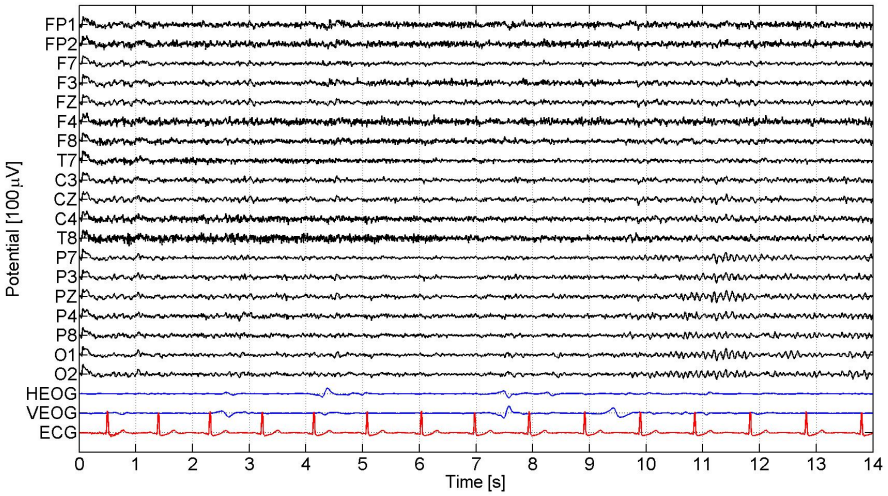

Figure 2.9: A 14 seconds long EEG recording with vertical and horizontal EOG and ECG signals after the preprocessing steps. All artifacts have been removed by using the explained approach.

2.3 Foundations

In this section, the mathematical background and some basic definitions used in the course of this work are presented. First, concepts of multivariate statistics are introduced and applied to EEG recordings [22]. This is followed by a short summary concerning eigenvalues, singular values, the Frobenius norm, the Rayleigh quotient and a popular application in this field [22], [23], [24] [25]. The last part provides a short introduction to multiple linear regression and the least squares method [22], [26]. As indicated above, this mathematical introduction is mainly based on [22], [23], [24] [25] and [26].

The measurements of a multichannel EEG indicate the electrical activity of the brain. At each time point, the measurements, also referred to as a *map*, describe a specific spatial potential distribution of the electric field. Therefore, each map can be written as a column vector

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} \tag{2.1}$$

with N components which refer to the number of electrodes used. Hence, it is possible to view the EEG recording as a sequence of maps of momentary spatial distributions of electric potential instead of waveshapes. This can be written as a $(N \times T)$ matrix V where each column represents a time point $t \in \{1, ..., T\}$ and each row $i \in \{1, ..., N\}$ an electrode

$$V = \begin{pmatrix} \mathbf{v}_1, & \mathbf{v}_2, & \ldots, & \mathbf{v}_T \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \ldots & v_{1T} \\ v_{21} & v_{22} & \ldots & v_{2T} \\ \vdots & \vdots & & \vdots \\ v_{N1} & v_{N2} & \ldots & v_{NT} \end{pmatrix}. \tag{2.2}$$

2.3.1 Descriptive Statistics

To analyze maps and for the often needed average-referencing, the *arithmetic mean* is introduced. It can be calculated for each column, thus each time point, of V and will be referred to as

$$\overline{v}_t = \frac{1}{N} \sum_{i=1}^{N} v_{it} \tag{2.3}$$

where t indicates the time point and i the electrode. So for the whole $(N \times T)$ data matrix V the *multivariate arithmetic mean* is the T dimensional vector

$$\overline{\mathbf{v}} = \left(\overline{v}_1, \quad \overline{v}_2, \quad \cdots, \quad \overline{v}_T \right) . \tag{2.4}$$

Another important parameter is the *empirical variance* which is defined for all time points $t \in \{1, ..., T\}$

$$q_{tt} = \frac{1}{N} \sum_{i=1}^{N} (v_{it} - \overline{v}_t)^2 . \tag{2.5}$$

The *empirical covariance*, which is a measure of the linear relation of two time points t and s, is given by

$$q_{ts} = \frac{1}{N} \sum_{i=1}^{N} (v_{it} - \overline{v}_t)(v_{is} - \overline{v}_s) . \tag{2.6}$$

Arranging the variance, along the main diagonal, and the covariance in a matrix, the *covariance matrix* Q is formed. It is symmetric since $q_{ts} = q_{st}$.

When normalizing the covariance, one obtains the so-called *correlation coefficient*

$$r_{ts} = \frac{\sum_{i=1}^{N} (v_{it} - \overline{v}_t)(v_{is} - \overline{v}_s)}{\sqrt{\sum_{i=1}^{N} (v_{it} - \overline{v}_t)^2} \sqrt{\sum_{i=1}^{N} (v_{is} - \overline{v}_s)^2}} \tag{2.7}$$

which ranges from -1 to 1.

2.3.2 Eigenvalues, Eigenvectors and Singular Value Decomposition

Definition 2.3.1 (Eigenvalues and Eigenvectors): *Let A be a $(T \times T)$ square matrix. Then*

$$p(\lambda) = |A - \lambda I| \tag{2.8}$$

is a polynomial of degree T in λ where $|.|$ denotes the determinant and I the identity matrix. The T roots of $p(\lambda)$, namely $\lambda_1, \lambda_2, ..., \lambda_T$, are called eigenvalues *of A. Then*

$$|A - \lambda_t I| = 0 \tag{2.9}$$

holds for all $t \in \{1, ..., T\}$*, so that the matrix* $A - \lambda_t I$ *is singular, which means not invertible. Hence, there is a vector* $\gamma_t \neq 0$ *with*

$$A\gamma_t = \lambda_t \gamma_t \quad \textit{for} \quad t \in \{1, ..., T\}. \tag{2.10}$$

This vector γ_t *is called* eigenvector *of* A *corresponding to* λ_t*.*

Theorem 2.3.1 (Singular Value Decomposition): *For any arbitrary matrix* $A \in \mathbb{R}^{m \times n}$ *there exist orthogonal matrices* $U \in \mathbb{R}^{m \times m}$*,* $O \in \mathbb{R}^{n \times n}$ *and a diagonal matrix* $\Sigma \in \mathbb{R}^{m \times n}$ *with* $\Sigma_{jk} = \sigma_j \delta_{jk}$ *such that*

$$A = U\Sigma O^\top \textit{ and } \sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{\min\{m,n\}} \geq 0 \tag{2.11}$$

where δ *denotes the Kronecker's delta. This factorization is called* singular value decomposition *(SVD) and all* σ_j *are called* singular values*. The matrix* Σ *is unique. The columns* u_i*,*$i \in \{1, ..., m\}$*, of* U *and* o_i*,* $i \in \{1, ..., n\}$*, of* O *are called* left singular vectors *and* right singular vectors *respectively.*

Remark 2.3.1. There are some useful relations between eigenvalues and singular values for a matrix $A \in \mathbb{R}^{m \times n}$:

- The eigenvectors of the matrix $AA^\top$ correspond to the left singular vectors of A.
- The eigenvectors of the matrix $A^\top A$ correspond to the right singular vectors of A.
- The matrices $A^\top A$ and $AA^\top$ have the same non-zero eigenvalues which correspond to the squared non-zero singular values of A.

The singular value decomposition has many applications in applied mathematics. One of it is the now presented matrix approximation which is stated in [25]. First, the Frobenius norm is introduced.

Definition 2.3.2 (Frobenius norm): *The* Frobenius norm *of a matrix* $A \in \mathbb{R}^{m \times n}$ *is defined as*

$$\|A\|_F := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}\,. \tag{2.12}$$

Remark 2.3.2. It is also possible to denote the Frobenius norm by using the SVD of $A = U\Sigma O^\top$ since

$$\begin{aligned} \mathsf{tr}(AA^\top) &= \mathsf{tr}((U\Sigma O^\top)(O\Sigma^\top U^\top)) = \mathsf{tr}(U\Sigma\Sigma^\top U^\top) \\ &= \mathsf{tr}(\Sigma\Sigma^\top) = \mathsf{tr}(\Sigma^2) = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2 \end{aligned} \tag{2.13}$$

where $\sigma_1, \sigma_2, \ldots, \sigma_{\min\{m,n\}}$ are the singular values of A. Hence,

$$\|A\|_F := \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{\min\{m,n\}}^2} \ . \tag{2.14}$$

Theorem 2.3.2 (Rank k approximation in the Frobenius norm): *Let $A \in \mathbb{R}^{m\times n}$ be a matrix of rank r. Then for all matrices $B \in \mathbb{R}^{m\times n}$ with rank $p < r$, there holds*

$$\|A - B\|_F \geq \|A - A_p\|_F \tag{2.15}$$

where

$$A_p = U\Sigma_p O^\top \tag{2.16}$$

and Σ_p is obtained from Σ of the singular value decomposition of A by setting all but its p largest singular values σ_i to zero.

Proof. Since $A = U\Sigma O^\top$,

$$\|A - B\|_F^2 = \|U\Sigma O^\top - UU^\top BOO^\top\|_F^2 = \|\Sigma - U^\top BO\|_F^2 \ .$$

Let $U^\top BO = C$. So this can be written as

$$\|\Sigma - C\|_F^2 = \sum_{i=1}^{n}(\sigma_i - c_{ii})^2 + \sum_{i\neq j} c_{ij}^2 \geq \sum_{i=1}^{n}(\sigma_i - c_{ii})^2 \ .$$

Thus, $\|A - B\|_F^2$ is minimized if $c_{ii} = \sigma_i$ for $i = 1, 2, \ldots, p$ and $c_{ij} = 0$ otherwise, i.e. $C = \Sigma_k$. For this approximation the error is

$$\|A - A_k\|_F = \sqrt{\sum_{i=p+1}^{r} \sigma_i^2} \ .$$

□

2.3.3 Rayleigh Quotient

Definition 2.3.3 (Rayleigh quotient): *Let $A \in \mathbb{R}^{m\times m}$ be a symmetric matrix and $x \in \mathbb{R}^m \setminus \{0\}$. Then*

$$r(x) = \frac{x^\top Ax}{x^\top x} \tag{2.17}$$

is called the Rayleigh quotient.

Remark 2.3.3. Note that, if $x \in \mathbb{R}^m$ is an eigenvector to the eigenvalue λ of $A \in \mathbb{R}^{m\times m}$, the following holds

$$r(x) = \frac{x^\top Ax}{x^\top x} = \frac{x^\top \lambda x}{x^\top x} = \lambda \ . \tag{2.18}$$

2.3.4 Multiple Linear Regression

A *linear regression model* with one dependent variable is of the general form

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_q x_q + \epsilon \ . \tag{2.19}$$

The unknown *regression coefficients* $\beta_0, \beta_1, \ldots, \beta_q$ declare the relation, in this case a linear one, between the *regressors* $x_1, x_2, \ldots, x_q$ and the dependent variable y including an error term ϵ. This error term is called *residual* and captures influences of other factors than the regressors. When analyzing n independent observations of y and the regressors $x_1, x_2, \ldots, x_q$, this results in a *multiple linear regression model* of the form

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_q x_{1q} + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_q x_{2q} + \epsilon_2 \\ &\vdots \quad\quad \vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_q x_{nq} + \epsilon_n \ . \end{aligned} \tag{2.20}$$

When denoting this by using matrices, one obtains

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{=:\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix}}_{=:\mathbf{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}}_{=:\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{=:\boldsymbol{\epsilon}} . \tag{2.21}$$

Here, the matrix $\mathbf{X} \in \mathbb{R}^{n \times (q+1)}$ is called *design matrix*. It is assumed that

1. $\mathbb{E}(\boldsymbol{\epsilon}) = 0$ and
2. $\mathbb{C}ov(\boldsymbol{\epsilon}) = \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \sigma^2 \mathbf{I}_n$

hold for the residuals. Here, σ^2 is the variance of the error term, which is the same for all components. Moreover, different components of the error term are uncorrelated.

The aim of regression analysis is to determine the unknown regression coefficients and σ^2. One approach is the *least squares method*. Thereby, the sum of the squared residuals

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_q x_{iq})^2 \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} \end{aligned} \tag{2.22}$$

is minimized. The specific $\boldsymbol{\beta}$ which minimizes $S(\boldsymbol{\beta})$ is the so-called *least squares estimator* $\hat{\boldsymbol{\beta}}$. Since all addends are scalars,

$$\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} = (\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y})^\top = \mathbf{y}^\top \mathbf{X} \boldsymbol{\beta} \tag{2.23}$$

holds and simplifying $\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}$ yields

$$\begin{aligned} \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} &= \mathbf{y}^\top \mathbf{y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} \\ &= \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} \ . \end{aligned} \tag{2.24}$$

When setting the partial derivative,

$$\frac{\partial(\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon})}{\partial \boldsymbol{\beta}} = 0 - 2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} \ , \tag{2.25}$$

to zero, one obtains the least squares estimator $\hat{\boldsymbol{\beta}}$ as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \ . \tag{2.26}$$

Here, it is necessary that $\mathbf{X}$ has full rank, since otherwise it is not possible to determine $(\mathbf{X}^\top \mathbf{X})^{-1}$. The estimated values $\hat{\mathbf{y}}$ and the estimated residuals $\hat{\boldsymbol{\epsilon}}$ can then be calculated by using $\hat{\boldsymbol{\beta}}$ which yields

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{=:\mathbf{H}} \mathbf{y} \tag{2.27}$$

and

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \ . \tag{2.28}$$

Theorem 2.3.3: *For a multiple linear regression model as in (2.21) with* $\mathbf{X} \in \mathbb{R}^{n \times (q+1)}$ *having full rank the following holds:*

- *The least squares estimator* $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ *is unbiased, i.e.* $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ *and* $\mathbb{C}ov(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.
- $\mathbb{E}(\hat{\boldsymbol{\epsilon}}) = 0$ *and* $\mathbb{C}ov(\hat{\boldsymbol{\epsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H})$
- $\hat{\boldsymbol{\epsilon}}$ *and* $\hat{\boldsymbol{\beta}}$ *are uncorrelated.*

Theorem 2.3.4 (Gauss-Markov): *For a multiple linear regression model as in (2.21) with* $\mathbf{X} \in \mathbb{R}^{n \times (q+1)}$ *having full rank, let the components of the residual vector* $\boldsymbol{\epsilon}$ *fulfill* $\mathbb{C}ov(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$*, then the following holds:*

- *The least squares estimator* $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ *is a unique, efficient and linear estimator of* $\boldsymbol{\beta}$ *and*

- $s^2 = \frac{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}{n-q-1} = \frac{1}{n-q-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ *is a unbiased estimator for the residual variance* σ^2 .

An efficient estimator has a covariance matrix which is smaller than the covariance matrix of all other linear, unbiased estimators. Therefore, the least squares estimator is also called *best linear unbiased estimator* (BLUE).

2.4 Global Field Power and Global Map Dissimilarity

The voltage potential field measured by the EEG has two main characteristics: *topography* and *strength*, also called *hilliness*. The topography is related to the location and orientation of the underlying neuronal activity and the strength describes the impact of simultaneously and synchronously working sources [27]. To analyze these parameters, the measures *Global Field Power* (GFP) and *Global Map Dissimilarity* (GMD) are introduced.

First proposed in [28], the GFP is a one-number statement for all electrode measurements respectively one map at each time point t. It is defined as the standard deviation of the measurements at all electrodes

$$GFP(\mathbf{v}_t) = \sqrt{q_{tt}} = \Big[\frac{1}{N}\sum_{i=1}^{N}(v_{it} - \overline{v_t})^2\Big]^{\frac{1}{2}} \tag{2.29}$$

where N is the number of electrodes, v_{it} is the measured potential at electrode i and time point t and $\overline{v_t}$ the arithmetic mean defined as in (2.3). Calculated as in (2.29), the GFP is rising with increasing hilliness of the electric field and turns zero for a flat field [29]. Since in many cases the recordings are average-referenced, meaning that they have zero mean, it can be simplified to

$$GFP(\mathbf{v}_t) = \Big[\frac{1}{N}\sum_{i=1}^{N} v_{it}^2\Big]^{\frac{1}{2}} \tag{2.30}$$

where N is the number of electrodes, v_{it} is the measured potential at electrode i and time point t.

The second measurement is the GMD. Analogously to the GFP, it is a one-number measure which estimates the difference of the spatial configuration of two maps $\mathbf{v}_t$ and $\mathbf{v}_{t+1}$. It is defined as the GFP of the difference map of these two maps $\hat{\mathbf{v}}_t$ and $\hat{\mathbf{v}}_{t+1}$, scaled to unity GFP

$$GMD(\hat{\mathbf{v}}_t, \hat{\mathbf{v}}_{t+1}) = \Big[\frac{1}{N}\sum_{i=1}^{N}[(\hat{v}_{it} - \overline{\hat{v}_t}) - (\hat{v}_{it+1} - \overline{\hat{v}_{t+1}})]^2\Big]^{\frac{1}{2}} \tag{2.31}$$

where $\hat{(\cdot)}$ indicates maps scaled to unity GFP, N the number of electrodes, $\hat{v}_{it}$ is the measured potential at electrode i and time point t and $\overline{\hat{v}_t}$ the arithmetic mean of the map scaled to unity GFP [27], [29]. To determine the range of the GMD, it is described in terms of empirical variances and covariances:

$$\begin{aligned} GMD^2(\hat{\mathbf{v}}_t, \hat{\mathbf{v}}_{t+1}) =& \frac{1}{N}\sum_{i=1}^{N}[(\hat{v}_{it} - \overline{\hat{v}_t}) - (\hat{v}_{it+1} - \overline{\hat{v}_{t+1}})]^2 \\ =& \frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{it} - \overline{\hat{v}_t})^2 - \frac{2}{N}\sum_{i=1}^{N}(\hat{v}_{it} - \overline{\hat{v}_t})(\hat{v}_{it+1} - \overline{\hat{v}_{t+1}}) + \\ & \frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{it+1} - \overline{\hat{v}_{t+1}})^2 \\ =& q_{tt} - 2q_{tt+1} + q_{t+1t+1} \end{aligned} \tag{2.32}$$

With (2.7) and the maps scaled to unity GFP, it follows that

$$GMD^2(\hat{\mathbf{v}}_t, \hat{\mathbf{v}}_{t+1}) = \underbrace{q_{tt}}_{=1} - 2q_{tt+1} + \underbrace{q_{t+1t+1}}_{=1} = 2 - 2\underbrace{q_{tt+1}}_{\in[-1,1]} \leq 4\ . \tag{2.33}$$

Therefore, the GMD ranges from 0 to 2, where 0 describes two topographically homogeneous maps and 2 the topographic inversion [27]. Using this measure on successive pairs of maps, GMD is a function of time indicating stable and unstable periods within an EEG recording [29].

2.5 The Microstate Model

As described in [30], the sequence of maps as in (2.2) can be segmented into so-called *microstates* which are characterized by quasi-stable spatial distributions, called *landscapes* of electric potential. These stable topographic voltage patterns are connected by rapid changes of the electric field. Since different spatial distributions of the potentials must have been generated by different neuronal activity in the brain, it is assumed that they belong to different functions in the brain. When analyzing these short stable configurations, the task is to determine epochs of variable length where the spatial configuration is stationary. Since the activity of neuronal generators is oscillatory and includes repetitive polarity reversals, the polarity of the maps is irrelevant. When comparing maps' landscapes, also the map strength can be disregarded [30]. There are different approaches of dealing with that problem. One is to segment the recording and analyze the sequences of the same spatial distribution. The other possibility is to cluster the given landscapes into a specific number of groups, so-called *microstate classes*.

2.6 Selection of Original Maps

For all further procedures, the data have to be average-referenced to enable comparison between the maps. This is followed by a selection of the maps with the most information, i.e. the best signal-to-noise ratio (SNR) [31]. During times of high GFP, the spatial configuration of maps is typically stable. The changing to other configurations is indicated with high GMD and occurs during times of low GFP. As a result, only data points of maximal GFP are representative for a whole stable epoch and are further analyzed [29]. These selected maps are called *original maps* of an EEG sequence [32]. In order to determine these local maxima of the GFP, both the GFP and GMD were calculated and visually inspected. Figure 2.10 shows both the GFP and GMD for a three second epoch of a patient.

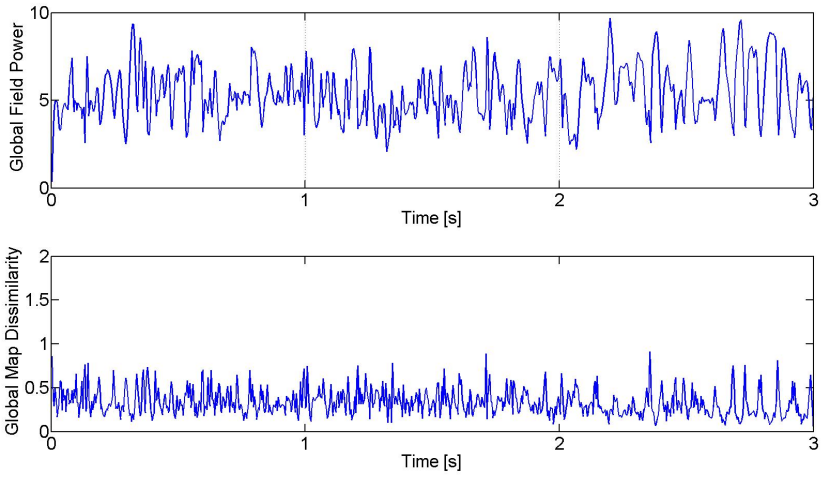

Figure 2.10: Plot of the GFP and GMD as functions of time for a three second EEG epoch of one patient in the study.

Foremost, all local maxima of the GFP function with a minimal peak distance of five time points were computed by using the MATLAB function `findpeaks` [33]. If at one time point the value of GFP is higher than both of its neighbors, the function assigns it as a local peak. Then, each of these maxima x_t^*, $t \in \{1, ..., T\}$ had to fulfill three criteria to be selected as original map:

$$x_t^* \text{ is original map if } \begin{cases} GFP(x_t^*) > z, z \in \mathbb{R} \\ GFP(x_t^*) > GFP(x_t^* + 5) \\ GFP(x_t^*) > GFP(x_t^* - 5) \end{cases} \tag{2.34}$$

where z denotes a specific threshold. It was set to the average GFP of the whole epoch and was found best by visual inspection. The second and third condition

assure that each maximum is compared to its neighborhood and not only the predecessor and the successor. This selection was based on the assumption that microstates stay in a stable condition within a short period of time which legitimates the elimination. The first condition was chosen to avoid low local maxima of GFP which implies high GMD, i.e. high dissimilarities, and the other two because of the possibility of small fluctuations and summits along the GFP function which are no relevant peaks for the selection of the original maps. In Figure 2.11, the selected maxima are shown for the same three second EEG epoch as in Figure 2.10. All further computations were only applied to these selected original maps.

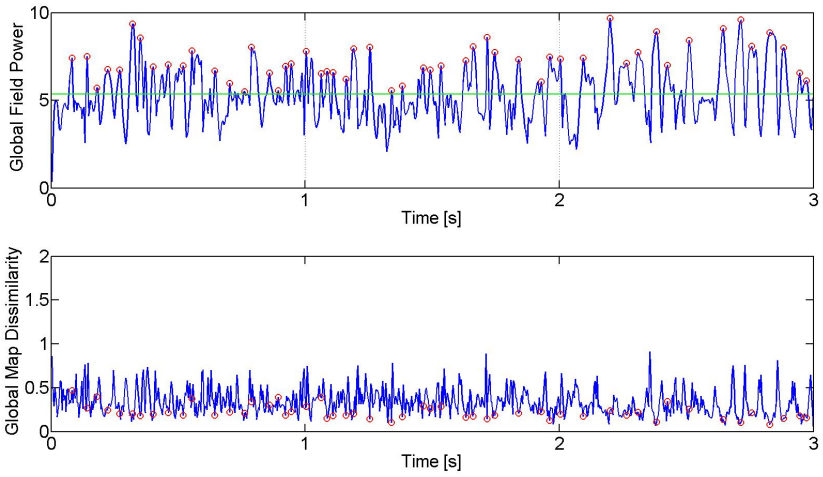

Figure 2.11: Plot of the GFP and GMD as functions of time for a three second EEG epoch of one patient in the study. In both graphs the selected original maps, e.g. local maxima of GFP, are marked by red circles. The green horizontal line indicates the mean of the GFP function. In the majority of cases, high GFP indicates low GMD.

2.7 Segmentation of EEG Recordings into Microstates

In this work, the goal of the segmentation is to create consecutive regions in which all data fulfill a certain homogeneity criterion which rates their similarity. Therefore, the matrix V, as in (2.2), is segmented into specific parts by comparing $\mathbf{v}_t$ and $\mathbf{v}_{t+1}$ for all $t \in \{1, ..., T-1\}$ by using an arbitrary homogeneity criterion H. By setting the first entry of the label vector, $L \in \mathbb{N}^N$, to $L_1 = 1$, the segmentation is started. If the homogeneity criterion applied to $\mathbf{v}_t$ and $\mathbf{v}_{t+1}$ is smaller than a given tolerance size, $L(\mathbf{v}_{t+1}) = L(\mathbf{v}_t)$. If the tolerance size is exceeded, $L(\mathbf{v}_{t+1}) = L(\mathbf{v}_t) + 1$ and a new segment begins. Repeating this procedure for all $t \in \{1, ..., T-1\}$ results in a segmentation marked by L. All maps which have the same entry in the label

vector are then seen as one microstate. This procedure leads to a simplification of the given data set.

2.7.1 Segmentation Procedure

For performing a segmentation of maps within an EEG recording, these maps have to be characterized by some specific features to rate their similarity. For the first approach, the maps selected as in the previous section are assessed by a strategy which results in two descriptors. Each map is characterized by two descriptors which are the locations of the centroids of the positive and negative potential areas of the map [10].

To be able to talk about centroid locations, the electrode setting is fitted into a reference frame using an Euclidean coordinate system [31]. In this system, the abscissa runs from left to right and the ordinate from anterior to posterior. The units used are *electrode distances* (ED). In the original 10-20 system, the actual size of one ED depends on the patient's head size. As a relative measure, the unit ED bears no metric unit. The reference frame in ED is necessary since different head sizes would make the absolute measurements incomparable [31]. Figure 2.12 shows this introduced coordinate system with a typical topographic distribution of potentials. For example, a position of $x = 3$ and $y = 4$ would indicate the electrode Pz, cf. Figure 2.5.

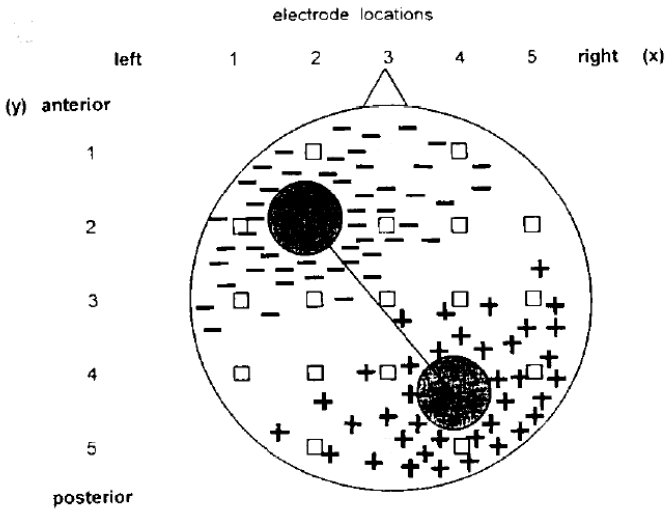

Figure 2.12: This figure shows the introduced coordinate system for indicating the descriptors of each map. The squares correspond to the 19 electrode positions, the dark circles to the centroid locations and the plus and minus signs to the electric potential measured in this area. This image was taken from [31].

Since every electrode is indicated by a point in the newly introduced coordinate system, each entry v_i of a map $\mathbf{v}$ corresponds to a specific point $(\ x_i,\ \ y_i\ )^\top$. Using these coordinates and the indicator function, the descriptors D_1 and D_2 of a map $\mathbf{v}_t$ can be calculated as

$$D_1(\mathbf{v}_t) = \frac{\sum_{i=1}^{N} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \cdot \mathbb{1}_{v_{it}>0}}{\sum_{i=1}^{N} 1 \cdot \mathbb{1}_{v_{it}>0}} \tag{2.35}$$

and

$$D_2(\mathbf{v}_t) = \frac{\sum_{i=1}^{N} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \cdot \mathbb{1}_{v_{it}<0}}{\sum_{i=1}^{N} 1 \cdot \mathbb{1}_{v_{it}<0}}\ . \tag{2.36}$$

For easier notation, $D_1(\mathbf{v}_t)$ and $D_2(\mathbf{v}_t)$ are written as $D_1^{\mathbf{v}_t}$ and $D_2^{\mathbf{v}_t}$ respectively.

By proceeding like this, all maps are reduced to and quantified by their descriptors. The descriptors of successive maps are then compared. An often observed phenomenon is that the positive and negative areas switch polarity back and forth while the centroid locations stay the same. That is a result of the oscillatory activity of the generating neurons and for this reason polarity is disregarded [10], [30], [31]. Therefore, each microstate can include standing waves with repeated polarity changes. A new microstate is only defined if the whole configuration of the potential map respectively the centroid locations change [31].

To include minor fluctuations of the electric field, small movements of the descriptors, within a small circular area, are allowed. These windows are called *topographic windows* of a specific *window size* given in ED [10], [31]. At the first map $\mathbf{v}_1$, two topographic windows are set around $D_1^{\mathbf{v}_1}$ and $D_2^{\mathbf{v}_1}$. Then the distances of $D_1^{\mathbf{v}_1}$ and $D_2^{\mathbf{v}_1}$ to both descriptors $D_1^{\mathbf{v}_2}$ and $D_2^{\mathbf{v}_2}$ of the second map $\mathbf{v}_2$ are calculated. The distance is calculated as the Euclidean distance between the two points. Either descriptor of map $\mathbf{v}_1$ is associated with the closest one of $\mathbf{v}_2$. This is done in order to disregard polarity as claimed before. The next step is to check if the corresponding descriptors of both maps can be accommodated within the allowed topographic window. If it is possible to set this circular window of the formerly fixed window size, given by a radius r (in ED), around the descriptors, the label of map $\mathbf{v}_2$ is the same as of $\mathbf{v}_1$. If not, the segment is terminated which means $L(\mathbf{v}_{t+1}) = L(\mathbf{v}_t) + 1$, and the rejected pair of descriptors serves as the starting points for the next iteration. If the segment goes on, the center of the already associated descriptors of both maps is used as the center of the new topographic window.

To generalize this procedure, it is presumed that the process has already proceeded to map $\mathbf{v}_t$ described by $D_1^{\mathbf{v}_t}$ and $D_2^{\mathbf{v}_t}$. Without loss of generality, one can

assume that there is no polarity reversal for this EEG recording, which means that $D_1^{\mathbf{v}_t}$ is always associated with $D_1^{\mathbf{v}_{t+1}}$ and $D_2^{\mathbf{v}_t}$ with $D_2^{\mathbf{v}_{t+1}}$for all $t \in \{1, ..., T-1\}$. So at time point t, it is already possible to set two topographic windows of radius r around

$$\begin{aligned} &(1)\ \ D_1^{\mathbf{v}_{t-k}},\ D_1^{\mathbf{v}_{t-k+1}}, ...,\ D_1^{\mathbf{v}_t} \quad \text{and} \\ &(2)\ \ D_2^{\mathbf{v}_{t-k}},\ D_2^{\mathbf{v}_{t-k+1}}, ...,\ D_2^{\mathbf{v}_t} \end{aligned}$$

for a certain $k \in \mathbb{N}$. This means that these maps have been assigned to the same label $L(\mathbf{v}_{t-k}) = L(\mathbf{v}_{t-k+1}) = ... = L(\mathbf{v}_t)$. In the next step, the algorithm checks if $D_1^{\mathbf{v}_{t+1}}$ and $D_2^{\mathbf{v}_{t+1}}$ also fit in the topographical window of (1) and (2) respectively. There are two possible cases:

- *Case 1*
  It is possible to fit both descriptors of map $\mathbf{v}_{t+1}$ to the topographical windows given by (1) and (2). Then $L(\mathbf{v}_{t-k}) = ... = L(\mathbf{v}_t) = L(\mathbf{v}_{t+1})$ and the next map is checked in the same way.

- *Case 2*
  It is not possible to fit both descriptors of map $\mathbf{v}_{t+1}$ to the topographical windows given by (1) and (2). Then $L(\mathbf{v}_{t-k})+1 = ... = L(\mathbf{v}_t)+1 = L(\mathbf{v}_{t+1})$ and a new segment starts. The procedure starts again by comparing $\mathbf{v}_{t+1}$ and $\mathbf{v}_{t+2}$.

This segmentation procedure using centroids as descriptors is presented in Figure 2.13.

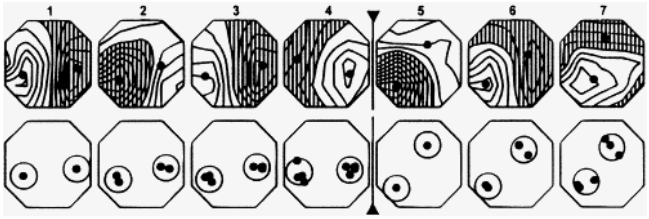

Figure 2.13: The figure illustrates the segmentation of a map series at seven time points of GFP maxima into microstates. The upper row shows maps with isopotential lines (white positive, hatched negative) and the lower row the corresponding segmentation steps. Black dots indicate centroid locations (descriptors). The vertical lines after map four indicate the beginning of a new microstate since the descriptors of map five cannot be fitted into the topographical window of the previous maps. This image was taken from [34].

2.7.2 Determination of the Optimal Window Size

To determine the optimal window size for the topographic windows in the segmentation process, the approach of [10] is used and summarized in this section. This data-driven determination is based on the two goals of each segmentation: the detection of similarity and the detection of dissimilarity between successive maps. There are four different conditions for one map depending on the separation between neighboring maps. The map can be separated from

- (1) both the preceding and the following map
- (2) only the preceding map
- (3) only the following map
- (4) from neither the preceding nor the following map.

If the segmentation process explained in Section 2.7.1 is executed for a wide range of possible window sizes, $r \in [0.1, 1.5]$, it is clear that for each r the maps meet different conditions. For the smallest window size all maps are separated from each other, i.e. all maps meet condition (1) and for the biggest all maps are labeled to the same microstate, i.e. all maps meet condition (4). The first and the last map are omitted since their conditions cannot be determined. Looking at the information each map and its condition contains, it is possible to conclude the following. A separate map (condition (1)) does not provide information about the capacity of the segmentation procedure to identify similarities, and a map which is not separated from the neighboring maps (condition (4)) does not offer information about the capacity to identify differences. But maps in conditions (2) or (3) contain information about both characteristics. So they share both goals of the procedure [10].

Therefore, it is concluded that the window size which maximizes the number of maps in condition (2) or (3), is optimal. Plotting the number of maps against the window size for each patient, called the *window-determining function*, and determining the maximum of maps in condition (2) or (3) leads to the optimal window size r^*. This is presented in Figure 2.14 for maps in condition (2). The corresponding plot for maps in condition (3) would be identical. If there is more than one maximum, r^* is defined as the mean of the maximal values [10].

2.7.3 Data Analysis and Microstate Classification

In a first run, the optimal window sizes for all patients are determined by using the data-driven method explained in Section 2.7.2. So for each patient an *individual*

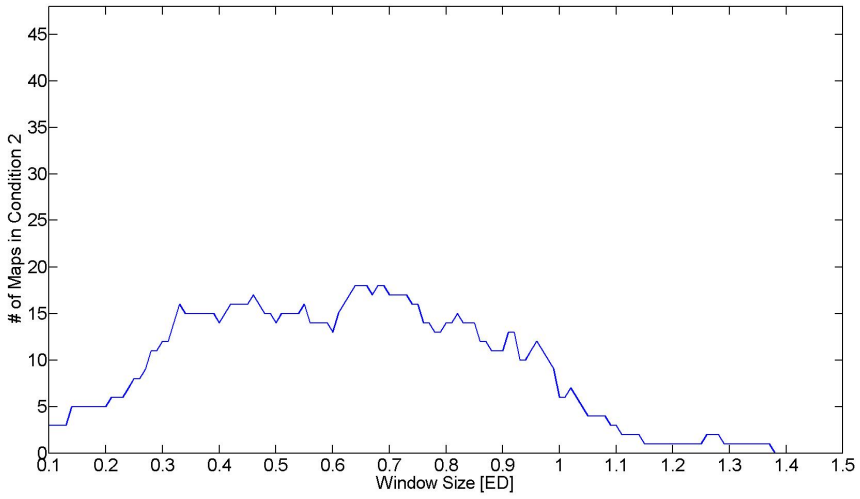

Figure 2.14: This figure shows the so-called window-determining function for one patient. It plots the window size given in ED against the number of maps in condition (2). The maximum of this function is reached for $r^* = 0.664$ and this value is used as the optimal window size of this patient.

window size is calculated [31]. For comparison between all individuals, a standardized window size has to be used since the window size directly affects microstate properties. Therefore, in a second run, the segmentation procedure is repeated with a fixed window size for all patients [31]. This *group window size* is given by the mean of all individual window sizes.

The topographical properties of the microstates are analyzed by the location of the microstate center of gravity [11]. This is the center of the two centroids, D_1 and D_2, which describe one microstate. For microstates which include more than one map the center of all descriptors is used. As a second quantity of a microstate the distance between the descriptors is examined [11].

The microstate duration is not calculated starting and ending at two GFP peaks but starting and ending at the midpoint in time between the last original map of the preceding microstate and the first original map of the following microstate [35]. Microstates consisting of one map are analyzed separately from microstates which consist of more than one map. The former are called *single map segments* or *single peak segments*, the latter *multiple map segments* [11]. This classification is done because of the time-stability which is assumed for the microstates in this model. Additionally the *longest microstate* is analyzed. These microstates are associated with the most prominent microstate in that EEG epoch and are therefore considered to be representative for the relevant mode of the working brain [11]. This results in three microstate classes which are examined separately.

For each of these three classes and each patient certain parameters are determined and analyzed according to the severity of AD [11], [31], [32], [35]:

- *Mean Microstate Duration* (ms): This is the mean microstate duration of all microstates in this class. For the longest microstate it is the duration of this specific microstate.

- *Ratio Covered*: This is the duration which is covered by one class divided by the duration of the epoch.

- *Centroid x-Position*: This is the mean x-position of the center of D_1 and D_2 of this microstate class according to the introduced coordinate system (cf. Figure 2.12).

- *Centroid y-Position*: This is the mean y-position of the center of D_1 and D_2 of this microstate class according to the introduced coordinate system (cf. Figure 2.12).

- *Distance* (ED): This is the distance between D_1 and D_2 of this microstate class.

Independent of microstate classes, the following parameters are analyzed [11], [31], [32], [35]:

- *Single Peak or Single Map Segments per Second*: This is the total number of single map segments of the epoch divided by the total time of the epoch in seconds.

- *GFP Peaks per Second*: This is the total number of original maps (peaks of GFP) of the epoch divided by the total time of the epoch in seconds.

- *Optimal Window Size* (ED): This is the window size calculated for each patient individually.

2.8 Clustering of EEG Recordings into Microstates

Cluster analysis is a particular way of data segmentation where the goal is to group a collection of objects into subsets or *clusters* [22], [36]. All objects within one cluster should be similar, i.e. of high homogeneity, and at the same time the dissimilarity between the clusters should be maximized. An object itself can be described by a set of measurements or relations. So each clustering method attempts to group the objects by a given definition of homogeneity [22], [36]. This measurement between two objects within the data set is often given by the Euclidean distance.

There are four types of classifications within the clustering methods [22]:

- **Overlapping Clustering**: The objects are allowed to belong to more than one cluster.
- **Partitioning Clustering**: This is a type of overlapping clustering, but the clusters are not allowed to overlap, i.e. all clusters are disjunct.
- **Quasihierarchical Clustering**: This classification is based on a sequence of overlapping clustering which leads to a hierarchy.
- **Hierarchical Clustering**: This classification is based on a sequence of partitioning clustering which leads to a hierarchy where each object only belongs to one cluster.

One of the most popular partitioning clustering methods is the *K-means algorithm*. This clustering method is a so-called *combinatorial algorithm* which means that it works directly on the observed data without a reference to an underlying probability model [36]. In this kind of algorithms, each observation is labeled by an integer $k \in \{1, ..., K\}$ with $K < N$ where K is the previously specified number of clusters. Since it is a partitioning method, each object is assigned to one and only one cluster [36].

The first run is started with a initial guess for the clusters, i.e. K objects are selected randomly. The remaining $N - K$ objects are than assigned to the cluster which is the most similar. Then for each cluster the new cluster center is calculated as a mean of the assigned objects. These steps are iterated until a certain error measure can not longer be reduced [36]. This procedure is shown in Figure 2.15.

2.8.1 A Modified K-Means Algorithm for EEG Microstate Analysis

When analyzing EEG microstates, a modified version of the classical K-means algorithm is used. This algorithm, called the *N-microstate algorithm*, and the underlying model was first presented in [14] by R. Pascual-Marqui and later used by a multitude of publications related to this research topic. This section is based on [14] and explains the mathematical method introduced by the authors.

To get an idea of the microstate model, it is easier to consider the case of only two electrode measurements for each time point. Then, it is possible to imagine a 2-D plane defined by the electric potential measurements of the two electrodes. One point on this plane corresponds to one electrode measurement. Using that coordinate system, microstates can be introduced by defining them as the coordinate vectors of a point located at unit distance from the origin. If a point is lying

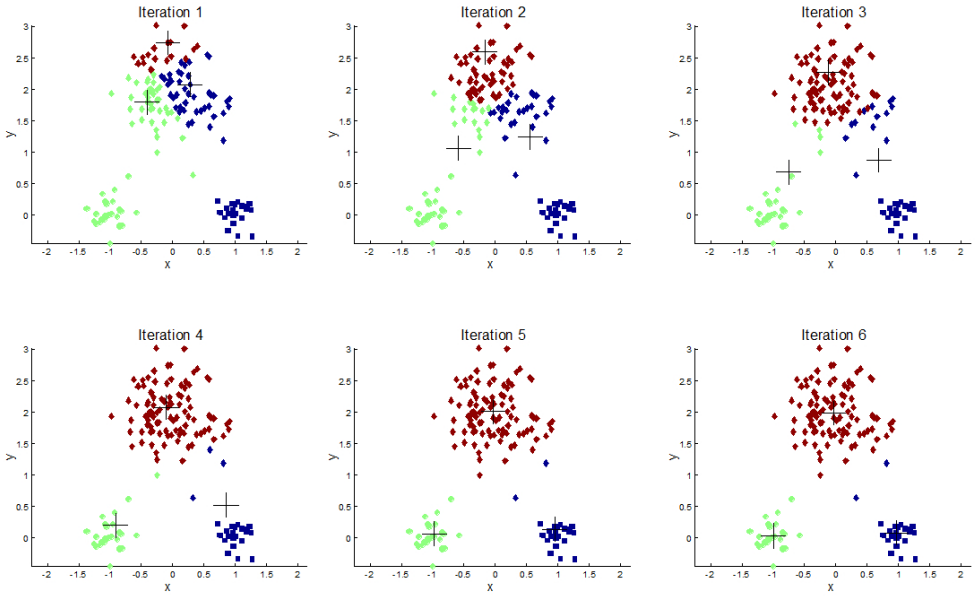

Figure 2.15: This figure shows a K-means algorithm which leads to a successful clustering within six iterations. The different clusters are marked with three different colors: red, green and blue. The cluster centers are tagged with crosses. This image was taken from http://apandre.wordpress.com/visible-data/cluster-analysis, accessed 21 November 2013.

on this line defined by a microstate, it belongs to that microstate. Hence, as long as successive measurements remain on that line, the brain remains in the same microstate. The distance of one point to the origin is then directly related to the intensity of the neuronal generators corresponding to that microstate.

Formulating this in a mathematical way and for all existing electrodes, this model results in

$$\mathbf{v}_t = \sum_{k=1}^{M} a_{kt}\Gamma_k \tag{2.37}$$

where M is the number of microstates, $\mathbf{v}_t$ is a map at a specific time point t, $t \in \{1, ..., T\}$, $\Gamma_k \in \mathbb{R}^N$ a normalized $(N \times 1)$ vector representing the k-th microstate with N being the number of electrodes and a_{kt} the intensity of the k-th microstate at time point t.

Since the microstates are always assumed to be non-overlapping at each time point t, there is only one a_{kt} unequal to zero. So at each time point the sum in (2.37) is reduced to a single nonzero term which means that only one microstate

is active. Expressing that by using constraints, the following must hold

$$\begin{array}{lll} (1) & a_{k_1 t} \cdot a_{k_2 t} = 0, & \text{for all } k_1 \neq k_2, \quad \text{for all } t \quad \text{and} \\ (2) & \sum_{k=1}^{M} a_{kt}^2 \geq 0, & \text{for all } t. \end{array} \tag{2.38}$$

In this model, Pascual-Marqui assumed that the measurements given by (2.37) are contaminated additively with zero mean random noise which is independent and identically distributed for all time points. Hence,

$$\mathbf{v}_t = \sum_{k=1}^{M} a_{kt} \Gamma_k + E_t \tag{2.39}$$

where all variables are defined as in (2.37) and E_t is a $(N \times 1)$ random vector. For the average-referenced data, it is further assumed that E_t has a covariance matrix given by $\sigma^2 H$ with $H = I - \iota \cdot \iota^\top / N$ where I denotes the identity matrix and $\iota = \begin{pmatrix} 1, & 1, & \ldots, & 1 \end{pmatrix}^\top$.

To reproduce that, it is necessary to go back to the original, not average referenced data, denoted by $\widehat{(.)}$. Each map is than given by

$$\widehat{\mathbf{v}}_t = \sum_{k=1}^{M} a_{kt} \widehat{\Gamma}_k + \widehat{E}_t \ . \tag{2.40}$$

Using the so-called linear average reference transformation matrix H, the average referencing can by written as

$$\mathbf{v}_t = \widehat{\mathbf{v}}_t - \iota \cdot \frac{1}{N} \cdot \iota^\top \cdot \widehat{\mathbf{v}}_t = \underbrace{(I - \iota \cdot \iota^\top \cdot \frac{1}{N})}_{=H} \cdot \widehat{\mathbf{v}}_t \ . \tag{2.41}$$

This results in

$$\mathbf{v}_t = H \cdot \widehat{\mathbf{v}}_t = \sum_{k=1}^{M} a_{kt} \underbrace{H \cdot \widehat{\Gamma}_k}_{=\Gamma_k} + \underbrace{H \cdot \widehat{E}_t}_{=E_t} \tag{2.42}$$

when applying H to (2.40). Looking at H, one can see that it is an orthogonal projection matrix with the properties:

- H is idempotent: $H^2 = H$
- H is self-adjoint : $H = H^\top$.

Using these properties. it is possible to calculate the covariance and the expected value of E_t. Since $\mathbb{E}(\widehat{E}_t) = 0$ by definition and $\mathbb{C}ov(\widehat{E}_t) = \sigma^2 I$,

$$\mathbb{E}(E_t) = H \cdot \mathbb{E}(\widehat{E}_t) = 0 \tag{2.43}$$

and

$$\begin{aligned}\mathbb{C}ov(E_t) &= \mathbb{C}ov(H \cdot \widehat{E}_t) = H \cdot \mathbb{C}ov(\widehat{E}_t) \cdot H^\top \\ &= \sigma^2 \cdot H \cdot H^\top = \sigma^2 \cdot H^2 = \sigma^2 H \ .\end{aligned} \tag{2.44}$$

Hence, E_t has a covariance matrix given by $\sigma^2 H$.

Given the number of microstates M, the model parameters can be estimated by minimizing the functional

$$F = \frac{1}{T(N-1)} \sum_{t=1}^{T} \left\| \mathbf{v}_t - \sum_{k=1}^{M} a_{kt} \Gamma_k \right\|^2 \tag{2.45}$$

with respect to all Γ_k and a_{kt} under the constraints given in (2.38) where all variables are defined as in (2.37).

The presented algorithm to determine these parameters and to minimize F consists of two basic steps which are alternated. For the first step, it is considered that the linearly independent and normalized microstates Γ_k are given for all $k = \{1, ..., M\}$. Then the orthogonal squared distance between each map $\mathbf{v}_t$ and each microstate Γ_k is computed

$$d_{kt}^2 = \mathbf{v}_t^\top \cdot \mathbf{v}_t - (\mathbf{v}_t^\top \cdot \Gamma_k)^2 \ . \tag{2.46}$$

This formula results from minimizing the distance between each measurement $\mathbf{v}_t$ and the possible microstates Γ_k, $k = \{1, ..., M\}$, with respect to a_{kt}

$$\|\mathbf{v}_t - a_{kt}\Gamma_k\|^2 \to 0 \ . \tag{2.47}$$

Simplifying this expression and using the fact that all microstates are normalized leads to

$$\begin{aligned}\|\mathbf{v}_t - a_{kt}\Gamma_k\|^2 &= (\mathbf{v}_t - a_{kt}\Gamma_k)^\top (\mathbf{v}_t - a_{kt}\Gamma_k) \\ &= \mathbf{v}_t^\top \mathbf{v}_t - \mathbf{v}_t^\top a_{kt}\Gamma_k - \Gamma_k^\top a_{kt}^\top \mathbf{v}_t + \Gamma_k^\top a_{kt}^\top a_{kt}\Gamma_k \\ &= \mathbf{v}_t^\top \mathbf{v}_t - a_{kt}(\mathbf{v}_t^\top \Gamma_k + \Gamma_k^\top \mathbf{v}_t) + a_{kt}^2 \Gamma_k^\top \Gamma_k \\ &= \mathbf{v}_t^\top \mathbf{v}_t - 2a_{kt}\mathbf{v}_t^\top \Gamma_k + a_{kt}^2 \ .\end{aligned} \tag{2.48}$$

When setting the derivative to zero, the minimum is given at $a_{kt} = \mathbf{v}_t^\top \Gamma_k = \Gamma_k^\top \mathbf{v}_t \in \mathbb{R}$. The minimum value of the norm in (2.47) is then obtained by

$$\begin{aligned}\min_{a_{kt}} \|\mathbf{v}_t - a_{kt}\Gamma_k\|^2 &= \mathbf{v}_t^\top \mathbf{v}_t - 2\mathbf{v}_t^\top \Gamma_k \mathbf{v}_t^\top \Gamma_k + (\mathbf{v}_t^\top \Gamma_k)^2 \\ &= \mathbf{v}_t^\top \mathbf{v}_t - (\mathbf{v}_t^\top \Gamma_k)^2 \ .\end{aligned} \tag{2.49}$$

Now, each map is associated with the microstate to which it is closest. Hence, the label is

$$L_t = \arg\min_k d_{kt}^2 \ . \tag{2.50}$$

To estimate the intensity given by $a_{\kappa t}$, where $\kappa = L_t$, it is calculated as follows

$$a_{\kappa t} = \mathbf{v}_t^\top \cdot \Gamma_\kappa \ . \tag{2.51}$$

The second step assumes the labels L_t to be given and the minimum of functional F with respect to Γ_k is obtained by minimizing the norm

$$\|V_k - \Gamma_k(a_{k1}, a_{k2}, \ldots, a_{km})\|_F^2 \to 0 \ . \tag{2.52}$$

Here, V_k is the matrix only including time points with $L_t = k$ and $m \in \{1, ..., T\}$ marks the number of these time points. Using Theorem 2.3.2, we can approximate V_k in the Frobenius norm by a rank $p = 1$ matrix

$$\min_{\Gamma_k} \|V_k - \Gamma_k(a_{k1}, a_{k2}, \ldots, a_{km})\|_F^2 = \|V_k - V_p\| \tag{2.53}$$

where

$$V_p = U\Sigma_p O^\top \tag{2.54}$$

and Σ_p is obtained from Σ of the singular value decomposition (cf. Theorem 2.3.1) of V_k by setting all but its p largest singular values σ_i to zero. Since $p = 1$,

$$\Gamma_k(a_{k1}, a_{k2}, \ldots, a_{km}) = U\Sigma_1 O^\top \tag{2.55}$$

which is in fact a multiple of the first singular vector u_1 of V_k. As stated in Remark 2.3.1, the eigenvectors of $V_k V_k^\top$, which can also be computed as

$$V_k V_k^\top = \sum_{t/L_t = k} \mathbf{v}_t \cdot \mathbf{v}_t^\top \tag{2.56}$$

where the summation only includes time points with $L_t = k$, correspond with the left singular vectors of V_k. Hence, microstate Γ_k can be obtained as the eigenvector of the largest eigenvalue of $V_k V_k^\top$. Since the matrix $V_k V_k^\top$ is positive semidefinit, all eigenvalues λ_i are greater or equal to zero. Because of that, it is enough to just claim the largest eigenvalue instead of the largest positive eigenvalue. Therefore, Γ_k is given by

$$\Gamma_k = \arg\max_x x^\top V_k V_k^\top x \quad \text{with} \quad \|x\| = 1 \tag{2.57}$$

when using the Rayleigh quotient as in Definition 2.3.3. If this is done for all microstates Γ_k, $k = \{1, ..., M\}$, one gets an estimate for the microstates in this iteration. The algorithm takes turns between these two basic steps to finally reach the minimum of functional Fgiven in (2.45). This is done by initially guess either the microstates Γ_k with $k = \{1, ..., M\}$ or the labels L_t with $t \in \{1, ..., T\}$. The algorithm is instructed to terminate if successive iterations of the functional F differ negligibly.

The minimum of F, in this case, is an estimator for the noise variance σ^2 introduced in (2.39) and given by

$$F_{min} = \sigma^2 = \sum_{t=1}^{T} (\mathbf{v}_t^\top \cdot \mathbf{v}_t - (\Gamma_\kappa^\top \cdot \mathbf{v}_t)^2)/(T \cdot (N-1)) \tag{2.58}$$

where $\kappa = L_t$. To measure the goodness of fit of this model, the squared correlation coefficient can be calculated as

$$R^2 = 1 - \sigma^2/\sigma_D^2 \tag{2.59}$$

where

$$\sigma_D^2 = \sum_{t=1}^{T} (\mathbf{v}_t^\top \cdot \mathbf{v}_t)/(T \cdot (N-1)) \tag{2.60}$$

is the data variance of the EEG signal.

As many clustering algorithms, this one presented in [14] does not necessarily find the absolute minimum of F since it is very sensitive to the selection of the starting points. So it is possible that it terminates at a local minimum or saddle point. To prevent these happenings, the algorithm is started several times using different initializations. These initial microstates are obtained by randomly selected normalized maps $\mathbf{v}_t$ from the given data. Using all possible maps as the first microstate and randomly select a certain number of other maps to complete the number of needed microstates, the algorithm is started T times since T is the number of time points, i.e. number of maps. The result which corresponds to the minimal value of F is then selected and used for further computations.

The segmentation of a whole four second epoch of spontaneous EEG by using the clustering algorithm is shown in Figure 2.16. Figure 2.17 describes the clustering algorithm for a simplified case of ten original maps and two microstate classes to get an idea of the working mode.

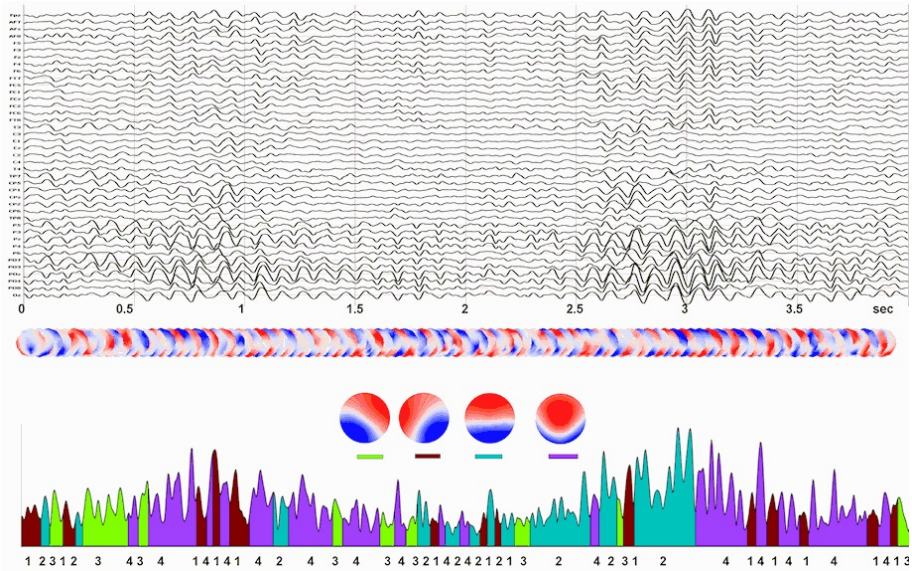

Figure 2.16: This figure shows the segmentation of a four second EEG epoch by using cluster analysis. The first part is an EEG recorded from 42 electrodes from a patient with closed eyes. The function at the bottom represents the GFP related to the EEG. Four dominant microstate classes were found by using a modified K-means clustering. These microstates are color-coded in the GFP curve and marked by numbers. This image was taken from [30].

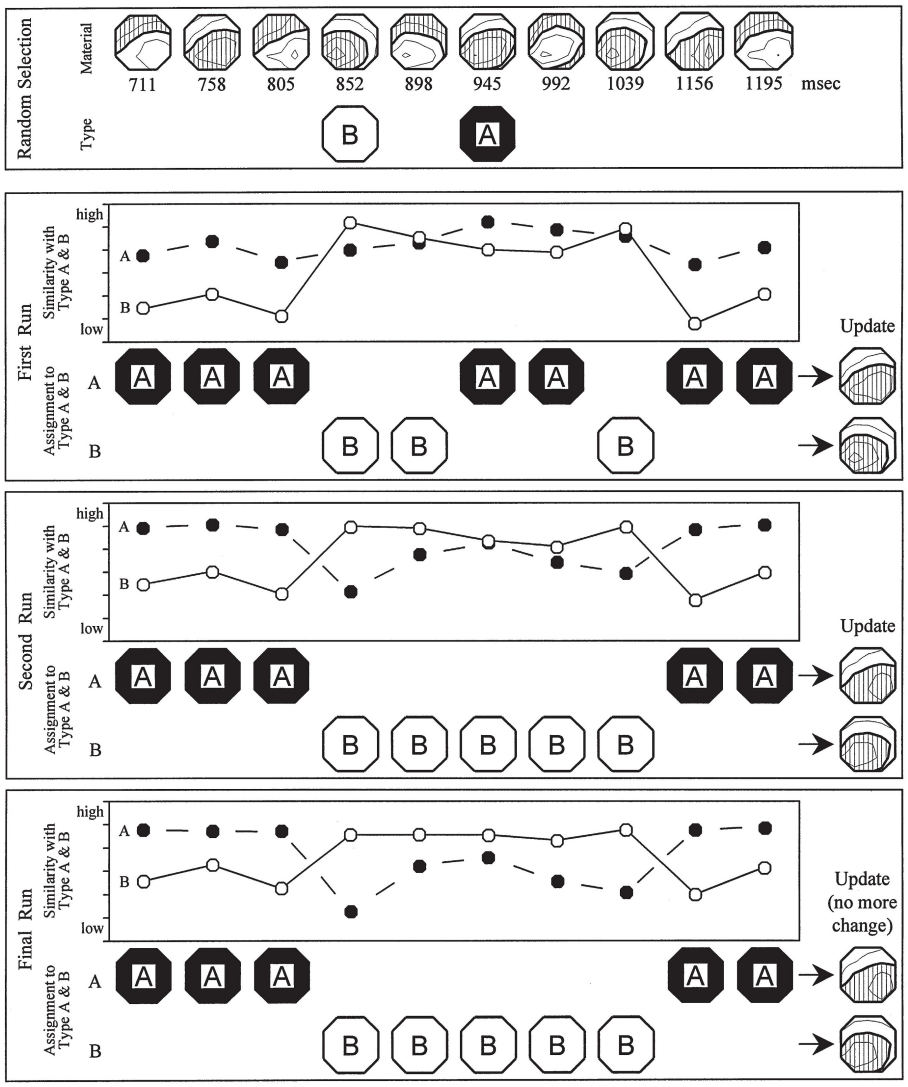

Figure 2.17: This figure describes the clustering of a sequence of ten original maps into two microstate classes. In the first box the maps are given with their potential landscape. Two maps have been randomly selected to be the initial prototype microstate classes. In the second box the first run of the algorithm is explained. Here, the similarity is calculated between each map and the microstate classes A (black) and B (white) using a arbitrary clustering criterion and plotted in the upper part of the box. The maps are assigned to the more similar class shown by black or white symbols. For each class, the prototypes are updated by combining the information given by all maps belonging to that class. In the third box the second iteration is shown. The same procedure as in box two is applied. It later runs until no more changes of the prototypes occur and these last prototypes represent the microstate classes of this selection of maps. This is given by the last box of this figure. This image was taken from [37].

2.8.2 Clustering Procedure

As explained in Section 2.8.1, the newly introduced modified K-means algorithm needs a prefixed number of microstates M or in this case microstate classes. The optimal number is chosen in dependence on [37]. In this paper, the optimal number of microstate classes is determined by the minimum of a cross-validation index. This index considers the number of classes used on the one hand and the percent variance explained by the class mean maps. The class mean maps are the mean of all maps within one class. This validation method was first presented in [14]. The optimal number of microstate classes was found to be four.

Following the procedure in [35], in a first run all original maps of each patient are clustered into four microstate classes using the algorithm in Section 2.8.1. Then *individual model maps* are computed for all patients. This is done by averaging all member maps of one class. Here, the polarity of the member maps has to be permuted to obtain minimal variance of the mean [35]. After that, four so-called *group model maps* are computed from the individual model maps. For this computation the modified K-means algorithm introduced in Section 2.8.1 is used again to cluster the individual model maps obtained in the previous step. It is constrained to produce one-to-one assignments of the patients' individual model maps to the new group model maps [35]. This is done to ensure that each individual model map of each subject is assigned to one of the newly calculated group model maps. These are named randomly as microstate class 'A', 'B', 'C' and 'D' following previous papers [35], [38]. In a final step, each original map is assigned to one of the four microstate classes which are represented by the group model maps. Successive original maps which are assigned to the same microstate class are then referred to as one single microstate.

2.8.3 Data Analysis of Microstate Classes

Following parameters are computed for all microstate classes for each patient [32], [35]:

- *Mean Microstate Duration* (ms): This is the mean duration of all microstates belonging to one class starting and ending at the midpoint in time between the last original map of the preceding microstate and the first original map of the following microstate.

- *Occurrences Per Second*: This is the number of occurrences of the microstates of one class per second.

- *Ratio Covered*: This is ratio of time covered by all microstates of a given class.

Additionally, global parameters are calculated for each individual [32], [35]:

- *GFP Peaks per Second*: This is the total number of original maps (peaks of GFP) of the epoch divided by the total time of the epoch in seconds.
- *Mean Microstate Duration All Classes* (ms): This is the mean duration of all microstates of all possible classes of one patient.
- *Duration of Longest Microstate* (ms): This is the duration of the longest microstate of this patient in the analyzed epoch.
- *Measure of Fit*: This is the measure of goodness of fit of the algorithm calculated as in (2.59). It is computed when determining the four microstate classes of each patient and a second time for the four group model maps.

The visualization of the electric potential landscape of the four microstate classes is done by using the software EEGLAB Toolbox for MATLAB [39].

2.9 Statistical Analysis of Calculated Parameter

In a first analysis, all parameters defined in Section 2.7.3 and Section 2.8.3 for all 96 patients were analyzed versus MMSE score by using a multiple linear least squares regression as described in Section 2.3.4. The MMSE score was introduced both as a quadratic and a linear regressor which results in the following model

$$y = \beta_0 + \beta_1 MMSE + \beta_2 MMSE^2 + \epsilon \ . \tag{2.61}$$

For this model the null hypothesis

$$H_0 : \ \beta_1 = \beta_2 = 0 \tag{2.62}$$

is tested by Fisher's F-test for a least squares quadratic regression model by using the MATLAB function `regstats` [40]. For each parameter of the analysis the significance p and the coefficient of determination R^2 were calculated. The significance obtained by the F-test is referred to as *significant* or *highly significant* if $p < 0.05$ and $p < 0.01$ respectively.

In a second approach, the parameters obtained for the 79 patients with more information given were analyzed versus MMSE score by using a multiple quadratic least squares regression including demographic variables. This is a common approach when dealing with medical data sets since these variables can have a strong influence on the measurements. For some patients, some of the demographic variables are missing, which finally results in 64 observations. The demographic variables age, sex, degree of education and duration of AD were used as additional

regressors. Age, duration and degree of education were applied by both linear and quadratic terms whereas sex was introduced only by a linear term. This results in

$$\begin{aligned} y =& \beta_0 + \beta_1 MMSE + \beta_2 MMSE^2 + \beta_3 Age + \beta_4 Age^2 + \\ & \beta_5 Sex + \beta_6 DE + \beta_7 DE^2 + \beta_8 DAD + \beta_9 DAD^2 + \epsilon \end{aligned} \tag{2.63}$$

where *DE* denotes the degree of education and *DAD* the duration of AD. The null hypothesis

$$H_0 : \ \beta_1 = \beta_2 = 0 \tag{2.64}$$

is tested by Fisher's F-test for a least squares quadratic regression model by using the MATLAB function `regstats` [40]. For this model the significance p and the coefficient of determination R^2 were calculated. The significance obtained by the F-test is referred to as *significant* or *highly significant* if $p < 0.05$ and $p < 0.01$ respectively. Subtracting the influences of the demographic variables, the dependent parameters were corrected and plotted versus their MMSE score.

Chapter 3

Results

This chapter is concerned with the results obtained by applying the different concepts and procedures explained in Chapter 2 to the EEG data provided by the PRODEM study. The first section presents the EEG measures versus MMSE scores using a standard regression model for the segmentation concept and the clustering procedure. The second part discusses the results obtained by a regression model using demographic variables like age, sex, duration of AD and degree of education. Again the first subsection is used to explain the results of the segmentation and the second one for the clustering. Each subsection starts with the main results and is followed by the significant parameters and their explanation.

3.1 Standard Regression Model

In this part, a common quadratic regression model as in (2.61) is used to obtain the statistical results for the whole database including 96 patients. For each parameter the significance p and the coefficient of determination R^2 is calculated. These are determined by Fisher's F-test for a least squares quadratic regression model. The significance obtained by the F-test is referred to as *significant* or *highly significant* which means $p < 0.05$ and $p < 0.01$ respectively. Since for all used approaches the selection of the original maps stays the same, the parameter GFP Peaks per Second is similar for segmentation and clustering. For this general parameter the regression parameters obtained by MATLAB are stated in Table 3.1. The results are also explained in Figure 3.1.

Parameter	Coef	StdError	tStat	pVal
Intercept	18.08700	0.76679	23.58800	$5 \cdot 10^{-41}$
MMSE	0.01426	0.31048	0.04594	0.96346
MMSE^2	-0.00859	0.02568	-0.33451	0.73875

Table 3.1: In this table, the regression parameters for the standard regression model, cf. (2.61), when analyzing the parameter GFP Peaks per Second are given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The regression resulted in $p = 0.5623$ and $R^2 = 0.0123$.

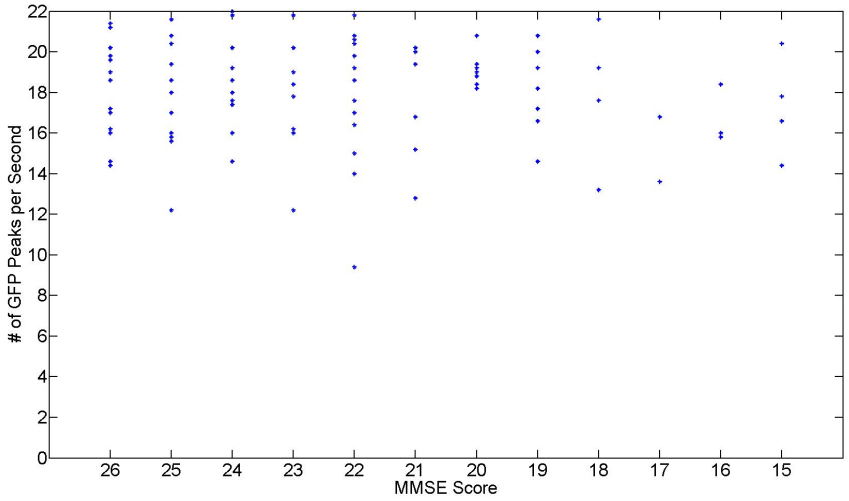

Figure 3.1: This figure shows the parameter GFP Peaks per Second versus the MMSE score for each patient. The average number of GFP peaks is 17.87 per second with a standard deviation of 2.55. A quadratic regression was fitted with $p = 0.5623$ and $R^2 = 0.0123$.

3.1.1 Segmentation of EEG Recordings into Microstates

In this section, the detailed results for segmenting EEG recordings into microstates are presented. As introduced in Section 2.7.3 the obtained microstates are divided in three classes: the longest microstate, single map segments and multiple map segments. For both, the individual window size and the group window size, the parameters are analyzed separately. The corresponding results are explained in the same order after discussing the main parameters relevant for both ways of examination. One is Optimal Window Size, the other two Single Map Segments per Second for both the individual and the group window size. All three are presented in Figures 3.2, 3.3, 3.4. The detailed results are given in Table 3.2.

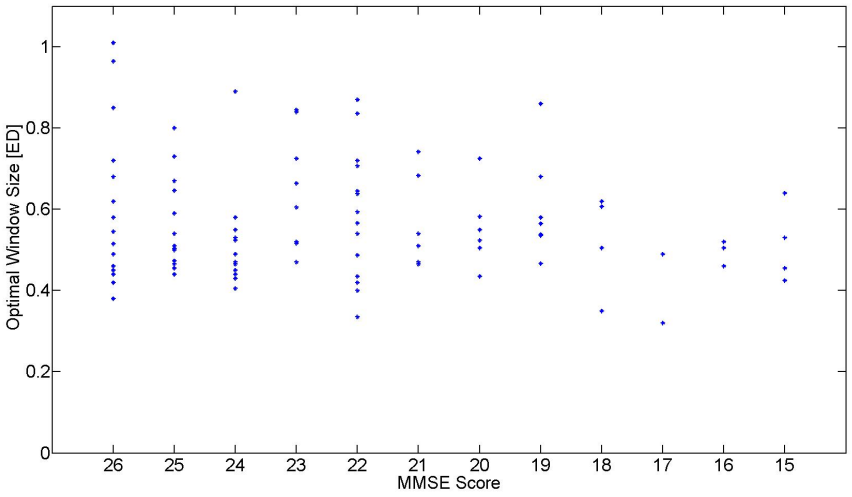

Figure 3.2: Here, the Optimal Window Size in ED versus the MMSE score for each patient is presented. The average size for all patients is 0.56 with standard deviation of 0.16 which corresponds to the group window size.

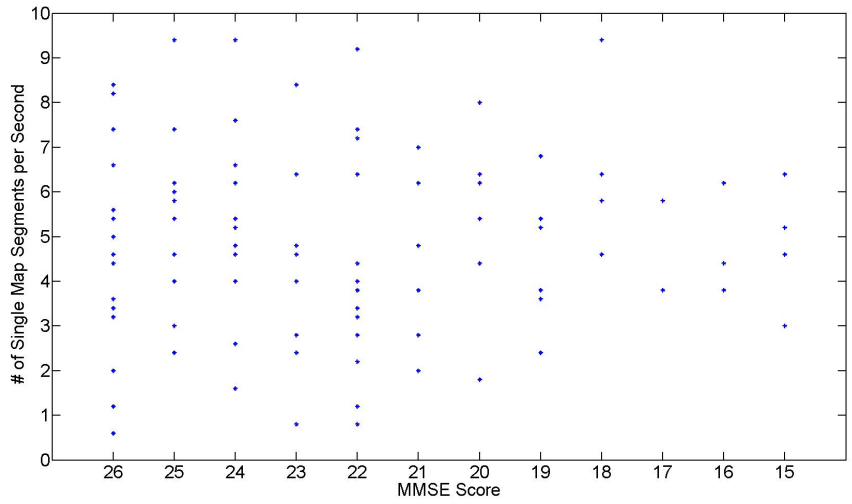

Figure 3.3: This figure shows the parameter Single Map Segments per Second versus MMSE score for each patient using the individual window size. The average number is 4.91 with a standard deviation of 2.05.

3.1.1.1 Individual Window Size

In the following, results calculated for each patient by using the individual window size are presented. The three classes longest microstate, single map segments and

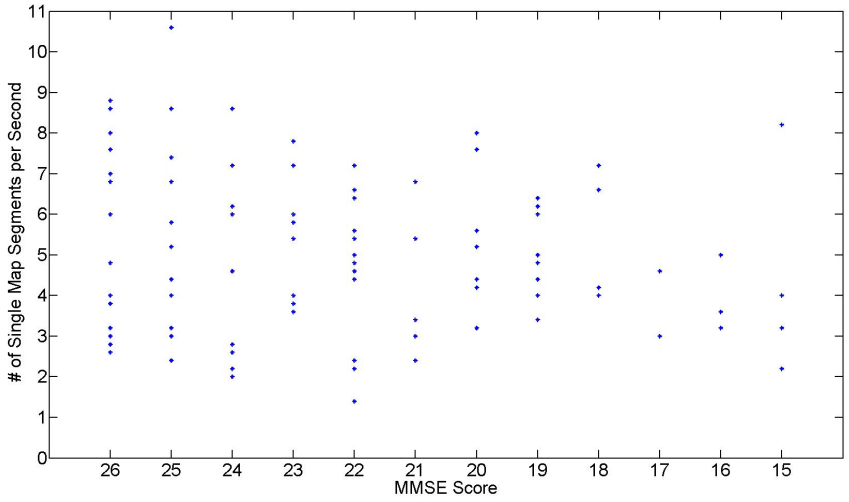

Figure 3.4: Here, the parameter Single Map Segments per Second versus MMSE score for each patient using the group window size is presented. The average number is 4.90 with a standard deviation of 1.97.

Parameter	p	R^2
Optimal Window Size	0.2586	0.0287
Single Peak Segments/s - Individual Window	0.9605	$9 \cdot 10^{-4}$
Single Peak Segments/s - Group Window	0.6033	0.0108

Table 3.2: In this table, the significance p and the coefficient of determination R^2 for the general parameters of the segmentation procedure are given. Starting with the Optimal Window Size in ED and then stating the Single Map Segments per Second for both the individual and the group window size.

multiple map segments are shown in Table 3.3 and the magnitude of the duration of all microstate classes is stated in Table 3.4.

Longest Microstate

Parameter	p	R^2
Mean Duration	0.7665	0.0057
Ratio Covered	0.7665	0.0057
Centroid x-Position	0.1506	0.0399
Centroid y-Position	0.9332	0.0015
Distance	0.6361	0.0097

Multiple Map Segments

Parameter	p	R^2
Mean Duration	0.5285	0.0136
Ratio Covered	0.9430	0.0013
Centroid x-Position	0.2684	0.0279
Centroid y-Position	0.0771	0.0536
Distance	0.4372	0.0176

Single Map Segments

Parameter	p	R^2
Mean Duration	0.2967	0.0258
Ratio Covered	0.7233	0.0069
Centroid x-Position	0.2617	0.0284
Centroid y-Position	0.1730	0.0370
Distance	0.7531	0.0061

Table 3.3: In this table, the significance p and the coefficient of determination R^2 for the parameters using the individual window size are presented.

Parameter	Mean	STD
Longest Microstate	570.78	319.88
Multiple Map Segments	212.69	75.51
Single Map Segments	55.50	9.12

Table 3.4: In this table, the mean and the standard deviation of the parameter Mean Microstate Duration are given in ms. This is stated for each of the three microstate classes for the individual window size.

3.1.1.2 Group Window Size

The results for each patient by using the group window size are presented. The three classes longest microstate, single map segments and multiple map segments are shown in the corresponding order in Table 3.5 and the magnitude of the duration of all microstate classes is stated in Table 3.6.

Longest Microstate

Parameter	p	R^2
Mean Duration	0.1064	0.0470
Ratio Covered	0.1064	0.0470
Centroid x-Position	0.0116 *	0.0915 *
Centroid y-Position	0.6877	0.0081
Distance	0.6417	0.0095

Multiple Map Segments

Parameter	p	R^2
Mean Duration	0.6684	0.0086
Ratio Covered	0.9326	0.0015
Centroid x-Position	0.1080	0.0467
Centroid y-Position	0.1892	0.0352
Distance	0.3654	0.0214

Single Map Segments

Parameter	p	R^2
Mean Duration	0.1325	0.0425
Ratio Covered	0.8720	0.0029
Centroid x-Position	0.3478	0.0225
Centroid y-Position	0.2615	0.0284
Distance	0.9164	0.0019

Table 3.5: In this table, the significance p and the coefficient of determination R^2 for the parameters using the group window size are presented. Significant results are marked with *.

Here, the results for one parameter are significant. This is the Centroid x-Position for the longest microstate class shown in Figure 3.5. The quadratic regression was characterized by an increase for MMSE scores between 26 and 20, and a slight decrease from 20 downwards. The regression parameters are given in Table 3.7.

Parameter	Mean	STD
Longest Microstate	620.28	244.04
Multiple Map Segments	254.69	92.73
Single Map Segments	55.62	8.88

Table 3.6: In this table, the mean and the standard deviation of the parameter Mean Microstate Duration are given in ms. This is stated for each of the three microstate classes for the group window size.

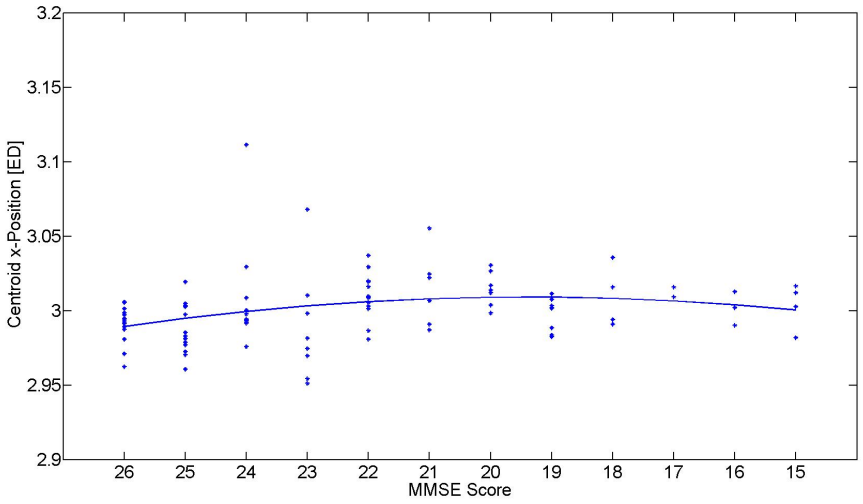

Figure 3.5: Here, the Centroid x-Position in ED for the longest microstate class of the group window approach versus the MMSE score is shown. The quadratic regression was fitted to the data with $p = 0.0116$ and $R^2 = 0.0915$.

Parameter	Coef	StdError	tStat	pVal
Intercept	2.98320	0.00658	453.38000	$2 \cdot 10^{-157}$
MMSE	0.00690	0.00266	2.58840	0.01119
MMSE^2	-0.00045	0.00022	-2.05180	0.04200

Table 3.7: In this table, the regression parameters for the standard regression model, cf. (2.61), when analyzing the parameter Centroid x-Position for the longest microstate class using the group window size are given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The quadratic regression was fitted to the data with $p = 0.0116$ and $R^2 = 0.0915$.

3.1.2 Clustering of EEG Recordings into Microstates

In this section, the results of the clustering procedure described in Section 2.8 are presented. First, a general parameter, the Data Variance as explained in (2.60), is plotted in Figure 3.6.

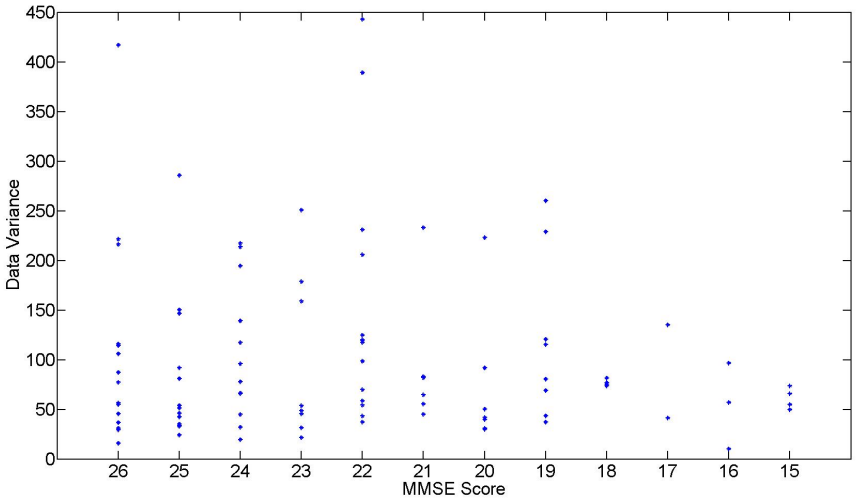

Figure 3.6: This figure shows the data variance computed as in (2.60) versus the MMSE score of each patient.

Now, the results for the clustering procedure itself are presented. The Measure of Fit, defined in (2.59), for the applied model when determining the four microstate classes of each patient is discussed first in Figure 3.7. The later on applying of the algorithm when computing the group model maps resulted in a squared correlation coefficient $R^2 = 0.5997$. Using EEGLAB, the group model maps are plotted for a better visualization. This is presented in Figure 3.8.

The detailed results for each parameter and microstate class are given in Table 3.8 and the magnitude of the duration of the microstate classes follows in Table 3.9.

Here, for one parameter the results are significant. This is the Ratio Covered by microstate class C shown in Figure 3.9. The quadratic regression was characterized by an decrease for MMSE scores between 26 and 20, and an increase from 20 downwards. The parameters obtained by the regression analysis are stated in Table 3.10.

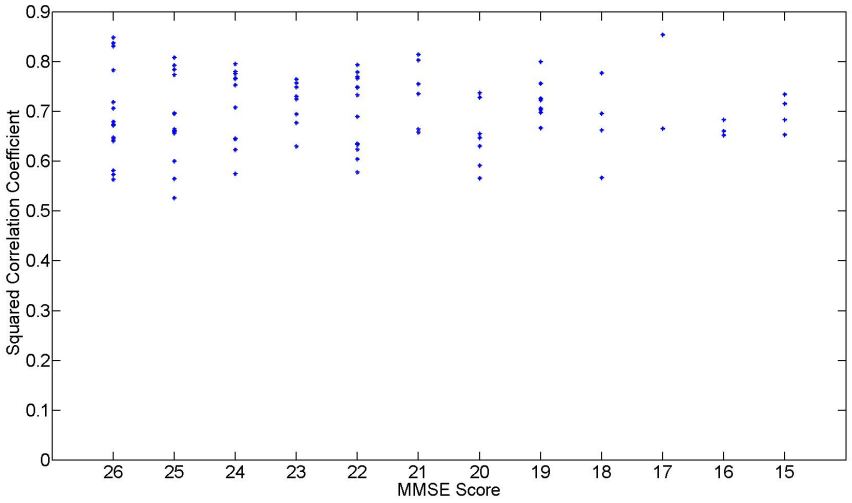

Figure 3.7: This plot illustrates the Measure of Fit of the applied model by using the squared correlation coefficient defined in (2.59) versus the MMSE score of all patients. The average value is 0.73 with a standard deviation of 0.07.

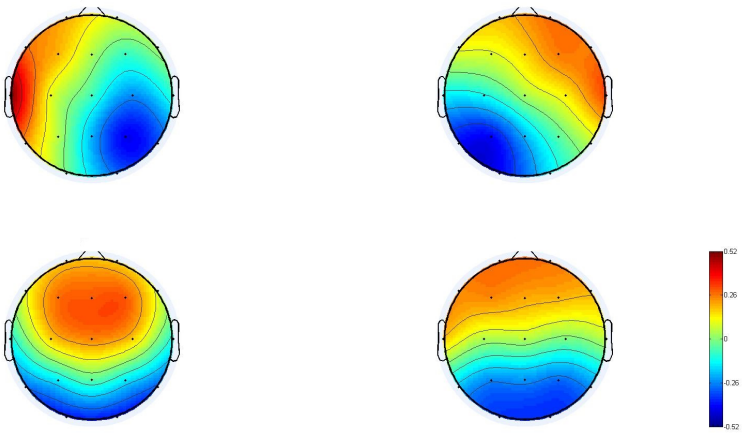

Figure 3.8: This figure shows the scalp maps given by the calculated group model maps. The four maps are named with Microstate Class 'A', 'B', 'C' and 'D' from left to right, starting with the left upper corner.

Class A

Parameter	p	R^2
Mean Duration	0.9697	0.0007
Ratio Covered	0.8471	0.0034
Occurrences/s	0.9348	0.0014

Class B

Parameter	p	R^2
Mean Duration	0.9081	0.0021
Ratio Covered	0.6420	0.0095
Occurrences/s	0.6961	0.0078

Class C

Parameter	p	R^2
Mean Duration	0.5615	0.0123
Ratio Covered	0.0431 *	0.0654 *
Occurrences/s	0.1910	0.0350

Class D

Parameter	p	R^2
Mean Duration	0.4433	0.0173
Ratio Covered	0.2372	0.0305
Occurrences/s	0.3360	0.0232

Table 3.8: In this table, the significance p and the coefficient of determination R^2 for the parameters of the clustering procedure are presented. Significant results are marked with *.

Parameter	Mean	STD
Class A	65.61	16.20
Class B	65.04	15.97
Class C	66.15	14.94
Class D	69.32	24.60

Table 3.9: In this table, the average Mean Microstate Duration and its standard deviation in ms are given for each patient and microstate class.

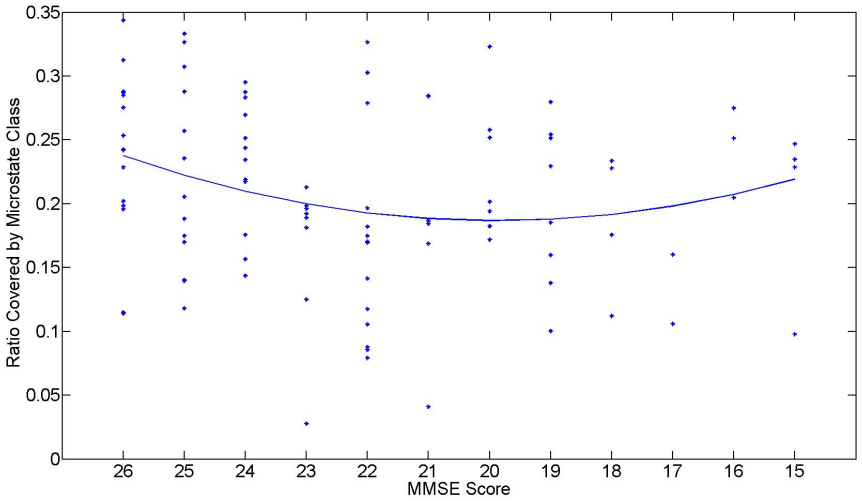

Figure 3.9: Here, the Ratio Covered by microstate class C versus the MMSE score is shown. The quadratic regression was fitted to the data with $p = 0.0431$ and $R^2 = 0.0654$.

Parameter	Coef	StdError	tStat	pVal
Intercept	0.25560	0.02032	12.57700	$9 \cdot 10^{-22}$
MMSE	-0.01937	0.00823	-2.35380	0.02069
MMSE^2	0.00136	0.00068	2.00130	0.04828

Table 3.10: In this table, the regression parameters for the standard regression model, cf. (2.61), when analyzing the parameter Ratio Covered by microstate class C using the group window size are given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The quadratic regression was fitted to the data with $p = 0.0431$ and $R^2 = 0.0654$.

3.2 Regression Model with Demographic Variables

When applying the second regression model as in (2.63), the demographic variables age, sex, degree of education and duration of AD are involved. Since for some of the 79 patients demographic variables are missing, the analysis results in only 64 observations. For each parameter the significance p and the coefficient of determination R^2 is calculated. These are determined by Fisher's F-test for a least squares quadratic regression model. The significance obtained by the F-test is referred to as *significant* or *highly significant* which means $p < 0.05$ and $p < 0.01$ respectively. Subtracting the influences of the demographic variables, the dependent parameters were corrected and plotted versus their MMSE score.

Since for all used approaches the selection of the original maps stays the same, the parameter GFP Peaks per Second is similar for segmentation and clustering. For this general parameter the regression parameters obtained by MATLAB are stated in Table 3.11. The results are also explained in Figure 3.10.

Parameter	Coef	StdError	tStat	pVal
Intercept	42.3650	26.97800	1.57040	0.12217
MMSE	-0.09231	1.66360	-0.05549	0.95596
$MMSE^2$	0.00997	0.03961	0.25167	0.80225
Age	-0.59893	0.62343	-0.96070	0.34099
Age^2	0.00358	0.00442	0.80996	0.42152
Sex	-1.04870	0.86726	-1.20920	0.23186
DE	-1.09320	1.17620	-0.92942	0.35681
DE^2	0.10106	0.16729	0.60408	0.54832
DAD	0.02157	0.05328	0.40482	0.68721
DAD^2	-0.00010	0.00048	−0.21342	0.83181

Table 3.11: In this table, the regression parameters for the regression model using demographic variables, cf. (2.63), when analyzing the parameter GFP Peaks per Second is given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The regression led to $p = 0.0564$ and $R^2 = 0.2153$.

3.2.1 Segmentation of EEG Recordings into Microstates

In this section, the detailed results for segmenting EEG recordings into microstates are presented. As introduced in Section 2.7.3, the obtained microstates are divided in three classes: the longest microstate, single map segments and multiple map segments. For both, the individual window size and the group window size, these parameters are analyzed separately. The corresponding results are explained in

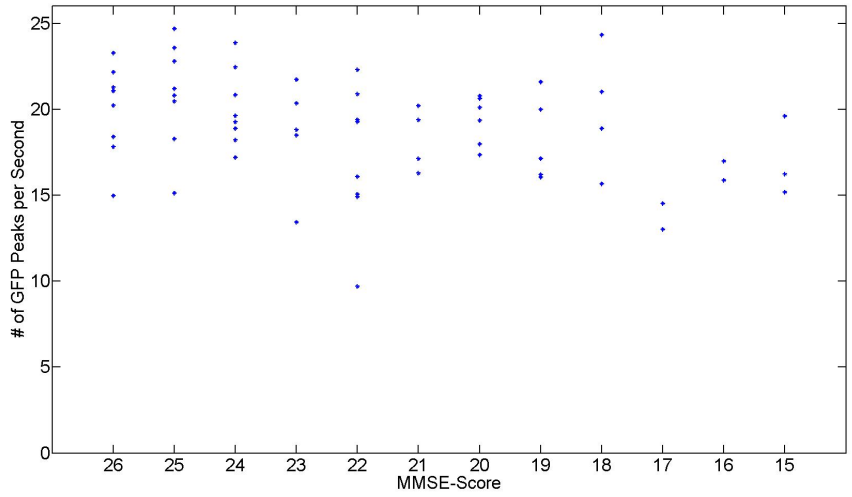

Figure 3.10: This figure shows the parameter GFP Peaks per Second versus the MMSE score for each patient. The average number of GFP peaks is 18.90 per second with a standard deviation of 3.00. A quadratic regression was fitted with $p = 0.0564$ and $R^2 = 0.2153$.

the same order after discussing the main parameters relevant for both ways of examination. One is Optimal Window Size, the other two Single Maps Segments per Second for both the individual and the group window size. All three are presented in Figures 3.11, 3.12, 3.13. The detailed results are given in Table 3.12.

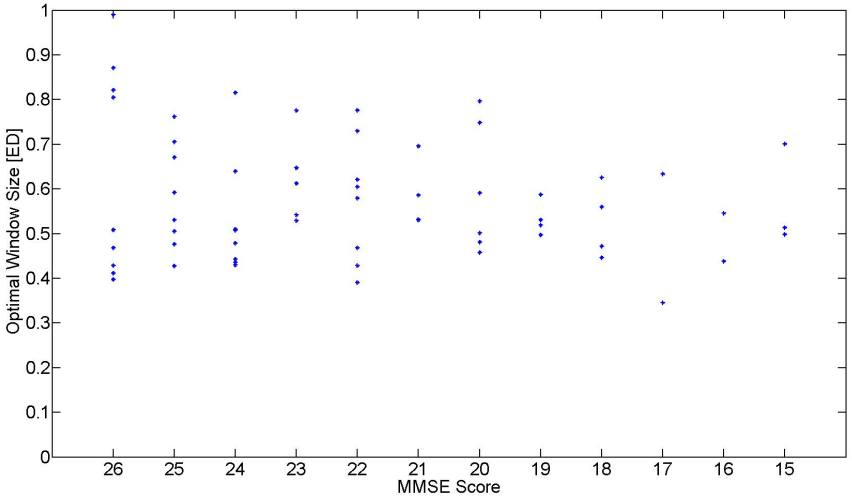

Figure 3.11: Here, the Optimal Window Size in ED versus the MMSE score for each patient is presented. The average optimal window size for all patients is 0.57 with standard deviation of 0.14 which corresponds to the group window size.

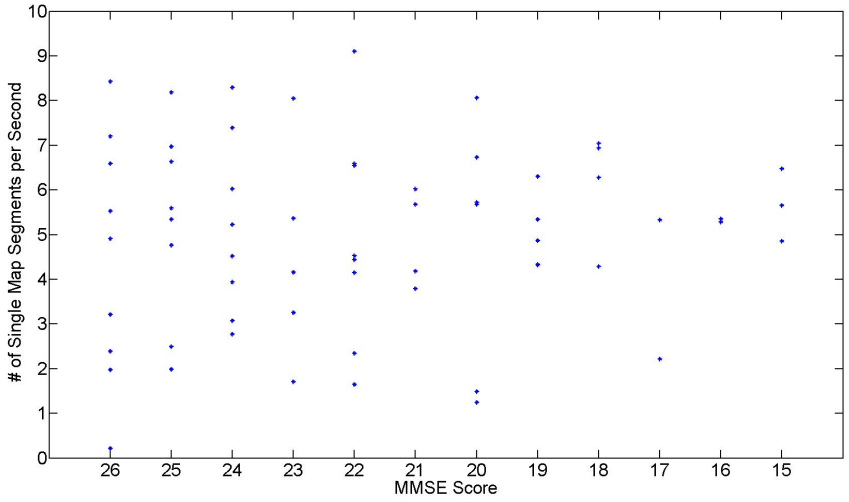

Figure 3.12: This figure shows the parameter Single Map Segments per Second versus MMSE score for each patient using the individual window size. The average number of single map segments is 4.98 with a standard deviation of 2.01.

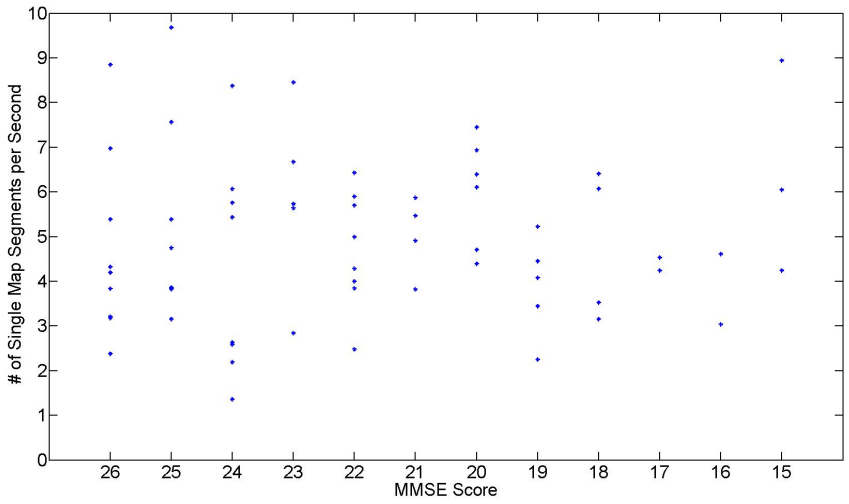

Figure 3.13: Here, the parameter Single Map Segments per Second versus MMSE score for each patient using the group window size is presented. The average number of single map segments is 4.94 with a standard deviation of 1.81.

3.2.1.1 Individual Window Size

In this part, the results calculated for each patient by using the individual window size are presented. The three classes longest microstate, single map segments and

Parameter	p	R^2
Optimal Window Size	0.4956	0.1577
Single Peak Segments/s - Individual Window	0.8499	0.1934
Single Peak Segments/s - Group Window	0.9574	0.1920

Table 3.12: In this table, the significance p and the coefficient of determination R^2 for the general parameters of the segmentation procedure are given. Starting with the Optimal Window Size in ED and then stating the Single Map Segments per Second for both the individual and the group window size.

multiple map segments are shown in the corresponding order in Table 3.13. Later on, the magnitude of the duration of all microstate classes is provided in Table 3.14.

For one parameter, namely the Mean Microstate Duration for the single map segment class, the results were significant. The corresponding plot is shown in Figure 3.14 and the regression parameters in Table 3.15. The regressions is characterized by an increase for more severe impairment.

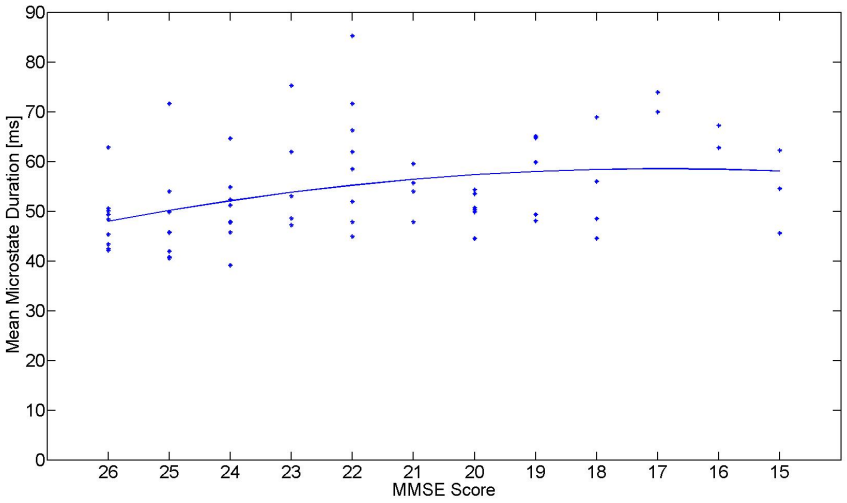

Figure 3.14: Here, the Mean Microstate Duration for the single map segment class of the individual window approach versus the MMSE score is shown. The quadratic regression was fitted to the data with $p = 0.0455$ and $R^2 = 0.1733$.

Longest Microstate		
Parameter	p	R^2
Mean Duration	0.8985	0.1064
Ratio Covered	0.8985	0.1064
Centroid x-Position	0.4918	0.1483
Centroid y-Position	0.4210	0.1537
Distance	0.3762	0.1750
Multiple Map Segments		
Parameter	p	R^2
Mean Duration	0.7135	0.0986
Ratio Covered	0.2440	0.1521
Centroid x-Position	0.2562	0.1497
Centroid y-Position	0.1747	0.2138
Distance	0.1726	0.1633
Single Map Segments		
Parameter	p	R^2
Mean Duration	0.0455 *	0.1733 *
Ratio Covered	0.1121	0.2825
Centroid x-Position	0.0669	0.1918
Centroid y-Position	0.4128	0.1666
Distance	0.9794	0.1540

Table 3.13: In this table, the significance p and the coefficient of determination R^2 for the parameters using the individual window size are presented. Significant results are marked with *.

Parameter	Mean	STD
Longest Microstate	563.44	301.45
Multiple Map Segments	207.51	77.78
Single Map Segments	54.12	9.95

Table 3.14: In this table, the mean and the standard deviation of the parameter Mean Microstate Duration are given in ms. This is stated for each of the three microstate classes for the individual window size.

3.2.1.2 Group Window Size

In this section, the results calculated for each patient by using the group window size are presented. The three classes longest microstate, single map segments and

Parameter	Coef	StdError	tStat	pVal
Intercept	0.00450	0.08881	0.05068	0.95977
MMSE	0.00434	0.00548	0.79316	0.43115
MMSE^2	-0.00013	0.00013	-0.98458	0.32922
Age	0.00012	0.00205	0.05649	0.95516
Age^2	$1 \cdot 10^{-7}$	0.00001	0.00954	0.99242
Sex	0.00221	0.00285	0.77579	0.44126
DE	0.00356	0.00387	0.91996	0.36169
DE^2	-0.00030	0.00055	-0.53817	0.59267
DAD	0.00003	0.00018	0.18851	0.85118
DAD^2	$-2 \cdot 10^{-7}$	$1 \cdot 10^{-6}$	-0.11928	0.90550

Table 3.15: In this table, the regression parameters for the regression model using demographic variables, cf. (2.63), when analyzing the parameter Mean Microstate Duration for the single map segment class and applying the individual window size are given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The regression led to $p = 0.0455$ and $R^2 = 0.1733$

multiple map segments are shown in the corresponding order in Table 3.16. Later on the magnitude of the duration of all microstate classes is provided in Table 3.17.

Again for one parameter, namely the Centroid x-Position for the longest microstate class, the statistical analysis led to significant results. The corresponding plot is shown in Figure 3.15. The regression parameters are stated in Table 3.18. Again the regression was characterized by a slight increase for more severe impairment.

Longest Microstate		
Parameter	p	R^2
Mean Duration	0.4904	0.1242
Ratio Covered	0.4904	0.1242
Centroid x-Position	0.0454 *	0.2245 *
Centroid y-Position	0.6439	0.0616
Distance	0.7171	0.0894

Multiple Map Segments		
Parameter	p	R^2
Mean Duration	0.8349	0.0543
Ratio Covered	0.3190	0.1990
Centroid x-Position	0.0853	0.1874
Centroid y-Position	0.3267	0.1777
Distance	0.0909	0.2182

Single Map Segments		
Parameter	p	R^2
Mean Duration	0.0590	0.1533
Ratio Covered	0.5122	0.2164
Centroid x-Position	0.2439	0.1006
Centroid y-Position	0.3249	0.1615
Distance	0.9824	0.1607

Table 3.16: In this table, the significance p and the coefficient of determination R^2 for the parameters using the group window size are presented. Significant results are marked with *.

Parameter	Mean	STD
Longest Microstate	616.12	233.50
Multiple Map Segments	252.51	117.60
Single Map Segments	54.61	9.91

Table 3.17: In this table, the mean and the standard deviation of the parameter Mean Microstate Duration are given in ms. This is stated for each of the three microstate classes for the group window size.

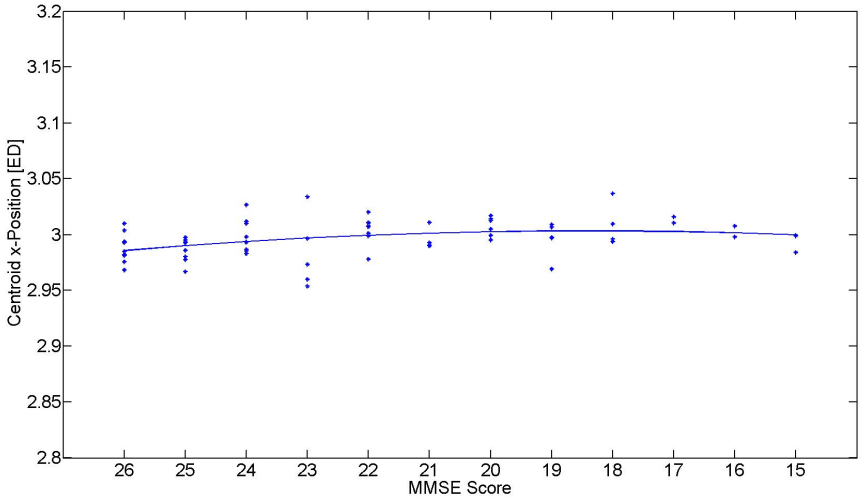

Figure 3.15: Here, the Centroid x-Position for the longest microstate of the group window approach versus the MMSE score is shown. The quadratic regression was fitted to the data with $p = 0.0454$ and $R^2 = 0.2245$.

Parameter	Coef	StdError	tStat	pVal
Intercept	2.66640	0.14697	18.14200	$1.2 \cdot 10^{-24}$
MMSE	0.01116	0.00906	1.23160	0.22342
MMSE^2	-0.00030	0.00022	-1.40660	0.16527
Age	0.00660	0.00340	1.94300	0.05724
Age^2	-0.00005	0.00002	-1.87990	0.06552
Sex	0.00586	0.00472	1.24060	0.22013
DE	0.00055	0.00641	0.08506	0.93253
DE^2	-0.00021	0.00091	-0.23484	0.81522
DAD	-0.00010	0.00029	-0.35752	0.72210
DAD^2	$-8 \cdot 10^{-8}$	0.00001	-0.32268	0.74819

Table 3.18: In this table, the regression parameters for the regression model using demographic variables, cf. (2.63), when analyzing the parameter Centroid x-Position for the longest microstate class and the group window size approach are given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The regression led to $p = 0.0454$ and $R^2 = 0.2245$

3.2.2 Clustering of EEG Recordings into Microstates

In this section, the results of the clustering procedure described in Section 2.8 are presented. First, a general parameter, the data variance as explained in (2.60), is plotted in Figure 3.16.

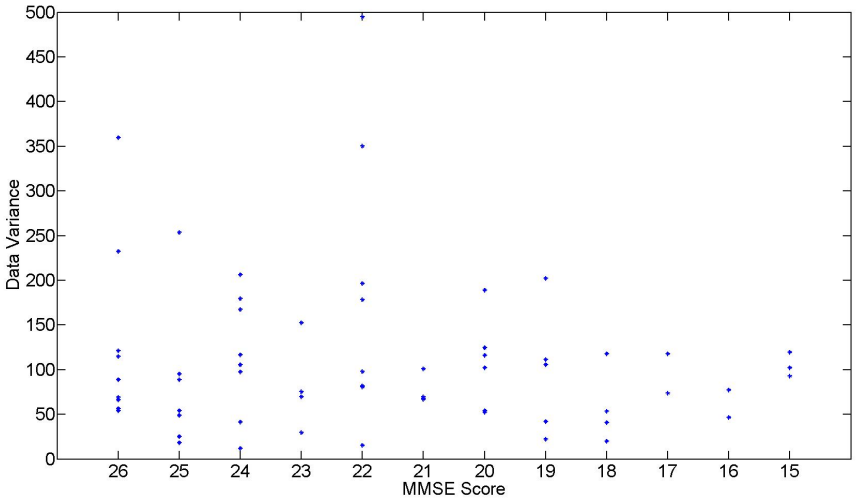

Figure 3.16: This figure shows the data variance computed as in (2.60) versus the MMSE score of each patient.

Now, the results for the clustering procedure itself are presented. The Measure of Fit, defined in (2.59), for the applied model when determining the four microstate classes of each patient is discussed first in Figure 3.17. The later on applying of the algorithm when computing the group model maps resulted in a squared correlation coefficient $R^2 = 0.6072$. Using EEGLAB, the group model maps are plotted for a better visualization. This is presented in Figure 3.18.

The detailed results for each parameter and microstate class are given in the Tables 3.19 and the magnitude of the duration of the microstate classes follows in Table 3.20.

For the parameter Ratio Covered by microstate class C the statistical analysis led to significant results. The corresponding plot is shown in Figure 3.19. The regression is indicated by a decrease for more severe impairment caused by AD. The parameters of the regression are shown in Figure 3.21.

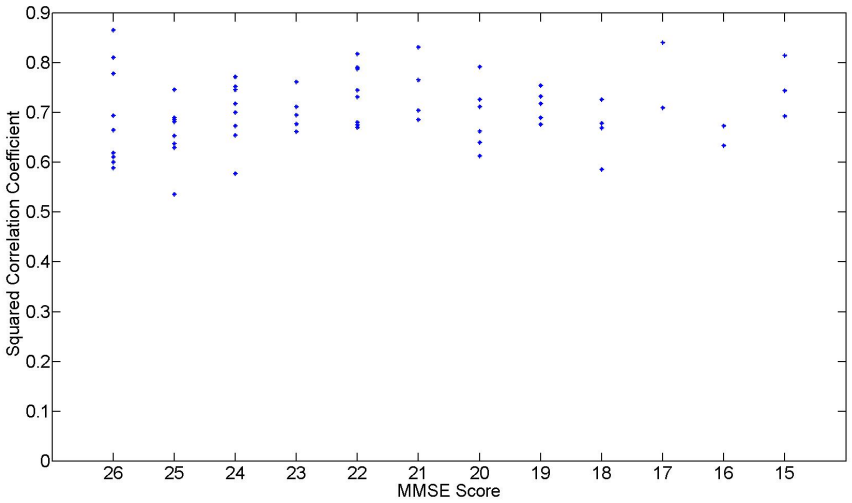

Figure 3.17: This plot illustrates the Measure of Fit of the applied model by using the squared correlation coefficient defined in (2.59) versus the MMSE score of all patients. The average value is 0.70 with a standard deviation of 0.08.

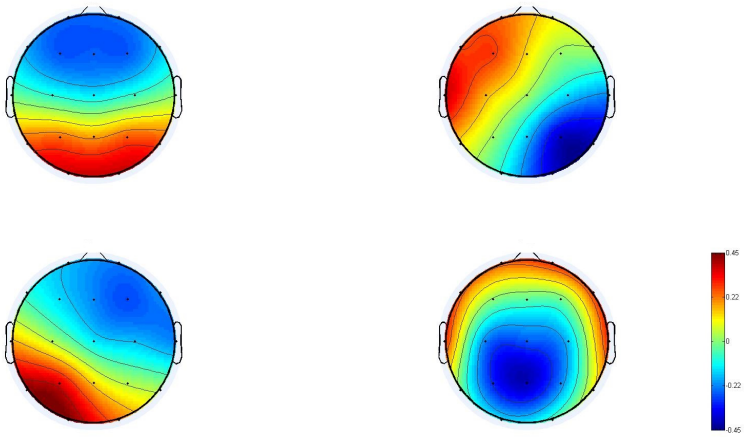

Figure 3.18: This figure shows the scalp maps given by the calculated group model maps. The four maps are named with Microstate Class 'A', 'B', 'C' and 'D' from left to right, starting with the left upper corner. These differ from the maps in Figure 3.8 since less patients were included in the calculation. Here, there are also two maps with diagonal isopotential lines, one with more or less horizontal ones as before. Only map D (lower right corner) differs more significantly from map C (lower left corner) in Figure 3.8.

Class A		
Parameter	p	R^2
Mean Duration	0.1557	0.1126
Ratio Covered	0.2558	0.1196
Occurrences/s	0.9173	0.1824

Class B		
Parameter	p	R^2
Mean Duration	0.2142	0.1697
Ratio Covered	0.3145	0.1927
Occurrences/s	0.9689	0.1226

Class C		
Parameter	p	R^2
Mean Duration	0.8531	0.1283
Ratio Covered	0.0247 *	0.2145 *
Occurrences/s	0.0572	0.1258
hline		

Class D		
Parameter	p	R^2
Mean Duration	0.3798	0.1132
Ratio Covered	0.4988	0.1398
Occurrences/s	0.3038	0.1681

Table 3.19: In this table, the significance p and the coefficient of determination R^2 for the parameters of the clustering procedure are presented. Significant results are marked with *.

Parameter	Mean	STD
Class A	72.10	23.42
Class B	65.32	16.21
Class C	65.52	14.58
Class D	63.61	15.35

Table 3.20: In this table, the average Mean Microstate Duration and its standard deviation in ms are given for each patient and microstate class.

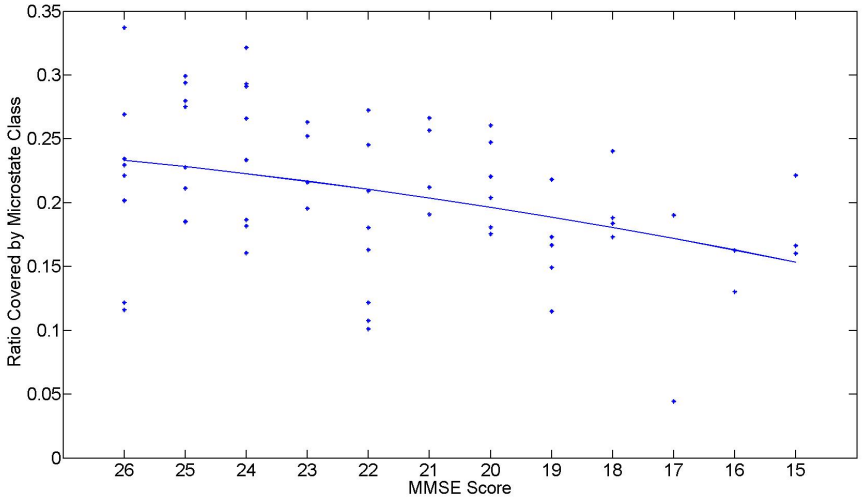

Figure 3.19: Here, the Ratio Covered by microstate class C versus the MMSE score is shown. The quadratic regression was fitted to the data with $p = 0.0247$ and $R^2 = 0.2145$.

Parameter	Coef	StdError	tStat	pVal
Intercept	0.82682	0.64362	1.28460	0.20440
MMSE	-0.01304	0.03969	-0.32859	0.74374
$MMSE^2$	0.00019	0.00095	0.19679	0.84473
Age	-0.01216	0.01487	-0.81746	0.41726
Age^2	0.00009	0.00011	0.82296	0.41415
Sex	-0.00097	0.02069	-0.04710	0.96261
DE	0.01771	0.02806	0.63122	0.53056
DE^2	-0.00211	0.00399	-0.53095	0.59763
DAD	-0.00018	0.00127	-0.14077	0.88858
DAD^2	-0.00001	0.00001	-0.45083	0.65392

Table 3.21: In this table, the regression parameters for the regression model using demographic variables, cf. (2.63), when analyzing the parameter Ratio Covered by microstate class C are given. *Coef* indicates the coefficient, *StdError* the standard error of the estimator, *tStat* and *pVal* the t statistics and p-values for the coefficients. The regression led to $p = 0.0247$ and $R^2 = 0.2145$

Chapter 4

Discussion

This chapter sums up and discusses the findings of this work provided in Chapter 3. The results are compared to other publications in the research field of EEG microstates and the present scientific knowledge. The weak points and problems of the used methods are analyzed and some critical thoughts are expressed. Again the terms *significant* and *highly significant* will be used for the results of the quadratic regression obtained by the F-test if the significance $p < 0.05$ and $p < 0.01$ respectively. Starting with the results of the first regression model and later on followed by the second one, the main results are now presented.

Standard Regression Model

- H_0^1: The EEG microstate duration shortens in the course of AD.

The analysis of the EEG microstates of the 96 probable AD patients revealed no significant changes in duration. For both, the segmentation and the clustering procedure, the results could not document a shortening of EEG microstate duration.

- H_0^2: The topography of the microstates' electric potential landscapes changes in the course of AD.

Generally, the examining of the topography of the microstates by using the segmentation procedure based on the locations of the positive and negative potential areas revealed no significant changes. There was only one significant regression for the Centroid x-Position parameter which was found by analyzing the longest microstate class of the group window size. The quadratic regression was characterized by an increase for MMSE scores between 26 and 20, and a slight decrease from 20 downwards.

- H_0^3: The characteristics, e.g. occurrences, duration or ratio, of specific microstate classes change in the course of AD.

Only one significant change could be found when using the clustering procedure. This was the Ratio Covered by Microstate Class C which was described by an decrease for MMSE scores between 26 and 20, and an increase from 20 downwards.

Regression Model with Demographic Variables

- H_0^1: The EEG microstate duration shortens in the course of AD.

The analysis of the EEG microstates of the 79 probable AD patients revealed no significant changes in duration. For both, the segmentation and the clustering procedure, the results could not document a shortening of EEG microstate duration.

- H_0^2: The topography of the microstates' electric potential landscapes changes in the course of AD.

Generally, the examining of the topography of the microstates by using the segmentation procedure based on the locations of the positive and negative potential areas revealed no significant changes. There was only one significant result analyzing the Centroid x-Position for the longest microstate class and the group window approach. The regression showed a slight increase for more severe impairment which remained stable after MMSE scores of 20.

- H_0^3: The characteristics, e.g. occurrences, duration or ratio, of specific microstate classes change in the course of AD.

There were two significant changes in the characteristics of certain microstate classes. For the individual window size approach of the segmentation procedure, this revealed an increase of duration for more severe impairment for the single map segment class. The regression obtained was characterized by an increase which sagged after MMSE score 20. For the clustering procedure, there was also one significant result for microstate class C. The Ratio Covered showed a strong decrease for more severe impairment.

Looking at the results obtained by the second regression model which takes the demographic variables age, sex, duration of AD and degree of education into account, this regression leads to higher R^2 values throughout all parameters which have been calculated. On the other hand, the inspection of the regression parameters obtained for the demographic parameters reveals no significance for them. Since the regression including demographic variables is the more common one in the medical research field, only the results obtained by this regression are compared with other papers and publications in the following part of the section.

There are only few papers which deal with EEG microstate alterations of subjects with AD. In all cases, the patients are compared with a group of healthy controls. These papers found that the microstate duration decreases with cogni-

tive impairment. In [11], the analysis resulted in a decreased microstate duration for the longest microstate class and an anteriorisation of the electric fields on the scalp for more severe AD. When observing the results for the resting condition, [12] revealed a shortened EEG microstate duration and an increase in number of single map segments for impaired or demented subjects. Another comparable paper found that the optimal window size gets bigger with more severe cognitive impairment. Also the microstate duration shortens and the potential areas tend to more anterior positions [13]. The magnitude of the parameters obtained by this thesis is in the same range as in the previous done research works.

Starting with the selection of the original maps, one has to consider the following problem. Since the GFP as a function of time is very different for some patients, the automatic selection by using an specifically instructed algorithm is hard to implement. For some patients the GFP changes very quickly and therefore results in a constantly increasing and decreasing function. This results in a high number of GFP peaks, i.e. local maxima, to be selected as original maps. Other patients' EEG signals have a quite steady GFP which increases and decreases slowly. These patients have a quite low number of GFP peaks. For that reason, the automatic selection is not always efficient. This also leads to a strongly varying number of GFP peaks within the study which does not allow to make statements about the changes of the number of GFP peaks in the course of AD. The highly different GFP functions of two patients are shown in Figure 4.1 and 4.2. The number of GFP peaks of the first patient is more than a double of the second one.

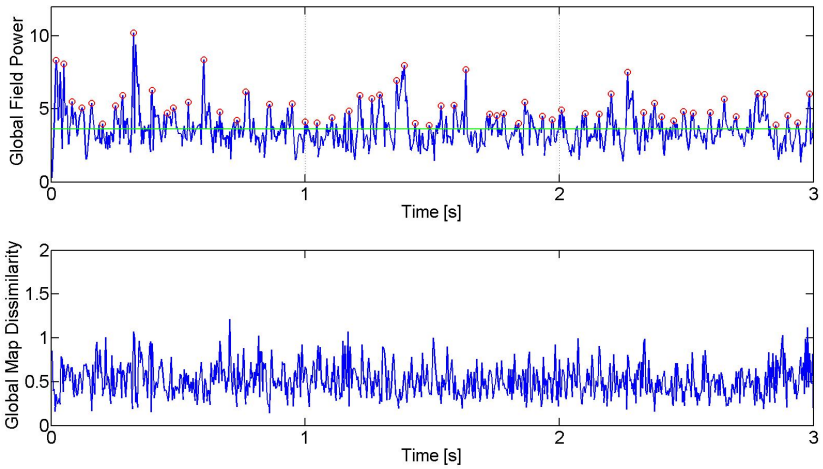

Figure 4.1: Plot of the GFP and GMD as functions of time for a three second EEG epoch of a patient in the study. The patient's GFP function varies a lot and results in 60 GFP peaks, marked with small, red circles for the whole three second epoch. The green horizontal line indicates the mean of the GFP function.

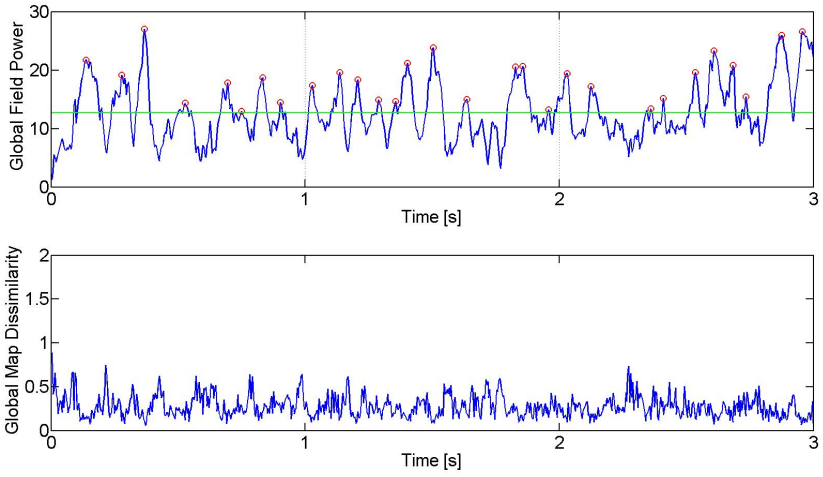

Figure 4.2: Plot of the GFP and GMD as functions of time for a three second EEG epoch of a patient in the study. The patient's GFP function varies slowly and results in 29 GFP peaks, marked with small, red circles for the whole three second epoch. The green horizontal line indicates the mean of the GFP function.

There is only one paper to compare the general result of the parameter GFP Peaks per Second, presented in Figure 3.10, with average of 18.90 ($\sigma = 3.00$), to. In this paper, the average number of GFP peaks per second of patients in rest with eyes closed is 22.7 ($\sigma = 3.3$) for demented (MMSE score under 21) and 22.2 ($\sigma = 3.1$) for impaired subjects (MMSE score under 24) [12].

Another problem occurs when using the segmentation approach related to the optimal window sizes. When looking at Figure 3.11, it is clear that these are varying very much and make a comparison between patients very difficult. The mean optimal window size is 0.57 ($\sigma = 0.14$). Compared to [13], where the optimal window size was smallest for healthy people and got larger for subjects with cognitive impairment, the results in this work show no significant differences related to the severity of AD.

Looking at the parameter Single Map Segments per Second for the group window size with a mean of 4.94 ($\sigma = 1.81$), cf. Figure 3.13, it is again possible to compare the results to these published in [12]. This previous research found that the average number of single map segments per second is 10.8 ($\sigma = 10.8$) for demented and 7.6 ($\sigma = 3.2$) for impaired subjects. Since the number of GFP peaks per second in [12] is also higher, it is quite the same ratio of number of single map segments to number of GFP peaks as in the results obtained here.

Generally, when evaluating the segmentation procedure, one should note that using centroid locations of positive and negative potential areas, the descriptors

always tend to central regions [31]. This fact makes this kind of descriptors quite imprecise and difficult to evaluate.

Now looking at the second way of analyzing microstates within EEG recordings, there were no directly comparable studies published. There are publications concerning EEG microstate analysis by clustering them into classes but not with subjects suffering from AD. Instead, patients with schizophrenia are examined for example in [35] and [32], sleep stages are analyzed in [41] and microstate changes within developmental stages and aging are presented in [38]. Concerning the topography, always four prominent microstate classes are stated, named from 'A' to 'D' and characterized as follows, cf. Figure 4.3:

- *Class A*: right-frontal to left-posterior
- *Class B*: left-frontal to right-posterior
- *Class C*: frontal to occipital
- *Class D*: mostly frontal and medial, slightly less occipital than in C

Comparing the four prominent microstate classes to the results in this work, the four classes found are more or less the same. Only class D is different - the image is about-faced, cf. Figure 3.18.

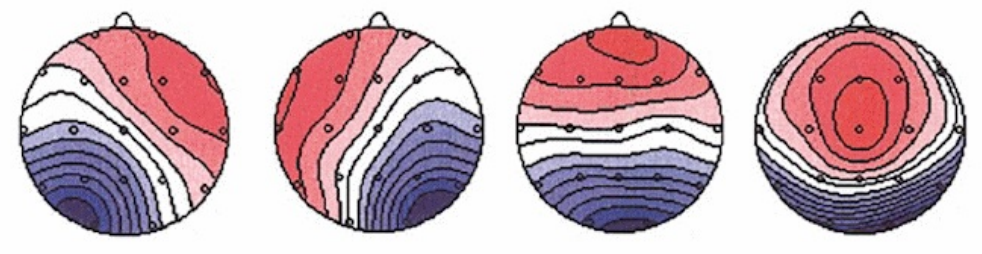

Figure 4.3: This plot presents the four popular microstate classes which have been found while researching on the field of EEG microstates. The human head is seen from above, nose up, red marks positive and blue negative potential areas. This figure was taken from [30].

The parameter Mean Microstate Duration of all four classes ranges from 63 to 73 ms. The analysis resulted in a mean microstate duration for all classes of 66.64 ms ($\sigma = 15.23$). Since there are no other publications concerning AD, the results of papers treating with schizophrenia are stated now. In the first one the control group had a mean microstate duration for all classes of 88.40 ms ($\sigma = 10.70$) and the patients 81.50 ms ($\sigma = 11.90$) [32]. Another paper dealing with that disease reported a mean microstate duration of controls of 89.90 ms ($\sigma = 12.00$) and patients of 84.50 ms ($\sigma = 13.70$) [35].

Chapter 5

Conclusion

Concluding the findings of this work, there is no significant relation between EEG microstates extracted either by clustering or the segmentation procedure using negative and positive potential areas and the severity of AD measured by the MMSE score. There were only scattered significant findings which are not relevant for the study. The usage of demographic variables in the regression model improved the results in terms of variety but the detailed evaluation of the regression parameters revealed no significance of the additionally used demographic variables. Future studies should think about a different procedure to extract original maps and therefore improve the information provided by them. Also, if segmenting the EEG recording, there should be ideas for better descriptors of the landscapes and their characteristics since these always tend to central regions. Another approach could be the splitting of EEG recordings into different frequency bands as it is often used in EEG studies. In most cases, the four bands δ, θ, α and β are used to describe to whole frequency range from approximately 2 to 30 Hz (frequency borders vary in literature). Also the impact of longitudinal studies could be interesting since the basic characteristics of EEG microstates of AD patients vary a lot. The comparing of past and more actual EEG recordings in relation to the worsening of the AD symptoms and the more severe impairment of one patient could therefore be revealing. This could improve the understanding of EEG microstates and their changing in the course of AD.

Bibliography

[1] Alzheimer's Disease International, "World Alzheimer Report 2010: The Global Economic Impact of Dementia," 2010.

[2] http://www.deutsche-alzheimer.de/die-krankheit/die-alzheimer-krankheit.html. Accessed 05 November 2013.

[3] Alzheimer's Disease International, "World Alzheimer Report 2009: The Global Prevalence of Dementia," 2009.

[4] R. Schmidt and et al., "Consensus statement 'Dementia 2010' of the Austrian Alzheimer Society," *Neuropsychiatrie*, vol. 24, no. 2, pp. 67–87, 2010.

[5] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, vol. 34, no. 7, pp. 939–944, 1984.

[6] K. Najarian and R. Splinter, *Biomedical Signal and Image Processing*. CRC Press Taylor & Francis Group, 2012.

[7] S. Sanei and J. Chambers, *EEG Signal Processing*. Wiley, 2007.

[8] I. Wellach, *Praxisbuch EEG - Einführung in die Befundung, Beurteilung und Differenzialdiagnose*. Georg Thieme Verlag, 2011.

[9] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing?," *Current Alzheimer Research*, vol. 7, no. 6, pp. 487–505, 2010.

[10] D. Lehmann and W. Strik, "Data-determined window size and space-oriented segmentation of spontaneous EEG map series," *Electroencephalography and Clinical Neurophysiology*, vol. 87, no. 4, pp. 169–174, 1993.

[11] W. Strik, R. Chiaramonti, G. Muscas, M. Paganini, T. Mueller, A. Fallgatter, A. Versari, and R. Zappoli, "Decreased EEG microstate duration and anteriorisation of the brain electrical fields in mild and moderate dementia of the Alzheimer type," *Psychiatry Research: Neuroimaging Section*, vol. 75, no. 3, pp. 183–191, 1997.

[12] A. Stevens and T. Kircher, "Cognitive decline unlike normal aging is associated with alterations of EEG temporo-spatial characteristics," *European Archives of Psychiatry and Clinical Neurosciences*, vol. 248, no. 5, pp. 259–266, 1998.

[13] T. Dierks, V. Jelic, P. Julin, K. Maurer, L. Wahlund, O. Almkvist, W. Strik, and B. Winblad, "EEG-microstates in mild memory imapirment and Alzheimer's disease: possible association with disturbed information processing," *Journal of Neuronal Transmission*, vol. 104, no. 4, pp. 483–495, 1997.

[14] R. Pascual-Marqui, C. Michel., and D. Lehmann, "Segmentation of Brain Electrical Activity into Microstates: Model Estimation and Validation," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 7, pp. 658–665, 1995.

[15] http://www.alzheimer-gesellschaft.at/index.php?id=27. Accessed 05 November 2013.

[16] M. Folstein, S. Folstein, and P. McHugh, "'Mini-Mental State'. A practical method of grading the cognitive state for patients for the clinician.," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[17] M. Waser, *Changes of EEG Synchrony in the Course of Alzheimer's Disease*. PhD thesis, Technical University of Vienna, 2013.

[18] http://www.alpha-trace.at. Accessed 05 November 2013.

[19] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[20] H.-J. Park, D.-U. Jeong, and K.-S. Park, "Automated detection and elimination of periodic ECG artifacts in EEG using the energy interval histogram method," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 12, pp. 1526–1533, 2002.

[21] M. Waser and H. Garn, "Removing Cardiac Interference from the Electroencephalogram Using a Modified Pan-Tompkins Algorithm and Linear Regression," *35th Annual International IEEE EMBS Conference*, 2013.

[22] P. Filzmoser, "Multivariate Statistik." University Lecture, 2010.

[23] D. Praetorius, "Numerische Mathematik." University Lecture, 2010.

[24] R. Plato, *Numerische Mathematik kompakt*. Vieweg+Teubner Verlag, 2010.

[25] G. Golub and W. Kahan, "Calculating the Singular Values and Pseudo-Inverse of a Matrix," *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965.

[26] J. Hartung and B. Elpelt, *Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik*. R. Oldenbourg Verlag, 1989.

[27] C. Michel and M. Murray, "Towards the utilization of EEG as a brain imaging tool," *NeuroImage*, vol. 61, no. 2, pp. 371–385, 2012.

[28] D. Lehmann and W. Skrandies, "Reference-free identification of components of checkerboard-evoked multichannel potential fields," *Electroencephalography and Clinical Neurophysiology*, vol. 48, no. 6, pp. 609–621, 1980.

[29] D. Lehmann, H. Ozaki, and I. Pal, "EEG alpha map series: brain micro-states by space-oriented adaptive segmentation," *Electroencephalography and Clinical Neurophysiology*, vol. 67, no. 3, pp. 271–288, 1987.

[30] D. Lehmann, R. D. Pascual-Marqui, and C. Michel, "EEG microstates," *Scholarpedia*, vol. 4, no. 3, p. 7632, 2009.

[31] A. Stevens, W. Lutzenberger, D. M. Bartels, W. Strik, and K. Lindner, "Increased duration and altered topography of EEG microstates during cognitive tasks in chronic schizophrenia," *Psychiatry Research*, vol. 66, no. 1, pp. 45–57, 1997.

[32] V. Strelets, P. Faber, J. Golikova, V. Novototsky-Vlasov, T. Koenig, L. Gianotti, J. Gurzelier, and D. Lehmann, "Chronic schizophrenics with positive symptomatology have shortened EEG microstate durations," *Clinical Neurophysiology*, vol. 114, no. 11, pp. 2043–2051, 2003.

[33] http://www.mathworks.de/de/help/signal/ref/findpeaks.html. Accessed 20 November 2013.

[34] D. Lehmann, W. Strik, B. Henggeler, T. Koenig, and M. Koukkou, "Brain electric microstates and momentary conscious mind states as building blocks of

spontaneous thinking: I. Visual imagery and abstract thoughts," *International Journal of Psychophysiology*, vol. 29, no. 1, pp. 1–11, 1998.

[35] D. Lehmann, P. Faber, S. Galderisi, W. Herrmann, T. Kinoshita, M. Koukkou, A. Mucci, R. Pascual-Marqui, N. Saito, J. Wackermann, G. Winterer, and T. Koenig, "EEG microstate duration and syntax in acute, medication-naï ve, first-episode schizophrenia: a multi-center study," *Psychiatry Research: Neuroimaging*, vol. 138, no. 2, pp. 141–156, 2004.

[36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2011.

[37] T. Koenig, D. Lehmann, M. Merlo, K. Kochi, D. Hell, and M. Koukkou, "A deviant EEG brain microstate in acute, neuroleptic-naive schizophrenics at rest," *European Archives of Psychiatry and Clinical Neurosciences*, vol. 249, no. 4, pp. 205–211, 1999.

[38] T. Koenig, L. Prichep, D. Lehmann, P. Sosa, E. Braeker, H. Kleinlogel, R. Isenhart, and E. John, "Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages," *NeuroImage*, vol. 16, no. 1, pp. 41–48, 2002.

[39] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[40] http://www.mathworks.de/de/help/stats/regstats.html. Accessed 20 January 2013.

[41] V. Brodbeck, A. Kuhn, F. von Wegner, A. Morzelewski, E. Tagliazucchi, S. Borisov, C. Michel, and H. Laufs, "EEG microstates of wakefulness and NREM sleep," *NeuroImage*, vol. 62, no. 3, pp. 2129–2139, 2012.